Journal of
Risk and Financial Management

Special Issue Reprint

Risk and Financial Consequences

Edited by
George Halkos

www.mdpi.com/journal/jrfm

MDPI

Risk and Financial Consequences

Risk and Financial Consequences

Editor

George Halkos

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editor
George Halkos
Department of Economics
University of Thessaly
Volos
Greece

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Journal of Risk and Financial Management* (ISSN 1911-8074) (available at: www.mdpi.com/journal/jrfm/special_issues/Risk_Finacial_Consequences).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-8369-3 (Hbk)
ISBN 978-3-0365-8368-6 (PDF)

Contents

About the Editor

George Halkos

George E. Halkos (BA, MSc, PhD) is a Professor in Economics of Natural Resources in the Department of Economics at the University of Thessaly, Greece. He has worked as team leader and research fellow in various research and academic institutions. He has participated and presented articles in international conferences, and he acts as a referee for many scientific journals.

His research interests are in the fields of applied statistics and econometrics, simulations of economic modelling, natural resource and environmental economics, applied micro-economics with an emphasis on welfare economics, air pollution, game theory, and mathematical models (such as non-linear programming).

Preface to ”Risk and Financial Consequences”

Risk is a term that exists in everyone’s life, not only when an unexpected event occurs, but in any decision someone has to make. Among those studying or working in the financial sector, there is widespread knowledge regarding the rational investor’s preferences. Regardless of the business sector in which they will choose to invest, an investor’s main goal is to maximize their profits. For this reason, investors are characterized in the literature as “risk averters”. Hedging and portfolio diversification appear to be efficient in reducing the potential loss of an investment. Considerable attention has been drawn to the advantages of portfolio diversification, while researchers have mentioned the ability of transferring the risk of investment through hedging, emphasizing that the size of the optimal hedging ratio is one of the main determinants used by decision makers, in addition to the financial situation of the corporation. However, there are some cases which cannot be predicted. Society, which refers to nature’s actions and human activities, is characterized by vulnerability and rapid changes. Nature acts independently; a common example of this independence is tectonic plate movement. Thus, natural actions that cannot be predicted may cause significant losses, both economic and life-related. Economic losses can be due to many factors, such as the partial or total destruction of homes or business premises that will lead to reduced, if not zero, productivity. In the case of businesses, reduced productivity may affect investors’perceptions, causing fluctuations in the share price or even volatility in the board of directors.

George Halkos
Editor

Journal of
Risk and Financial
Management

MDPI

Article

Is Investors' Psychology Affected Due to a Potential Unexpected Environmental Disaster?

George Halkos * and Argyro Zisiadou

Laboratory of Operations Research, Department of Economics, University of Thessaly, 38333 Volos, Greece; zisiadou@uth.gr

* Correspondence: halkos@econ.uth.gr; Tel.: +30-2421-074-664

Received: 9 June 2020; Accepted: 10 July 2020; Published: 12 July 2020

Abstract: The purpose of this paper is to approach the way investors perceive the risk associated with unexpected environmental disasters. For that reason, we examine certain types of natural and technological disasters, also known as "na-tech". Based on the existing relevant literature and historical sources, the most common types of such disasters are geophysical and industrial environmental disasters. After providing evidence of the historical evolution of the na-tech events and a brief description of the events included in the sample, we estimate the systematic risk of assets connected to these events. The goal is to capture possible abnormalities as well as to observe investors' psychology of risk after the occurrence of an unexpected event. Finally, we examine whether macroeconomic factors may affect those abnormalities. The empirical findings indicate that the cases we examined did not cause significant cumulative abnormal returns. Moreover, some events caused an increase in systematic risk while surprisingly some others reduced risk, showing that investors tend to support a country and/or corporation due to their reputation.

Keywords: na-tech; systematic risk; market reaction; unexpected events; investing

---

1. Introduction

Among those studying or working in the financial sector there is a well-spread knowledge regarding the rational investors' preferences. Regardless of the business sector they invest in, investors' main goal is to maximize their profits, being characterized as "risk averters" (Merton 1969; Benartzi and Thaler 1999; Campbell and Cochrane 1999; Aıt-Sahalia and Lo 2000; Jackwerth 2000; Rosenberg and Engle 2002; Brandt and Wang 2003; Gordon and Pascal 2004; Bliss and Panigirtzoglou 2004; Bollerslev et al. 2011; Halkos et al. 2017). Investors are usually assumed to be rational, so if we ignore the arbitrage case, they tend to choose more "safe" investments which will allow them to maximize their profits, or in other words minimize potential risk they receive by investing (Cohn et al. 1975; Benartzi and Thaler 1995; Haigh and List 2005).

Hedging and portfolio diversification may appear to be efficient in reducing the potential loss of an investment. Great attention has been drawn about the advantages of portfolio diversification (Bugár and Maurer 2002). Graham and Jennings (1987) mentioned the ability of transferring the risk of investment through hedging, while Bond and Thompson (1985) highlighted that the size of the optimal hedging ratio is one of the main determinants used by decision makers apart from cash position of the corporation.

Based on available information, such as credit rating, stability of the corporation or government, whether we are working with stocks or bonds, as well as investors' preferences, diversification of portfolios and assurance of investors' capital may be achieved. The first researcher who extended the Markovitz theoretical idea of the modern portfolio selection was Grubel (1968). We can use options or future derivatives concerning our predictions regarding, for instance, the value of exchange rates.

However, there are some cases which cannot be predicted. The act of nature is such a case (Halkos and Zisiadou 2018).

Nature acts independently, and a common example of that independence is the tectonic plate movement (Halkos and Zisiadou 2018). Distinguished sciences, such as Geology and Seismology, have the techniques to monitor, observe, and examine the geophysical events caused by those tectonic plate movements. However, even those specialized sciences cannot predict the occurrence, and more specifically the exact place, epicenter, and intensity of an upcoming event. Thus, those actions which cannot be predicted may cause significant losses, both economic and life-related. Regarding economic losses, they can be due to many factors such as partial or total destruction of homes or business premises that will lead to reduced, if not zero, productivity.

Concerning the country that may be affected by such a situation, reduced productivity can cause a drastic decline in gross domestic product (GDP)[1], as well as affect a country's borrowing capacity and reliability, which tend to be depicted in its credit rating. As we have already mentioned, the credit rating of a country or corporation is one of the most common rates that mirror the potential risk the investor is about to perceive by investing in this specific bond or stock. In the case of businesses, reduced productivity may affect investors' perceptions causing fluctuations in the share price or even volatility in its board of directors. Although disasters are associated with risk, investors tend to have a different perspective regarding the source of the disaster[2].

More specifically, if a country is facing a natural disaster, where no one can be blamed, foreign investors who may hold this country's bonds will continue to trust the country due to the "innocence" of the country. On the other hand, when a firm causes a technological disaster, such as a nuclear power plant explosion, investors will "punish" the firm by selling its shares at any price to avoid a bigger loss, if this corporation is publicly traded. In such cases, corporations may lose trustworthiness. Nevertheless, in some cases, the possible technological disaster is not a firm's fault.

In the literature, these cases tend to be called "na-tech", a term that actually depicts the source of the disaster. Sometimes, one natural disaster, caused by a tectonic plate movement, can lead to another natural or even technological disaster. For instance, a ground movement may lead to another earthquake[3] at the seabed, known as a tsunami, or an earthquake, in general, may cause a dysfunction in a factory installation which may therefore cause an industrial disaster. A characteristic example is the case of Fukushima Daiichi Power Plant disaster, which will be analyzed in the following section.

The purpose of this paper is to approach the way investors perceive the risk associated with unexpected environmental disasters. More specifically, it examines certain types of natural and technological disasters which tend to be associated and are listed under the categories of geophysical and industrial disasters. The goal is to capture possible abnormalities as well as to observe investors' psychology of risk after the occurrence of an unexpected event. Ultimately, we examine whether macroeconomic factors may affect those abnormalities. The structure of the paper is as follows. Section 2 reviews the existing relevant literature, while Section 3 provides the methodology and data used. In Section 4 the results of the analysis are presented and discussed. Finally, Section 5 concludes the study providing important statements and fundamentals for further research.

---

1 Athukorala et al. (2018) investigated the impact caused by natural disasters regarding residential property pricing.

2 A detailed review of terminology regarding disasters and all criteria taking into consideration when characterizing an event as a disaster is given in Halkos and Zisiadou (2018).

3 Earthquakes are divided into two types of events, the "ground movement" which is the movement on the land caused by the tectonic plate movement, and the "tsunami" which is the waves caused by the movements at the seabed (Halkos and Zisiadou 2018).

2. Literature Review

2.1. Geophysical and Industrial Hazards

Geophysical phenomena are not unexpected processes in terms of appearance and frequency. More specifically, since ancient times, the existence of these phenomena was known and commonly detected at a higher frequency in certain areas across the globe. The continuous movement of the Earth's parts combined with weather condition changes have shaped the present image of the planet. Islands have been created or destroyed by volcanic eruptions, and landscapes have undergone changes from ground movements and tidal waves; however, the intensity of the event is the main factor affecting the final outcome.

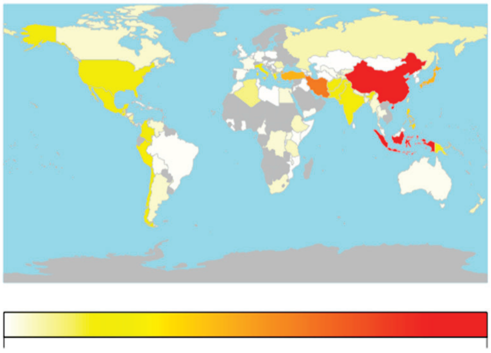
Additionally, natural disasters cause a great number of fatalities as well as supreme national catastrophes (Viscusi 2009). An extended literature considering the terminology and high-risk areas was presented by Halkos and Zisiadou (2018). Based on the CRED (Centre for Research on the Epidemiology of Disaster) database, 1621 geophysical events have been recorded since 1900 causing 2,678,022 fatalities and economic damages, which aggregately exceeded 781.5 billion United States Dollar (USD) (EMDAT 2017; Halkos and Zisiadou 2018). The significance of a natural disaster to the economy was also emphasized by Lee et al. (2018), who described how possible natural disasters tend to cause volatility on stock markets.

Technological accidents, on the other hand, do not have similarities with the geophysical phenomena regarding expectancy. Nowadays, technology takes up more and more space in our lives, not only for professional but also for personal purposes. Of course, when it comes to technological accidents and disasters, the first thing that comes to mind is industrial accidents. What is important to mention is that, technological disasters include all types of accidents that may occur with technology as one of the main factors. The three main categories of technological accidents are industrial, miscellaneous, and transport accidents (Halkos and Zisiadou 2019).

Once again using the CRED database, we can come to the conclusion that industrial hazards are not the most frequent, however, they are the most disastrous. Over the last 117 years (1900–2016), 1434 industrial events have caused 57,619 fatalities and almost 43.1 billion USD economic damages (EMDAT 2017; Halkos and Zisiadou 2019). Industrial hazards, or even disasters, include all cases that may cause production disruption or even fatalities involving industrial buildings, such as chemical spills, collapses, explosions, fires, gas leaks, oil spills, poisoning, and radiation. What is really important to mention is that the number of fatalities became greater during the last years due to population increases in these high-risk areas (Kunreuther 1996).

Although there is a perception that natural phenomena are unexpected and occur randomly, there is evidence to suggest that, partially, the assertion of randomness is not valid. To be more specific, there is a proven regional distribution regarding geophysical events, initially mentioned by Bolt (1988) as the "Ring of Fire". Based on the CRED database (EMDAT 2017) and with the use of R-studio packages and routines, maps of occurrence have been created both for geophysical and industrial hazards (Halkos and Zisiadou 2018, 2019). Figure 1a represents the space concentration of geophysical hazards in the high-risk area called the "Ring of Fire".

In other words, although the exact place and time of an upcoming geophysical event cannot be predicted, based on evidence, we know a priori, which regions are more prone to face another disaster. Due to the possibility of a new catastrophe, governments should pay more attention to those high-risk areas as an attempt to reduce possible losses (Viscusi 2006). What is interesting though is that, although we were expecting a space concentration regarding natural events, the assertion of regional distribution is also observed in the case of industrial hazards. As seen in Figure 1b, East Asia is the most suffered region regarding industrial disasters. Although the reasons for such a space concentration are not known, based on evidence, researchers have at their disposal data that provide them with a first illustration of the riskiest areas.



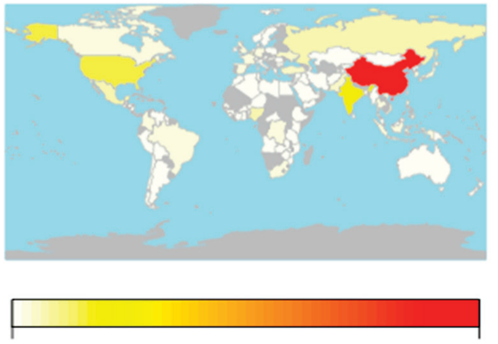


Figure 1. Maps of occurrence. (**a**) Geophysical hazards and (**b**) industrial hazards. Note: The scale of maps represents cumulative observations of event occurrences in each country over the years since 1900.

As already mentioned, investors are primarily oriented to avoid most risk, or at least try to protect themselves from it. If they know, therefore, in advance, the risks they adopt by investing in those regions, they may be able to fully diversify their portfolios. Before analyzing the cases that will be used in our modeling, it is important to understand the basic concepts related to the seriousness of incidents included in our sample. The first basic requirement for sample creation is the date that each event occurred, as we included events from 2000 onwards for reasons of availability of stock data. The second, and equally important, reason is the intensity of each event.

Regarding the events we examine, there are four different intensity scales. For earthquakes, either ground movements or tsunamis, there are two different scales, the Moment Magnitude (M_w), also known as the Richter Scale, and the Intensity Scale, also known as the Mercalli Scale. These scales are most commonly used while measuring the intensity of a natural geophysical event. Regarding industrial events, to our knowledge, there is only one scale that is used as a tool to rate an industrial disaster. This scale is the International Nuclear and Radiological Event Scale (INES) which was created by the International Atomic Energy Agency (IAEA) and the Nuclear Energy Agency of the Organization for Economic Co-operation and Development (OECD/NEA) in 1990[4]. The values of each scale and their impact are quite useful when describing an event and explaining the reason for inclusion in the sample.

2.2. Events of the Research

The survey sample contains 25 events, including 12 earthquakes, nine volcanic eruptions, and just four industrial disasters. The initial list of events that occurred since 2000 included more industrial accidents, however most cases were related to non-publicly listed corporations. These corporations have no available share price data which automatically excludes them from the sample of examination. Additionally, events before 2000 were also excluded due to the fact that open source databases barely have available historical data before that date. A significant number of industrial disasters has been excluded due to the fact that most of the corporations are not publicly listed, thus there are no stock prices available for market reactions estimation. Moreover, regarding geophysical disasters, the 2006 Indonesia earthquake was excluded from the research after observing anomalies in its behavior, and more specifically, negative systematic risk.

The first event of the analysis (Event 1) is the Denali earthquake in Alaska, USA on 3 November 2002 (Dunham and Archuleta 2004; Eberhart-Phillips et al. 2003; Jibson et al. 2006; Freed et al. 2006), with a M_w = 7.9 and a maximum intensity IX causing approximately 56 million USD in economic

[4] International Atomic Energy Agency (1998).

losses. This earthquake caused an estimated total damage of 20–56 million USD and one injury, while it triggered several landslides, with the worst of them causing a collision of 30 million m^3 of rocks and ice (Eberhart-Phillips et al. 2003). The 2004 Indian Ocean earthquake, which took place in Thailand on 26 December 2004 (Event 2), with a M_w = 9.1–9.3 and a maximum intensity IX, caused an estimated total damage of 15 billion USD and 227,838 fatalities. The tsunami was created after the ground movement reached a 51 m wave. According to Telford and Cosgrave (2007), the most interesting evidence from the earthquake, that occurred over the Burma and Indian plate joint, was the immediate funding response across the globe. Wang and Liu (2006) characterized this geophysical event as one of the most devastating over the last 100 years. Moreover, they mentioned that the initial magnitude estimation was 9.0 which afterwards was updated to 9.1–9.3 while the waves created by the earthquake travelled at the speed of 700 Km/h.

In Pakistan, the Kashmir earthquake took place on 8 October 2005 (Event 3) with a M_w = 7.6 and a maximum intensity VII causing in economic damage as reported by the Asian Development Bank and the World Bank. The casualties of this disaster were 28 million displaced citizens, an estimation of 86,000–87,351 fatalities, and 6900–75,266 injuries citizens (Avouac et al. 2006). According to Kamp et al. (2008), the Kashmir earthquake triggered severe landslides, however, the majority of the fatalities were caused due to the inappropriate design of buildings and poor quality of construction materials. The aftermath of natural disasters regarding losses can be reduced by respecting and following the construction building codes (Priest 1996).

Based on Liu-Zeng et al. (2009), the 2008 Sichuan earthquake (12.05.2008) in China (Event 4) was a geophysical disaster with a M_w = 8.0 which devastated the western rim of Sichuan Basin. The maximum intensity of that earthquake was XI and the estimated total damage was 150 billion USD, with 87,587 fatalities, 374,643 injured, and 18,392 missing citizens. One of the most known and disastrous geophysical events of the new millennium was the tsunamigenic[5] earthquake in Tohoku, Japan on 11 March 2011 (Event 5) which led to the greatest nuclear accident in recent years (after the case of Chernobyl) at the Fukushima Daiichi Nuclear Power Plant. The earthquake with M_w = 9.0–9.1 and maximum intensity IX created waves up to 40.5 m as well as landslides. The estimated total damage was 360 billion USD. The industrial case of the Daiichi disaster is described further in detail as Event 25. The estimated economic damage that occurred due to this disastrous event was approximately 360 billion USD, from which, 210 billion USD was recorded as due to the industrial disaster.

Moving forward, on 22 February 2011 in Christchurch in New Zealand, a M_w = 7.2 earthquake, which also led to a tsunami and landslides, caused 115 fatalities and 1500–2000 injuries. The maximum intensity of this event (Event 6) was IX. Bradley and Cubrinovski (2011) explained that New Zealand is located on the joint of the Pacific and Australian plates, two active tectonic plates with lateral sedimentations. This earthquake, together with the one that occurred on 23 December 2011 in New Zealand, caused an estimated 15–30 billion USD in damage.

Less than eight years after the devastating Indian Ocean earthquake in 2004, another M_w = 8.6 earthquake struck on 11 April 2012 in the Indian Ocean, however, this time the affected country was Indonesia (Event 7). Pollitz et al. (2012), mentioned that this earthquake, with a maximum intensity VII, 10 fatalities, and 12 injured citizens, was by far the largest strike-slip event, causing almost 9 billion USD in economic damage. The 2012 Indian Ocean earthquake was a tsunami associated event, as most of the cases examined in this paper, showing that natural events, or even disasters, are not individual incidents—nature interacts. The Illapel case was a M_w = 8.3 earthquake which affected both Chile and Argentina on 16 September 2015 with maximum intensity IX; it also created a tsunami. The aftermath of this earthquake in Chile was 15 fatalities and six missing citizens and 800 million USD in economic losses, while in Argentina there was only one fatality and few injured citizens (Event 8). Ruiz et al. (2016) mentioned that since the earthquake of Malue in 2010, there is an extensive post

---

5 Kanamori (1972), all earthquakes that can create tsunamis can be classified as tsunamigenic earthquakes.

seismic distortion. Based on Heidarzadeh et al. (2016), the Illapel case raised a lot of attention and has been observed by the Pacific Tsunami Warning Center and the Japan Meteorological Agency. The tsunami, which was created by the earthquake in South America that affected Chile and Argentina, reached the coastlines of Japan, Hawaii, New Zealand, Vanuatu, and Australia.

Meanwhile, on 25 October 2010, Indonesia faced all three cases of geophysical disasters (Event 9). The Mentawai earthquake in Sumatra was a $M_w = 7.8$ earthquake, which led to a tsunami as well, that caused 408 fatalities and 303 missing citizens, but reported no economic losses. As Newman et al. (2011) mentioned, the Mentawai earthquake was characterized as a rare slow-source tsunami earthquake. On the same day, in the region near Java in Indonesia, one of the most active and hazardous volcanoes globally, Mount Merapi, erupted (Jousset et al. 2012). This volcanic eruption caused a chaotic situation regarding air traffic to the point that 2000 flights were cancelled. At the end of 2011, and more specifically, on 23 December 2011, New Zealand experienced another earthquake event (Event 10). As Bannister and Gledhill (2012) described, two ground movements took place in Christchurch 10 km and 15 km east from the city center, respectively, only 10 months after the first earthquake of the same year. The cumulative economic losses were 15–30 billion USD as previously mentioned.

On 21 July 2014, a lake tsunami occurred in Iceland and more specifically in Aksja, caused by volcanic activity (Event 11); no fatalities nor economic damage was recorded. Gylfadóttir et al. (2017) emphasized the unique phenomenon of a tsunami into a lake due to the rockslide that was released from the inner Askja caldera. The last earthquake included in the sample of analysis is the one that took place in Kaikoura, New Zealand on 14 November 2016 (Event 12). Once again, it was the joint of the Pacific and Australian plates, two active tectonic plates with lateral sedimentations, that created the $M_w = 7.8$ earthquake event with a maximum intensity IX which afterwards led to a tsunami causing two fatalities and 57 injured citizens (Hollingsworth et al. 2017). The economic impact of this event reached 613 million New Zealand Dollar (NZ$).

Moving forward, the next events included in the analysis are related to volcanic activity since 2000. On 7 August 2008, Kasatochi volcano (Event 13) in the USA erupted unexpectedly. Based on Waythomas et al. (2010), this specific volcano had no significant eruptions since then, however, the eruption of 2008 received a level 4 rating on the Volcanic Explosivity Index (VEI) scale. Almost a year later, another eruption received a level 4 rating. On 11 June 2009, in Russia, the Sarychev Peak erupted (Event 14). As Urai and Ishizuka (2011) mentioned, the Sarychev Peak is not monitored with ground-based instruments, however, the great eruption of 2009, which lasted almost eight days was captured by satellites. Iceland is very famous regarding volcanic activities. On 20 March 2010 and for 39 days (Event 15), a level 4 volcanic eruption took place in Eyjafjallajokull (Gudmundsson et al. 2012). Moreover, the next year, on 22 May 2011 (Event 16), another level 4 eruption occurred at the most active volcano in Iceland, Grimsvotn, located beneath the Vatnajokull ice sheet (Sigmarsson et al. 2013).

Scollo et al. (2014) and Viccaro et al. (2015) examined the Etna eruption on 05 March 2013 (Event 17) which received a level 3 rating on the VEI scale. In a period of two years, Etna in Italy produced 38 basaltic lava fountains. According to Viccaro et al. (2015), the volcanic activity started after eight months of rest. Kato et al. (2015) analyzed the Mount Ontake, Japan, volcanic eruption on 27 September 2014 (Event 18), that caused the deaths of 57 climbers. The number of fatalities increased since six more missing climbers were assumed to be dead. Kaneko et al. (2016) characterized this eruption as a small eruption with a short period; since it received a level 3 rating, it has not been examined thoroughly, thus the causes remain unknown. The volcanic eruption of Kelud, Indonesia on 13 February 2014 (Event 19), raised a lot of attention as it was characterized as the most powerful eruption of the decade (Caudron et al. 2015) causing 185 million USD in economic damage compared to all previous volcanic activities that had no reported economic impacts. A historic eruption was the one of Calbuco, Chile on 22 April 2015 (Event 20) since it did not have any recorded eruption in the last 43 years (Van Eaton et al. 2016). Due to the eruption, volcanic ash was dispersed in Chile, Argentina, and Uruguay. Ivy et al. (2017) observed a change in the ozone hole caused by that eruption, which also reported 600 million USD in economic losses. The last volcanic eruption included in the sample is that of Sinabung (Event 21) on 22 May 2016. Sinulingga and Siregar (2017) mentioned that Sinabung is one of 130 volcanoes in Indonesia; it lies on the Ring of Fire, which is a high-risk area concerning earthquakes.

The final category of the events included in the sample are related to industrial accidents. Only four events were included in the analysis due to the fact that the rest of the corporations, which caused industrial disasters, are not publicly listed. Three of those events are oil spills while the last event is the greatest nuclear disaster in recent years, and one of the two greatest disasters in history. Event 22 is the Prudhoe Bay oil spill caused by BP on 2 March 2006. Kurtz (2010), this specific oil spill was the largest pipeline incident in the history of the operating system. The reputation of BP suffered from this incident and four years after the first oil spill of the new millennium, a new oil spill, this time in the Gulf of Mexico, occurred and aggravated the existing situation. Moreover, BP was forced to pay a 25 million USD fine for the environmental disaster. On 20 April 2010, Deepwater Horizon (Event 23), caused 11 fatalities and 17 injuries, as well as an environmental disaster due to the 2.1 million gallons of dispersants on the surface and wellhead of the Gulf of Mexico (Kujawinski et al. 2011). This time the economic impact for BP was dramatically increased compared to the previous event. The corporation had to pay 70 billion USD in fines and cleanup costs, while at the same time the market value of the corporation faced a sharp decrease. Three years later, on 29 March 2013, the Mayflower oil spill (Event 24) caused by Exxon Mobil, released more than 5000 barrels of crude oil. Droitsch (2014), 1.36 million gallons of crude oil have proven very difficult to clean up. The fine that Exxon Mobil was forced to pay equaled 5 million USD.

Last, but not least, is the nuclear disaster of the Fukushima Daiichi Nuclear Power Plant that occurred on 11 March 2011, after the Tohoku earthquake and tsunami that reached the coast of Japan. Due to the earthquake 11 nuclear power plants stopped their operations. The cooling system of Fukushima's power plant also stopped operating, causing the mot catastrophic nuclear accident after the one in Chernobyl. As already mentioned, 210 billion USD of the 360 billion USD recorded as economic damage were due to this industrial disaster and not due to the tsunami. Initially, the Fukushima Daiichi disaster received a level 5 rating on the International Nuclear and Radiological Event Scale (INES), but after the reassessment of the situation, Fukushima disaster received a level 7 rating. Till that day, only Chernobyl had a level 7 rating (Norio et al. 2011).

2.3. Proposed Methodologies

The process of modeling has never been easy. Of course, it may become even more lax and chaotic if qualitative variables are included in the study, which may not be measurable (such as the investor's psychology, credibility of a government, and reputation of a business). Of course, with appropriate econometric methods and use of specific variables, we can partially integrate qualitative variables in our models. Although one factor that cannot be modeled is randomness. From a theoretical point of view, it is expected that we cannot model and therefore predict the "unexpected" because then it would cease to be a random event. Randomness, and consequently uncertainty, are what characterize markets. However, significant efforts have been made to evaluate models that can determine the expected value of an asset.

The most known models are the market model, arbitrage pricing theory (APT), and capital asset pricing model (CAPM)[6]. Regarding cases where unexpected events or announcements occur, event study analysis initially proposed by MacKinlay (1997) is the most common method estimating

---

[6] Another well-known pricing model is the Fama–French three-factor model. The Fama–French three-factor model is a widespread asset pricing model that expands on the capital asset pricing model by adding size risk and value risk factors to the market risk factors. One of the factors included, though, is the "HML factor" which represents the "high minus low" book-to-market ratio. This ratio can be estimated for corporations' shares, however, to our knowledge, it is not applicable to government bond cases. Our attempt mainly included government bonds, which as a result lacks information regarding the HML factor. Supplementary research to this topic with the inclusion of more corporations will definitely incorporate the Fama–French three-factor asset pricing model.

abnormality on returns (Prabhala 1997; Binder 1998; Maloney and Mulherin 2003; Gaspar et al. 2005; Karolyi and Martell 2010; Charles and Darné 2006; Walker et al. 2006; Arin et al. 2008; Brounen and Derwall 2010; Carpentier and Suret 2015; Halkos et al. 2017).

Of these three approaches, the preferred one is the CAPM. Most financial advisors as well as many researchers tend to use it in order to estimate systematic risk of each stock or bond (Strong 1992; Faff 1991; Fernald and Rogers 2002; Chen 2003; Womack and Zhang 2003; Fernandez 2006; Bruner et al. 2008; Adrian and Francesco 2009). By estimating systematic risk, we can therefore predict expected returns of the asset examined, as well as abnormal returns using the actual value of return. When an unexpected event occurs, the question raised is whether those abnormal returns tend to be significant, showing the reaction, either positive or negative, of investors. This is the main path to follow in our analysis.

3. Methodology

3.1. Hypotheses and Data

Carter and Simkins (2004) decided to investigate airline stocks after the terrorist attacks on 11 September 2001. Their findings provide information regarding the USA capital market and the returns of airline corporations. They found that after the attack, statistically significant negative abnormal returns were observed for the examined airlines. Based on that outcome, we intend to observe if the under-investigation assets follow the same path (Hypothesis 1). Another significant finding by Carter and Simkins (2004) is that the results indicated a rapid drop of stock prices which led to a shock of the USA capital markets. Based on that finding, we seek to examine whether an unexpected disaster can have a similar impact on the government's bond price or the share price of a corporation[7] (Hypothesis 2).

Moving forward, the psychological impact of an unexpected event, which in Carter and Simkins' (2004) case is the September 11th attack, may trigger the rationality that characterizes investors and leads them to react immediately causing pricing volatility. However, it is proven by evidence that larger airline corporations took advantage of that event, while smaller airline corporations did not have that opportunity. On the same path, we shall observe if such a condition is feasible at a country level (Hypothesis 3). Finally, Carter and Simkins (2004) investigated the impact of corporations' size to market reaction giving us the priming to include countries' economic status and its impact on the investors' psychology (Hypothesis 4).

The first step of our analysis is to put the underlying assumptions to determine both the course of analysis and the time interval and variables to be used. The main hypotheses that will be examined here as follows:

Hypothesis 1 (**H1**)**.** *There is no significant abnormal return after an unexpected na-tech disaster.*

Hypothesis 2 (**H2**)**.** *The systematic risk of an asset remains unaffected by an unexpected na-tech disaster.*

Hypothesis 3 (**H3**)**.** *Macroeconomic factors of the country suffering from an unexpected na-tech disaster cannot influence investors' psychology and decisions.*

Hypothesis 4 (**H4**)**.** *The status of an economy does not affect investors' decision.*

For examining the above hypotheses, both financial and macroeconomic data are used. Regarding financial data, it is important to mention that daily stock and bond prices have been derived from

[7] It refers to the industrial disasters, which may have been caused by anthropogenic factors.

open-source databases[8] with a time-span of 125 days before the occurrence of the event as well as three days after the occurrence of the event to capture the possible return abnormality. When the event of analysis belongs to a natural disaster, the asset of examination is the country's government bond, while the market index used is the corresponding Government Bond Index. In order to collect the bond data, we searched for data related to the government bond with longer time-to-maturity of each country facing a disaster. However, in some cases, the longer time-to-maturity bond had stable (same) bond price values, which would have given us bond returns equal to zero. In those cases, the exact previous bond was selected and included in our analysis. Restrictions regarding the open source data, possible exclusion of events due to lack of data, and the unavailability of dividend yields and/or stock splits have undoubtedly affected our final estimations. When the event of analysis belongs to a technological disaster, the asset of examination is the corporation's stock price, while the market index used is the corresponding market index which the corporation is listed in.

Concerning the risk-free asset that is necessary for the CAPM approach, the assumption of Barro and Misra (2016) was used; they underlined that gold can be considered as a risk-free asset since it cannot be used as a hedge against macroeconomic declines and its expected real rate of return should be close to risk-free. As already mentioned, some events have been excluded from the analysis due to lack of information, mainly because corporations are not publicly listed, or due to overlapping cases, where the examination window of one event overlaps with the estimation window of another in the same country. Using this open-source database, confronts us with the main limitation of the research, in terms of time span. This is the main reason we chose na-tech disasters which occurred since 2000. For the macroeconomic factors' variables, the reliable and recognized database of the World Bank[9] was used.

3.2. Event Study Analysis

The most widespread method for analysis of the market reactions is the event study analysis as described by MacKinlay (1997). The initial step for the following analysis is to set the estimation and the event windows. As an event window, we used a seven-day period (−3, +3) centered to the event day[10] and including three days before and three days after the event, in an attempt to capture market reaction to the disaster. This event window will be used to estimate the expected return of the asset as well as the possible abnormality. As an estimation window, we used a 120-day period (−124, −4), which should not include the days of the event window. By establishing a wider estimation window, compared to the time span proposed by MacKinlay (1997), we estimated the systematic risk before the occurrence of the event with higher accuracy. This approach allowed us to predict more precisely the expected returns of the assets on the seven-day event window and these expected returns provided more accurate abnormal returns. After calculating systematic risk, we moved forward to the event window to approach abnormal returns.

The final step was to compute the cumulative abnormal return (CAR) which was examined for its significance. Moreover, as an extension of the proposed methodology, we decided to examine the abnormal returns of the seven-day event window for all 25 events and how they were influenced

8 The source of data is the website Investing.com: www.investing.com (accessed on 1 October 2018). We are familiar with that fact that this database is not the most accurate source due to the fact that provides data for delisted stocks nor it is adjusted for splits and dividends; however, to our knowledge, it is the only open source which provides the majority of the needed information. Sources such as Bloomberg and/or Thomson Reuters DataStream are preferable, however, no access was granted. The non-inclusion of dividend yield and/or stock spilt event certainly has an impact on our estimations. These non-adjustments may cause under/overestimation of the systematic risk. Further research would preferably include a more detailed data source which will give us the ability to include those adjustments in our estimations.

9 Source of data is the website of the World Bank: https://data.worldbank.org (accessed on 5 October 2018).

10 Many events occurred over multiple days. However, we consider the first day of the event as day zero due to the fact that at this moment the event was recognized as unexpected, while the aftershocks on the following days are assumed to be expected. Moreover, due to the time span of the 70-day ex-ante analysis, we included the possible reaction due to the multiple day occurrences that followed the first day of the events.

by macroeconomic factors. The initial methodology proposed by MacKinlay (1997) examined the cumulative abnormal returns using cross-sectional data analysis. However, we decided to observe separately all abnormal returns instead of their aggregations. The other extension included in the analysis was the inclusion of a dummy variable receiving the value 1 for the day of the event occurrence and the three-day span after the occurrence, and zero otherwise. The products (X_jD) will shed more light on the post event reaction.

Specifically, as the abnormal return (*AR*) we set the actual ex post return of the security over the event window after extracting the normal return of security over the same period. The normal return equals to the expected return without occurrence of the unexpected event. For each case *i* and during period *t* the abnormal return is given by (1), where AR_{it}, RA_{it}, $E(RA_{it}|X_t)$, RM_{it}, RF_{it}, and e_{it} stand for abnormal, actual, normal returns, return of market and risk-free assets and residuals, respectively during the period *t* and X_t refers to the conditioning information (MacKinlay 1997):

$$AR_{it} = e_{it} = RA_{it} - E(RA_{it}\big|X_t) \tag{1}$$

Based on CAPM specification, systematic risk known as β_i, is defined as the covariance of RA_{it} with RM_{it} over some estimation period (Cov(RA_{it}, RM_{it}) divided by the variance of RM_{it} over the same period (Var(RM_{it})) (Jagannathan and Wang 1993; Armitage 1995).

$$\begin{gathered} E[RA_{it}] = RF + \beta_i[E(RM_{it}) - RF_{it}] + \varepsilon_{it} \\ E(\varepsilon_{it}) = 0 \quad Var(\varepsilon_{it}) = \sigma^2_{\varepsilon_{it}} \end{gathered} \tag{2}$$

where ε_{it} is the disturbance term with the usual properties.

We next built on this result and considered aggregation of abnormal returns as shown in (3) (Campbell and Wesley 1993). That is, the cumulative abnormal returns are given as:

$$CAR_t = \sum_{t=-3}^{3} AR_t \tag{3}$$

What is important to mention though, is that, to the best of our knowledge, similar papers examining market reactions using event study analysis, do not examine the model specification for the ordinary least squares (OLS) hypotheses violations regarding time series analysis. In other words, and since we are dealing with time series data, it is crucial to evaluate whether our estimation outputs for autocorrelation and autoregressive conditional heteroskedasticity (ARCH) effect possible problems, and if any assumption is violated, to correct the model before forecasting. There is no need for specification error diagnostics since we are using an established model.

3.3. Pre-Event and Post-Event Comparison

Moving forward, we re-examined our events under a second hypothesis, that is related to the comparison of systematic risk before and after the event. This approach has one similarity with the event study analysis regarding the estimation window, however, the contrast comes to the period after the event. Firstly, we set the pre-event estimation window which in this case had a time span of 70 days. The pre-event estimation window begins just the day before the event (−70, −1).

The next step was to create the post-event estimation window using the same technique and setting the time span to (+1, +70). Day 0 is the day of the disaster occurrence and it has been excluded from both estimation windows. The estimation period in this analysis is limited, compared to the event study analysis described in the previous section, to 70 days before and after the occurrence of an event. Our attempt is to capture the immediate impact of the systematic risk change in a time span close to the event occurrence. Once again, all appropriate diagnostic tests were considered. This procedure provided us with different systematic risks (betas) before and after the occurrence which we assume will provide us useful information regarding investors' perspectives.

3.4. Pooled OLS Regressions

In an attempt to understand investors' possible reaction after an unexpected hazard, we tried to investigate the causes or factors that may influence this possible abnormality. Thus, the final part of the analysis evaluated all results of possible abnormal returns and combined them with macroeconomic factors. The main idea was to observe if there are specific macroeconomic factors that may influence investors to react positively or negatively to the asset price after the event. The idea behind the macroeconomic factors derives from the credit rating methodology, which uses fundamental variables of each economy to rate its creditability and reliability. As already mentioned, credit rating is one of the main elements investors use to diversify their portfolios. Consequently, the question raised is "Does the economic status of a country affect the final decision?"

For that purpose and due to small panel data with even within country differentiations, pooled OLS regressions specifications were used of the form

$$Y_{it} = \alpha_0 + \alpha_1 X_{1it} + \ldots + \alpha_k X_{kit} + \beta_1 X_{1it} D_{1t} + \ldots + \beta_k X_{kit} D_{kt} + u_{it} \tag{4}$$

where Y_{it}, X_{it}, D_i, and u_{it} are the dependent variable, independent variables, dummy, and disturbance term (with the usual properties), respectively. As dependent variable, we set the abnormal returns that occurred after an unexpected event. For the calculation of the abnormal returns we used the beta estimations computed using a 120-day estimation window. These betas were then used for a seven-day forecast, in which the abnormality was then estimated. In other words, each case of examination included abnormal returns of seven days. The whole dataset used for the estimation has 175 observations (25 events $\times$ 7 days abnormal returns). Although the number of observations per event are equal among all events, the period of the occurrence differs, meaning that each event occurred in a separate historical moment, making dynamic cross-sectional panel estimations a non-appropriate approach of estimation.

4. Results and Discussion

The presentation and discussion of results follow the same flow as the methodology from the previous section.

4.1. Event Study Analysis

As described, our initial attempt was to estimate systematic risk during the estimation window which will afterwards be used for computing expected returns and abnormalities. Table A1 (in Appendix A) presents the estimated systematic risk per event of analysis. Specifically, columns (1)[11] and (2) refer to two different periods before the events with estimations (1) including a 120-day estimation window and estimations (2) including a 70-day estimation window. Parentheses oppose the t-statistics of the systematic risks, while the brackets oppose the probability values for the coefficients. The purpose of such an attempt is to observe the sensitivity of systematic risk even on non-risky periods.

It is crucial to mention, based on Table A1, that only three betas of the total (3 $\times$ 25 events) fall below the value of 0.40, while the majority is concentrated between values of 0.60 and 0.99 (44 out of the 75 estimations). Moreover, 11 cases exceeded the value of 1, indicating the greater risk the investors

[11] It is crucial to test our estimations for autocorrelation and ARCH effect. In cases with such problems, appropriate correction methods have been used in solving them. Moreover, in case of ARCH effects, various estimations of the ARCH-family have been used and the most appropriate have been chosen based on AIC. Final estimates corrected for any econometric problems are presented in the tables with initial estimations and diagnostic tests available on request. The results strengthen our belief that all estimations must be tested for all possible violations of the hypotheses of OLS estimations.

adopt. Finally, 17 of the 75 estimations ranged between 0.40 and 0.59, with the majority of them being concentrated in the range of 0.5 to 0.59 (12 out of the 75 estimated cases). As is obvious, most of the cases have a mediocre to high risk, especially if we take into consideration the fact that the assets examined are government bonds.

The results in most cases indicate that there is a possible underestimated systematic risk when less observations are included. This may be crucial advice for portfolio managers who diversify portfolios and provide alternative investing options to investors. Although the difference between estimated betas may be slight, it may lead to large gains/losses if we consider the volume of capital placed on each investment.

Based on the 120-day CAPM analysis, after the corrections needed due to OLS hypotheses violations, all final betas are statistically significant at a 95% level of significance. Having *p*-values as a benchmark, 24 events were statistically significant at 99% level of significance with a probability value lower than 0.01 and only one event (Event 14) was statistically significant at 95% level of significance with a p-value = 0.0174.[12] Although we would like to be able to compare our findings with outcomes from similar researches, to our knowledge, there is limited research done regarding the market reactions after an unexpected disaster. Moreover, in most cases diagnostic tests on estimations were not applied, thus the estimations tend to be biased and do not reflect the actual systematic risk of an investment. Therefore, we assume that a potential comparison would not be applicable in our case.

After estimating the final systematic risk for each event, we computed both the abnormal returns in the event window as well as the cumulative abnormal return (CAR). Using simple hypothesis testing (Table 1, first two columns), we did not reject H_0 of the test as the *p*-value was greater than the usual significance levels (α = 0.1, 0.05 or 0.01); thus, we cannot conclude whether events caused an impact on bond/stock returns or not and we cannot give a clear answer on the H_A of our research, so the hypothesis is still debatable.

Table 1. Simple hypothesis testing results.

Cumulative Abnormal Return	t-Statistics	Change of Systematic Risk	t-Statistics
Capital Asset Pricing Model (CAPM)	−0.694372 (0.4941)	$\Delta(\beta)$	1.312378 (0.2018)

p-values in brackets.

4.2. Pre-Event and Post-Event Comparison

Our next attempt is to examine the H_B, and whether there is a significant difference between the systematic risk estimators before and after an unexpected event. For that purpose, we are going to use the analysis described before in a methodology review based on the event study analysis being able to observe abnormalities that appear on systematic risk following an unexpected event as well as examine whether cumulative reaction has a significant impact on systematic risk. The fact that this abnormality is observed over the event window, which in our case includes a seven-day time span, captures a SR (Short-Run) investor's reaction.

Figure 2 visualizes the under-examination difference of the estimators. As is obvious, in most cases the systematic risk increases after the occurrence of the disaster. Based on the na-tech events, we observed 25 events, for 70 days before the occurrence of the event (ex-ante) as well as 70 days after the occurrence (ex-post). We estimated the systematic risks before and after the event, which once

12 Relying on the coefficient of determination (R^2), 5 out of 25 regressions have a really low goodness of fit (0.00–0.20), showing that less than 20% of the market returns can explain returns of assets. Similarly, 12 out of 25 regressions have low goodness of fit (0.20–0.50), and 5 out of 25 regressions have a mediocre goodness of fit (0.50–0.80), showing that less than 50% and 80%, respectively of market returns can explain assets returns. Finally, 3 out of 25 regressions have a high goodness of fit (over 0.80) with more than 80% of market returns explaining assets returns.

again were diagnosed for all possible OLS hypotheses violations and if any occurred, was solved using the appropriate econometric approaches. The systematic risk results from the ex-ante and ex-post analysis are presented in Table A1 (in Appendix A) as well as in Figure 2 as mentioned. In most cases the systematic risk increases after the occurrence of the disaster. More specifically, from the 25 events of the analysis, the 14 events present a greater systematic risk after the event and 11 events received a lower systematic risk. What is also obvious is the fact that in some cases low values of estimated betas are reported, however, even these low values have sense and may cause a great impact if we consider the volume of capital someone may invest.

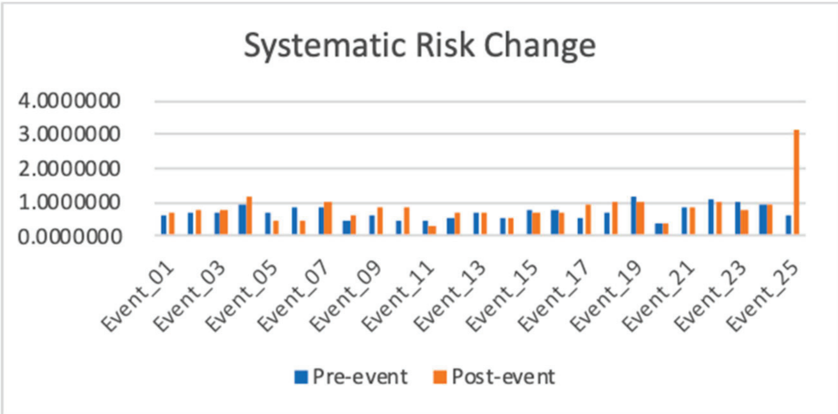


Figure 2. Pre-event and post-event estimations.

Although, we cannot jump to conclusions based on a simple histogram or the percentage of change description. For that reason, we compute the $\Delta(\beta)$[13] which is the change of the systematic risk. Using simple hypothesis testing, and setting the condition of examination as H_0: $\mu = 0$, (Table 1, last two columns), we do not reject H_0 of the test as the *p*-value is greater than usual levels of significance ($\alpha = 0.1$, 0.05 or 0.01); thus we cannot come to conclusions whether change of systematic risk is significant. With that in mind, we cannot give a clear answer if systematic risk of an asset remains unaffected by an unexpected na-tech disaster, so the hypothesis is still debatable.

To be more analytical let us consider each event in turn. Although the Denali earthquake in Alaska in 2002 terrified investors causing an increase of the systematic risk from 0.604627 to 0.692331, the 14.50% change is assumed to be low compared to other higher changes. However, we should always bear in mind that all these changes are multiplied by a great amount of capital investments and may cause huge losses. The Indian Ocean region belongs to the Ring of Fire, giving us the a priori information that there is an 80% perception of an earthquake occurrence; the $M_w = 9.3$ earthquake that took place in Thailand in 2004, which then led to a tsunami causing 227,838 fatalities and 15 billion USD total damage, also led to a 35.42% increase in the systematic risk of the country's government bond from 0.627643 to 0.850764, giving us the belief that investors were scared that Thailand would not be able to cover their requirements.

Though the great earthquake of 2005 in Pakistan caused a remarkable number of fatalities and injuries, the systematic risk of the government's bond decreased by 19.80%. More specifically, the systematic risk before the earthquake occurrence was 0.995803, however after the unexpected event the value of the systematic risk dropped to 0.798632 showing the investors' attempt to support the country and keeping their trust against the Pakistani Government. The Chinese earthquake on 12 May 2008 appears to have the same flow as the previous event. Once again, the number of causalities and economic losses is remarkable, and systematic risk mentions a 12.83% decrease from 0.954242 to 0.831756.

[13] $\Delta(\beta) = \beta_{\text{post-event}} - \beta_{\text{pre-event}}$.

The next na-tech analyzed is the Tohoku earthquake connected to the Fukushima Daiichi Power Plant disaster. The most interesting part of this analysis is the fact that the systematic risk of the Japanese Government mentioned a 25.84% decrease (from 0.677096 to 0.477908) giving the belief that investors showed trust in the government's reputation to possibly overcome this. Moreover, Japan is located on the Ring of Fire, a region with high earthquake occurrence. On the other hand, the systematic risk of the corporation shares dramatically increased from 0.657127 to 3.078681. The 368.50% increase shows investors' tendency to sell the corporation's shares at any cost, in an attempt to avoid further losses. In that way, investors show their disappointment against the firm, or in other words, punish the corporation for its actions. However, it is important to mention that in this case, the disaster did not occur due to the firm's fallacy; however, it is the most devastating nuclear disaster of the new millennium.

On the same path as the Japanese government's bond after the earthquake occurrence, we can find the systematic risk of New Zealand's case after the earthquake on 22 February 2011, which led to a 35.32% decrease, from 0.663115 to 0.428907. New Zealand kept its trustworthiness and persuaded the investors to support the country, thus securing their capitals.

Moving forward, the next four events present mentionable increases in the systematic risks. More specifically, the two earthquakes in Indonesia (Event 7 and Event 9), as well as the earthquake in Argentina. led to systematic risk increases reaching 17.36%, 47.47%, and 50.80% in positive change. Both regions belong to the Ring of Fire, and the fact that earthquakes are a common phenomenon in those countries probably terrifies investors. Instead of being informed and prepared for a possible upcoming earthquake, they may assume that an earthquake, which may follow, will be even worse and probably devastating. The first volcanic eruption of the analysis is the one that occurred in New Zealand in 2011. Some volcanic activities appear to have a great impact on investors' psychology. Probably the fact that a volcanic eruption is not as common as a ground movement, terrifies citizens. In addition, outcomes after volcanic eruptions are more disastrous compared to a high intensity earthquake. An example is New Zealand's case in 2011, where systematic risk sharply increased from 0.500411 to 0.854470 (70.75%).

A remarkable case is the unique phenomenon of a tsunami into a lake, which was also connected to volcanic activity. This Icelandic case, however, recorded a negative change on the systematic risk of Iceland's government bond. The beta decreased from 0.466809 to 0.193990 (58.44%). The past volcanic activity experience in Iceland and the fact that they can take an advantage of such a case in that country, may have influenced the investors to support the country after the event's occurrence. That theory is also supported by the following events (Event 15 and Event 16), which indicates that Iceland tends to record negative change (decrease) on the systematic risk of its government's bonds after an unexpected volcanic eruption. The next volcanic activity in New Zealand, which occurred five years after the previous events, found the investors more prepared and the beta of the bond recorded a 45.54% decrease. On the other hand, some unexpected eruptions, such as Events 13 and 14, may have caused a small-scale reaction with a decrease in the systematic risk (5.05% and 9.31%, respectively).

Italy on the other hand, although it has a huge history regarding volcanic activity, such as Mount Vesuvius and Etna, faced a dramatic systematic risk increase from 0.532475 to 0.949284 (78.27%) after Etna's unexpected eruption in 2013. Similar reactions are also reported in cases of Japan and Chile with a 32.94% and 17.21% increase, respectively, after the volcanic eruption occurrence (Event 18 and Event 20). Indonesia's systematic risks on the other hand, tend to record negative changes after a volcanic eruption such as the 2014 and 2016 cases where the betas decreased by 14.14% and 33%, respectively.

The last category analyzed on na-tech events was the technological disasters, and more specifically the three oil spills and a nuclear disaster. We already analyzed the Daiichi nuclear disaster in this section, mentioning the remarkable systematic risk increase. The other three cases, though, do not appear to have a similar impact on the investors' actions. Initially, the two oil spills that occurred in the Gulf of Mexico by BP did not influence the corporation's shares in a negative way. The systematic risk decreased 4.85% and 22.47%, respectively with the firm's announcements trying to save the corporation's reputation and investors supporting the corporation's trustworthiness. The huge environmental disaster that occurred in the ecosystem did not influence investors' beliefs and actions since the corporation announced they would "clean" the oil spill from the Gulf, ignoring the already existing damage. The Exxon Mobil oil spill case slightly increased the corporation's beta from 0.949111 to 0.966894, and once again, investors tended to ignore the devastating environmental result, due to the fact that the petroleum industry is highly lucrative. As can be seen, some events caused an increase in the systematic risk after the occurrence of the event, however, there are some cases where a decrease in the beta indicates a possible support for the country and/or corporation. This support may be due to the reputation of the country or corporation.

To conclude, although earthquakes are a really common phenomenon, in most cases considered they tended to have a moderate-to-high increase in the systematic risk of the bonds analyzed after the occurrence of each event. Regarding the moderate cases, five events caused a moderate increase in the systematic risk and were observed in countries with known tectonic plate movement activity such as the USA (Event 1), Indian Ocean (Event 2), Pakistan (Event 3), China (Event 4), and New Zealand (Event 12). Moving forward, there were four more cases recording a high-to-significantly high increase in the systematic risk, also observed in countries with high risk of occurrence. In Indonesia, which also lays on the Indian Ocean, Events 7 and 9 caused two significantly high increase of betas, while the other two countries were Chile (Event 8) and New Zealand (Event 10). As it appears, although high-risk areas exist and frequent earthquake activity is recorded, investors tend to have an immediate reaction after those events. A potential new earthquake may increase the risk of a country (or even a region) facing a possible new disaster.

It is also important to mention that there are three cases where the occurrence of an earthquake reduced beta. More specifically, Japan (Event 5), New Zealand (Event 6), and Iceland (Event 12) caused negative changes on the betas. These results raise great interest. Initially, the case of Japan was connected to the Fukushima Daiichi nuclear disaster. The earthquake which caused a tsunami leading to an industrial accident has raised a lot of attention from the media. However, the systematic risk of the Japanese bond revealed a significant decrease. In other words, the investors kept supporting the country which faced a natural disaster and a devastating nuclear hazard at the same time. What is crucial though is the Fukushima Daiichi share price faced an unprecedented shock (Event 25). The systematic risk of the stock dramatically increased giving the belief that the investors "punished" the corporation for causing the largest historical nuclear disaster of the new millennium. The ruined reputation of the corporation as well as its uncertain future probably scared the investors who reacted rapidly.

The last two negative cases on earthquake reactions were observed in New Zealand (Event 6) and Iceland (Event 11). The fact that in some cases New Zealand has negative changes and in other cases has high positive changes may be because of the possible expectance of such an event due to previous smaller earthquakes. Iceland is also observing negative changes to the systematic risk of the government bonds. Most countries analyzed regarding earthquakes are placed on the Ring of Fire area, a well-known area that concentrates the majority of earthquakes annually. Although these unexpected events are more likely to occur in these countries, investors are not prepared for such cases and immediately react.

Moving forward to the volcanic eruptions, it is surprisingly interesting to observe that the majority of the unexpected events caused negative changes in the systematic risk. Volcanic eruptions in most cases do not raise a lot of attention. The two cases that raised a lot of attention regarding the volcanic eruptions occurred in Italy (Event 17) and Japan (Event 18) and caused significantly high change on the ex-post analysis. Initially, the Etna case may have caused such a reaction because two years prior, this specific volcano recorded 38 basaltic fountains, which possibly increased the probability of a greater volcanic explosion. Finally, the Japanese case may have caused a significant reaction due to the fact that it is the only volcanic eruption which encountered fatalities. More specifically, 63 people lost their lives due to that eruption. Such outcomes may have influenced the behavior of the investors.

Last but not least is the oil spill disasters. The research included three oil spill events that occurred since 2000. Although, an increase of the systematic risk was expected due to the environmental disaster caused from those oil spills, investors appeared to decrease the betas in two cases—Prudhoe Bay (Event 22) and Deepwater Horizon (Event 23)—and a slight increase in the case of Mayflower (Event 24). Based on those results, we may assume that the investors, knowing that the oil industry is very profitable, tended to ignore the environmental impact of such disasters.

4.3. Pooled OLS Regressions

Moving to the final part of the analysis, we sought to determine the abnormal returns of the 25 events by notable macroeconomic factors. The importance of this analysis is to examine whether widely known and available macroeconomics can influence investors' actions and either support an investment or not. In other words, a well-stated economy that exudes reliability and credibility can positively affect investors, which will then lead to lower and probably insignificant abnormal returns. In this part of the analysis, the dummy variable included indicates with 1 the days affected by the unexpected event, and zero otherwise.

Table A2 (in Appendix A) presents the results from the pooled panel regression. As we can see, most of the macroeconomic variables have a significant impact on the abnormal returns. By including variables, such as tourism expenditures and tourist arrivals, we assume that similar independent variables may affect the dependent in the same way. However, as observed in Table A2 (in Appendix A), tourism expenditures and tourism arrivals have different impact on abnormal returns. More specifically, the former appears to have a negative sign, which actually means that when the number of arrivals increases, the abnormal returns of the asset examined decrease. This gives us the feeling that investors tend to react less to countries with increased tourism. However, expenditures of tourism appear to have the exact opposite impact on abnormal returns.

What is interesting to observe is how the aftermath of an unexpected event may affect abnormalities. For that purpose, we used the products of estimation with dummies, where 1 is for the days after the event occurrence. As observed, tourism arrivals have a positive sign to abnormal returns. In other words, more tourism arrivals lead to more abnormality probably due to the increased level of uncertainty. Investors may recognize each country as a risky place to visit due to a potential outbreak of a new disaster. On the other hand, increased revenues from tourists (tourism expenditures) decrease the abnormality, possibly due to the fact that more revenues may allow the country to pay the bond coupons as well as keep stability and credibility.

The purpose of GDP growth inclusion is the fact that the percentage of GDP growth through the years may be used as a proxy to the economic growth of a country. Based on the results, we can see that there is a positive relation between economic growth (GDP growth) and abnormal returns, which means that the more GDP growth, the greater the abnormal returns. Based on our assumptions, we were expecting the exact opposite relation due to the fact that a higher GDP growth will make investors believe that the country can cope with the disaster. What was interesting to observe was the results referring to the period after the occurrence of an event (GDP growth*D). The expected negative sign of the explanatory variable that represents the reduction of abnormalities caused by investors was proven in the case of the dummy influence. In other words, while increasing the growth of an economy, the country becomes more trustworthy and the investors are less negatively influenced.

Similar to GDP growth, we included GDP/c as an explanatory variable expecting to observe a negative impact to abnormalities. The GDP/c is a variable that is weighted by population. Higher levels of GDP/c represent better economic conditions for the population of a nation and probably a better economic status as a total. Once again, the interest would have been gathered on the product of dummy variable (GDP/c*D), however, this is a statistically insignificant variable, possibly implying for the cases considered that abnormal returns and therefore investors, are not influenced by such a factor.

In addition, there was an increase both in inflation and imports to abnormal returns after the events' occurrence. The increase in imports may be affected by the need for supplies for the suffered regions, even for basic everyday goods. If the country needs more imports this may indicate a difficulty in covering their basic needs, placing the country in an unstable condition and increasing the risk for investors. A possible increase in inflation indicates a decrease of the value of the country's currency which once again increases the risk for investors. Based on those results, investors tend to be influenced by macroeconomic factors regarding the decision to keep investing on a bond/stock after a na-tech event.

5. Conclusions

Risk aversion is the main determinant that influences investors' choices. Thus, when investment advisors diversify a portfolio, they should always take into consideration the investors' preferences and their tolerance to risk as well as any aspect that may lead to money loss. Uncertainty, on the other hand, is the main characteristic of capital markets, with the idea that "the more you risk, the more you gain". Although analysis techniques for stock performances exist, these models cannot capture the potential risk of "unexpected". Environmental hazards, which are assumed to be random, have a significant impact on society and influence everyone's life. In this research, we decided to examine four hypotheses regarding the influence that certain unexpected events may have on investors' decisions.

Initially, we proved that neither natural, and more specifically geophysical, nor technological, and more specifically industrial, hazards are random. Based on our previous research (Halkos and Zisiadou 2018, 2019) both cases have high- and low-risk areas, so the probability of occurrence may be predicted up to a certain point. If investment advisors know a priori the possibility of an environmental hazard, natural or technological, they may be able to diversify portfolios of high-risk areas by including assets from low-risk areas as well. The first under-examination hypothesis is not rejected, showing that the abnormal returns occurred after an unexpected environmental hazard are not statistically significant. In other words, a na-tech occurrence does not seem to affect investors' psychology given they tend to keep an unchangeable strategic investment plan when an unexpected environmental disaster occurs.

Additionally, as examined by the second hypothesis, we have proven that although there is a change in systematic risk after the occurrence of an event, in comparison with the one before the occurrence, this change is not significant; thus the question whether a potential disaster may affect systematic risk remains debatable. What is also crucial to mention is that the avoidance of diagnostic tests may under/overestimate systematic risk which is differentiated when there are violations of the basic hypotheses in OLS specifications with no adequate corrections in the final estimations.

The final part of the analysis was based on the macroeconomic influence on investors' behavior. With the use of pooled OLS regressions, we examined the third and fourth hypotheses of research proving that there are several macroeconomic factors, like GDP growth, tourism factors, inflation, and imports that may affect abnormal returns after an event occurrence. The third hypothesis has been accepted indicating that macroeconomic factors, influencing the investors' point of view, exist. More specifically, the abnormal returns after the occurrence of an unexpected events tend to decrease if the country of examination records an increase in tourism, both arrivals and expenditures. On the other hand, if the inflation and/or the imports of the suffered country increase then the abnormality recorded will also increase, due to the positive sign of their coefficients. Probably, the increased inflation and/or imports may settle the country in a needy condition, making it even riskier and probably unable to cope with investors' requirements. The fourth hypothesis was examined with the inclusion of GDP/c and GDP growth. The GDP/c in our research gives statistically insignificant coefficients indicating a non-affectionate behavior, however, the GDP growth coefficient is statistically significant and has a negative sign. This sign indicates that a possible increase in GDP growth will lead to a decrease in the abnormal returns when an unexpected environmental disaster occurs. This reaction may likely be influenced by a country's trustworthiness. In other words, increased economic growth influences the reliability of a country which therefore influences investors in a positive way. Generally, it is stated in the literature that investors tend to support a country, which has faced a natural disaster, while at the same time they tend to "punish" a corporation that caused a technological disaster, which may have led to economic losses and adverse impacts on flora, fauna, and the environment in general. What should raise more attention though is the fact that the oil companies that caused huge environmental disasters after the occurrence of the oil spills reached lower values of systematic risk. In other words, investors tended to support those companies and the answer may be hidden in the great profit those companies recorded.

With all this information beforehand, we believe that governments may have the opportunity to create better security and rescuing plans, as well as preparedness systems, while at the same time be able to cover possible damages with emergency payments from their annual budgets. Investment advisors, as we have already mentioned, may help their clients to diversify or hedge in a way that will minimize potential risk, without avoiding investment on specific corporations or countries. Furthermore, our paper provides evidence that some macroeconomic factors may have an effect on investors' psychology regarding their investment decisions after the occurrence of an unexpected environmental disaster. We can infer that the reputation of a country or corporation may be a decisive factor.

Supplementary research could be done by including events from other categories of unexpected events such as transport accidents, meteorological hazards, etc., or by expanding the time span of the event analysis. Although we would like to include cases such as Chernobyl nuclear accident or Three Mile Island, the lack of available information reduced our sample. Moreover, further research could be carried out with inclusion of more macroeconomic factors or with the use of other advanced econometric methods on modeling such issues. Inclusion of other explanatory variables like governmental announcements, such as bankruptcy, or the rise of an extremist political party may also be useful. Likewise, announcements of the downgrade/upgrade of countries or corporations from credit rating agencies may lead to useful knowledge on how investors may weight their risk based on available information.

Author Contributions: Conceptualization, G.H.; data curation, A.Z.; methodology, G.H. and A.Z.; supervision, G.H.; writing—original draft, A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The research received no external funding.

Acknowledgments: This work was supported by the General Secretariat for Research and Technology and the Hellenic Foundation for Research and Innovation (HFRI). Thanks are due to the Academic Editor and to three anonymous reviewers for their helpful and constructive comments. Any remaining errors are solely the authors' responsibility.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Beta estimations.

	120 Days Pre-Event β	**70 Days Pre-Event β**	**70 Days Post-Event β**
	(1)	**(2)**	**(3)**
Event_01	0.566306 (4.77) (0.0000)	0.604627 (3.406) (0.0011)	0.692331 (5.311) (0.0000)
Event_02	0.716023 (12.823) (0.0000)	0.678660 (7.131) (0.0000)	0.804563 (11.004) (0.0000)
Event_03	1.043605 (5.381) (0.0000)	0.704594 (4.282) (0.0000)	0.798632 (15.329) (0.0000)
Event_04	0.852833 (170.56) (0.0000)	0.954242 (8.879) (0.0000)	1.141257 (10.723) (0.0000)
Event_05	0.878074 (7.850) (0.0000)	0.677096 (3.793) (0.0003)	0.477908 (3.557) (0.0007)
Event_06	0.670526 (9.984) (0.0000)	0.860940 (11.912) (0.0000)	0.428907 (4.494) (0.0000)

Table A1. *Cont.*

	120 Days Pre-Event β	70 Days Pre-Event β	70 Days Post-Event β
	(1)	(2)	(3)
Event_07	0.841452 (7.272) (0.0000)	0.848343 (5.158) (0.0000)	0.995659 (7.570) (0.0000)
Event_08	0.514983 (7.269) (0.0000)	0.429804 (4.228) (0.0001)	0.648179 (6.242) (0.0000)
Event_09	0.629387 (4.866) (0.0000)	0.585545 (3.295) (0.0016)	0.863526 (3.040) (0.0034)
Event_10	0.579811 (7.378) (0.0000)	0.500411 (5.212) (0.0000)	0.854470 (7.653) (0.0000)
Event_11	0.310581 (6.565) (0.0000)	0.466809 (6.443) (0.0000)	0.322860 (3.867) (0.0001)
Event_12	0.601126 (30.66) (0.0000)	0.552577 (5.528) (0.0000)	0.658889 (2.639) (0.0103)
Event_13	0.529411 (4.383) (0.0000)	0.732459 (6.698) (0.0000)	0.695440 (5.570) (0.0000)
Event_14	0.670650 (2.411) (0.0174)	0.572725 (2.663) (0.0096)	0.519378 (1.876) (0.0649)
Event_15	0.922181 (24.736) (0.0000)	0.805822 (10.030) (0.0000)	0.721883 (7.294) (0.0000)
Event_16	0.818089 (6.929) (0.0000)	0.759383 (6.654) (0.0000)	0.731661 (6.516) (0.0000)
Event_17	0.595667 (5.679) (0.0000)	0.532475 (3.175) (0.0022)	0.949284 (10.404) (0.0000)
Event_18	0.855654 (9.661) (0.0000)	0.683732 (5.517) (0.0000)	1.017484 (13.920) (0.0000)
Event_19	1.311919 (11.012) (0.0000)	1.206142 (6.078) (0.0000)	1.035517 (8.432) (0.0000)
Event_20	0.517310 (7.865) (0.0000)	0.354545 (4.329) (0.0001)	0.415583 (6.222) (0.0000)
Event_21	0.856874 (11.889) (0.0000)	0.880564 (9.712) (0.0000)	0.840760 (8.921) (0.0000)
Event_22	1.077069 (15.344) (0.0000)	1.060922 (13.241) (0.0000)	1.009471 (20.085) (0.0000)
Event_23	1.010974 (15.654) (0.0000)	0.975651 (9.299) (0.0000)	0.756432 (2.879) (0.0053)
Event_24	0.975986 (23.390) (0.0000)	0.949111 (16.630) (0.0000)	0.966894 (28.721) (0.0000)
Event_25	0.802265 (8.134) (0.0000)	0.657127 (5.586) (0.0000)	3.111465 (3.207) (0.0013)

t-stat. and *p*-values parentheses.

Table A2. Pooled regression results.

Variable	Pooled AR	Pooled AR
Constant	−0.024887 (−6.604322) (0.0000)	−0.027235 (−9.871251) (0.0000)
GDP growth	0.005188 (10.20047) (0.0000)	0.005497 (13.52677) (0.0000)
GDP/c	8.44×10^{-7} (4.507790) (0.0000)	8.36×10^{-7} (6.603492) (0.0000)
Population Density	-5.64×10^{-5} (−3.643325) (0.0003)	-6.41×10^{-5} (−5.781233) (0.0000)
FDI	3.66×10^{-13} (5.470146) (0.0000)	4.12×10^{-13} (9.534101) (0.0000)

Table A2. *Cont.*

Variable	Pooled AR	Pooled AR
Household Consumption	-6.21×10^{-14} (−7.121643) (0.0000)	-6.46×10^{-14} (−10.71255) (0.0000)
Imports	1.02×10^{-13} (3.938239) (0.0001)	1.10×10^{-13} (5.931378) (0.0000)
Inflation	−0.000355 (−0.842450) (0.3996)	
Tourism Expenditures	4.76×10^{-12} (8.670022) (0.0000)	4.87×10^{-12} (11.61634) (0.0000)
Tourism Arrivals	-4.52×10^{-10} (−2.937924) (0.0033)	-4.10×10^{-10} (−3.352551) (0.0008)
Exports	-9.29×10^{-14} (−6.229301) (0.0000)	-9.85×10^{-14} (−8.532868) (0.0000)
Gov. Health Expenditures/c	-4.89×10^{-6} (−2.065084) (0.0390)	-4.61×10^{-6} (−3.023661) (0.0025)
GDP growth*D	−0.001366 (−2.419998) (0.0156)	−0.001745 (−5.958745) (0.0000)
GDP/c*D	-6.04×10^{-8} (−0.270809) (0.7866)	
Population Density*D	−0.000116 (−5.693055) (0.0000)	-9.82×10^{-5} (−9.961197) (0.0000)
FDI*D	7.64×10^{-14} (0.900765) (0.3678)	
Household Consumption*D	2.14×10^{-14} (1.925287) (0.0543)	2.67×10^{-14} (6.228393) (0.0000)
Imports*D	1.60×10^{-13} (4.740005) (0.0000)	1.44×10^{-13} (9.023375) (0.0000)
Inflation*D	0.004192 (9.399820) (0.0000)	0.004057 (13.28318) (0.0000)
Tourism Expenditures*D	-4.92×10^{-12} (−7.071382) (0.0000)	-5.17×10^{-12} (−12.58543) (0.0000)
Tourism Arrivals*D	1.63×10^{-9} (8.021313) (0.0000)	1.61×10^{-9} (10.86971) (0.0000)
Exports*D	-7.00×10^{-14} (−3.673465) (0.0002)	-6.02×10^{-14} (−5.329575) (0.0000)
Gov. Health Expenditures/c*D	-5.09×10^{-6} (−1.651309) (0.0988)	-5.48×10^{-6} (−9.441656) (0.0000)
R^2	0.28813	0.288651
F-statistics	75.03277 (0.000000)	98.15229 (0.000000)
AIC	−4.088925	−4.090998

t-stat in parentheses and *p*-values in brackets.

References

Adrian, Tobias, and Franzoni Francesco. 2009. Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance* 16: 537–56. [CrossRef]

Aıt-Sahalia, Yacine, and Andrew W. Lo. 2000. Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94: 9–51. [CrossRef]

Arin, K. Peter, Davide Ciferri, and Nicola Spagnolo. 2008. The price of terror: The effects of terrorism on stock market returns and volatility. *Economics Letters* 101: 164–67. [CrossRef]

Armitage, Andrew. 1995. *Comparing the Policy of Aboriginal Assimilation: Australia, Canada, and New Zealand*. Vancouver: UBC Press.

Athukorala, Wasantha, Wade Martin, Prasad Neelawala, Darshana Rajapaksa, Jeremy Webb, and Clevo Wilson. 2018. Impact of natural disasters on residential property values: Evidence from Australia. In *World Scientific Book Chapters*. Singapore: World Scientific Publishing Co. Pte. Ltd., pp. 147–79. [CrossRef]

Avouac, Jean-Philippe, Francois Ayoub, Sebastien Leprince, Ozgun Konca, and Don V. Helmberger. 2006. The 2005, Mw 7.6 Kashmir earthquake: Sub-pixel correlation of ASTER images and seismic waveforms analysis. *Earth and Planetary Science Letters* 249: 514–28. [CrossRef]

Bannister, Stephen, and Ken Gledhill. 2012. Evolution of the 2010–2012 Canterbury earthquake sequence. *New Zealand Journal of Geology and Geophysics* 55: 295–304. [CrossRef]

Barro, Robert J., and Sanjay Misra. 2016. Gold returns. *The Economic Journal* 126: 1293–317. [CrossRef]

Benartzi, Shlomo, and Richard H. Thaler. 1995. Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics* 110: 73–92. [CrossRef]

Benartzi, Shlomo, and Richard H. Thaler. 1999. Risk aversion or myopia? Choices in repeated gambles and retirement investments. *Management Science* 45: 364–81. [CrossRef]

Binder, John. 1998. The event study methodology since 1969. *Review of Quantitative Finance and Accounting* 11: 111–37. [CrossRef]

Bliss, Robert R., and Nikolaos Panigirtzoglou. 2004. Option-implied risk aversion estimates. *The Journal of Finance* 59: 407–46. [CrossRef]

Bollerslev, Tim, Michael Gibson, and Hao Zhou. 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* 160: 235–45. [CrossRef]

Bolt, Bruce A. 1988. *Earthquakes*. New York: W.H Freeman and Company.

Bond, Gary E., and Stanley R. Thompson. 1985. Risk aversion and the recommended hedging ratio. *American Journal of Agricultural Economics* 67: 870–72. [CrossRef]

Bradley, Brendon A., and Misko Cubrinovski. 2011. Near-source strong ground motions observed in the 22 February 2011 Christchurch earthquake. *Seismological Research Letters* 82: 853–65. [CrossRef]

Brandt, Michael D., and Kevin Q. Wang. 2003. Time-varying risk aversion and unexpected inflation. *Journal of Monetary Economics* 50: 1457–98. [CrossRef]

Brounen, Dirk, and Jeroen Derwall. 2010. The impact of terrorist attacks on international stock markets. *European Financial Management* 16: 585–98. [CrossRef]

Bruner, Robert F., Wei Li, Mark Kritzman, Simon Myrgren, and Sébastien Page. 2008. Market integration in developed and emerging markets: Evidence from the CAPM. *Emerging Markets Review* 9: 89–103. [CrossRef]

Bugár, Gyöngyi, and Raimond Maurer. 2002. International equity portfolios and currency hedging: The viewpoint of German and Hungarian investors. *ASTIN Bulletin: The Journal of the IAA* 32: 171–97. [CrossRef]

Campbell, John Y., and John H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107: 205–51. [CrossRef]

Campbell, Cynthia J., and Charles E. Wesley. 1993. Measuring security price performance using daily NASDAQ returns. *Journal of Financial Economics* 33: 73–92. [CrossRef]

Carpentier, Cécile, and Jean-Mark Suret. 2015. Stock market and deterrence effect: A mid-run analysis of major environmental and non-environmental accidents. *Journal of Environmental Economics and Management* 71: 1–18. [CrossRef]

Carter, David A., and Betty J. Simkins. 2004. The market's reaction to unexpected, catastrophic events: The case of airline stock returns and the September 11th attacks. *The Quarterly Review of Economics and Finance* 44: 539–58. [CrossRef]

Caudron, Corentin, Benoit Taisne, Milton Garcés, Le Pichon Alexis, and Pierrick Mialle. 2015. On the use of remote infrasound and seismic stations to constrain the eruptive sequence and intensity for the 2014 Kelud eruption. *Geophysical Research Letters* 42: 6614–21. [CrossRef]

Charles, Amélie, and Olivier Darné. 2006. Large shocks and the September 11th terrorist attacks on international stock markets. *Economic Modelling* 23: 683–98. [CrossRef]

Chen, Ming-Hsiang. 2003. Risk and return: CAPM and CCAPM. *The Quarterly Review of Economics and Finance* 43: 369–93. [CrossRef]

Cohn, Richard A., Wilbur G. Lewellen, Ronald C. Lease, and Gary G. Schlarbaum. 1975. Individual investor risk aversion and investment portfolio composition. *The Journal of Finance* 30: 605–20. [CrossRef]

Droitsch, Danielle. 2014. Tar Sands Crude Oil: Health Effects of a Dirty and Destructive Fuel. National Resources Defense Council Issue Brief (Feb 2014). Available online: https://www.nrdc.org/sites/default/files/tar-sands-health-effects-IB.pdf (accessed on 15 April 2018).

Dunham, Eric M., and Ralph J. Archuleta. 2004. Evidence for a supershear transient during the 2002 Denali fault earthquake. *Bulletin of the Seismological Society of America* 94: S256–S268. [CrossRef]

Eberhart-Phillips, Donna, Peter J. Haeussler, Jeffrey T. Freymueller, Arthur D. Frankel, Charles M. Rubin, Patricia Craw, Natalia A. Ratchkovski, Greg Anderson, Gary A. Carver, Anthony J. Crone, and et al. 2003. The 2002 Denali fault earthquake, Alaska: A large magnitude, slip-partitioned event. *Science* 300: 1113–18. [CrossRef] [PubMed]

EMDAT. 2017. The International Disaster Database, Centre for Research on the Epidemiology of Disaster—CRED. Available online: http://www.emdat.be (accessed on 5 May 2017).

Faff, Robert W. 1991. A likelihood ratio test of the zero-beta CAPM in Australian equity returns. *Accounting & Finance* 31: 88–95. [CrossRef]

Fernald, John, and John H. Rogers. 2002. Puzzles in the Chinese stock market. *Review of Economics and Statistics* 84: 416–32. [CrossRef]

Fernandez, Viviana. 2006. The CAPM and value at risk at different time-scales. *International Review of Financial Analysis* 15: 203–19. [CrossRef]

Freed, Andrew M., Roland Bürgmann, Eric Calais, Jeff Freymueller, and Sigrún Hreinsdóttir. 2006. Implications of deformation following the 2002 Denali, Alaska, earthquake for postseismic relaxation processes and lithospheric rheology. *Journal of Geophysical Research: Solid Earth* 111. [CrossRef]

Gaspar, José-Miguel, Massimo Massa, and Pedro Matos. 2005. Shareholder investment horizons and the market for corporate control. *Journal of Financial Economics* 76: 135–65. [CrossRef]

Gordon, Stephen, and St-Amour Pascal. 2004. Asset returns and state-dependent risk preferences. *Journal of Business & Economic Statistics* 22: 241–52. [CrossRef]

Graham, David, and Robert Jennings. 1987. Systematic risk, dividend yield and the hedging performance of stock index futures. *The Journal of Futures Markets* 7: 1.

Grubel, Herbert G. 1968. Internationally diversified portfolios: Welfare gains and capital flows. *The American Economic Review* 58: 1299–314.

Gudmundsson, Magnus T., Thorvaldur Thordarson, Ármann Höskuldsson, Gudrún Larsen, Halldór Björnsson, Fred J. Prata, Björn Oddsson, Eyjólfur Mangússon, Thórdis Högnadóttir, Gudrun Nina Petersen, and et al. 2012. Ash generation and distribution from the April-May 2010 eruption of Eyjafjallajökull, Iceland. *Scientific Reports* 2: 1–12. [CrossRef]

Gylfadóttir, Sigríður Sif, Jihwan Kim, Jón Kristinn Helgason, Sveinn Brynjólfsson, Ármann Höskuldsson, Tómas Jóhannesson, Carl Bonnevie Harbitz, and Finn Løvholt. 2017. The 2014 Lake Askja rockslide-induced tsunami: Optimization of numerical tsunami model using observed data. *Journal of Geophysical Research: Oceans* 122: 4110–22. [CrossRef]

Haigh, Michael S., and John A. List. 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. *The Journal of Finance* 60: 523–34. [CrossRef]

Halkos, George, and Argyro Zisiadou. 2018. Examining the Natural Environmental Hazards over the Last Century. *Economics of Disasters and Climate Change* 3: 119–50. [CrossRef]

Halkos, George, and Argyro Zisiadou. 2019. An Overview of the Technological Environmental Hazards over the last century. *Economics of Disaster and Climate Change* 4: 411–28. [CrossRef]

Halkos, George, Shunsuke Managi, and Argyro Zisiadou. 2017. Analyzing the determinants of terrorist attacks and their market reactions. *Economic Analysis and Policy* 54: 57–73. [CrossRef]

Heidarzadeh, Mohammad, Satoko Murotani, Kenji Satake, Takeo Ishibe, and Aditya Riadi Gusman. 2016. Source model of the 16 September 2015 Illapel, Chile, Mw 8.4 earthquake based on teleseismic and tsunami data. *Geophysical Research Letters* 43: 643–50. [CrossRef]

Hollingsworth, James, Lingling Ye, and Jean-Philippe Avouac. 2017. Dynamically triggered slip on a splay fault in the Mw 7.8 2016 Kaikoura (New Zealand) earthquake. *Geophysical Research Letters* 44: 3517–25. [CrossRef]

International Atomic Energy Agency. 1998. Available online: https://www.iaea.org/topics/emergency-preparedness-and-response-epr/international-nuclear-radiological-event-scale-ines (accessed on 10 October 2018).

Ivy, Diane J., Susan Solomon, Doug Kinnison, Michael J. Mills, Anja Schmidt, and Ryan R. Neely. 2017. The influence of the Calbuco eruption on the 2015 Antarctic ozone hole in a fully coupled chemistry-climate model. *Geophysical Research Letters* 44: 2556–61. [CrossRef]

Jackwerth, Jens Carsten. 2000. Recovering risk aversion from option prices and realized returns. *The Review of Financial Studies* 13: 433–51. [CrossRef]

Jagannathan, Ravi, and Zhenyu Wang. 1993. *The CAPM is Alive and Well (No. 165)*. Minneapolis: Federal Reserve Bank of Minneapolis.

Jibson, Randall W., Edwin L. Harp, William Schulz, and David K. Keefer. 2006. Large rock avalanches triggered by the M 7.9 Denali Fault, Alaska, earthquake of 3 November 2002. *Engineering Geology* 83: 144–60. [CrossRef]

Jousset, Philippe, John Pallister, Marie Boichu, Fabrizia M. Buongiorno, Agus Budisantoso, Fidel Costa, Supriyati Andreastuti, Fred Prata, David Schneider, Lieven Clarisse, and et al. 2012. The 2010 explosive eruption of Java's Merapi volcano—A '100-year'event. *Journal of Volcanology and Geothermal Research* 241: 121–35. [CrossRef]

Kamp, Ulrich, Benjamin J. Growley, Ghazanfar A. Khattak, and Lewis A. Owen. 2008. GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology* 101: 631–42. [CrossRef]

Kanamori, Hiroo. 1972. Mechanism of tsunami earthquakes. *Physics of the Earth and Planetary Interiors* 6: 346–59. [CrossRef]

Kaneko, Takayuki, Fukashi Maeno, and Setsuya Nakada. 2016. 2014 Mount Ontake eruption: Characteristics of the phreatic eruption as inferred from aerial observations. *Earth, Planets and Space* 68: 72. [CrossRef]

Karolyi, Andrew G., and Rodolfo Martell. 2010. Terrorism and the Stock Market. *International Review of Applied Financial Issues & Economics* 2: 285–314.

Kato, Aitaro, Toshiko Terakawa, Yoshiko Yamanaka, Yuta Maeda, Shinichiro Horikawa, Kenjiro Matsuhiro, and Takashi Okuda. 2015. Preparatory and precursory processes leading up to the 2014 phreatic eruption of Mount Ontake, Japan. *Earth, Planets and Space* 67: 111. [CrossRef]

Kujawinski, Elizabeth B., Melissa C. Kido Soule, David L. Valentine, Angela K. Boysen, Krista Longnecker, and Molly C. Redmond. 2011. Fate of dispersants associated with the Deepwater Horizon oil spill. *Environmental Science & Technology* 45: 1298–306. [CrossRef]

Kunreuther, Howard. 1996. Mitigating disaster losses through insurance. *Journal of Risk and Uncertainty* 12: 171–87. [CrossRef]

Kurtz, Rick S. 2010. Oil pipeline regulation, culture, and integrity: The 2006 BP North Slope spill. *Public Integrity* 13: 25–40. [CrossRef]

Lee, Kuo-Jung, Su-Lien Lu, and You Shih. 2018. Contagion effect of natural disaster and financial crisis events on international stock markets. *Journal of Risk and Financial Management* 11: 16. [CrossRef]

Liu-Zeng, Jing, Zhihui Zhang, Li Wen, Paul E. Tapponnier, Jing Sun, Xiucheng Xing, Guyue Hu, Qiang Xu, Lingsen Zeng, Lin Ding, and et al. 2009. Co-seismic ruptures of the 12 May 2008, Ms 8.0 Wenchuan earthquake, Sichuan: East–west crustal shortening on oblique, parallel thrusts along the eastern edge of Tibet. *Earth and Planetary Science Letters* 286: 355–70. [CrossRef]

MacKinlay, Craig A. 1997. Event studies in economics and finance. *Journal of Economic Literature* 35: 13–39.

Maloney, Michael T., and Harold J. Mulherin. 2003. The complexity of price discovery in an efficient market: The stock market reaction to the Challenger crash. *Journal of Corporate Finance* 9: 453–79. [CrossRef]

Merton, Robert C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics* 51: 247–57. [CrossRef]

Newman, Andrew W., Gavin Hayes, Yong Wei, and Jaime Convers. 2011. The 25 October 2010 Mentawai tsunami earthquake, from real-time discriminants, finite-fault rupture, and tsunami excitation. *Geophysical Research Letters* 38. [CrossRef]

Norio, Okada, Tao Ye, Yoshio Kajitani, Peijun Shi, and Hirokazu Tatano. 2011. The 2011 eastern Japan great earthquake disaster: Overview and comments. *International Journal of Disaster Risk Science* 2: 34–42. [CrossRef]

Pollitz, Fred F., Ross S. Stein, Volkan Sevilgen, and Roland Bürgmann. 2012. The 11 April 2012 east Indian Ocean earthquake triggered large aftershocks worldwide. *Nature* 490: 250. [CrossRef] [PubMed]

Prabhala, Nagpurnanand R. 1997. Conditional methods in event studies and an equilibrium justification for standard event-study procedures. *The Review of Financial Studies* 10: 1–38. [CrossRef]

Priest, George L. 1996. The government, the market, and the problem of catastrophic loss. *Journal of Risk and Uncertainty* 12: 219–37. [CrossRef]

Rosenberg, Joshua V., and Robert F. Engle. 2002. Empirical pricing kernels. *Journal of Financial Economics* 64: 341–72. [CrossRef]

Ruiz, Sergio, Emilie Klein, Francisco del Campo, Efrain Rivera, Piero Poli, Marianne Metois, Vigny Christophe, Juan Carlos Baez, Gabriel Vargas, Felipe Leyton, and et al. 2016. The seismic sequence of the 16 September 2015 M w 8.3 Illapel, Chile, earthquake. *Seismological Research Letters* 87: 789–99. [CrossRef]

Scollo, Simona, Michele Prestifilippo, Emilio Pecora, Stefano Corradini, Luca Merucci, Gaetano Spata, and Mauro Coltelli. 2014. Eruption column height estimation of the 2011–2013 Etna lava fountains. *Annals of Geophysics* 57: 0214. [CrossRef]

Sigmarsson, Olgeir, Baptiste Haddadi, Simon Carn, Séverine Moune, Jónas Gudnason, Kai Yang, and Lieven Clarisse. 2013. The sulfur budget of the 2011 Grímsvötn eruption, Iceland. *Geophysical Research Letters* 40: 6095–100. [CrossRef]

Sinulingga, Emerson, and Baihaqi Siregar. 2017. Remote Monitoring of Post-eruption Volcano Environment Based-On Wireless Sensor Network (WSN): The Mount Sinabung Case. *Journal of Physics: Conference Series* 801: 012084.

Strong, Norman. 1992. Modelling abnormal returns: A review article. *Journal of Business Finance & Accounting* 19: 533–53. [CrossRef]

Telford, John, and John Cosgrave. 2007. The international humanitarian system and the 2004 Indian Ocean earthquake and tsunamis. *Disasters* 31: 1–28. [CrossRef] [PubMed]

Urai, Minoru, and Yoshihiro Ishizuka. 2011. Advantages and challenges of space-borne remote sensing for Volcanic Explosivity Index (VEI): The 2009 eruption of Sarychev Peak on Matua Island, Kuril Islands, Russia. *Journal of Volcanology and Geothermal Research* 208: 163–68. [CrossRef]

Van Eaton, Alexa R., Daniel Bertin Álvaro Amigo, Larry G. Mastin, Raúl E. Giacosa, Jerónimo González, Oscar Valderrama, Karen Fontijn, and Sonja A. Behnke. 2016. Volcanic lightning and plume behavior reveal evolving hazards during the April 2015 eruption of Calbuco volcano, Chile. *Geophysical Research Letters* 43: 3563–71. [CrossRef]

Viccaro, Marco, Rosario Calcagno, Ileana Garozzo, Marisa Giuffrida, and Eugenio Nicotra. 2015. Continuous magma recharge at Mt. Etna during the 2011–2013 period controls the style of volcanic activity and compositions of erupted lavas. *Mineralogy and Petrology* 109: 67–83. [CrossRef]

Viscusi, Kip W. 2006. Natural disaster risks: An introduction. *Journal of Risk and Uncertainty* 33: 5–11. [CrossRef]

Viscusi, Kip W. 2009. Valuing risks of death from terrorism and natural disasters. *Journal of Risk and Uncertainty* 38: 191–213. [CrossRef]

Walker, Thomas J., Kuntara Pukthuanthong, and Sergey S. Barabanov. 2006. On the stock market's reaction to major railroad accidents. *Journal of the Transportation Research Forum* 45. [CrossRef]

Wang, Xiaoming, and Philip L.-F. Liu. 2006. An analysis of 2004 Sumatra earthquake fault plane mechanisms and Indian Ocean tsunami. *Journal of Hydraulic Research* 44: 147–54. [CrossRef]

Waythomas, Christopher F., William E. Scott, Stephanie G. Prejean, David J. Schneider, Pavel Izbekov, and Christopher J. Nye. 2010. The 7–8 August 2008 eruption of Kasatochi Volcano, central Aleutian Islands, Alaska. *Journal of Geophysical Research: Solid Earth* 115. [CrossRef]

Womack, Kent L., and Ying Zhang. 2003. Understanding Risk and Return, the CAPM, and the Fama-French Three-Factor Model. Available online: https://ssrn.com/abstract=481881 (accessed on 22 March 2017).

Journal of
Risk and Financial Management

MDPI

Article

Risk-Sharing and the Creation of Systemic Risk

Viral V. Acharya [1], Aaditya M. Iyer [2,*] and Rangarajan K. Sundaram [1]

1 Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, USA; vacharya@stern.nyu.edu (V.V.A.); rsundara@stern.nyu.edu (R.K.S.)
2 Man Group, Man Investments Inc., 452 Fifth Avenue, 27th floor, New York, NY 10018, USA
* Correspondence: aiyer@stern.nyu.edu

Received: 24 June 2020; Accepted: 7 August 2020; Published: 17 August 2020

Abstract: We address the paradox that financial innovations aimed at risk-sharing appear to have made the world riskier. Financial innovations facilitate hedging idiosyncratic risks among agents; however, aggregate risks can be hedged only with liquid assets. When risk-sharing is primitive, agents self-hedge and hold more liquid assets; this buffers aggregate risks, resulting in few correlated failures compared to when there is greater risk sharing. We apply this insight to build a model of a clearinghouse to show that as risk-sharing improves, aggregate liquidity falls but correlated failures rise. Public liquidity injections, for example, in the form of a lender-of-last-resort can reduce this systemic risk ex post, but induce lower ex-ante levels of private liquidity, which can in turn aggravate welfare costs from such injections.

Keywords: banking; clearinghouses; systemic risk

JEL Classification: G21; G22; G31

---

1. Introduction

Over much of the last several decades, instruments and contracts aimed at facilitating risk-sharing between financial agents have seen explosive growth. By and large (although not uniformly) over this period, regulators viewed this as a benign and even beneficial development, a point of view perhaps best exemplified by Federal Reserve Board Chairman Alan Greenspan's remarks at the American Bankers Association Annual Convention in October 2004.[1] Referring to the unbundling and transfers of risks facilitated by derivatives, Greenspan suggested that, as a consequence

> ... not only have individual financial institutions become less vulnerable to shocks from underlying risk factors, but also the financial system as a whole has become more resilient.

The advent of the financial crisis in 2007–2008 led to a sharp reappraisal of these views. Today, it is widely acknowledged that the very markets and contracts designed with the ostensible purpose of facilitating risk-sharing among financial intermediaries—including derivatives such as credit default swaps and securitization instruments such as collateralized debt obligations—played a central role in fanning the crisis, nearly bringing about a collapse of the entire financial system.

How could the facilitation of risk-sharing increase systemic risk and systemic fragility? In this paper, we suggest an explanation of this apparently paradoxical situation. The intuition behind our formal analysis is rather simply expressed. We begin with the observation that while risk-sharing arrangements are effective at hedging against idiosyncratic shocks (such shocks are by

---

1 http://www.federalreserve.gov/BOARDDOCS/Speeches/2004/20041005/default.htm (accessed on 17 July 2020).

definition uncorrelated across agents), they have, at best, limited potential to provide a hedge against aggregate shocks. When an aggregate shock hits the economy, everyone is affected by the common shock and any existing risk-sharing arrangements face increased likelihood of counterparty default. The only effective ex-ante defense against systemic shocks is to, so to speak, save for the rainy day, viz., to increase holdings of safe, liquid assets (“cash”) and reduce investment in risky and illiquid assets.

Now consider an agent who faces shocks that could be idiosyncratic or aggregate. Aggregate shocks are those that hit all agents and firms; they may be far less likely (they may be “tail risks”) than the idiosyncratic risk exposures, and as such, the agent may find it privately costly to hold liquidity purely for withstanding such aggregate shocks. However, when risk-sharing opportunities are absent or scarce, the more-likely idiosyncratic shocks also need to be hedged, providing the agent with an incentive to hold liquid assets to withstand the shocks and making it privately optimal to save to hedge the risk exposures. As a consequence, in the (perhaps low-likelihood) event that an aggregate risk materializes, the system has adequate liquidity to deal with it, so systemic crises are averted or unlikely.

In contrast, when financial innovations allow agents and firms to share idiosyncratic risks with each other, the hedging motive for holding liquidity to withstand these shocks is reduced. Risk-sharing allows the financial system to economize on its pool of liquidity. Now, the principal benefit of holding liquidity arises in case of aggregate shocks, but if these are low-likelihood events, it will not generally be in the interest of individual agents or firms to hedge against these states. As a consequence, while idiosyncratic risks are shared and risk-sharing promises are honored when idiosyncratic shocks arise, there is inadequate liquidity and insufficient savings for the rainy day when an aggregate shock hits, making it difficult or impossible to honor risk-sharing promises. The result is financial fragility or systemic risk as an equilibrium outcome.

In short, financial innovations facilitating risk-sharing may increase systemic vulnerability by inducing agents to invest resources more fully in high-return risky projects by reducing liquidity for dealing with low-risk aggregate shocks. To be sure, the increased systemic fragility could be very costly for the economy when an aggregate shock arises since a meltdown of the financial sector could produce an elongated period of financial disruption affecting payment and settlement systems, transactional services, and savings schemes.

We deliver this insight in a series of models. Section 3 introduces the foundation on which we build, the problem of an individual firm seeking to hedge against total risk exposure in the absence of risk-sharing possibilities. (We call this no-risk-sharing setting one of “autarky.”) We show that if the risk exposure is sufficiently large, the autarkic firm finds it privately optimal to fully hedge against the exposure by holding adequate liquidity. In Sections 4 and 5, we embed this one-firm setting in different multi-firm contexts (with each firm’s risk exposure consequently broken into idiosyncratic and aggregate parts), and study the consequences of permitting risk-sharing among the firms.

Section 4 looks at the simplest and most transparent of these extensions, that of a two-firm setting in which the firms can write a risk-sharing contract to make transfers to each other in idiosyncratic risk states. When risk-sharing is disallowed, each firm is, of course, in the autarky setting of Section 3, and for the reasons explained in the previous paragraph, there is adequate liquidity in the system to handle the aggregate risk state. However, when risk-sharing is allowed, we show that the firms optimally choose to economize on liquidity and increase their investment in risky assets; and, as a consequence, when the (possibly low-likelihood) aggregate risk state does materialize, the system-wide liquidity proves inadequate to meet the shock, and both firms fail.

Section 5 then generalizes the setting to a setting with a large number of firms and a richer description of uncertainty, and looks at risk-sharing via a co-insurance arrangement akin to a clearinghouse. The clearinghouse collects premiums up-front to build its reserves but it can further also make ex-post “capital calls” on liquidity-surplus banks to aid (co-insure) the liquidity-short firms. Clearinghouse insolvency and systemic failure occur if the reserves and capital calls are collectively insufficient to meet the needs of liquidity-short firms. In this case, a fire sale takes place and liquidity-surplus banks acquire the assets of liquidity-short banks. The fire sale is costly since banks do not realize full returns on acquired assets.

We show that in this case too our general insight prevails: risk-sharing arrangements can incentivize firms to take on greater risk at the individual level, reduce overall liquidity, and increase systemic fragility. The level of correlation in firm risks plays a key role. When the correlation between underlying risks of firms is low, risk-sharing arrangements are enabled that do not require high upfront premiums, and the clearinghouse setting provides risk-sharing of good quality in most states of the world, offering a clear benefit over a world without risk-sharing. However, when a large aggregate shock materializes, this arrangement fails, creating greater systemic fragility than in a world without risk-sharing. Conversely, when underlying risks of firms are highly correlated, upfront premiums required by insurance and clearinghouse rise correspondingly, reducing the attractiveness of risk-sharing. At sufficiently high correlations, premiums become so high that autarky is preferable and aggregate outcomes mirror the no risk-sharing setting.

While the clearinghouse leads to systemic risk, it also economizes ex-ante on liquidity held by banks. In other words, while there is increased risk of aggregate failure in the presence of a clearinghouse, the expected losses arising from this risk is countered by the expected gains to banks in making risky investments. This result breaks down, however, in the presence of a lender-of-last-resort (LOLR) that can choose to inject public liquidity into the clearinghouse at a cost. We model this extension in Section 6. While LOLR interventions lower the incidence of fire sales ex post, there is a countervailing ex-ante effect: banks internalize the LOLR intervention and hold even less private liquidity than is optimal under the first-best scenario. This increases welfare losses from LOLR interventions, even as fire sale costs are reduced, and creates the rationale for regulations such as minimum margin requirements on clearinghouse members.

Indeed, consistent with our model, Alessandri and Haldane (2009) document that capital ratios maintained by financial sectors in the United States and the United Kingdom have been declining steadily over time in the presence of a lender of last resort (the Federal Reserve Banks in the US and the Bank of England in the UK). This is illustrated in Figure 1.

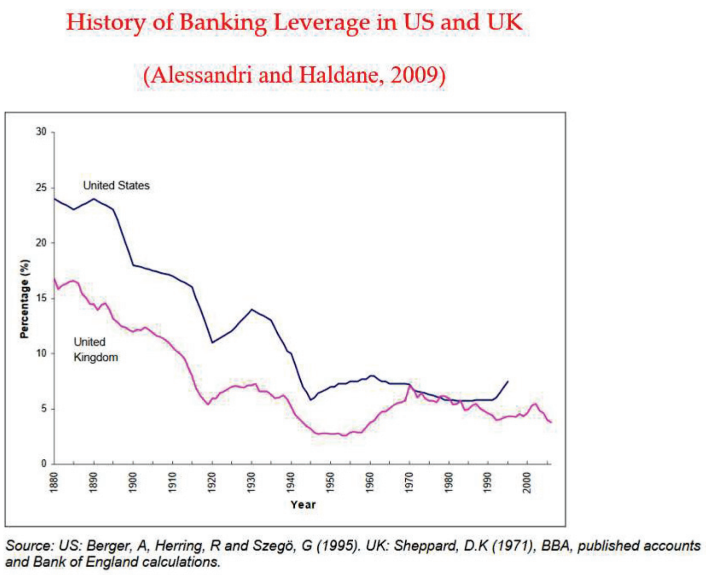


Figure 1. Bank capital ratios. This figure is taken from Chart 2 of Alessandri and Haldane (2009), and plots the evolution of capital ratios of US and UK banks over time.

These models help illustrate our main point that aggregate risk translates into systemic risk or collective failures depending on the liquidity choices of agents which are *endogenous* to the available risk-sharing opportunities.[2]

The ideas in this paper may have general applicability beyond the immediate concerns analyzed here. Consider any setting in which much of the activity is "routine" but there is an occasional need to face a complex unexpected scenario. (Most organizations can, in fact, be described in this fashion as largely requiring a set of routine activities which in principle may be implemented following protocol or "rules," but occasionally an unexpected—and consequential—query arises that cannot be addressed within the rules. In our model, the routine is the management of idiosyncratic risk, the dramatic is that of systemic risk.) With primitive technology, the same resources are used to manage both sets of activities, the routine and the rare, so, as a consequence, there are generally adequate resources available to address crises when they arise. However, as technology improves facilitating specialized handling of activities, the management of the routine can be separated from the management of the unexpected; the consequent specialization reduces resources available to address the rare hit (it may no longer be economically viable to maintain resources just to address the rare complex event), and so makes the organization more vulnerable to derailment from tail events. Note, however, that while this is a general idea, this paper models and understands the implications of this idea in the context of risk-sharing between banks (and more generally financial intermediaries).

Section 2 discusses the related literature. Sections 3 and 4 contain our analysis for the cases of autarky and two banks respectively. Section 5 generalizes this framework to many banks forming a clearinghouse and discusses the characteristics of equilibria that occur, comparing them to first-best outcomes. Section 6 incorporates a lender of last resort in the clearinghouse model. Finally, Section 7 concludes. Proofs not contained in the main body of the paper may be found in the appendices.

2. Related Literature

As noted in the introduction, prior to the financial crisis, the view that risk-sharing and derivatives enhanced systemic stability was widely held, but there were important notes of dissent. Speaking in Jackson Hole, Wyoming, in August 2005, Rajan (2005), then Chief Economist of the IMF, commented that

> While the system now exploits the risk-bearing capacity of the economy better by allocating risks more widely, it also takes on more risks than before. ... [T]he linkages between markets, and between markets and institutions, are more pronounced. While this helps diversify the system against small shocks, it also exposes the system to large systemic shocks.[3]

Rajan's (qualitative) case for why financial innovations may have made the world riskier focuses on tail-risk seeking that is driven by short-termism of financial sector pay and incentives. Our paper offers a formalization that is complementary and that leads to similar conclusions that the growth of markets for risk-sharing could result in greater risk-taking and make the system more vulnerable to "large systemic" shocks.

Acharya et al. (2010) document in detail the process of "manufacturing tail risks" that took hold during 2003–2007 and the reasons behind it, in particular, excessive seeking of aggregate risk due to the presence of government guarantees, imperfect regulation and its arbitrage, and inadequate internalization by the financial sector of externalities from collective failures. Our paper offers a viewpoint that is complementary to these. It suggests that improvements in financial risk-sharing technology and innovations, and the resulting moral-hazard effect on private liquidity choices,

---

2 As the models will make clear, we refer to aggregate risk as the risk arising from a bad realization of an aggregate shock that affects multiple banks and requires them to refinance. Systemic risk refers to the risk that multiple banks, that require refinancing, may simultaneously fail.

3 See Cassidy (2009, chp. 1, p. 21).

may be the subtle underlying force behind the growing inadequacy of private reserves to withstand aggregate shocks.[4]

Several papers consider the effect of risk-sharing on risk-taking and are related to our paper. In an early contribution, Bhattacharya and Gale (1987) consider how inter-bank contracts to share liquidity can lead to free-riding on the common pool of liquidity as individual banks, privately informed about own liquidity shocks, aim to transfer risks to counterparties. Acemoglu and Zilibotti (1997) provide a model in which risk-averse agents when offered diversification opportunities are willing to take on risks that were privately too risky to undertake otherwise. Allen and Carletti (2006) model how ability to transfer risks from banking to the insurance sector can create risk of contagion from one sector to the other.

In more recent work on this theme, Acharya and Bisin (2009) consider how managerial incentives to take on aggregate or idiosyncratic risks are altered by the ability to privately hedge these risks by trading in capital markets. Zawadowski (2013) considers a market for insurance in an "entangled" financial network and shows how the presence of network externalities can lead in equilibrium to inefficiently little insurance purchase against low-risk events. Yorulmazer (2013) considers how the access to credit default swaps (insurance against default risk) can lead the insured and the insurer to "herd" as they collectively prefer correlated underlying risks to transfer risks outside of the financial sector, e.g., on to the regulator and the taxpayers.

While sharing some similarity with each of these papers linking risk-sharing to risk-taking, our primary insight is somewhat different in that ability to share risks alters the cost-benefit tradeoff in holding liquidity to hedge against aggregate risk states and this alteration can endogenously transform such states into collective failure or systemic risk states.

Finally, the moral hazard effect of risk-sharing that we propose and model provides a novel way of understanding the risks that are shared, and more importantly, risks that remain, from co-insurance arrangements such as clearinghouses.

Much of the existing literature on clearinghouses is focused on somewhat different issues: determining whether risks will be adequately pooled by single versus multiple clearinghouses (Duffie and Zhu 2011) and whether clearinghouses can resolve incomplete information about agents' positions (Leitner 2012; Acharya and Bisin 2014). Pirrong (2009) argues that costs relating to information asymmetry are higher with a clearinghouse and may potentially outweigh the benefits associated with the mutualization of risks. Menkveld (2017) identifies the systemic risk arising from crowded trades by clearinghouse members and proposes a margin methodology to account for it. In contrast, a recent speech by Tucker (2014) highlights the moral hazard consequences of clearinghouse arrangements and resonates with our analysis of clearinghouse arrangements (the creation of systemic risk relative to autarky).

There has been some recent work on the optimal design of clearinghouses and on the welfare implications of their introduction and use. Koeppl et al. (2012) consider the tradeoff faced by a clearinghouse in providing liquidity to members, while ensuring incentives are in place for carrying out and settling transactions. Paddrik and Zhang (2020) study the optimal default waterfall design to limit clearinghouse losses given a set of exposures. Biais et al. (2012) consider whether centralized clearing makes trading parties better off and whether it eliminates counterparty risk of its members. They build a model in which buyers of protection (who are risk-averse) trade with risk-neutral sellers who may default. To reduce this counterparty risk, buyers have an incentive to expend effort and trade with a "good" protection seller, with low counterparty risk. In line with our findings, the authors show that—if risks are idiosyncratic—the clearinghouse is able to fully insure agents against possible future

[4] This, in turn, may also explain the explosive growth of the financial sector over the past two decades (Philippon and Reshef 2012) relative to other sectors of the economy; the growth occurred at the cost of grave risk of future crises whose costs were partly borne by sectors other than the financial sector.

losses. If there is aggregate risk, however, agents expend effort to search for good counterparties, and are fully insured only in the case when effort is observable and contractible.

Our model, however, is different in that it focuses on the assumption of risk undertaken by every bank (and hence the aggregate) when risk sharing is and is not possible. We show that risk sharing endogenously incentivizes the agents in our model to take on more risk leading to an increased likelihood of systemic failure, and efficiency losses resulting from a fire sale. The fire sale is modeled along the same lines as in Acharya et al. (2011), who study losses associated with transfers of control from domestic to less efficient foreign firms in a crisis.

Rochet and Roger (2016) build a general contracting theory of "risky utilities"—firms that provide public benefits while still at the risk of default, such as clearinghouses, large banks and infrastructure companies. They show that the optimal regulatory contract is implemented with a capital requirement, which, if breached, results in restructuring and expropriation. In contrast, we show that a simple margin requirement can serve to increase welfare when clearinghouse members do not internalize the cost of systemic risk, and that this regulation is at its most effective when the correlation between the shocks to member banks is high.

Castiglionesi et al. (2017) argue how increased financial integration can lead to a less stable banking system in the presence of aggregate shocks. They apply this idea in the contet of the inter-bank lending market and show how interest rates can be more skewed in the presence of greater financial integration. By contrast, our paper applies this idea to clearinghouses, from which we obtain different policy implications. Additionally, our paper studies how the presence of a Lender of Last Resort can affect bank choices and welfare outcomes.

3. The One-Bank Setting

3.1. Model and Payoffs

In this section, we introduce the foundational one-bank model that we embed in multi-bank settings in Sections 4 and 5. We consider a model with three dates $t \in \{0, 1, 2\}$. At date 0, a bank has initial investable capital of \$1. Banks have access to two investment opportunities. The first is a safe asset ("cash"). The safe asset has a net return of zero; an investment of ℓ at time t returns ℓ at time $t+1$.

The second investment opportunity is in a risky asset or risky project. Figure 2 summarizes the consequences of investing in this opportunity. Investment in this risky project at time 0 yields returns at time 2. The risk in the project comes from the potential need to refinance the investment at time 1. There is a probability α that such refinancing will be required. Refinancing is a zero-one decision; partial refinancing of the project is not possible. The refinancing amount required is $\$\varphi$ per \$1 that was invested in the project. For specificity, we take $\varphi = 1$; i.e., if refinancing is required, the refinancing need is 100% of the initial invested amount in the risky project.[5]

If the project does not require refinancing at time 1, or if it requires refinancing and is refinanced, then the expected return at time 2 is $R > 1$ per unit investment in the project at time 0. If refinancing is needed and the project is not refinanced, the investment returns zero with certainty at time 2.

[5] This is for expository purposes to more easily illustrate our main point that increased risk sharing can lead to greater systemic risk. Appendix B contains a more detailed discussion on bank behavior given a more general value of φ for both the 1-bank and the 2-bank model. Our results—that there exists a risk-sharing equilibrium that leads to greater systemic risk than in autarky—continue to hold.

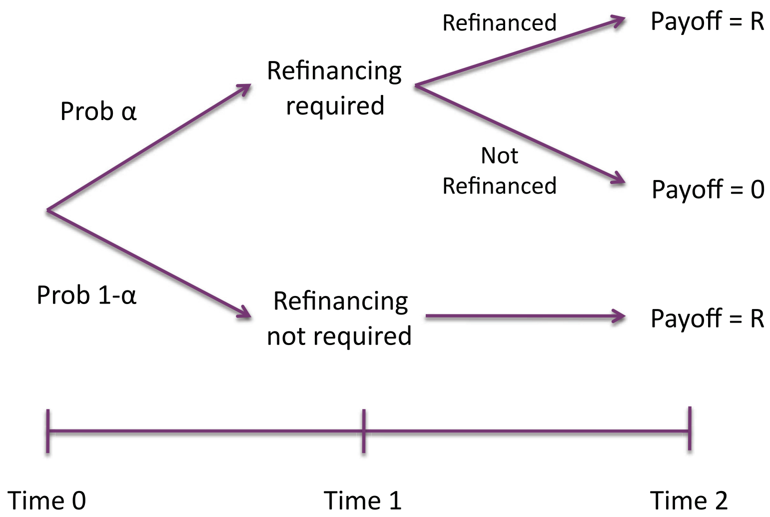


Figure 2. The risky investment opportunity. This figure summarizes the salient features of the risky investment opportunity described in Section 3.

Since $R > 1$, it is always worth refinancing the project at time 1 should such refinancing be required (and should it be feasible). We make two assumptions concerning the refinancing. First, we assume that the refinancing need is observable and verifiable. This assumption plays no role in the current section, but is used in a later section when we consider allowing the bank to use a contingent claim to guard against this state. Second, we assume that the cash flows from the risky project are not pledgeable, so the bank cannot raise outside financing against these cash flows. For the time being (until we introduce risk-sharing possibilities), this means the project may only be refinanced using the bank's own internal resources, which in turn means it can only be refinanced using the amount the bank has invested in the riskless asset.

3.2. The Bank's Optimal Investment Decision

The tradeoff driving the bank's investment decision is a simple one: the riskless asset's return is lower than that from the risky project, but investing (a suitable amount) in the riskless asset provides the bank with the ability to refinance the project should refinancing be required—which happens with probability α. We show that if α is small, then it is not worthwhile for the bank to divert investment away from the risky project; there is no investment in the riskless asset and the risky project fails with probability α. As α increases past a critical level, then refinancing becomes optimal; the bank fully self-insures and invests enough in the riskless asset to meet the refinancing need (should it arise), and the risky project never fails. Finally, at very high levels of α, investment in the risky project becomes unprofitable and all investment is made in the riskless asset.

Specifically, let $(\ell, 1-\ell)$ denote the investment strategy in which the bank chooses to invest ℓ in the riskless asset and $1-\ell$ in the risky technology. We prove the following result:

Proposition 1. *Let $\lambda(R)$ and $\nu(R)$ be defined by*

$$\lambda(R) \;=\; \frac{R-1}{2R-1}; \qquad \nu(R) = R-1.$$

Then:

1. *If $\alpha \in [0, \lambda(R))$, then the optimal action is $\ell^* = 0$. Since there is no investment in the riskless asset, the risky project is never refinanced so fails with probability α. The bank's expected payoff is $(1-\alpha)R$.*

2. *If $\alpha \in [\lambda(R), \nu(R)]$, the optimal action is $\ell^* = 1/2$. There is always exactly enough to refinance the risky project if this is required. The risky project is always refinanced, so never fails. The bank's expected payoff is*

$$\frac{1}{2}[R + 1 - \alpha]. \tag{1}$$

3. *If $\alpha > \nu(R)$, the optimal action is $\ell^* = 1$. There is no investment in the risky project and the bank's expected payoff is* 1.

Remark 1. *Note that Case 3 is irrelevant when $R \geq 2$ since $\alpha \leq \nu(R)$ is always satisfied in this case.*

Proof. See Appendix A. □

In the sections following we extend this model in several directions. Section 4 considers a basic extension to a two-bank setting and looks at the consequences of permitting risk sharing in this set-up. Section 5 develops a many-firms model and look at the consequences of permitting risk-sharing via a third-party insurance-style arrangement and a clearinghouse structure, respectively.

4. Two-Banks: Risk-Sharing and Increasing Systemic Risk

Consider a two-bank setting in which each bank faces exactly the same investment opportunities as in the previous section with exactly the same payoffs, but in which the refinancing needs may be correlated. Thus, there are four states of the world that are possible:

State	Probability	Risky Projects Outcome
1	p_1	Neither project requires refinancing
2	p_2	Only Bank 2 needs refinancing
3	p_3	Only Bank 1 needs refinancing
4	p_4	Both banks need refinancing

We refer to States 2 and 3 as the "idiosyncratic" risk states and to State 4 as the one of "aggregate" or "systemic" risk. The total probability of Bank 1 needing refinancing ($p_3 + p_4$) and the total probability of Bank 2 needing refinancing ($p_2 + p_4$) are each taken to equal α, where α is the probability in the one-bank setting of a refinancing need. For an arbitrary correlation ρ in refinancing needs, the probabilities p_i take on the form [6]

$$\begin{aligned} p_1 &= \rho(1-\alpha) + (1-\rho)(1-\alpha)^2 \\ p_2 &= (1-\rho)\alpha(1-\alpha) \\ p_3 &= (1-\rho)\alpha(1-\alpha) \\ p_4 &= \rho\alpha(1-\alpha) + \alpha^2 \end{aligned} \tag{2}$$

Note that the banks are ex-ante identical. We focus on the case $\alpha \in [\lambda(R), \nu(R)]$. As noted in Proposition 1, the bank under autarky then invests in the risky project but fully (self-)insures this investment so there is no failure.

We allow the banks to "risk-share" by trading in contingent claims that pay off in specified states of the world. When contingent claim trading is disallowed, each bank faces precisely the one-bank model of Section 3 the solution to which is identified in Proposition 1; we refer to this as the model of "autarky." Of course, even if contingent claim trading is allowed, each bank has the option to not

[6] These probabilities can be derived by using the properties of Bernoulli random variables. For example, if Z_1 is an indicator variable equal to 1 if Bank 1 fails, and Z_2 is an indicator variable equal to 1 if bank 2 fails, then the probability that both banks fail is just $E(Z_1 Z_2) = Cov(Z_1, Z_2) + E(Z_1)E(Z_2)$. $E(Z_1) = E(Z_2) = \alpha$ and $Cov(Z_1, Z_2) = \rho\alpha(1-\alpha)$, from which we get the results. The other probabilities are computed similarly.

participate in the contingent claim market; it will elect to participate if and only if participation raises its expected payoff beyond that in the autarkic solution.

Risk-sharing affords each bank the possibility of lowering its investment in the safe asset while increasing its investment in the risky asset and relying instead on additional funds from the contingent claims should refinancing of its risky project be required. There is a chance, of course, that the refinancing shock may be "systemic," i.e., both banks may require refinancing and there may not be enough cash to meet the combined refinancing needs.[7] Thus risk-sharing involves a trade-off between the larger returns obtained from investing in the risky project and its failure in the systemic state.

Observe that risk-sharing benefits accrue in the idiosyncratic-risk States 2 and 3 but not in the systemic-risk State 4. The likelihood of the idiosyncratic-risk states decreases, and that of the systemic-risk state increases, as the correlation ρ in refinancing needs increases. Intuitively, therefore, one would expect that there is a critical cut-off level of ρ below which risk-sharing may dominate the autarkic outcome but above which autarky should dominate. Proposition 2 establishes precisely this to be the case:

Proposition 2. *Fix any $R > 1$. There is a critical level of correlation $\rho^* > 0$ given by*

$$\rho^* \; = \; \frac{1}{2R}$$

such that

1. *For $\rho < \rho^*$:*
   (a) *There is an interval of values $[\lambda(R), \alpha^*)$ of α under any of which there is a risk-sharing equilibrium in which each bank invests an amount $2/3$ in the risky project and an amount $1/3$ in the riskless asset.*
   (b) *Outcomes in this risk-sharing equilibrium strictly dominate outcomes under autarky.*
   (c) *There is a positive probability given by $\rho\alpha(1-\alpha) + \alpha^2$ in this risk-sharing equilibrium of systemic failure (i.e., of both projects needing refinancing and neither obtaining it).*
2. *For $\rho > \rho^*$, the outcomes under autarky dominate those under risk-sharing.*

Proof. See Appendix B. □

The risk-sharing equilibrium described in Proposition 2 has the following properties:

1. There is greater individual risk-taking in the form of investment in the risky project by each bank under risk-sharing than under autarky (2/3 compared to 1/2), and correspondingly lower total liquidity/cash reserves held in the system under risk-sharing than under autarky (2/3 compared to 1).
2. A systemic failure (State 4 with no accompanying refinancing) happens with positive probability in the risk-sharing equilibrium whereas, for the same parameters, there is no systemic failure under autarky; indeed, there are no individual failures under autarky because each bank maintains adequate reserves to insure itself.
3. Thus, expected systemic shortfall is zero under autarky but, under risk-sharing, each bank falls 2/3 short of its refinancing need in the systemic failure state, so expected systemic shortfall is equal to
$$2 \times \frac{2}{3} \times (\rho\alpha(1-\alpha) + \alpha^2) \; = \; \frac{4}{3}\,(\rho\alpha(1-\alpha) + \alpha^2).$$

[7] Note that if there was enough cash to meet combined refinancing need, each bank must have invested enough in the riskless asset to meet at least its own refinancing need, so we are effectively in the autarky solution.

The results of this section have been derived in a setting that is simple and transparent but also obviously restrictive (only two firms sharing risks, etc.). In the sections that follow, we describe a richer framework and examine the consequences of risk-sharing in a clearinghouse-style co-insurance arrangement (Section 5).

A final comment is relevant. The results of this section do not, in themselves, carry any welfare implications, in particular, that risk-sharing is "bad" in the sense of being welfare-decreasing. Indeed, if there are no externalities from systemic failures, then welfare is (trivially) increased by allowing for risk-sharing. However, if there are social and macroeconomic costs of bank bankruptcy and systemic failure (resulting, for example, from disruptions in the payment system), risk-sharing may well be welfare-reducing also.

5. Clearinghouse: A Co-Insurance Model

In this section, we move to a setting with a large number of banks and a richer description of uncertainty. Refinancing needs are driven by a common factor and also by idiosyncratic factors. When viewed in isolation, each individual bank in this model continues to face the same decision problem as in the one-bank setting of Section 3; thus, if the bank decides to remain in autarky and not participate in risk-sharing, its optimal actions are those specified by Proposition 1.

5.1. Description of Uncertainty

Suppose there are n banks. If Bank i invests an amount $y > 0$ in the risky project, it receives an interim signal $A_i \sim \Phi(0,1)$ concerning the returns from the project, where $\Phi(\cdot)$ denotes the cumulative standard normal distribution. If $A_i \leq c_i$ (where c_i is a given critical cut-off), then the project needs to be refinanced, i.e., an additional amount y needs to be invested in the project. If the project is not refinanced, it returns zero with certainty. If it is refinanced, it returns $R > 1$ with certainty.

The banks are ex-ante identical, and in particular $c_1 = \cdots = c_n (= c$, say). We let $\alpha = \text{Prob}(A_i \leq c) = \Phi(c)$ be the ex-ante probability that a bank will need refinancing. We assume that the signals A_i satisfy[8]

$$A_i \;=\; \sqrt{\rho}\, A + \sqrt{1-\rho}\,\epsilon_i$$

where $A, \epsilon_i \sim \Phi(0,1)$, and A and the ϵ_i's are independent. Under this condition, banks' refinancing needs have a pairwise correlation of $\rho \in (0,1)$. We refer to A as the common or aggregate factor in determining refinancing needs and to the ϵ_i's as bank-specific or idiosyncratic factors. Conditional on $A = a$, note that the probability $p(a)$ that a generic bank needs refinancing is

$$\begin{aligned} p(a) &= \text{Prob}\{\sqrt{\rho}\, a + \sqrt{1-\rho}\,\epsilon_i \leq c\} \\ &= \text{Prob}\left\{\epsilon_i \leq \frac{c - \sqrt{\rho}\, a}{\sqrt{1-\rho}}\right\} \;=\; \Phi\left(\frac{c - \sqrt{\rho}\, a}{\sqrt{1-\rho}}\right). \end{aligned} \tag{3}$$

We assume that there is a "clearinghouse" wholly owned by the banks that provides insurance to member banks from counterparty credit risk. The clearinghouse is funded with up-front premium or "margin" payments of size k from each of the participating banks; in addition, the clearinghouse also has the right to make ex-post capital calls on its members. Specifically, in period 1, when banks' refinancing needs are known, these are first funded out of the margins collected by the clearinghouse and liquidity carried over by all member banks. Should this prove insufficient, the clearinghouse can then pledge a fraction τ of future payoffs realized by the investments of those banks that did not experience a refinancing need. However, if the initial margins and these capital calls prove collectively inadequate to meet the total refinancing needs, then the clearinghouse becomes insolvent and no bank's refinancing insurance contract is honored. In this event, we assume that the clearinghouse

8 The uncertainty structure is similar to that introduced in Vasicek (2002) in a different context.

uses its margins to refinance as many banks as possible. We also assume that the refinancing choice is made randomly, i.e., ex-ante, every insolvent bank has the same probability of being refinanced by the clearinghouse. Those banks which have not been refinanced will sell their assets to solvent banks in a fire sale that we describe later in this section.

Banks take into account both the possibility of ex-post transfers and the chance of participating in the fire sale in their decision to deposit margin in the clearinghouse. Banks choose the level of margin they contribute to the clearinghouse, as well as the liquidity they choose to carry over for purchasing assets in the fire sale in case they remain solvent.

Conditional on insurance remaining solvent, the size of the transfers from the successful banks to the failed banks is a function of the fraction of banks requiring refinancing. This fraction (or more accurately, the distribution of this fraction) depends on the size of the aggregate shock. We look to characterize the equilibria that result in this setting.

As mentioned, k denotes the up-front margin paid by each bank to the clearinghouse. Let ℓ denote the liquidity carried by each bank. Per dollar invested in the risky project, the bank obtains a return of R if it is successful on the investment. Each bank is subject to an aggregate shock a as well as an idiosyncratic shock ϵ_i and needs to refinance if $\sqrt{\rho}\,a + \sqrt{1-\rho}\,\epsilon_i < c$. Banks are ex-ante identical. Conditional on the aggregate shock a, let $f(a)$ be the fraction of banks requiring refinancing. For expositional simplicity, we write f for $f(a)$ in the steps that follow.

Conditional on the aggregate shock a, the aggregate revenue of the banks not requiring refinancing is $R(1-f)(1-k-\ell)$, while the total refinancing need is $f(1-k-\ell)$. Given the initial margin collection of k, and the initial liquidity holdings of ℓ, this leaves a deficit of $f(1-k-\ell)-k-\ell$ to be shared between those banks that were successful, i.e., did not experience a refinancing need. Recall that the clearinghouse can pledge a fraction τ of solvent banks' payoff (if required) to refinance those that have failed. Thus, the share $\eta(f)$ of its realized revenues that each successful bank must give up to the clearinghouse is

$$\eta(f) \;=\; \frac{f(1-k-\ell)-k-\ell}{\tau R(1-k-\ell)(1-f)}$$

If $\eta(f) > 1$, then the clearinghouse cannot meet the refinancing needs even with the ex-post capital calls, so becomes insolvent. Note that we may also have $\eta(f) < 0$; this simply means that the clearinghouse has a surplus left over after meeting refinancing needs and returns this amount to the successful banks.

Therefore, we now have two cases:

- $f > [\tau R(1-k-\ell)+k+\ell]/[(1-k-\ell)(1+\tau R)]$, i.e., $\eta(f) > 1$: The clearinghouse fails and uses its margin k to refinance a fraction of the failed banks. The remaining insolvent banks sell their assets in a fire sale. Solvent banks do not need refinancing and realize the return on their risky investment. They also purchase assets if a fire sale takes place.
- $f \le [\tau R(1-k-\ell)+k+\ell]/[(1-k-\ell)(1+\tau R)]$: Failed banks have their refinancing needs honored, realizing (after the bailout) a payoff of $R(1-k-\ell)$. Successful banks also lose a fraction $\eta(f)$ of the pledgeable portion of their payoff to fund the bailout.

Recall that

$$f \;=\; f(a) \;=\; \Phi\left(\frac{c-\sqrt{\rho}a}{\sqrt{1-\rho}}\right).$$

Let $a_0(k,\ell)$ be such that $f(a_0(k,\ell)) \;=\; [\tau R(1-k-\ell)+k+\ell]/[(1-k-\ell)(1+\tau R)]$. Then the clearinghouse fails and a fire sale takes place if $a < a_0(k,\ell)$, and is solvent for $a \ge a_0(k,\ell)$.

The next subsections outline the fire sale process before characterizing the equilibrium.

5.2. The Fire Sale

We will assume that the fire sale of firms' assets takes place according to the following rules:

(1) Each solvent bank takes prices as given, and submits a demand function to acquire assets—which are perfectly divisible—of the insolvent banks.
(2) Given the price, demand for each bank is constrained by the amount of liquidity it has at time 1.
(3) The initial margin the clearinghouse collects, k, is used to bail out some of the firms that need refinancing.
(4) There is full liquidation of failed firms' assets. So the payoff for an insolvent bank is just the price at which it sells its assets, along with any liquidity it may carry over into time 1.
(5) Solvent banks can pledge a fraction τ of payoffs to acquire assets. These banks can only generate a return of $R - \Delta$ from acquired assets.

Let us assume that the clearinghouse has failed and that there will be a fire sale of assets of the insolvent firms. As before, let $f(a)$ be the number of firms that have failed, where a is the value of the aggregate shock. Recall that

$$f(a) = \Phi((c - \sqrt{\rho}a)/\sqrt{1-\rho})$$

where $\Phi(\cdot)$ is the CDF of the shocks.

The clearinghouse fails and the fire sale takes place when

$$f(a) > f(a_0(k,\ell)) = \frac{\tau R(1-k-\ell) + k + \ell}{(\tau R + 1)(1-k-\ell)}$$

Let the number of firms refinanced by the clearinghouse be given by $g(k,\ell)$, where k is the clearinghouse margin and ℓ is the liquidity carried over by each bank from time 0. Since each failed bank requires $(1-k-\ell)$ to refinance, we have that $g(k,\ell) = k/(1-k-\ell)$.

The number of banks selling assets in the fire sale equals $f(a) - g(k,\ell)$ while the number of banks buying assets equals $1 - f(a)$, if bailed out banks do not participate in the fire sale. Let $y(p,k,\ell)$ denote the demand function submitted by each acquiring bank in the fire sale.

The market clearing condition is given by

$$y(p,k,\ell)[1 - f(a)] = (1-k-\ell)[f(a) - g(k,\ell)]$$

The demand curves also satisfy

$$y(p,k,\ell) = \frac{(\ell + \tau R(1-k-\ell) - y(p,k,\ell))^+}{p} \text{ where } x^+ = \max(x,0)$$

To understand the above expression, note that—given a price p—the firm has liquidity it carries over, together with the payoff realized at time 1. These realized payoffs equal a fraction τ of future payoffs given by $R(1-k-\ell)$. The firm also has to refinance projects it acquires, however, and has to pay a unit amount per asset acquired.

The fire sale price is therefore given by

$$p(k,\ell) = \max(0, -1 + [\ell + \tau R(1-k-\ell)]\frac{(1-f(a))}{(1-k-\ell)(f(a)-g(k,\ell))})$$

Acquiring banks profit from the fire sale when p is less than $(R - \Delta - 1)$. This corresponds to the condition $f > f_1(a)$ where f_1 is given by

$$f_1(k,\ell) = \frac{(\ell + \tau R(1-k-\ell)) + k(R-\Delta)}{\ell + \tau R(1-k-\ell) + (1-k-\ell)(R-\Delta)}$$

For $f > f_1$, $p < (R - \Delta - 1)$. In addition, to prevent prices being negative, we must have that $p = 0$ when $f > \underline{f}(k, \ell)$, where

$$\underline{f}(k, \ell) = \frac{k + \bar{\ell} + \tau R(1 - k - \bar{\ell}))}{(1 - k - \bar{\ell}) + (\bar{\ell} + \tau R(1 - k - \bar{\ell}))}$$

Note that the fire sale only happens when insolvent banks cannot refinance themselves, i.e., $\ell < (1 - k - \ell)$. If ℓ were greater than $(1 - k - \ell)$, then banks would always choose to refinance since the gains from refinancing and obtaining the return of R on risky assets would always exceed the payoff from selling assets in the fire sale and a price lower than or equal to $(R - \Delta - 1)$.

5.3. Prices and Profits

In this table, we summarize the model in the above section and explicitly list the fire sale prices and the demand for each bank in the fire sale. Insolvent firms obtain the proceeds from the fire sale as payoff. This payoff equals the fire sale price multiplied by the assets requiring refinancing. Solvent firms get the payoff they generate from acquired assets, $(R - \Delta - 1)$, net of fire sale price, as profit per unit of asset acquired in the fire sale. The total profit for acquiring banks is listed in the fourth column of the table. For simplicity, let us denote $f(a_0(k, \ell))$ by f_0. Recall that a fire sale happens only when the clearinghouse becomes insolvent and $f > f^*$

Region	Fire sale price	Fire sale demand	Profits (for acquiring firms)
$f \in [f_0, f_1)$	$(R - \Delta - 1)$	$y(R - \Delta - 1, k, \ell)$	0
$f \in [f_1, \underline{f}]$	$p(k, \ell)$	$y(p, k, \ell)$	$(R - \Delta - 1 - p)y(p, k, \ell)$
$f \in [\underline{f}, 1]$	0	$y(0, k, \ell)$	$(R - \Delta - 1)y(0, k, \ell)$

Let a_1 and $\underline{a}$ denote the corresponding value of the aggregate shock if the number of failures equals f_1 and $\underline{f}$ respectively. Then, from the above table, we see that for surviving firms, profit from the fire sale is 0 if $a > a_1$, while insolvent firms get a price $(R - \Delta - 1)$ for their assets. Since each insolvent firm sells $(1 - k - \ell)$ of assets in the fire sale, its profit from the fire sale is just $(1 - k - \ell) \times p$, where p is the fire sale price. Further, insolvent firms carry over the liquidity, ℓ from time 0. If $a > a_0(k, \ell)$, the clearinghouse is solvent. There are no fire sales and the profit is given by total payoff net of transfer $\eta(k, \ell)$.

When computing expected profits for a given bank, we will assume that the clearinghouse sets the margin k, and all the other banks carry over liquidity $\bar{\ell}$. The bank takes $\bar{\ell}$ as given, and chooses ℓ accordingly. We will derive profits as a function of k, ℓ and $\bar{\ell}$.

Since all other banks carry over $\bar{\ell}$, and since each bank is atomistic, economy-wide variables such as prices, thresholds and transfers will be given by $\bar{\ell}$ instead of ℓ. In particular, the fire sale price is given by $p(k, \bar{\ell})$. Clearing house failure is determined by the threshold $a_0(k, \bar{\ell})$, while transfers (in the case that the clearinghouse is solvent) is given by

$$\eta(f) = \frac{f(1 - k - \bar{\ell}) - k - \bar{\ell}}{\tau R(1 - k - \bar{\ell})(1 - f)}$$

Therefore, the payoff of a bank if the clearinghouse is solvent and if it carries liquidity ℓ is $\tau R[1 - k - \ell][1 - \eta(f)] + (1 - \tau R)(1 - k - \ell) + \ell$ if it does not require refinancing, and $R[1 - k - \ell] + \ell$ otherwise.

For notational ease, let $\underline{m} = \min(\underline{a}, a_0)$ and $m_1 = \min(a_0, a_1)$.

Integrating over a, the ex-ante expected profit of a representative bank as a function of k and ℓ is, therefore (where $\phi(\cdot)$ is the PDF of aggregate shock a):

$$
\begin{aligned}
\mathbb{E}\Pi(k,\ell,\bar{\ell}) \quad = \quad & \\
& \int_{-\infty}^{a_0(k,\bar{\ell})} [R(1-k-\ell)+\ell](1-f(a)+g(k,\bar{\ell}))\phi(a)da \\
& + \int_{-\infty}^{\underline{m}} [(R-\Delta-1)y(0,k,\ell)(1-f(a))+\ell(f(a)-g(k,\ell))]\phi(a)da \\
& + \int_{\underline{m}}^{m_1} [(R-\Delta-1-p)y(p,k,\ell)(1-f(a))+[p\times(1-k-\ell)+\ell](f(a)-g(k,\bar{\ell}))]\phi(a)da \\
& + \int_{m_1}^{a_0(k,\bar{\ell})} [(R-\Delta-1)(1-k-\ell)+\ell)(f(a)-g(k,\bar{\ell})]\phi(a)da \\
& + \int_{a_0(k,\bar{\ell})}^{\infty} ([(\tau R(1-\eta(f(a)))+(1-\tau R))(1-f(a))+Rf(a)](1-k-\ell))\,\phi(a)da \qquad (4)
\end{aligned}
$$

The initial margin k is chosen by the mutually-owned clearinghouse to maximize this profit. The choice of k will evidently depend on the correlation ρ and other factors. We look to characterize the optimal outcome, focusing in particular on the questions of whether risk-sharing can dominate autarky; whether risk-sharing leads to greater individual risk-taking than autarky; and whether risk-sharing leads to greater systemic failure (here insolvency of the clearinghouse) than autarky.

5.4. Properties of Equilibrium

As mentioned before, when solving for the equilibrium, we assume that the clearinghouse sets the margin level, k paid by each bank. Banks can then choose liquidity ℓ to carry over from time 0. When making this choice, banks take as given the liquidity carried over by the other banks. More formally, the equilibrium quantities k^* and ℓ^* solve the following system:

$$k^* = \arg\max_k \;\; \mathbb{E}\Pi(k,\ell^*(k)) \;\; \text{where}$$

$$\ell^*(k) = \arg\max_\ell \;\; \mathbb{E}\Pi(k,\ell,\bar{\ell}) \;\; \text{and} \;\; \ell^* = \bar{\ell}$$

The above equation states that banks choose liquidity ℓ^* taking as given liquidity $\bar{\ell}$ carried over by other banks (which affects the probability of failure of the clearinghouse). In a symmetric equilibrium, $\ell^* = \bar{\ell}$.

Characterizing the Equilibrium

It is clear from the expression earlier that expected profits are linear in ℓ, i.e., expected profits can be written as

$$\mathbb{E}\Pi(k,\ell,\bar{\ell}) = \alpha_0(k,\bar{\ell}) + \alpha_1(k,\bar{\ell})\ell$$

Since profit is linear in ℓ, there is a bang-bang solution to the bank's choice of ℓ. Therefore we have three cases:

Case 1: $\alpha_1(k,\bar{\ell}) < 0$ In this case, the bank chooses $\ell^*(k,\bar{\ell}) = 0$. For a symmetric equilibrium to exist in this case, we must have that $\bar{\ell} = 0$, and for consistency, we must also have that $\alpha_1(k,0) < 0$. This situation corresponds to the case where the bank carries over no liquidity from time 0.

Case 2: $\alpha_1(k,\bar{\ell}) = 0$ In this case, the bank is indifferent to the choice of ℓ. For a symmetric equilibrium, the bank chooses $\ell^*(k,\bar{\ell}) = \bar{\ell}$. The clearinghouse then implements

$$k^* = \arg\max_k \;\; \mathbb{E}\Pi(k,\ell^*(k)) \;\; \text{where}$$

$$\ell^*(k) \;\; \text{solves} \;\; \alpha_1(k,\ell^*(k)) = 0$$

Case 3: $\alpha_1(k, \bar{\ell}) > 0$ In this case, the bank chooses $\ell^*(k, \bar{\ell}) = 1$ and in equilibrium, $\ell^* = \bar{\ell} = 1$. There is no systemic risk or investment in the risky asset and the clearinghouse never fails. This is an uninteresting case.

Proposition 3.

1. *For every ρ and for every k, there exists a unique $\ell^*(k)$ such that $(k, \ell^*(k))$ is an equilibrium.*
2. *If (k, ℓ) is an equilibrium for a given ρ, then $(k + \ell, 0)$ is an equilibrium also.*
3. *For every ρ, $\ell^*(k^*) = 0$, where $k^* = \arg\max_k \; \mathbb{E}\Pi(k, \ell^*(k))$*

Proof. See Appendix C. □

The above results imply that at the optimal margin (for each value of ρ) chosen by the clearinghouse, banks do not carry over any liquidity. To better understand the gains from the clearinghouse and risk sharing, it is important to obtain the payoffs for each bank without the clearinghouse. In the absence of the insurance provided by the clearinghouse, banks may still choose to hold liquidity. Their liquidity holdings will allow them to purchase assets in a fire sale if they remain solvent, or may be large enough to help them refinance if they become insolvent, in which case a fire sale never takes place. In the proposition below, we show that the second situation always occurs, i.e., banks carry enough liquidity to refinance in the event of insolvency. Banks are therefore autarkic, meaning that they always refinance themselves if becoming insolvent.

Proposition 4. *In the absence of the clearinghouse, banks choose liquidity levels $\ell^* = 1/2$ for every ρ. There are no fire sales and banks have the same expected payoffs as under autarky. There is no systemic risk.*

Proof. See Appendix D. □

Our main result is that co-insurance is always strictly better than autarky in that it implies a higher expected profit for banks. However, for every possible equilibrium under coinsurance, there is higher systemic risk and greater investment in the risky asset than under autarky. Further, liquidity held by banks is sharply curtailed in the presence of a clearinghouse. This also means that if the clearinghouse fails, banks that do not get bailed out by the clearinghouse are worse off since other banks do not carry liquidity, and likely will have to sell their assets at a lower price.

The three panels of Figure 3 illustrate the equilibrium with the clearinghouse and fire sale for two separate values of τ. The figure considers $R = 1.5$, and $\alpha = 0.4$. Under autarky, there is no systemic risk and payoffs are given by $(1 + R - \alpha)/2 = 1.05$.

The upper-left panel presents a plot of the initial margin as correlation varies. When correlation is very low, the optimal initial margin is such that the clearinghouse only fails with low probability. When τ is high, the margin is very low at low values of ρ since the clearinghouse can pledge a large fraction of solvent banks' payoffs to refinance insolvent banks. On the other hand, when τ is low, it is optimal for banks to deposit a large margin into the clearinghouse to prevent failure. As correlation increases, more firms fail together, so reliance on successful investments alone is insufficient and the initial margin level becomes positive, eventually rising towards the autarky level of 0.50. The investment in the risky asset under coinsurance is therefore equal to 1, the maximum possible amount, at low levels of correlation but decreases towards the autarky level of 1/2 as correlation increases. As τ—the fraction of future payoffs that can be pledged—goes up, the optimal margin paid to the clearinghouse falls. As τ rises, it becomes more probable that the clearinghouse will not fail and, if it does, that the fire sale will happen at no arbitrage, i.e., with a price of $(R - \Delta - 1)$, and insolvent banks will be better off. Banks internalize this and respond by depositing less in margin to the clearinghouse.

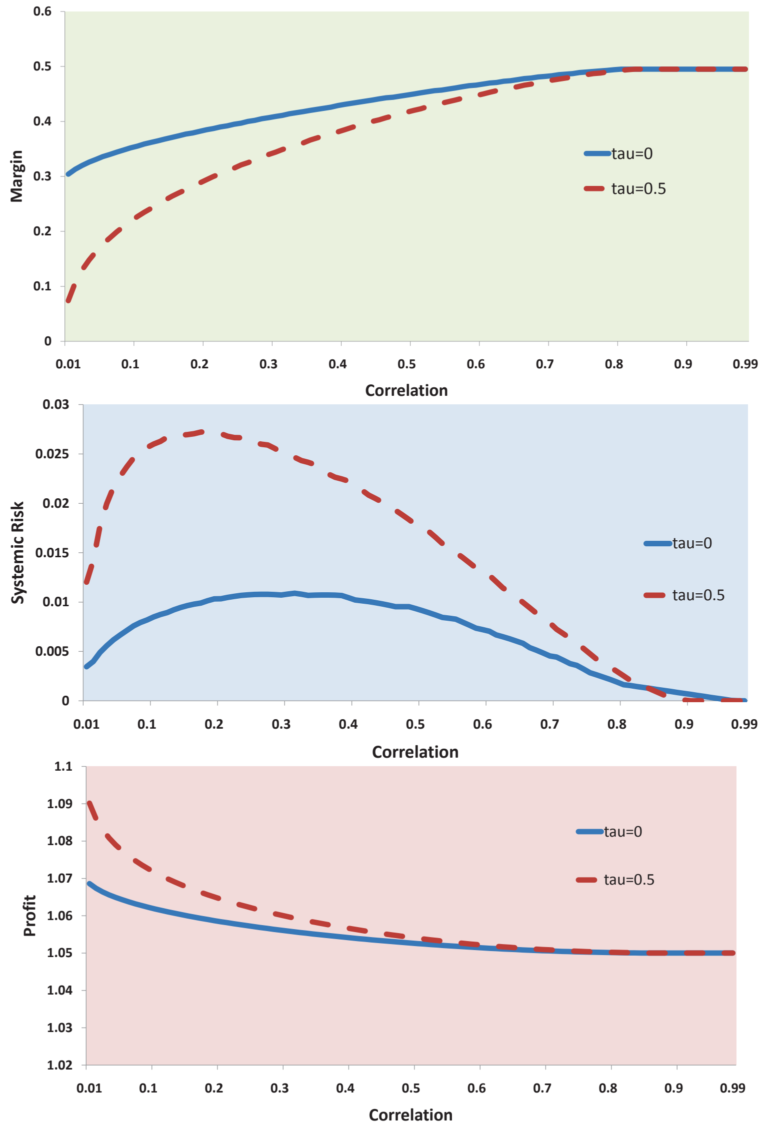


Figure 3. Equilibrium under co-insurance and fire sale. The three panels of this figure plot various properties of equilibrium by varying τ in the model of Section 5. In all the figures, the return from the risky project is set to $R = 1.50$ and the value of α (the probability that refinancing of the risky project will be required) is set to 0.4. Panel 1 (top left) plots aggregate systemic liquidity in equilibrium as correlation changes. Panel 2 (top right) plots systemic risk as correlation changes. Lastly, Panel 3 (bottom) plots the payoff to a generic bank as correlation changes.

The upper right panel plots systemic risk in the model as a function of correlation. We define systemic risk as the ex-ante probability that a bank will fail and have to sell its assets inefficiently in a fire sale. For the initial low levels of correlation where the optimal initial margin is zero, this probability of failure increases with correlation. Beyond this point, the collection of initial margin causes the probability of failure to fall back down, eventually reaching the zero probability level of autarky. We see that systemic risk sharply rises as τ goes up. Since an increase in τ is accompanied by a lowering of the liquidity held by the clearinghouse, an inefficient fire sale becomes more likely leading to a rise in systemic risk. This is another illustration of the fact that as markets become more efficient, systemic risk may rise.

The lower panel presents a typical plot of payoffs in the clearinghouse model as correlation varies. The payoff with the clearinghouse is at least as large as that under autarky. As correlation increases, the benefits of risk-sharing decrease, so co-insurance payoffs fall towards the autarky payoff.

Not surprisingly, the payoff rises with τ since the pledgeability of the clearinghouse and banks to acquire assets in the fire sale increases, and hence banks are strictly better off.

Recall that, for any given value of ρ, and any equilibrium pair (k^*, ℓ^*) chosen by banks, the thresholds $a_0(k,\ell)$, $a_1(k,\ell)$ and $\underline{a}(k,\ell)$ determine whether the clearinghouse fails and—if so—the fire sale prices and payoffs for participating banks. We have been assuming that the aggregate shock a follows a standard normal distribution, i.e., $a \sim \Phi(0,1)$. Therefore, we can obtain the 95th and 5th percentile realization of a, and obtain profits and systemic risk at these realizations using Equation (5), which shows the ex-ante expected payoffs for a bank according to the realization of the aggregate shock a, and whether there is a fire sale or not.

These are shown in the two panels of Figure 4. The dotted line in the first panel of the figure shows profits with a bad realization of aggregate shock a. For low values of ρ, equilibrium quantities (k^*, ℓ^*) are such that the clearinghouse does not fail for a 5th percentile realization of a. At ρ of around 0.05, this no longer holds, which is why we see a sudden jump downwards in profits and rise in systemic risk. The interpretations of profits and systemic risk with bad realizations of a is similar to the interpretation of value at risk measures, i.e., they indicate how much fragility is present in the system if there is a tail event (in this case, a bad draw of a).

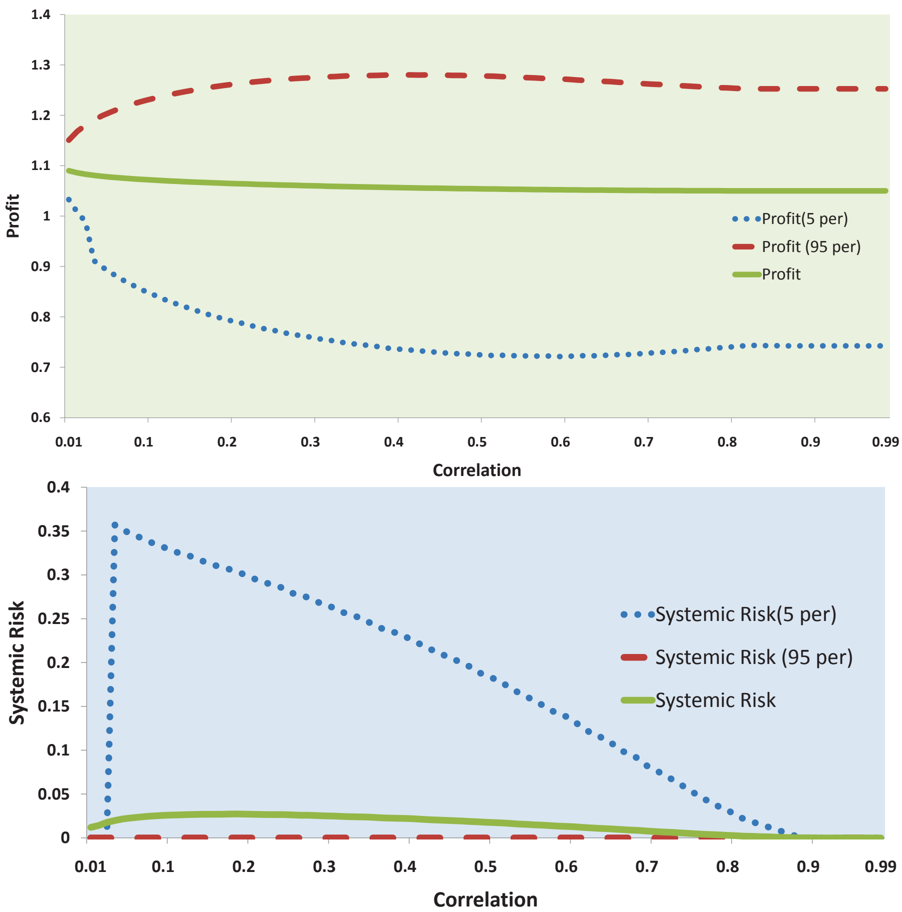


Figure 4. Dependence of profits and systemic risk on aggregate shock. The two panels of this figure indicate how profit and systemic risk is affected if "good" and "bad" aggregate shocks hit the economy. For a given value of ρ, "good" and "bad" shocks are defined as the 95th and 5th percentile realizations of the aggregate shock drawn from the distribution defined by ρ. As before, R is set to 1.5 and α is set to 0.4. τ is set to 0.5. Panel 1 (top left) shows how profits change if a "good" or "bad" shock is realized, while Panel 2 (top right) plots the corresponding change in systemic risk.

5.5. Regulation and the First-Best Outcome

The paper has so far assumed that banks take as given the liquidity choices of other banks and respond accordingly. In this section, we will consider the case where a regulator can mandate levels of margin and liquidity for each bank to carry. We will examine whether the solution to the regulators' problem differs from the equilibrium described in the previous section, and whether systemic risk is higher or lower as a result.

Regulator's Problem:

The regulator chooses margin (k^{FB}) and liquidity (ℓ^{FB}) to maximize bank profits. In particular, k^{FB} and ℓ^{FB} solve

$$(k^{FB}, \ell^{FB}) = \arg\max_{k,\ell} \ \mathbb{E}\Pi^{FB}(k, \ell)$$

where (as before $\underline{m} = \min(\underline{a}, a_0)$ and $m_1 = \min(a_0, a_1)$)

$$\begin{aligned}
\mathbb{E}\Pi^{FB}(k,\ell) \quad = & \\
& \int_{-\infty}^{a_0(k,\ell)} [R(1-k-\ell)+\ell](1-f(a)+g(k,\ell))\phi(a)da \\
& + \int_{-\infty}^{\underline{m}} [(R-\Delta-1)y(0,k,\ell)(1-f(a))+\ell(f(a)-g(k,\ell))]\phi(a)da \\
& + \int_{\underline{m}}^{m_1} [(R-\Delta-1-p)y(p,k,\ell)(1-f(a))+[p\times(1-k-\ell)+\ell](f(a)-g(k,\ell))]\phi(a)da \\
& + \int_{m_1}^{a_0(k,\ell)} [(R-\Delta-1)(1-k-\ell)+\ell)(f(a)-g(k,\ell)]\phi(a)da \\
& + \int_{a_0(k,\ell)}^{\infty} ([(\tau R(1-\eta(f(a)))+(1-\tau R))(1-f(a))+Rf(a)](1-k-\ell))\,\phi(a)da
\end{aligned}$$

The expression above for Π^{FB} is similar to the expression for profit in the previous section. However, while earlier, banks chose liquidity ℓ taking as given the liquidity chosen by other banks ($\bar{\ell}$), here, the regulator solves for the optimal (k^{FB}, ℓ^{FB}) imposing as a condition that banks carry the same liquidity and contribute the same margin.

The question we are interested in is whether and to what extent regulatory requirements can improve welfare and lower systemic risk. This leads us to the following result:

Proposition 5.

1. *For every value of correlation ρ, $k^{FB}(\rho) = k^*(\rho)$ and $\ell^{FB}(\rho) = \ell^*(k^*, \rho) = 0$*
2. *For every value of ρ, $\mathbb{E}\Pi^{FB}(\rho) = \mathbb{E}\Pi(\rho)$, and systemic risk is as large under the first-best outcome as under equilibrium.*

Proof. See Appendix E. □

We see that the optimal values of margin and liquidity chosen by the regulator coincide with those obtained in equilibrium. Put another way, the equilibrium obtained in the previous section is ex-ante efficient and cannot be improved upon. In both the equilibrium and in the first-best case, liquidity carried over by banks equals zero. Regulators find it optimal to make banks deposit any non-risky investment as margin in the clearinghouse rather than carry as liquidity. The intuition for this result is that a dollar deposited in the clearinghouse as margin can be used to refinance banks without any loss in efficiency, i.e., assets refinanced by the clearinghouse generate a return of R. In contrast, liquidity carried over by banks helps to purchase assets in the fire sale, but this only generates a return of $(R-\Delta-1)$, i.e., there is some efficiency loss to acquiring new assets. It is therefore ex-ante optimal for regulators to mandate higher margins but no "carried-over" liquidity.

As discussed earlier, it is also a property of equilibrium that at the optimal margin chosen by banks, liquidity carried over is zero. Thus, the first-best and equilibrium problems become identical if ℓ is zero, and optimal margins in both settings coincide. This in turn leads to identical profits and systemic risk in both cases.

6. Model with Lender of Last Resort

6.1. Setting and Payoffs

Let us assume that there is a government or a central bank that acts as a lender of last resort to the clearinghouse. There has been a precedent for a lender of last resort to bail out a clearinghouse. For example, after the 1987 stock market crash, the Federal Reserve extended credit to members of the Chicago Mercantile Exchange clearinghouse (CME).

Let the lender of last resort inject funds $g(a)$ into the clearinghouse, where a is the realization of the aggregate shock. The cost to the government is given by $c(g) = a_{gov}g^2$. This cost can be thought of as arising from an increased tax levied by the government on the economy to fund the bailout.

If each bank submits margin k and carries liquidity ℓ, then each bank needs $(1-k-\ell)$ to refinance. The total number of banks that get refinanced by the government is $g(a)/(1-k-\ell)$.

Each bank that gets refinanced by the government sees a benefit of Δ per unit of asset refinanced[9]. Therefore, the total benefit to each bank from the government is $\Delta(1-k-\ell)$.

Since there are $g(a)/(1-k-\ell)$ banks that need refinancing, the total benefit to the economy through the government injecting liquidity is $\Delta g(a)$.

The cost function to the government is convex, and the government only injects funds when the marginal benefit is greater than the marginal cost, i.e.,

$$g(a) < g^*, \text{ where } c'(g^*) = \Delta$$

Let us assume that if the government acts as a lender of last resort, it still requires the clearinghouse to make capital calls on its members and refinance as many banks as possible. In the earlier version of the model, the clearinghouse could bail out a maximum of $f_0(k,\ell)$ banks, where

$$f_0(k,\ell) = \frac{\tau R(1-k-\ell)+k+\ell}{(1+\tau R)(1-k-\ell)}$$

The government can act as a lender of last resort to a maximum of $g^*/(1-k-\ell)$ banks.

Therefore, there is a fire sale only when the number of failures $f(a) > f_g = f_0 + g^*/(1-k-\ell)$. In the case there is a fire sale, the clearinghouse uses margin k to bail out $k/(1-k-\ell)$ banks, while the government acts as a lender of last resort for $g^*/(1-k-\ell)$ banks.

As before, if there is a fire sale, there are failure thresholds $\underline{f}$ and f_1 such that if $f > \underline{f}$, the fire sale price is zero, while if $f < f_1$, the price is $(R-\Delta-1)$. The thresholds are given by

$$\underline{f} = \frac{\bar{\ell}+k+g^*+\tau R(1-k-\bar{\ell})}{(1-k-\bar{\ell})+(\bar{\ell}+\tau R(1-k-\bar{\ell}))}\ ;\ f_1 = \frac{\bar{\ell}+\tau R(1-k-\bar{\ell})+(R-\Delta)(k+g^*)}{\bar{\ell}+\tau R(1-k-\bar{\ell})+(R-\Delta)(1-k-\bar{\ell})}$$

Since the government acts as a lender of last resort, banks have to repay the government after getting refinanced. Therefore, if $f(a) > f_g$, there is a fire sale, the government injects g^*, and this is repaid to the government after some banks get refinanced.

If $f(a) \in (f_0, f_g]$, the government injects funds $g(a) = (1-k-\ell)\times(f(a)-f_0)$.

If $f(a) < f_0$, the clearinghouse remains solvent, and the government plays no role in the economy.

---

9 This is because this is the loss that results if the bank participates in a fire sale.

6.2. Expected Profits and Welfare

Let $a_0, a_1, \underline{a}$ and a_g be values of aggregate shock associated with failures $f_0, f_1, \underline{f}$ and f_g respectively. In addition, for notational ease, let $\underline{m} = \min(\underline{a}, a_g)$ and $m_1 = \min(a_g, a_1)$.

Let $g_1(k, g^*, \bar{\ell}) = (k + g^*)/(1 - k - \bar{\ell})$ be the number of banks either bailed out or refinanced by both the clearinghouse and the government if the clearinghouse fails.

Expected profits to banks in the presence of government acting as lender of last resort is given by

$$\begin{aligned} \mathbb{E}\Pi(k,\ell,\bar{\ell}) \quad = \quad & \\ & \int_{-\infty}^{a_g(k,\bar{\ell})} \left([R(1-k-\ell)+\ell](1-f(a)+g_1(k,g^*,\bar{\ell})) - g^*\right) \phi(a)da \\ & + \int_{-\infty}^{\underline{m}} [(R-\Delta-1)y(0,k,\ell)(1-f(a)) + \ell(f(a)-g_1(k,g^*,\bar{\ell}))]\phi(a)da \\ & + \int_{\underline{m}}^{m_1} [(R-\Delta-1-p)y(p,k,\ell)(1-f(a)) + [p\times(1-k-\ell)+\ell](f(a)-g_1(k,g^*,\bar{\ell}))]\phi(a)da \\ & + \int_{m_1}^{a_g(k,\bar{\ell})} [(R-\Delta-1)(1-k-\ell)+\ell)(f(a)-g_1(k,g^*,\bar{\ell})]\phi(a)da \\ & + \int_{a_g(k,\bar{\ell})}^{a_0(k,\bar{\ell})} [(1-\tau)R(1-f(a)) + Rf(a) - (f(a)-f_0)]\,(1-k-\bar{\ell})\phi(a)da \\ & + \int_{a_0(k,\bar{\ell})}^{\infty} \left([(\tau R(1-\eta(f(a))) + (1-\tau R))(1-f(a)) + Rf(a)](1-k-\ell)+\ell\right)\phi(a)da \end{aligned} \tag{5}$$

This expression is similar to our original expression for profit, except for the fact that more banks can get refinanced (g_1 has changed), and banks have to repay the government after being refinanced.

The first term is the expected profit of banks from their own assets (without participating in the fire sale) if a fire sale takes place. For each realization of the aggregate shock a, $(1 - f(a))$ banks remain solvent and do not need refinancing. g_1 insolvent banks do get refinanced using the margins present in the clearinghouse and injected government liquidity. Thus the number of banks that either do not need refinancing (and receive full return on their risky holdings) or do require refinancing and are refinanced is $(1 - f(a) + g_1(k, g^*, \bar{\ell}))$.

Finally, banks repay the government the injected amount g^* after getting refinanced, giving the first term of the expression above.

The second, third and fourth terms capture expected payoffs from participating in the fire sale for both acquiring and selling banks. In the second term, the fire sale price is 0, and so only acquiring banks realize a profit. In the third term, the fire sale price is some $p \in (0, R - \Delta - 1)$. Both acquiring banks and banks that sell in the fire sale realize a payoff. Acquiring banks get $(R - \Delta - 1 - p)$ per unit of asset bought in the fire sale, while selling banks get price p for the $(1 - k - \ell)$ assets sold in the fire sale. The fourth term corresponds to profits when the price in the fire sale is $(R - \Delta - 1)$. In this case, acquiring banks do not realize a profit in the fire sale.

The fifth term corresponds to expected profits when the clearinghouse has implemented maximum transfers ($\eta(f) = 1$) and the government still steps in to inject liquidity of $(f(a) - f_0)(1 - k - \ell)$. Finally, the last term corresponds to the case where the clearinghouse remains solvent and there is no role for the government.

Welfare is given by

$$W(k,\ell,\bar{\ell}) = \mathbb{E}\Pi(k,\ell,\bar{\ell}) - \Phi(a_g(k,\bar{\ell}))c(g^*) - \int_{a_g(k,\bar{\ell})}^{a_0(k,\bar{\ell})} a_{gov}\,[f(a)-f_0]^2\,(1-k-\bar{\ell})^2 \tag{6}$$

where there is a welfare loss of $c(g^*) = a_{gov}(g^*)^2$ if a fire sale happens (with probability $\Phi(a_g(k, \bar{\ell}))$) and the government uses g^* to bail out banks, while there is a welfare loss of $a_{gov}\,[f(a) - f_0]^2\,(1 - k - \bar{\ell})^2$ if $a \in [a_0, a_g]$, since in this region, the government has injected $g(a) = (1 - k - \ell) \times (f(a) - f_0)$.

As before, k_{pub} is chosen to maximize welfare. First best liquidity will still always be zero.

Let k_{pub} be the margin mandated by the government to maximize welfare, and let k_{pvt} be the optimal margin chosen by banks.

$$k_{pub} = \arg\max_k W(k, \bar{\ell}(k))\ ;\ k_{pvt} = \arg\max_k \mathbb{E}\Pi(k, \bar{\ell}(k))$$

The optimal margins chosen by the government and individual banks differ in that the government internalizes the costs of bank failures. This naturally makes the government more cautious than banks and we would expect the government to mandate higher margins than would be chosen privately. Systemic risk would correspondingly fall. This leads us to the following proposition:

Proposition 6.

$$k_{pub} > k_{pvt}\ ;\ S.risk_{pub} < S.risk_{pvt}$$

Proof. See Appendix F. □

The top left panel of Figure 5 shows how the margin changes with correlation for four separate cases—autarky, equilibrium in the absence of a government (acting as a lender of last resort), equilibrium in the presence of a government, and the first-best case.

The margin under autarky is fixed at 0.5. Banks opt to self-hedge and hold enough liquidity to refinance their project if the need arises. The rising solid line and (respt.) the dashed line plot optimal margins under the equilibrium without and (respt.) with the government. We note that margins are higher in the absence of the government. In the plots, the maximal amount of funds the government injects into the economy is $g^* = 0.5$. This means that margins in the presence of the government never exceeds $(1/2 - g^*) = 0.45$. Banks know that they will be bailed out if needed and choose to correspondingly carry less in margin. The difference in margins carried with and without the government is highest when correlation is low. Margin under the first-best case is shown as the dotted line in the figure. This is always higher than margin under the private equilibrium in the presence of government, illustrating the proposition stated earlier.

The top right panel of the figure shows systemic risk for the four different cases discussed earlier. Systemic risk under autarky is just zero, while systemic risk with and without the government acting as lender of last resort are almost identical. While margins are lower with the government, the injection of public funds into the economy can lower the incidence and expected losses resulting from fire sales. Systemic risk is much lower, however, in the first-best outcome, which reflects the fact that margins are higher. The bottom panel shows how welfare varies with correlation for each of the four outcomes. Welfare under autarky is fixed at $(1 + R - \alpha)/2$, and, by definition, is highest under the first-best outcome. Further, welfare in the equilibrium with the government is higher than without.

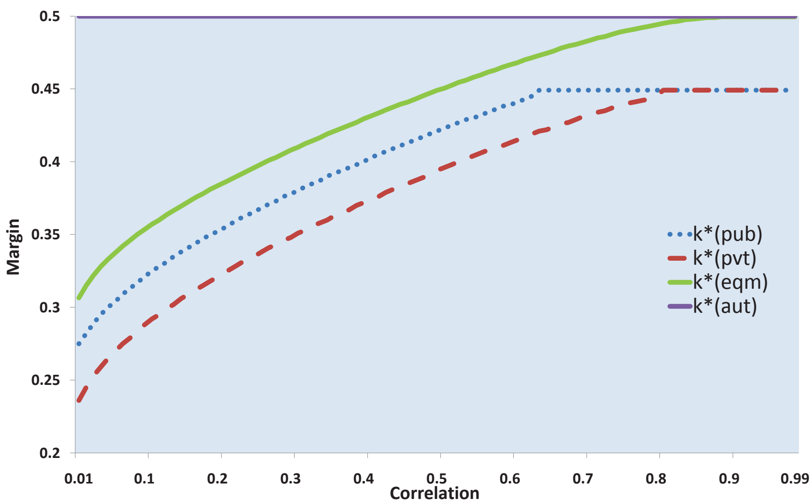
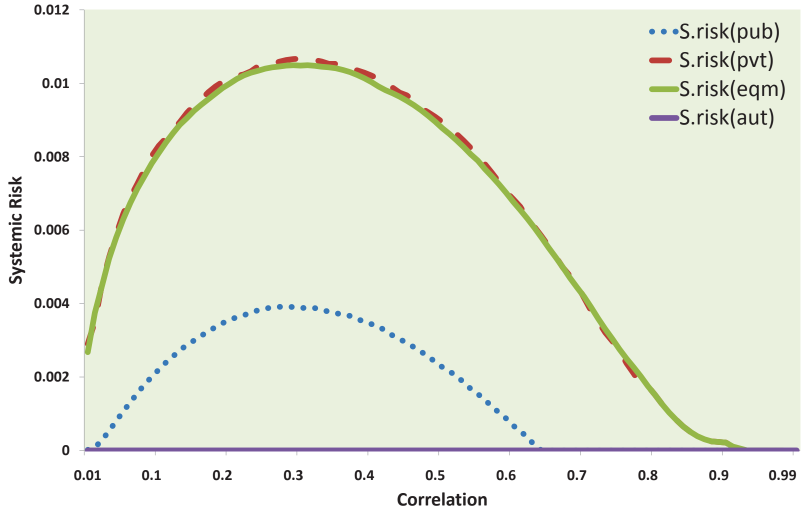


Figure 5. *Cont.*



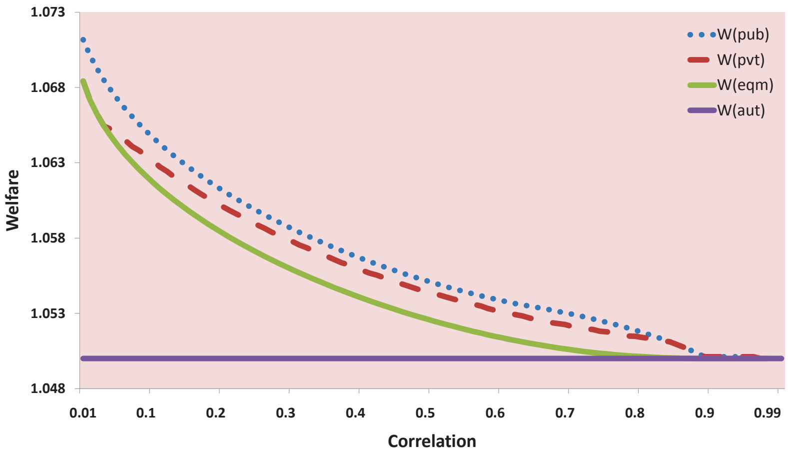


Figure 5. Outcomes with lender of last resort. The three panels of this figure plot outcomes under four different environments - autarky, equilibrium with and without the lender of last resort, and first-best. In all the figures, the return from the risky project is set to $R = 1.50$ and the value of α (the probability that refinancing of the risky project will be required) is set to 0.4. Panel 1 (top left) plots margin, k as correlation changes. Panel 2 (top right) plots systemic risk, while Panel 3 (bottom) plots the payoff to a generic bank as correlation changes.

7. Conclusions

This paper makes the fundamental point that whether aggregate shocks lead to systemic risk or collective failure of financial firms depends on the liquidity choices of firms but that these liquidity choices are intimately shaped by the risk-sharing opportunities in the economy. When risk-sharing opportunities are abundant, holding liquidity to hedge against low-risk aggregate shocks is privately costly resulting in excessive investment and too little reserves in the system for dealing with such shocks. This novel insight helps resolve the paradox that risk-sharing appears to have led to greater systemic risk and incidence of financial crises. Interestingly, however, risk-sharing can lead to first-best outcomes and is efficient, even though it leads to increased systemic risk. In the presence of a lender of last resort, however, this is no longer true, and banks choose lower liquidity than is optimal under first-best. Important future work that remains is to understand how such regulation would affect the price of liquidity and the return on investments, both of which we took as given in this paper.

Author Contributions: Data curation, V.V.A. and A.M.I.; Formal analysis, R.K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We are grateful to Stew Myers, participants at the Derivatives Research Project Day of the NYU Stern School of Business, participants at the "Challenges in Interconnected Financial Systems" at the Sveriges Riksbank, the Systemic Risk Center at LSE, and the Commodities Futures Trading Commission for their feedback.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Proposition 1

Pick a strategy $(\ell, 1-\ell)$. Since refinancing costs are a dollar per dollar of investment in the risky asset, the total refinancing need (should a refinancing need arise) will be $(1-\ell)$. Since only internal resources may be used to refinance, this means the bank can refinance the risky investment if and only if $\ell \geq (1-\ell)$, i.e., if and only if $\ell \geq 1/2$.

If the bank chooses $\ell < 1/2$, then refinancing is impossible, and investment in the risky project yields a return of R (per dollar investment) with probability $1-\alpha$ and 0 with probability α. Since the riskless project has a return of 1 (per dollar investment) with certainty, the expected payoff from the strategy $(\ell, 1-\ell)$ is

$$\ell + (1-\alpha)(1-\ell)R \;=\; \ell[1 - R + \alpha R] + (1-\alpha)R. \tag{A1}$$

On the other hand, if the bank chooses $\ell \geq 1/2$, then refinancing is always feasible. Since refinancing is optimal when it is feasible, and since refinancing costs of a dollar (per dollar investment) are incurred with probability α, the expected payoff per dollar of investment in the risky strategy is $R-\alpha$. So the payoff from the strategy $(\ell, 1-\ell)$, conditional on choosing $\ell \geq 1/2$, is

$$\ell + (1-\ell)R - \alpha(1-\ell) \;=\; \ell[1 - R + \alpha] + R - \alpha. \tag{A2}$$

To identify the optimal strategy for the bank, we compare the maximized values of the payoffs (A1) and (A2), i.e., (i) the maximal payoff conditional on $\ell < 1/2$, and (ii) the maximal payoff conditional on $\ell \geq 1/2$. Consider expression (A1) first. It is apparent from visual inspection that the optimal action is to choose $\ell^* = 0$ if $1 - R + \alpha R \leq 0$, and to choose ℓ^* as high as possible (subject to $\ell < 1/2$) if $1 - R + \alpha R > 0$. Thus, the maximum value of the payoff (or the supremum since there is no maximum in the latter case) is

$$\begin{cases} (1-\alpha)R, & \text{if } 1 - R + \alpha R \leq 0 \\ \dfrac{1}{2}[1 + R - \alpha R], & \text{if } 1 - R + \alpha R > 0 \end{cases}$$

Now consider the choice of ℓ that maximizes expression (A2) subject to $\ell \geq 1/2$. Again, simple visual inspection shows that the optimal action is the smallest possible value of ℓ (which is $\ell^* = 1/2$) if $1 - R + \alpha \leq 0$ and the largest possible value of ℓ (which is $\ell^* = 1$) if $1 - R + \alpha \geq 0$. Thus, the maximized value of the payoff is

$$\begin{cases} \dfrac{1}{2}[1 + R - \alpha], & \text{if } 1 - R + \alpha \leq 0 \\ 1, & \text{if } 1 - R + \alpha > 0 \end{cases}$$

Since $R > 1$, this gives this three cases to consider:

- **Case 1.** $0 \;\geq\; 1 - R + \alpha R \;>\; 1 - R + \alpha.$

In this case, the maximized payoff under $\ell < 1/2$ is $(1-\alpha)R$, which corresponds to a choice of $\ell^* = 0$, while the maximized payoff under $\ell \geq 1/2$ is

$$\frac{1}{2}\left[1 + R - \alpha\right] \tag{A3}$$

The former is greater if and only if $2(1-\alpha)R > 1 + R - \alpha$, i.e., if and only if

$$\alpha \; < \; \frac{R-1}{2R-1} \; = \; \lambda(R)$$

Thus, for $\alpha < \lambda(R)$, the optimal action is $\ell^* = 0$ leading to a payoff of $1 - \alpha)R$ if $\alpha < \lambda(R)$, while the optimal action is $\ell^* = 1/2$ leading to a payoff of (A3) if $\alpha \in [\lambda(R), (R-1)/R]$. This establishes Part 1 of the proposition.

- **Case 2.** $1 - R + \alpha R \; > \; 0 \; > \; 1 - R + \alpha$.

  Note that $1 - R + \alpha \;\; < \;\; 0 \;\; \Longleftrightarrow \;\; \alpha \; < \; \nu(R)$ as defined in the statement of Proposition 1. The supremum of the payoffs (there is no maximum) subject to $\ell < 1/2$ here is

$$\frac{1}{2}\left[1 + R - \alpha R\right], \tag{A4}$$

  while the maximized value of the payoffs subject to $\ell \geq 1/2$ (corresponding to the action $\ell^* = 1/2$) is the same as given by (A3). Since $R > 1$, the latter is always greater, so it remains optimal to choose $\ell^* = 1/2$ and thence to obtain the expected payoff (A3). This establishes Part 2 of the proposition.
- **Case 3.** $1 - R + \alpha R \; > \; 1 - R + \alpha \; \geq \; 0$.

  In this case, the supremum of the payoffs (there is no maximum) subject to $\ell < 1/2$ continues to be given by (A4), while the maximized payoff subject to $\ell \geq 1/2$, corresponding to the optimal action $\ell^* = 1$, is 1. The latter is greater as long as

$$1 + R - \alpha R \; < \; 2 \; \Longleftrightarrow \; R(1-\alpha) \; < \; 1$$

  However, this always holds since we are considering the case $1 - R + \alpha R > 0$. Thus, the optimal action in this case is to set $\ell^* = 1$ and to receive an expected payoff of 1. This establishes Part 3 and completes the proof of the proposition.

Appendix B. Proof of Proposition 2

Recall that when $\alpha \in [\lambda(R), \nu(R)]$ (the case we are considering), under autarky each bank invests an amount of $1/2$ in the riskless asset and $1/2$ in the risky project, and receives an ex-ante expected payoff of

$$\frac{1}{2}\left(1 + R - \alpha\right). \tag{A5}$$

Consider the following strategies for the two banks that corresponds to a "mutualization" of risk (equivalently, to each bank buying a specific contingent claim from the other bank):

1. Each bank contributes an amount $\ell^* = 1/3$ to a common insurance "pool." The pool is invested in the riskless asset.
2. Each bank invests its remaining funds $1 - \ell^* = 2/3$ in its risky project.
3. Either bank can claim the amount in the pool if its risky project requires refinancing. (Recall that the refinancing need is observable and verifiable, so false claims cannot be made.)
4. If more than one bank makes a claim at the same time, then the banks share the available money equally.

5. If neither bank experiences a refinancing need, the available money in the pool is equally divided between the two banks at time 2.

Since each bank invests $2/3$ in the risky project, and there is a total of $2/3$ invested in the riskless asset, there is exactly enough to refinance one project but not both. So if only one bank needs refinancing (which happens in States 2 and 3), the project can be refinanced, but if both banks need refinancing (State 4), refinancing is impossible and there is "systemic" failure. Note too that if either bank deviates unilaterally to a lower contribution to the pool, there will not be enough cash to refinance even one project, so the only possible deviation that could dominate the chosen allocation is autarky, a point we will return to shortly.

The state-wise payoff from the specified strategies for each bank is:

State	Payoff
1	$1/3 + 2R/3$
2	$2R/3$
3	$2R/3$
4	$1/3$

So the expected payoff to each bank from the described strategies is

$$\frac{1}{3}(p_1 + p_4) + \frac{2}{3}(1 - p_4)R, \tag{A6}$$

or, substituting for the probabilities p_i,

$$\frac{1}{3}[1 - 2(1-\rho)\alpha(1-\alpha)] + \frac{2}{3}(1-\alpha)(1+\alpha-\alpha\rho)R \tag{A7}$$

This risk-sharing payoff (A7) dominates the autarky payoff (A5) as long as

$$\left\{\frac{1}{3}[1 - 2(1-\rho)\alpha(1-\alpha)] + \frac{2}{3}(1-\alpha)(1+\alpha-\alpha\rho)R\right\} \geq \frac{1}{2}(1+R-\alpha)$$

A little bit of algebra shows that this condition is equivalent to the condition that

$$(R-1)[1 - 4\alpha(\alpha+\rho-\alpha\rho)] \geq \alpha \tag{A8}$$

When does (A8) hold? As $\rho \downarrow 0$, this reduces to the condition that

$$(R-1)(1-4\alpha^2) \geq \alpha \tag{A9}$$

Now, as α approaches its lower bound $\lambda(R) = (R-1)/(2R-1)$, the left-hand side of (A9) tends to

$$\left(\frac{R-1}{2R-1}\right)\left(\frac{4R-3}{2R-1}\right)$$

while the right-hand side approaches $(R-1)/(2R-1)$. Therefore, in the limit, the left-hand side is strictly larger as long as

$$\frac{4R-3}{2R-1} > 1,$$

which always holds since $R > 1$. Since the inequality holds strictly, it follows that there exists an interval $[\lambda(R), \alpha^*)$ such that for α in this interval, risk-sharing strictly dominates the autarky payoff.

It is easy to see from this that for any such α the proposed strategies constitute an equilibrium. From the dominance, neither party will prefer to revert to an autarkic equilibrium. Neither party will also wish to deviate to a lower (but positive) contribution to the pool because there will then not be enough resources in the pool to refinance even one project. Neither party will wish to deviate to a zero contribution since payoff from a zero contribution leads, at best, to the autarkic payoff which is dominated by the proposed risk-sharing strategies.

By continuity in ρ, it follows that for all suitably small ρ, there is an interval $[\lambda(R), \alpha^*)$ of possible values of α such that the risk-sharing equilibrium dominates autarky for all α in this interval. The upper-bound α^* depends on ρ (and indeed, declines in ρ, since it is easily checked that the payoff (A7) from risk-sharing is declining in ρ).[10]

At high correlations, however, the situation reverses itself, i.e., autarky dominates the payoff under risk-sharing:

$$\frac{1}{3}\,[1-2(1-\rho)\alpha(1-\alpha)] + \frac{2}{3}\,(1-\alpha)(1+\alpha-\alpha\rho)R \;<\; \frac{1}{2}\,(1+R-\alpha). \tag{A10}$$

To see (A10), note that as $\rho \uparrow 1$, the left-hand side tends to the limit

$$\frac{1}{3} + \frac{2}{3}\,(1-\alpha)R,$$

and this is strictly less than the right-hand side of (A10) if

$$2 + 4(1-\alpha)R \;<\; 3(1+R-\alpha)$$

i.e., if $\alpha \;>\; (R-1)/(4R-3)$. However, $\alpha > (R-1)/(2R-1)$ by hypothesis, and $(2R-1) < (4R-3)$, so certainly $\alpha > (R-1)/(4R-3)$, so the strict inequality (A10) holds in the limit. Once again, by continuity in ρ, it follows that for sufficiently high ρ, autarky dominates risk-sharing.

Finally, as observed earlier, the expected payoff from risk-sharing (A7) is strictly decreasing in ρ: differentiating this payoff with respect to ρ, we obtain

$$\frac{2}{3}\,\alpha(1-\alpha) - \frac{2}{3}\,\alpha(1-\alpha)R \;=\; \frac{2}{3}\,\alpha(1-\alpha)(1-R) \;<\; 0 \tag{A11}$$

It follows that there is a unique ρ^* such that (a) for $\rho > \rho^*$, autarky dominates risk-sharing for all $\alpha \geq \lambda(R)$, while (b) for $\rho < \rho^*$, there is an interval $[\lambda(R), \alpha^*)$ of possible values of α such that risk-sharing dominates autarky for all α in this interval. Since α^* declines in ρ, we can solve explicitly for ρ^* by taking α equal to $\lambda(R)$ and setting the two sides of (A8) equal:

$$(R-1)\left[1 - 4\,\frac{(R-1)^2}{(2R-1)^2} - 4\rho^*\,\frac{R-1}{2R-1}\left(1-\frac{R-1}{2R-1}\right)\right] \;=\; \frac{R-1}{2R-1}$$

or, rearranging and canceling common terms,

$$\left[1 - 4\,\frac{(R-1)^2}{(2R-1)^2} - 4\rho^*\,\frac{R(R-1)}{(2R-1)^2}\right] \;=\; \frac{1}{2R-1}$$

Multiplying through by $2R-1$,

$$2R - 1 - 4\,\frac{(R-1)^2}{2R-1} - 4\rho^*\,\frac{R(R-1)}{2R-1} \;=\; 1$$

So

$$2R - 2 \;=\; 4\,\frac{(R-1)^2}{2R-1} + 4\rho^*\,\frac{R(R-1)}{2R-1}$$

[10] How large can α^* be? It must clearly be strictly less than $1/2$, since (A9) clearly fails for $\alpha \geq 1/2$.

Multiplying through again by $(2R-1)$ and re-arranging, we have

$$2(R-1)(2R-1) - 4(R-1)^2 \;=\; 4\rho^* R(R-1)$$

The left-hand side of this expression is $2(R-1)[(2R-1) - 2(R-1)] \;=\; 2(R-1)$, so we obtain $2(R-1) \;=\; 4\rho^* R(R-1)$, which finally gives us

$$\rho^* \;=\; \frac{1}{2R} \tag{A12}$$

This completes the proof of Proposition 2. □

Note on Choice of φ *:*

A natural question to ask is whether the results in the above proposition continue to hold if $\varphi < 1$. It turns out that they do, and that there continues to exist an equilibrium in which risk sharing can lead to greater systemic risk. In particular, $\lambda(R,\varphi)$ and $\nu(R,\varphi)$ in the 1-bank model become $\varphi(R-1)/(R+\varphi(R-1))$ and $(R-1)/\varphi$, respectively.

In the 2-bank model, the optimal choice of liquidity carried in the case of risk-sharing, $\ell^* = \varphi/(2+\varphi)$. With these expressions in place, we can verify that the results from the previous proposition, namely:

1. For small enough ρ there always exists an α^* such that for $\alpha \in [\lambda(R,\varphi), \alpha^*]$, risk sharing dominates autarky, and
2. $\exists \rho^*$ such that for $\rho < \rho^*$, there is an equilibrium where risk-sharing dominates autarky

continue to hold.

Appendix C. Proof of Proposition 3

Let us define

$$T_1(a,k,\bar{\ell}) = -R(1 - f(a) + g(k,\bar{\ell}))$$

$$T_2(a,k,\bar{\ell}) = (R - \Delta - 1)(1 - \tau R)(1 - f(a))$$

$$T_3(a,k,\bar{\ell}) = \frac{(R-\Delta)(1-k-\bar{\ell})[f(a) - g(k,\bar{\ell})](1-\tau R)}{[\bar{\ell} + \tau R(1-k-\bar{\ell})]} - \frac{(1-k)}{(1-k-\bar{\ell})}(1 - f(a)) + f(a) - g(k,\bar{\ell})$$

$$T_4(a,k,\bar{\ell}) = -(R - \Delta - 1)(f(a) - g(k,\bar{\ell}))$$

$$T_{solv}(a,k,\bar{\ell}) = -[R - f(a) + \frac{k+\bar{\ell}}{1-k-\bar{\ell}}]$$

Then,

$$\begin{aligned}\alpha_1(k,\bar{\ell}) = &\int_{-\infty}^{a_0(k,\bar{\ell})} T_1(a,k,\bar{\ell})\phi(a)da + \int_{-\infty}^{\underline{m}} T_2(a,k,\bar{\ell})\phi(a)da + \int_{\underline{m}}^{m_1} T_3(a,k,\bar{\ell})\phi(a)da \\ &+ \int_{m_1}^{a_0(k,\bar{\ell})} T_4(a,k,\bar{\ell})\phi(a)da + \int_{a_0(k,\bar{\ell})}^{\infty} T_{solv}(a,k,\bar{\ell})\phi(a)da + 1\end{aligned}$$

The expressions for T_1, T_2, T_4 and T_{solv} are immediate. To see why the expression for T_3 must be true, note that $T_3(k,\bar{\ell})$ is the coefficient of ℓ in the expression $[(R-\Delta-1-p)y(p,k,\ell)(1-f(a)) + [p \times (1-k-\ell) + \ell](f(a) - g(k,\ell))]$ Now,

$$y(p,k,\ell) = \frac{\ell + \tau R(1-k-\ell)}{(1+p)}, \text{ and}$$

$$p(k,\bar{\ell}) = \max(0, -1 + [\bar{\ell} + \tau R(1-k-\bar{\ell})]\frac{(1-f(a))}{(1-k-\bar{\ell})(f(a) - g(k,\bar{\ell}))})$$

This means that the coefficient of ℓ in $(R-\Delta)y(p,k,\ell)(1-f(a))$ is $(R-\Delta)(1-f(a))(1-\tau R)/(1+p)$, which in turn equals $(R-\Delta)(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})](1-\tau R)/[\bar{\ell}+\tau R(1-k-\bar{\ell})]$

Furthermore, the coefficient of ℓ in $(1+p)y(p,k,\ell)(1-f(a))+p(k,\bar{\ell})(1-k-\ell)(f(a)-g(k,\bar{\ell}))$ equals $(1-\tau R)(1-f(a))+p(k,\bar{\ell})(f(a)-g(k,\bar{\ell}))$.

After doing some algebra, and substituting the value of $p(k,\bar{\ell})$, we find that this reduces to

$$(1-\tau R)(1-f(a))+p(k,\bar{\ell})(f(a)-g(k,\bar{\ell})) = (1-f(a))\frac{(1-k)}{(1-k-\bar{\ell})}+f(a)-g(k,\bar{\ell})$$

Now we are in a position to derive comparative statics on $\alpha_1(k,\bar{\ell})$.

Claim: $\partial\alpha_1(k,\bar{\ell})/\partial\bar{\ell} < 0$

Proof. First note that $\partial g(k,\bar{\ell})/\partial\bar{\ell} > 0$. This is immediate from the definition of $g(k,\bar{\ell}) = k/(1-k-\bar{\ell})$.

This means that $\partial T_1(a,k,\bar{\ell})/\partial\bar{\ell} < 0$ and that $\partial T_4(a,k,\bar{\ell})/\partial\bar{\ell} < 0$. Further, it is clear that $\partial T_{solv}(a,k,\bar{\ell})/\partial\bar{\ell} < 0$

We will show that $\partial T_3(a,k,\bar{\ell})/\partial\bar{\ell} < 0$. First, note that

$$\begin{aligned}
&\frac{d}{d\bar{\ell}}\frac{(R-\Delta)(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})](1-\tau R)}{[\bar{\ell}+\tau R(1-k-\bar{\ell})]} \\
&= \frac{d}{d\bar{\ell}}[\frac{f(a)-g(k,\bar{\ell})}{\bar{\ell}/(1-k-\bar{\ell})+\tau R}] = \frac{d}{d\bar{\ell}}[\frac{(1-k-\bar{\ell})f(a)-k}{\bar{\ell}+\tau R(1-k-\bar{\ell})}] \\
&= \frac{-\left[\bar{\ell}+\tau R(1-k-\bar{\ell})\right]f(a)-\left[(1-k-\bar{\ell})f(a)-k\right](1-\tau R)}{\left(\bar{\ell}+\tau R(1-k-\bar{\ell})\right)^2} \\
&= \frac{k(1-\tau R)-(1-k)f(a)}{\left(\bar{\ell}+\tau R(1-k-\bar{\ell})\right)^2}
\end{aligned}$$

Now, for $a \in [\underline{m}, \underline{m_1}]$, we have that $f(a) > f_1(k,\bar{\ell})$. Therefore,

$$f(a) > \frac{(\bar{\ell}+\tau R(1-k-\bar{\ell}))+k(R-\Delta)}{\bar{\ell}+\tau R(1-k-\bar{\ell})+(1-k-\bar{\ell})(R-\Delta)} > \frac{k}{(1-k-\bar{\ell})}$$

This implies that $k(1-\tau R)-(1-k)f(a) < 0$, and so

$$\frac{d}{d\bar{\ell}}\frac{(R-\Delta)(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})](1-\tau R)}{[\bar{\ell}+\tau R(1-k-\bar{\ell})]} < 0$$

Therefore,

$$\frac{\partial T_3(a,k,\bar{\ell})}{\partial\bar{\ell}} = \frac{d}{d\bar{\ell}}\frac{(R-\Delta)(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})](1-\tau R)}{[\bar{\ell}+\tau R(1-k-\bar{\ell})]} - \frac{d}{d\bar{\ell}}\frac{(1-k)(1-f(a))+k}{(1-k-\bar{\ell})} < 0$$

We have now shown that $\partial T_i/\partial\bar{\ell} < 0$ for $1 \le i \le 4$. We also know that $\partial T_{solv}/\partial\bar{\ell} < 0$.

In our representation earlier, $\alpha_1(k,\bar{\ell})$ has been written as the sum of a series of definite integrals. By the Leibniz rule, $\partial\alpha_1(k,\bar{\ell})/\partial\bar{\ell} = p_1(k,\bar{\ell}) + p_2(k,\bar{\ell})$ where $p_1(k,\bar{\ell})$ is the sum of the integrals of the derivatives of the integrands, while $p_2(k,\bar{\ell})$ is the derivative of the integral limit, multiplied by the value of the integrand at this limit.

Therefore, in this case,

$$p_2(k,\bar{\ell}) = \tfrac{\partial a_0}{\partial\bar{\ell}}[T_1(a_0,k,\bar{\ell}) + T_2(a_0,k,\bar{\ell})\mathbb{1}_{\underline{m}=a_0} + T_3(a_0,k,\bar{\ell})\mathbb{1}_{\underline{m}<a_0 and m_1=a_0} + T_4(a_0,k,\bar{\ell})\mathbb{1}_{\underline{m},m_1<a_0} - T_{solv}(a_0,k,\bar{\ell})]$$

Now

$$
\begin{aligned}
T_1(a,k,\bar{\ell}) + T_4(a,k,\bar{\ell}) - T_{solv}(a,k,\bar{\ell}) &= -R(1 - f(a) + g(k,\bar{\ell}) - (R-1-\Delta)(f(a) - g(k,\bar{\ell})) \\
&\quad + R - f(a) + g(k,\bar{\ell}) \\
&= \Delta(f(a) - g(k,\bar{\ell})) \\
&> 0
\end{aligned}
$$

In addition, recall that $T_3(a,k,\bar{\ell})$ can be written as

$$
\begin{aligned}
T_3(a,k,\bar{\ell}) &= \frac{(R-\Delta)(1-f(a))(1-\tau R)}{(1+p)} - (1-\tau R)(1-f(a)) - p(k,\bar{\ell})(f(a) - g(k,\bar{\ell})) \\
&> -p(k,\bar{\ell})(f(a) - g(k,\bar{\ell}) \\
&> -(R-\Delta-1)(f(a) - g(k,\bar{\ell}) \\
&= T_4(a,k,\bar{\ell})
\end{aligned}
$$

where both inequalities follow from the fact that fire sale price $p(k,\bar{\ell})$ is always less than or equal to $(R-\Delta-1)$.
By the same token, we can write

$$
\begin{aligned}
T_3(a,k,\bar{\ell}) &= \frac{(R-\Delta)(1-f(a))(1-\tau R)}{(1+p)} - (1-\tau R)(1-f(a)) - p(k,\bar{\ell})(f(a) - g(k,\bar{\ell})) \\
&< (R-\Delta)(1-f(a))(1-\tau R) - (1-\tau R)(1-f(a)) \\
&= (R-\Delta-1)(1-f(a))(1-\tau R) \\
&= T_2(a,k,\bar{\ell})
\end{aligned}
$$

where the first inequality follows since price $p(k,\bar{\ell}) \geq 0$

Putting it all together, we have that $T_4(a,k,\bar{\ell}) < T_3(a,k,\bar{\ell}) < T_2(a,k,\bar{\ell}$, and so

$$T_1(a,k,\bar{\ell}) + T_2(a,k,\bar{\ell}) - T_{solv}(a,k,\bar{\ell}) \geq 0$$

and

$$T_1(a,k,\bar{\ell}) + T_3(a,k,\bar{\ell}) - T_{solv}(a,k,\bar{\ell}) \geq 0$$

Finally, $\partial a_0(k,\bar{\ell})/\partial\bar{\ell} < 0$ implies that $p_2(k,\bar{\ell}) < 0$. We have already shown that $p_1(k,\bar{\ell}) < 0$, and so

$$\frac{\partial \alpha_1(k,\bar{\ell})}{\partial \bar{\ell}} < 0,$$

This shows that $\alpha_1(k,\bar{\ell})$ is a decreasing function. To prove uniqueness of equilibrium, all we have to do is show that there exists some limit $\tilde{\ell}$ at which $\alpha_1(k,\tilde{\ell} < 0$. This would mean that either there exists a unique $\ell^* \in (0,\tilde{\ell})$ for which $\alpha_1(k,\ell^*) = 0$, or it would mean that $\alpha_1(k,x) < 0$ for all $x < \tilde{\ell}$, in which case $\ell^* = 0$ is an equilibrium, as discussed earlier.

Let us choose $\tilde{\ell}$ such that $k/(1-k-\tilde{\ell}) = 1$. Then, the clearinghouse never fails. $a_0(k,\tilde{\ell}) = -\infty$, and it is easy to see that $\alpha_1(k,\tilde{\ell}) < 0$. □

Proof of part b. Here, we will show that if $(k,\bar{\ell})$ is an equilibrium, then $(k+\bar{\ell},0)$ is an equilibrium also. To prove this, we will show that $\mathcal{L}(\alpha_1(k,\bar{\ell}) < 0$, where $\mathcal{L}(f(a,b)) = \partial f/\partial a - \partial f/\partial b$. Note that proving this fact will imply the result since if $(k,\bar{\ell})$ is an equilibrium (and $\bar{\ell} > 0$), it means that $\alpha_1(k,\bar{\ell}) = 0$, and the result would imply that $\alpha_1(k+\bar{\ell},0) < 0$, which would imply $(k+\bar{\ell},0)$ is an equilibrium.

The outline of the proof is similar to before. First, we will prove that $\mathcal{L}(T_i) < 0$ for $1 \leq i \leq 4$, and also that $\mathcal{L}(T_{solv}) < 0$. We will then use Leibniz' rule to complete the proof.

Note that $\mathcal{L}(f(k,\bar{\ell})) = \partial f/\partial k - \partial f/\partial\bar{\ell} = \partial f/\partial k|_{k+\bar{\ell}=c}$ where c is a constant. This interpretation will help us in some of the steps to follow.

It is clear that $\mathcal{L}(g(k,\bar{\ell})) = \mathcal{L}(k/(1-k-\bar{\ell})) > 0$. This immediately implies that

$$\mathcal{L}(T_1) < 0\ ;\ \mathcal{L}(T_4) < 0\ ;\ \mathcal{L}(T_{solv}) < 0$$

Since $\mathcal{L}(T_2) = 0$, it only remains to prove that $\mathcal{L}(T_3) < 0$.

Recall that

$$T_3(a,k,\bar{\ell}) = \frac{(R-\Delta)(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})](1-\tau R)}{[\bar{\ell}+\tau R(1-k-\bar{\ell})]} - \frac{(1-k)}{(1-k-\bar{\ell})}(1-f(a)) + f(a) - g(k,\bar{\ell})$$

We have that

$$\mathcal{L}\left(-\frac{(1-k)}{(1-k-\bar{\ell})}(1-f(a)) + f(a) - g(k,\bar{\ell})\right) = -\mathcal{L}\left(\frac{(1-k)(1-f(a))+k}{1-k-\bar{\ell}}\right) < 0$$

In addition,

$$\begin{aligned}\mathcal{L}\left(\frac{(R-\Delta)(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})](1-\tau R)}{[\bar{\ell}+\tau R(1-k-\bar{\ell})]}\right) &= \lambda\mathcal{L}\left(\frac{[f(a)-g(k,\bar{\ell})]}{[\bar{\ell}+\tau R(1-k-\bar{\ell})]}\right)\\ &= \lambda\mathcal{L}\left(\frac{f(a)(1-k-\bar{\ell})-k}{\bar{\ell}+\tau R(1-k-\bar{\ell})}\right) < 0\end{aligned}$$

where $\lambda = (R-\Delta)(1-k-\bar{\ell})(1-\tau R) > 0$ Therefore, $\mathcal{L}(T_3) < 0$.

This implies that $\mathcal{L}(T_i) < 0$ for $1 \le i \le 4$, and $\mathcal{L}(T_{solv}) < 0$.

Furthermore,

$$\mathcal{L}(f_0) = \mathcal{L}\left(\frac{R(1-k-\bar{\ell})+k}{(R+1)(1-k-\bar{\ell})}\right) > 0 \implies \mathcal{L}\left(a_0(k,\bar{\ell})\right) < 0$$

Therefore,

$$\mathcal{L}(a_0)[T_1(a_0,k,\bar{\ell}) + T_2(a_0,k,\bar{\ell})\mathbb{1}_{\underline{m}=a_0} + T_3(a_0,k,\bar{\ell})\mathbb{1}_{\underline{m}<a_0 and m_1=a_0} + T_4(a_0,k,\bar{\ell})\mathbb{1}_{\underline{m},m_1<a_0} - T_{solv}(a_0,k,\bar{\ell})] < 0$$

Together, this result, along with the fact that $\mathcal{L}(T_i) < 0$ for $1 \le i \le 4$, and $\mathcal{L}(T_{solv}) < 0$ means we can prove the result by Leibniz's Rule, i.e., $\mathcal{L}(\alpha_1(k,\bar{\ell})) < 0$. $\square$

Proof of part c. Let us denote $\mathbb{E}\Pi(a,b,c)$ to be the expected profit for a given bank that carries liquidity b when the margin is a and all other banks carry liquidity c.

Claim: Let k and $\bar{\ell}$ be such that both $(k,\bar{\ell})$ and $(k+\bar{\ell},0)$ are possible equilibria in the context of the definition in Section 4. Then,

$$\mathbb{E}\Pi(k,\ell,\bar{\ell}) < \mathbb{E}\Pi(k+\bar{\ell},0,0)$$

i.e., it is always optimal for banks to not carry over any liquidity and only risk-share in the form of the margin. $\square$

Proof. Recall from Section 3 that

$$\begin{aligned}\mathbb{E}\Pi(k,\ell,\bar{\ell}) \quad = & \\ & \textstyle\int_{-\infty}^{a_0(k,\bar{\ell})}[R(1-k-\ell)+\ell](1-f(a)+g(k,\bar{\ell}))\phi(a)da \\ & + \textstyle\int_{-\infty}^{\min(\underline{a},a_0)}[(R-\Delta-1)y(0,k,\ell)(1-f(a))+\ell(f(a)-g(k,\bar{\ell}))]\phi(a)da \\ & + \textstyle\int_{\min(\underline{a},a_0)}^{\min(a_0,a_1)}[(R-\Delta-1-p)y(p,k,\ell)(1-f(a))+[p(1-k-\ell)+\ell](f(a)-g)]\phi(a)da \\ & + \textstyle\int_{\min(a_0,a_1)}^{a_0(k,\bar{\ell})}[(R-\Delta-1)(1-k-\ell)+\ell)(f(a)-g(k,\bar{\ell})]\phi(a)da \\ & + \textstyle\int_{a_0(k,\bar{\ell})}^{\infty}([R(1-k-\ell)(1-\eta(f(a)))(1-f(a))+R(1-k-\ell)f(a)])\,\phi(a)da\end{aligned}$$

Now, let us assume that $(k,\bar{\ell})$ is an equilibrium. Then, by definition, $\mathbb{E}\Pi(k,\ell,\bar{\ell}) = \mathbb{E}\Pi(k,\bar{\ell},\bar{\ell})$ We have that

$$\begin{aligned}&(R-\Delta-1-p(k,\bar{\ell}))y(p,k,\bar{\ell})(1-f(a))+p(k,\bar{\ell})(1-k-\bar{\ell})(f(a)-g(k,\bar{\ell})) \\ &\quad = (R-\Delta)\frac{[f(a)-g(k,\bar{\ell})]}{1-f(a)}(1-k-\bar{\ell})(1-f(a))-(1+p(k,\bar{\ell}))y(p,k,\bar{\ell})(1-f(a)) \\ &\qquad + (1+p(k,\bar{\ell}))(1-k-\bar{\ell})(f(a)-g(k,\bar{\ell}))-(1-k-\bar{\ell})(f(a)-g(k,\bar{\ell})) \\ &= (R-\Delta)[f(a)-g(k,\bar{\ell})](1-k-\bar{\ell})-(1-f(a))[\bar{\ell}+\tau R(1-k-\bar{\ell})]+(1-f(a))[\bar{\ell}+\tau R(1-k-\bar{\ell})] \\ &\qquad -(1-k-\bar{\ell})[f(a)-g(k,\bar{\ell})] \\ &\quad = (R-\Delta-1)[f(a)-g(k,\bar{\ell})](1-k-\bar{\ell}) = (R-\Delta-1)f(a)(1-k-\bar{\ell})-k(R-\Delta-1)\end{aligned}$$

where the third step follows from the fact that $(1+p(k,\bar{\ell}))y(p,k,\bar{\ell}) = \bar{\ell}+\tau R(1-k-\bar{\ell})$. Plugging this in to the expression for profit, we get that

$$\begin{aligned}\mathbb{E}\Pi(k,\bar{\ell},\bar{\ell}) \quad = & \\ & \int_{-\infty}^{a_0(k,\bar{\ell})}[R(1-k-\bar{\ell})+\bar{\ell}](1-f(a)+g(k,\bar{\ell}))\phi(a)da \\ & + \int_{-\infty}^{\min(\underline{a},a_0)}((R-\Delta-1)[\bar{\ell}+\tau R(1-k-\bar{\ell})](1-f(a))+\bar{\ell}[f(a)-g(k,\bar{\ell})])\phi(a)da \\ & + \int_{\min(\underline{a},a_0)}^{\min(a_0,a_1)}((R-\Delta-1)[f(a)-g(k,\bar{\ell})](1-k-\bar{\ell})+\bar{\ell}[f(a)-g(k,\bar{\ell})])\phi(a)da \\ & + \int_{\min(a_0,a_1)}^{a_0(k,\bar{\ell})}[(R-\Delta-1)(1-k-\bar{\ell})+\bar{\ell})(f(a)-g(k,\bar{\ell})]\phi(a)da \\ & + \int_{a_0(k,\bar{\ell})}^{\infty}\left(R(1-k-\bar{\ell})[1-\frac{f(a)}{R}+\frac{k}{R(1-k-\bar{\ell})}]\right)\phi(a)da\end{aligned}$$

From now on, let us denote $\mathbb{E}\Pi(k,\bar{\ell},\bar{\ell})$ by $\mathbb{E}\Pi(k,\bar{\ell})$.

To prove the proposition, let us first calculate $\partial\mathbb{E}\Pi(k,\bar{\ell})/\partial k|_{k+\bar{\ell}=c}$. This is the derivative of the profit function with respect to k, keeping $k+\bar{\ell}$ fixed (and equal to c. Put another way, this equals $\partial\mathbb{E}\Pi(k,\bar{\ell})/\partial k - \partial\mathbb{E}\Pi(k,\bar{\ell})/\partial\bar{\ell}$.

Now, $\partial\mathbb{E}\Pi(k,\bar{\ell})/\partial k|_{k+\bar{\ell}=c}$ If this is positive, then clearly $\mathbb{E}\Pi(k+\bar{\ell},0) > \mathbb{E}\Pi(k,\bar{\ell})$. Note that $\mathbb{E}\Pi(k,\bar{\ell})$ is a sum of different integrals, where the limits of integration themselves depend on k and $\bar{\ell}$. Then, by the Leibniz Rule, we have that $\partial\mathbb{E}\Pi(k,\bar{\ell})/\partial k|_{k+\bar{\ell}=c} = k_1+k_2$, where k_1 is the integral of the derivative of the functions inside the integrals, keeping the limits of integration fixed. k_2 is the derivative of the integral limits, multiplied by the function value at those limits.

Noting that $g(k,\bar{\ell}) = k/(1-k-\bar{\ell})$, we have that

$$k_1 = \int_{-\infty}^{a_0} R\phi(a)da + \int_{-\infty}^{\min(\underline{a},a_0)} -(R-\Delta-1)(1-f(a))\phi(a)da + \int_{\min(\underline{a},a_0)}^{\min(a_0,a_1)} -(R-\Delta-1)\phi(a)da + \int_{\min(a_0,a_1)}^{a_0} -(R-\Delta-1)\phi(a)da - \int_{-\infty}^{a_0}\phi(a)da$$

Simplifying, this implies that

$$k_1 = \int_{-\infty}^{a_0}(R-1)\phi(a)da - (R-\Delta-1)\int_{-\infty}^{a_0}\phi(a)da + (R-\Delta-1)\int_{-\infty}^{\min(\underline{a},a_0)} f(a)\phi(a)da = \Delta\int_{-\infty}^{a_0}\phi(a)da + (R-\Delta-1)\int_{-\infty}^{\min(\underline{a},a_0)} f(a)\phi(a)da > 0$$

Now, let us assume that $(R-\Delta) > 1$, and let us define the operator $\mathcal{L}(f(k,\ell)) = \partial f/\partial k - \partial k/\partial \ell$.

In addition, let $m_0 = \min(\underline{a}, a_0)$ and $m_1 = \min(a_0, a_1)$.

Then, by the Leibniz Rule, k_2 is given by

$$\begin{aligned}
k_2 = {} & \mathcal{L}(a_0)[R(1-k-\bar{\ell})+\bar{\ell}](1-f(a_0)+g(k,\bar{\ell}))\phi(a_0) + \mathcal{L}(m_0)(R-\Delta-1)[\bar{\ell}+\tau R(1-k-\bar{\ell})](1-f(m_0))\phi(m_0) \\
& +(R-\Delta-1)\mathcal{L}(m_1)\left([f(m_1)-g(k,\bar{\ell})](1-k-\bar{\ell})+\bar{\ell}[f(m_1)-g(k,\bar{\ell})]\right)\phi(m_1) \\
& -(R-\Delta-1)\mathcal{L}(m_0)\left([f(m_0)-g(k,\bar{\ell})](1-k-\bar{\ell})+\bar{\ell}[f(m_0)-g(k,\bar{\ell})]\right)\phi(m_0) \\
& +[(R-\Delta-1)(1-k-\bar{\ell})+\bar{\ell}][\mathcal{L}(a_0)(f(a_0)-g(k,\bar{\ell}))\phi(a_0) - \mathcal{L}(m_1)(f(m_1)-g(k,\bar{\ell}))\phi(m_1)] \\
& -\mathcal{L}(a_0)\left(R(1-k-\bar{\ell})[1-\tfrac{f(a_0)}{R}+\tfrac{k}{R(1-k-\bar{\ell})}]+\bar{\ell}\right)\phi(a_0)
\end{aligned}$$

After some algebra, this becomes

$$\begin{aligned}
k_2 = {} & \mathcal{L}(a_0)[R(1-k-\bar{\ell})+\bar{\ell}](1-f(a_0)+g(k,\bar{\ell}))\phi(a_0) + \mathcal{L}(m_0)(R-\Delta-1)[\bar{\ell}+\tau R(1-k-\bar{\ell})](1-f(m_0))\phi(m_0) \\
& +[(R-\Delta-1)(1-k-\bar{\ell})+\bar{\ell}][\mathcal{L}(a_0)(f(a_0)-g(k,\bar{\ell}))\phi(a_0) - \mathcal{L}(m_0)(f(m_0)-g(k,\bar{\ell}))\phi(m_1)] \\
& -\mathcal{L}(a_0)\left(R(1-k-\bar{\ell})[1-\tfrac{f(a_0)}{R}+\tfrac{k}{R(1-k-\bar{\ell})}]+\bar{\ell}\right)\phi(a_0)
\end{aligned}$$

and then

$$\begin{aligned}
k_2 = {} & \mathcal{L}(a_0)([R(1-k-\bar{\ell})+\bar{\ell}](1-f(a_0)+g(k,\bar{\ell})) + [(R-\Delta-1)(1-k-\bar{\ell})+\bar{\ell}](f(a_0)-g(k,\bar{\ell})) \\
& - R(1-k-\bar{\ell})[1-\frac{f(a_0)}{R}+\frac{k}{R(1-k-\bar{\ell})}]+\bar{\ell})\phi(a_0) \\
& + \mathcal{L}(m_0)\left((R-\Delta-1)[\bar{\ell}+\tau R(1-k-\bar{\ell})](1-f(m_0)) - (1-k-\bar{\ell})(f(m_0)-g(k,\bar{\ell}))]\right)\phi(m_1)
\end{aligned} \quad \text{(A13)}$$

□

Lemma A1. $\mathcal{L}(m_0)$; $\mathcal{L}(a_0) < 0$

Proof. First, recall that $\underline{f}$, f_0 and f_1 are defined such that $f > \underline{f}$ whenever $a < \underline{a}$ and so on. Then, from earlier work,

$$f_1(k,\ell) = \frac{(\bar{\ell}+\tau R(1-k-\ell)) + k(R-\Delta)}{\bar{\ell}+\tau R(1-k-\bar{\ell}) + (1-k-\bar{\ell})(R-\Delta)} \; ; \; \underline{f}(k,\ell) = \frac{k+\bar{\ell}+\tau R(1-k-\bar{\ell}))}{(1-k-\bar{\ell}) + (\bar{\ell}+\tau R(1-k-\bar{\ell}))}$$

$$f_0(k,\ell) = \frac{k+R(1-k-\bar{\ell})}{(1+R)(1-k-\bar{\ell})}$$

Then,

$$\mathcal{L}(\underline{f}) = \frac{\underline{f}}{(1-k)+\tau R(1-k-\bar{\ell})} \; ; \; \mathcal{L}(f_1) = \frac{f_1 + (R-\Delta-1)}{\bar{\ell}+\tau R(1-k-\bar{\ell}) + (1-k-\bar{\ell})(R-\Delta)}$$

$$\mathcal{L}(f_0) = \frac{1}{(1+R)(1-k-\bar{\ell})}$$

In general, for any aggregate shock a, the number of failures $f(a)$ is given by $f(a) = \Phi((c - \sqrt{\rho}a)/\sqrt{1-\rho})$. Therefore,

$$\mathcal{L}(a) = -\frac{\sqrt{1-\rho}}{\sqrt{\rho}} \times \frac{1}{\phi(N^{-1}f(a))} \times \mathcal{L}(f)$$

and the result follows since $\mathcal{L}(f) > 0$ for $f = \underline{f}, f_0$, and $f(a_0) > R/(R+1) > 1/2$.

Now, note that for any $k, \bar{\ell}$,

$$\begin{aligned}[R(1-k-\bar{\ell})+\bar{\ell})(1-f(a_0)+g(k,\bar{\ell}))+(R-\Delta-1)(1-k-\bar{\ell})(f(a_0)-g(k,\bar{\ell}))+\bar{\ell}(f(a_0)-g(k,\bar{\ell}))\\ = R(1-k-\bar{\ell})f(a_0)+\bar{\ell}\end{aligned}$$

In addition, $f(a_0)$ satisfies $f(a_0) = [1 - f(a_0)/R + k/R(1-k-\bar{\ell})]$

Therefore, Equation (1) implies that

$$k_2 \;=\; \mathcal{L}(m_0)\left((R-\Delta-1)[\bar{\ell}+\tau R(1-k-\bar{\ell})](1-f(m_0)) - (1-k-\bar{\ell})(f(m_0)-g(k,\bar{\ell}))]\right)\phi(m_1)$$

Since we know that $\mathcal{L}(m_0) < 0$ and $\phi(m_0) > 0$, we just have to show that

$$[\bar{\ell}+\tau R(1-k-\bar{\ell})](1-f(m_0)) \leq (1-k-\bar{\ell})(f(m_0)-g(k,\bar{\ell}))$$

Since $g(k,\bar{\ell}) = k/(1-k-\bar{\ell})$, and doing some algebra, we must show that

$$f(m_0) \geq \frac{k+\bar{\ell}+\tau R(1-k-\bar{\ell})}{1-k+\tau R(1-k-\bar{\ell})} \;\; \textit{i.e.,} \;\; f(m_0) \geq \underline{f}$$

However, recall that $m_0 = \min(a_0, \underline{a})$. This means that $m_0 \leq \underline{a}$, and so $f(m_0) \geq \underline{f}$. □

Appendix D. Proof of Proposition 4

We first derive the expected profit for banks if there is no clearinghouse, and banks carry liquidity ℓ. If $\ell \geq 1/2$, banks self hedge and refinance themselves if required. There is no fire sale. If all banks carry liquidity ℓ (determined in equilibrium), then a fire sale might well occur. Let us restrict our attention to the case where all banks carry liquidity $\bar{\ell} < 1/2$, and derive the profit for a bank carrying liquidity $\ell < 1/2$. Then,

$$\begin{aligned}\mathbb{E}\Pi(\ell,\bar{\ell}) \;=\; & \\ & \int_{-\infty}^{\infty}[R(1-\ell)+\ell](1-f(a))\phi(a)da \\ & + \int_{-\infty}^{\underline{a}}[(R-\Delta-1)y(0,\ell)(1-f(a))+\ell(f(a))]\phi(a)da \\ & + \int_{\underline{a}}^{a_1}[(R-\Delta-1-p\bar{\ell}))y(p,\ell)(1-f(a))+[p(\bar{\ell})\times(1-\ell)+\ell]f(a)]\phi(a)da \\ & + \int_{a_1}^{\infty}[((R-\Delta-1)(1-\ell)+\ell)f(a)]\phi(a)da\end{aligned}$$

If $\bar{\ell}$ is an equilibrium, then for any ℓ, $\mathbb{E}\Pi(\ell,\bar{\ell}) = \mathbb{E}\Pi(\bar{\ell},\bar{\ell})$. We can substitute $\ell = \bar{\ell}$ in the equation above to get

$$
\begin{aligned}
\mathbb{E}\Pi(\bar{\ell},\bar{\ell}) \quad = & \\
& \int_{-\infty}^{\infty}[R(1-\bar{\ell})+\bar{\ell}](1-f(a))\phi(a)da \\
& +\int_{-\infty}^{\underline{a}}((R-\Delta-1)[\bar{\ell}+\tau R(1-\bar{\ell})](1-f(a))+\bar{\ell}f(a))\phi(a)da \\
& +\int_{\underline{a}}^{a_1}((R-\Delta-1)f(a)(1-\bar{\ell})+\bar{\ell}f(a))\phi(a)da \\
& +\int_{a_1}^{\infty}[(R-\Delta-1)(1-\bar{\ell})+\bar{\ell}]f(a)\phi(a)da
\end{aligned}
$$

The integrand in the third integral is lower than the one in the fourth integral since $f(a) = 1$. In addition, the integrand in the second integral is a special case of the one in the third integral with $p(\bar{\ell})$ set to zero. Therefore, the integrands in the second and third integrals in the expression above are both lower than the integrand in the fourth integral.

So, we have that

$$
\begin{aligned}
\mathbb{E}\Pi(\bar{\ell},\bar{\ell}) \quad \leq \quad & \int_{-\infty}^{\infty}[R(1-\bar{\ell})+\bar{\ell}](1-f(a))\phi(a)da \\
& +\int_{-\infty}^{\infty}[(R-\Delta-1)(1-\bar{\ell})+\bar{\ell})]f(a)\phi(a)da
\end{aligned}
$$

Simplifying the above, we get that

$$\mathbb{E}\Pi(\bar{\ell},\bar{\ell}) \leq \bar{\ell}+(1-\bar{\ell})(R-\alpha) < (1+R-\alpha)/2$$

where we used the fact that $\int_{-\infty}^{\infty} f(a)da = \alpha$.

This completes the proof of the proposition.

Appendix E. Proof of Proposition 5

The proof is immediate using the results already obtained in earlier proofs. In particular, note that

$$\mathbb{E}\Pi^{FB}(k,\bar{\ell}) = \mathbb{E}\Pi(k,\bar{\ell},\bar{\ell})$$

We have already shown that $\mathbb{E}\Pi(k,\bar{\ell},\bar{\ell}) < \mathbb{E}\Pi(k+\bar{\ell},0,0)$. Therefore, it becomes clear that $\ell^{FB} = 0$.

However, conditioning on $\ell^{FB} = 0$, the first-best problem becomes formally identical to the equilibrium problem when $\ell = \bar{\ell} = 0$, which we know to be true at the optimal equilibrium k^*. Therefore, $k^{FB} = k^*$.

Since margins and liquidity are identical under the first-best measure as under equilibrium, there is similar equivalence of profits and systemic risk.

This completes the proof of the proposition.

Appendix F. Proof of Proposition 6

Let $\mathcal{C}(k,\bar{\ell}) = \Phi(a_g(k,\bar{\ell}))c(g^*) + \int_{a_g(k,\bar{\ell})}^{a_0(k,\bar{\ell})} a_{gov}\,[f(a)-f_0]^2\,(1-k-\bar{\ell})^2$ Then,

$$\frac{d\mathcal{C}(k,\bar{\ell})}{dk} = c(g^*)\phi(a_g)\frac{\partial a_g}{\partial k} + 2\int_{a_g}^{a_0}(1-k-\bar{\ell})a_{gov}[f(a_g)-f_0]^2 + \frac{\partial a_g}{\partial k}a_{gov}[f(a_g)-f_0]^2(1-k-\bar{\ell})^2 < 0$$

Therefore, $dW(k,\bar{\ell}(k))/dk = d\mathbb{E}\Pi(k,\bar{\ell}(k))/dk - d\mathcal{C}(k,\bar{\ell}(k))/dk > d\mathbb{E}\Pi(k,\bar{\ell}(k))/dk$

In addition, as $k \to 1/2$, $a_g \to a_0$ and $\mathcal{C}(k, \bar{\ell}(k)) \to 0$.

Now, let k^* solve $k^* = \arg\max \mathbb{E}\Pi(k, \bar{\ell}(k))$. Then, for $k < k^*$, $\mathbb{E}\Pi(k, \bar{\ell}(k) < \mathbb{E}\Pi(k^*, \bar{\ell}(k^*)$, and so, since $\mathcal{C}(k, \bar{\ell}(k)) > 0$, and $d\mathcal{C}(k, \bar{\ell}(k))/dk < 0$, we must have that $W(k, \bar{\ell}(k) < W(k^*, \bar{\ell}(k^*)$. Since, by definition, $k_{pvt} = k^*$, this means that $k_{pub} \geq k_{pvt}$, completing the proof.

Appendix G. Note on Asymmetric Equilibria

Let us now relax our assumption that all banks carry the same liquidity $\bar{\ell}$ and instead assume that there is an asymmetric equilibrium where a fraction w_1 of the banks carry over liquidity $\bar{\ell}_1$, while a fraction $w_2 = (1 - w_1)$ of the banks carry over liquidity $\bar{\ell}_2$.

As before, a fraction τ of future payoffs can be pledged by banks in the fire sale, so the fire sale demand function for a bank with liquidity ℓ is given by

$$y(p, k, \ell) = \frac{\ell + \tau R(1 - k - \ell) - y(p, k, \ell)}{p}$$

where p is the price of assets in the fire sale.

For convenience, let us refer to banks that carry liquidity $\bar{\ell}_1$ as "Type 1 banks", and banks that carry liquidity $\bar{\ell}_2$ as "Type 2 banks". Let $\bar{y}_1$ and $\bar{y}_2$ be the total demand submitted by Type 1 banks and Type 2 banks respectively. Then,

$$\bar{y}_1 = \frac{\bar{\ell}_1 + \tau R(1 - k - \bar{\ell}_1) - \bar{y}_1}{p} \; ; \; \bar{y}_2 = \frac{\bar{\ell}_2 + \tau R(1 - k - \bar{\ell}_2) - \bar{y}_2}{p} \tag{A14}$$

As before, let g be the total number of banks bailed out by the clearinghouse (with margin k) if it is insolvent. The clearinghouse bails out the failed banks with its margin, using a total amount of $w_1 k$ to bail out Type 1 banks, and $w_2 k$ to bail out Type 2 banks. Then the total number of banks bailed out by the clearinghouse is given by

$$g(k, w_1, w_2, \bar{\ell}_1, \bar{\ell}_2) = \frac{w_1 k}{(1 - k - \bar{\ell}_1)} + \frac{w_2 k}{(1 - k - \bar{\ell}_2)}$$

Let $g_1 = w_1 k/(1 - k - \bar{\ell}_1)$ and $g_2 = w_2 k/(1 - k - \bar{\ell}_2)$ denote the number of banks of Type 1 and Type 2 bailed out.

Then, the supply of assets in the fire sale is given by

$$S = (1 - k - \bar{\ell}_1)\,[w_1 f(a) - g_1] + (1 - k - \bar{\ell}_2)\,[w_2 f(a) - g_2]$$

The demand of assets in the fire sale is given by

$$D = w_1(1 - f(a))\bar{y}_1 + w_2(1 - f(a))\bar{y}_2$$

where $w_1(1 - f(a))$ is the number of banks that survived that are of Type 1, while $w_2(1 - f(a))$ is the number of banks that survived that are of Type 2.

Therefore, the market clearing condition is

$$w_1\bar{y}_1 + w_2\bar{y}_2 = \frac{(1 - k - \bar{\ell}_1)[w_1 f(a) - g_1] + (1 - k - \bar{\ell}_2)[w_2 f(a) - g_2]}{1 - f(a)} \tag{A15}$$

Further, Equation (A14) implies that

$$p(k, w_1, w_2, \bar{\ell}_1, \bar{\ell}_2) = -1 + \frac{\bar{\ell}_1 + \tau R(1 - k - \bar{\ell}_1)}{\bar{y}_1} = -1 + \frac{\bar{\ell}_2 + \tau R(1 - k - \bar{\ell}_2)}{\bar{y}_2} \tag{A16}$$

Equation (A15) implies that

$$\bar{y}_1 w_1(1-f) + \bar{y}_2 w_2(1-f) - f(1-k) + w_1 f \bar{\ell}_1 + w_2 f \bar{\ell}_2 + k = 0$$

Together, Equations (A15) and (A16) imply values for $\bar{y}_1$ and $\bar{y}_2$.

$$\bar{y}_2 = \frac{f(1-k) - k - f[w_1\bar{\ell}_1 + w_2\bar{\ell}_2]}{(1-f)[\eta w_1 + w_2]}\ ;\ \bar{y}_1 = \frac{f(1-k) - k - f[w_1\bar{\ell}_1 + w_2\bar{\ell}_2]}{(1-f)[w_1 + w_2/\eta]}$$

where $\bar{y}_1 = \eta\bar{y}_2$ and

$$\eta = \frac{\bar{\ell}_1 + \tau R(1-k-\bar{\ell}_1)}{\bar{\ell}_2 + \tau R(1-k-\bar{\ell}_2)}$$

Defining thresholds:

As before, let $\eta(f)$ denote the portion of payoffs that successful banks transfer to failed banks if the clearinghouse remain solvent. Then,

$$\eta(f) = \frac{w_1 f(1-k-\bar{\ell}_1) + w_2 f(1-k-\bar{\ell}_2) - k}{R(1-k-\bar{\ell}_1)(1-f)w_1 + R(1-k-\bar{\ell}_2)(1-f)w_2}$$

If f is such that $\eta(f) > 1$, then the clearinghouse becomes insolvent. Let f^* denote this unique threshold, and a_0 denote the level of aggregate shock a at which $f = f^*$.

As before, let f_1 denote the number of failures above which there is a fire sale with price less than $(R - \Delta - 1)$. f_1 is determined from Equation (A16). Let a_1 denote the level of aggregate shock a at which $f = f_1$.

Expected profits are given by

$$\begin{aligned}
\mathbb{E}\Pi(k,\ell,\bar{\ell}_1,\bar{\ell}_2) \quad = \quad & \\
& \textstyle\int_{-\infty}^{a_0(k,\bar{\ell}_1,\bar{\ell}_2)} [R(1-k-\ell) + \ell](1 - f(a) + g_1 + g_2)\phi(a)da \\
& + \textstyle\int_{-\infty}^{\min(a_0,a_1)} [(R-\Delta-1-p)y(p,k,\ell)(1-f(a)) + [p(1-k-\ell)+\ell](f(a)-g)]\phi(a)da \\
& + \textstyle\int_{\min(a_0,a_1)}^{a_0(k,\bar{\ell})} [(R-\Delta-1)(1-k-\ell) + \ell)(f(a) - g_1 - g_2)]\phi(a)da \\
& + \textstyle\int_{a_0(k,\bar{\ell})}^{\infty} ([R(1-k-\ell)(1-\eta(f(a)))(1-f(a)) + R(1-k-\ell)f(a)] + \ell)\,\phi(a)da
\end{aligned}$$

Bailing out Firms in the Asymmetric Equilibrium Model

One issue that becomes important when writing out the asymmetric equilibrium model is how failed banks are bailed out with the margin available in the clearinghouse (k). If k_1 of the margin is used to bail out Type 1 banks and k_2 of the margin is used to bail out Type 2 banks, then the total number of banks bailed out is given by

$$g(k_1,k_2,\bar{\ell}_1,\bar{\ell}_2) = \tfrac{k_1}{1-k-\bar{\ell}_1} + \tfrac{k_2}{1-k-\bar{\ell}_2}$$

Let g_1 and g_2 be the proportion of Type 1 and Type 2 banks (respt.) bailed out. Then,

$$g_1 = \frac{k_1}{1-k-\bar{\ell}_1}\ ;\ g_2 = \frac{k_2}{1-k-\bar{\ell}_2}$$

The supply of assets in the fire sale is given by

$$\mathcal{S} = (1-k-\bar{\ell}_1)[w_1 f(a) - g_1] + (1-k-\bar{\ell}_2)[w_2 f(a) - g_2] = (1-k-\bar{\ell}_1)\times w_1 f(a) + (1-k-\bar{\ell}_2)\times w_2 f(a) - k_1 - k_2$$

This implies that

$$\mathcal{S} = (1-k-\bar{\ell}_1)\times w_1 f(a) + (1-k-\bar{\ell}_2)\times w_2 f(a) - k$$

which is independent of k_1 and k_2. Thus the total supply of assets in the fire sale is independent of the way the clearinghouse bails out banks of each type.

Earlier, I had assumed that $k_1 = w_1 k$ and $k_2 = w_2 k$, i.e., the portion of margin used to bail out each bank type were proportional to the number of banks of each type. Instead, let

$$k_1 = \frac{w_1(1-k-\bar{\ell}_1)k}{w_1(1-k-\bar{\ell}_1) + w_2(1-k-\bar{\ell}_2)}\ ;\ k_2 = \frac{w_2(1-k-\bar{\ell}_2)k}{w_1(1-k-\bar{\ell}_1) + w_2(1-k-\bar{\ell}_2)}$$

Then,

$$g_1 = \frac{w_1 k}{w_1(1-k-\bar{\ell}_1) + w_2(1-k-\bar{\ell}_2)}\ ;\ g_2 = \frac{w_2 k}{w_1(1-k-\bar{\ell}_1) + w_2(1-k-\bar{\ell}_2)}$$

In the above, $g_1/g_2 = w_1/w_2$, and so this is consistent with a framework where each dollar of margin in the clearinghouse is used to bail out a bank at random, whether it is a Type 1 or a Type 2 bank. The law of large numbers implies that the proportion of each type bailed out will be w_1 and w_2.

Claim: If g_1 and g_2 are defined in this way, then $\mathbb{E}\Pi(k, \ell, \bar{\ell}_1, \bar{\ell}_2)$ is a function of $w_1\ell_1 + w_2\ell_2$. We can then write $\mathbb{E}\Pi(k, \ell, \bar{\ell}_1, \bar{\ell}_2) = \alpha_0(k, w_1\ell_1 + w_2\ell_2) + \ell\alpha_1(k, w_1\ell_1 + w_2\ell_2)$.

Further, for any asymmetric equilibrium $(\ell_1^*, \ell_2^*, w_1, w_2)$, $\exists$ a unique symmetric equilibrium $\ell^*(k)$ that generates the same profits for banks.

Proof. The proof is immediate from the above result. Set $\ell^*(k) = w_1\ell_1^* + w_2\ell_2^*$. Then $(\ell^*, y, 1, 0)$ is a symmetric equilibrium for any y, since $\alpha_1(\ell^*, y, 1, 0) = 0$. Further, $\alpha_0(\ell^*, y, 1, 0) = \alpha_0(\ell_1^*, \ell_2^*, w_1, w_2)$, and so profits for the asymmetric equilibrium and the symmetric equilibrium are the same. □

References

Acemoglu, Daron, and Fabrizio Zilibotti. 1997. Was Prometheus Unbound by Chance? Risk, Diversification, and Growth. *Journal of Political Economy* 105: 709–51. [CrossRef]

Acharya, Viral, and Alberto Bisin. 2014. Counterparty Risk Externality: Centralized versus Over-the-counter Markets. *Journal of Economic Theory* 149: 153–82. [CrossRef]

Acharya, Viral, and Alberto Bisin. 2009. Managerial Hedging, Equity Ownership, and Firm Value. *RAND Journal of Economics* 40: 47–77. [CrossRef]

Acharya, Viral, Hyun Song Shin, and Tanju Yorulmazer. 2011. Fire-sale FDI. *Korean Economic Review*. 27: 163–202. [CrossRef]

Acharya, Viral, Thomas Cooley, Matthew Richardson, and Ingo Walter. 2010. Manufacturing Tail Risk: A Perspective on the Financial Crisis of 2007–09. *Foundations and Trends in Finance* 4: 247–325. [CrossRef]

Allen, Franklin, and Elena Carletti. 2006. Credit Risk Transfer and Contagion. *Journal of Monetary Economics* 53: 89–111. [CrossRef]

Alessandri, Piergiorgio, and Andrew Haldane. 2009. *Banking on the State*. London: Bank of England.

Bhattacharya, Sudipto, and Douglas Gale. 1987. Preference shocks, Liquidity and Central Bank Policy. *New Approaches to Monetary Economics* 35: 69–88.

Biais, Bruno, Florian Heider, and Marie Hoerova. 2012. Clearing, Counterparty Risk and Aggregate Risk. *IMF Economic Review* 60: 193–222. [CrossRef]

Cassidy, John. 2009. *How Markets Fail: The Logic of Economic Calamities*. London: Penguin.

Castiglionesi, Fabio, Fabio Feriozzi, and Guido Lorenzoni. 2017. Financial Integration and Liquidity Crises. *Management Science* 65: 955–1453. [CrossRef]

Duffie, Darrell, and Haoxiang Zhu. 2011. Does a Central Clearing Counterparty Reduce Counterparty Risk? *Review of Asset Pricing Studies* 1: 74–95. [CrossRef]

Koeppl, Thorston, Cyril Monnet, and Ted Temzelides. 2012. Optimal Clearing Arrangements for Financial Trades. *Journal of Financial Economics* 103: 189–203. [CrossRef]

Leitner, Yaron. 2012. Inducing Agents to Report Hidden Trades: A Theory of an Intermediary. *Review of Finance* 16: 1013–42. [CrossRef]

Menkveld, Albert. 2017. Crowded Positions: An Overlooked Systemic Risk for Central Clearing Counterparties. *Review of Asset Pricing Studies* 7: 209–42. [CrossRef]

Paddrik, Mark, and Simpson Zhang. 2020. Central Counterparty Default Waterfalls and Systemic Loss. Working Paper, Office of Financial Research, Washington, DC, USA.

Philippon, Thomas, and Ariell Reshef. 2012. Wages and Human Capital in the U.S. Financial Industry: 1909–2006. *Quarterly Journal of Economics* 127: 1551–609. [CrossRef]

Pirrong, Craig. 2009. The Economics of Clearing in Derivatives Markets: Netting, Asymmetric Information, and the Sharing of Default Risks Through a Central Counterparty. Working Paper, University of Houston, Houston, TX, USA.

Rajan, Raghuram. 2005. Has Financial Development made the World Riskier? Paper presented at a Symposium Sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, WY, USA, August 26. pp. 313–69.

Rochet, Jean-Charles, and Guillaume Roger. 2016. Risky Utilities. *Economic Theory* 62: 361–82. [CrossRef]

Tucker, Paul. 2014. Are Clearing Houses the new Central Banks? Paper presented at Over-the-Counter Derivatives Symposium, Chicago, IL, USA, April 11.

Vasicek, Oldrich. 2002. The Distribution of Loan Portfolio Value. *Risk* 15: 160–62.

Yorulmazer, Tanju. 2013. Has Financial Innovation Made the World Riskier? CDS, Regulatory Arbitrage and Systemic Risk. Working Paper, Federal Reserve Bank of New York, New York, NY, USA.

Zawadowski, Adam. 2013. Entangled Financial Systems. *Review of Financial Studies* 26: 1291–323. [CrossRef]

Journal of
Risk and Financial Management

MDPI

Article

Capabilities and Reputation Risks Towards Firm Performance

Noraznira Abd Razak [1,*], Najihah Hanisah Marmaya [2], Mohd Zailani Othman [3,*], Idris Osman [4], Suhailah Kassim [4], Fatin Aqilah Maskuri [4] and Nik Kutina Mat Tahir [5]

1 Risk and Insurance Department, Faculty of Business and Management, University Technology MARA, Malacca Main Campus, Alor Gajah 78000, Malacca, Malaysia
2 Management Department, Faculty of Business, Economy and Accountancy, University Malaysia Sabah, Kota Kinabalu 88400, Sabah, Malaysia
3 Management Department, Faculty of Business and Management, University Technology MARA, Malacca City Campus, Malacca City 75300, Malacca, Malaysia
4 Human Resource Department, Faculty of Business and Management, University Technology MARA, Malacca City Campus, Malacca City 75300, Malacca, Malaysia
5 Head, Risk Management, Islamic Banking and Finance Institute Malaysia (IBFIM), Kuala Lumpur 50480, Kuala Lumpur, Malaysia
* Correspondence: noraznira@uitm.edu.my (N.A.R.); mzothman@uitm.edu.my (M.Z.O.)

Abstract: The effects of firm-specific resources on firm performance has been a quest of many and widely studied worldwide. In today's business environment, arguments suggesting the relative importance of firm-specific resources in explaining firm performance variation are said to be of the greatest influence on the study of firm behavior. On the other hand, firms with strong, positive reputations can attract and retain crucial talent and often have loyal customers likely to buy a broader range of products and services. It can lead to higher sales generated by satisfied customers and their referrals and can potentially raise capital and share price, and improve the firm performance. An empirical study such as this attempts to investigate the combinations of resources of the firm and focus on reputational risk management concerning firm performance. As such, this study involves variables partially adopted from Donabedian Theory, such as intangible resources, namely capability as an exogenous construct towards endogenous construct and firm performance, as well as proposing a mediation model to analyze the mediated relationship of reputational risk in accelerating the relationship between capabilities and firm performance. This study applies variance-based structural equation modeling via Smart PLS to a sample of 161 listed firms in Malaysia as respondents. A judgment purposive sampling technique has been adopted as the respondents are derived from listed firms under Malaysian Bourse. Overall, the findings of this study reveal how firms may gain competitive advantages in terms of their reputation and eventually be able to sustain their firm's performances by implementing an integrative model of intangible resources such as capabilities and in their routines and processes within the firms.

Keywords: capabilities; reputation risk; firm performance; mediation model

Citation: Razak, Noraznira Abd, Najihah Hanisah Marmaya, Mohd Zailani Othman, Idris Osman, Suhailah Kassim, Fatin Aqilah Maskuri, and Nik Kutina Mat Tahir. 2023. Capabilities and Reputation Risks Towards Firm Performance. *Journal of Risk and Financial Management* 16: 125. https://doi.org/10.3390/jrfm16020125

Academic Editor: George Halkos

Received: 11 October 2022
Revised: 11 January 2023
Accepted: 12 January 2023
Published: 15 February 2023

1. Introduction

It has been argued by Foss (1996, 1998) and Foss and Knudsen (2003) that capacities are among the most prominent dominant characteristics of the resources pool, which have the largest effect on business performance. It was also argued by Galbreath and Galvin (2008) that there is a substantial linkage between competencies and company performance. Erdil et al. (2010) claimed that their research showed a connection between fundamental personnel qualities, organizational capacities, and the success of a company.

Because Malaysia is working toward achieving sustainable economic development in which knowledge and know-how become the primary drivers of economic growth, it is particularly important to evaluate the significant aspect of know-how as a core skill among managers. This is because Malaysia is working toward achieving sustainable economic development (Majlis Inovasi Negara 2007). On the other hand, Galbreath (2005) discovered that the misallocation of resources or the inability to fully employ the resources of the organization likewise had a substantial influence on the firm's performance. The possibility of suffering damage due to a damaging reputation event is, without a doubt, very real. This damage can manifest in various ways, including a decline in consumer confidence in the brand, an effect on revenue and earnings, and increased oversight from government officials. However, there is also the possibility to learn from how a reputation incident is handled, both to lessen the immediate impact of the event and to acquire long-term understanding that may be used to better respond to future situations. Therefore, an effective allocation of resources has to be established after giving careful attention to several factors, given the level of risk that is associated with a particular project or investment made by the company itself (Razak et al. 2016). In general, enterprise risk cannot be considered in a silo approach since the management of risks has several specialized hazards, including reputation risks. These risks fall under the umbrella of enterprise risk (Bhanot 2011; Heidinger and Gatzert 2018; Kim et al. 2021; Pretty 2018; Voskovskaya et al. 2022; Razak et al. 2016).

A stronger reputation and higher status are associated with greater access to essential resources and better organizational performance, according to the findings of a variety of studies that were conducted in the past (Kim et al. 2021). Furthermore, reputation is important to corporate practice since it is a valuable intangible asset that may contribute to a competitive advantage; hence, this research gives the necessary viewpoints (Veh et al. 2018).

There has been a lot of research done on how firm resources might affect a company's success. However, research studying the combinations of business resources and management of enterprise risks, particularly reputation risk, as they relate to firm performance, is scarce and currently barely scratches the surface (Razak et al. 2013). The proliferation of the internet and social media has led to an increase in the significance of reputation risk for businesses (Scott and Walsham 2005; Magnus Boyd n.d.; Walsh et al. 2016; Szyntar and Heijmeskamp 2020). In this environment, unwelcome news can particularly spread more quickly. One of the most important questions that must be answered by businesses is whether or not the state of their corporate reputation or the occurrence of reputation-harming events (also known as "crisis events") affects the financial performance of the company (Gatzert 2015).

The following portion will present a short review of the direct relation of capabilities to firm performance, as well as the roles of reputation risk as a mediating impact between the two variables.

2. Literature Review

2.1. Firm Performance

Previous research has shown that different articulations may be found across fields of study regarding the definition of performance. As a result, the notion of performance assessment inside academic borders changes distinctly (Venkatraman and Ramanujam 1986). Previous research on company performance may be broken down into three main groups. To begin, some studies outline the measurements utilized in entrepreneurship and strategic management research, such as (Brush and Vanderwerf 1992; Murphy et al. 1996; Carton 2005; Carton and Hofer 2010). Studies that focus on the need for multidimensional measures of organizational financial performance, such as (Rawley and Lipson 1985), (Jantunen 2005) and (Venkatraman and Ramanujam 1987), and finally, studies that seek to determine the "best" measures of organizational financial performance, such as (Robinson and Mcdougall 2001). Venkatraman and Ramanujam (1986) were certain that the success of a company should be evaluated using a wider range of metrics, including financial and operational considerations. Because of its connection to accounting measurements and the economy's success, financial performance analysis indicators such as the growth of sales, earnings per share, and profitability.

Nevertheless, operational success or performance not based on financial metrics still considers product quality, market share, and marketing efficiency. In addition, the use of numerous performance indicators in empirical research was stressed by several academics, including (Demirbag et al. 2006). Previous empirical research also showed that the performance construct may be broken down into a number of different aspects (Venkatraman and Ramanujam 1986). However, no ultimate consensus has been presented in previous research on entrepreneurship and strategic management on the best or even a sufficient set of organizational performance measurements because there has been no research done on these topics. The vast majority of theorists have concluded that the nature of organizational performance is multidimensional (Carton and Hofer 2010). Accordingly, to ensure firm performance is measured accurately, Dess and Davis (1984) recommend that firms employ a composite measure by utilizing multiple indicators as it is more informative than relying on a single indicator only. Prior empirical research has demonstrated multiple dimensions of the performance construct (Venkatraman and Ramanujam 1986).

Venkatraman and Ramanujam (1987) empirically demonstrated that growth and profitability were different performance measures. Overall performance was measured with three perceptual items, including sales turnover and profitability as financial performance and market share as indicators of operating performance (Spanos and Lioukas 2001; Venkatraman and Ramanujam 1987). Venkatraman and Ramanujam (1986) also emphasized that business performance can always be measured by financial performance or operational performance, or even both, as the sources of performance data can either be primary (e.g., questionnaire survey) or secondary (e.g., published data). As such, this study follows the approaches of Venkatraman and Ramanujam (1986), Spanos and Lioukas (2001), and Galbreath and Galvin (2008) concerning the dimension of performance construct.

2.2. Capabilities

The idea of something being intangible refers to it being either imperceptible or unquantifiable. In contrast to physical resources, intangible resources are considerably harder to quantify due to the very nature of the resources themselves (Blair and Wallman 2001). According to Galbreath and Galvin (2008), intangible resources consist of many components not typically accounted for in the balance sheet. Intangible resources are described as "nonphysical factors that contribute to or are used in producing goods or providing services, or that are expected to generate future productive benefits for the individuals or firms that control the use of those factors," according to the definition provided by (Blair and Wallman 2001).

The conceptual definitions in the literature (Hall 1992, 2002; Fahy and Smithee 1999; Hoopes et al. 2003; Ray et al. 2004) cover a wide range of topics, so it is difficult to say definitively whether some intangible resources are, in fact, assets or capabilities. However, there appears to be a fine line between the two. On the other hand, the technique proposed by Hall (1992, 2002) is used, which stipulates that intangible resources be classified as either assets (what the company has) or skills (what the firm does). The categories of resources that will be discussed and utilized in this investigation were decided upon because they have been referenced in a wide variety of previous research, such as the general management, strategic management, marketing, and economics literature, and because they are of interest to a large number of academics. The know-how and knowledge capacity of the company are reflected in its capabilities (Grant 1996; Galbreath 2005; Galbreath and Galvin 2008).

According to Amlt and Schoemaker (1993), the term "capabilities" refers to a company's ability to deploy resources, often in combination, via organizational procedures, to achieve a certain goal. They are information-based processes that might be physical or intangible, are unique to the company, and are formed through time via the intricate interactions of its resources. In contrast to the other aspects of a company's resources, capabilities are predicated on the firm's human capital being able to create information, convey information, and exchange information with other people. According to Fahy and Smithee (1999), capabilities have been referred to using several names, such as talents, invisible assets, and intermediate products. They are also referred to as what an organization "does" as opposed to what it "has", and they often include the routines and interactions that take place inside the organization. Another way of describing them is as "what an organization "has". A company's capabilities include both the individual skills of its workers and the resources that emerge from the various interactions and routines that take place within the company itself. These interactions and routines can occur anywhere within the company, such as within teams, between workers and managers, or between personnel and tangible assets. They are distinguished by the presence of major obstacles to duplication, which may take the form of the tacitness inherent in the capabilities of people or groups or the intricacy and distinctiveness of organizational procedures.

These organizational capacities develop in tandem with organizational knowledge cycles, which begin with peripheral inducements and are then integrated via networking cycles to produce acquired knowledge inside the organization (Collis 1994; Collis and Montgomery 1995; Collis and Montgomery 1998). Before being used to address possible issues, this information is first made apparent and then transforms as it passes through several stages of internal selection, appraisal, and legitimization. In other words, the information is used and kept inside an organization by being ingrained in procedures and put into practice to generate retained knowledge via the accumulation of relevant experience (Collis 1994; Collis and Montgomery 1998). According to the theory put forward by HassabElnaby et al. (2012), organizational capabilities are a business's capacities to carry out a set of activities using the resources available to the company. Companies cultivate and manage their organizational capabilities to acquire a competitive advantage by fostering organization-categorical competencies. This process is known as organizational capability development and management. The longer talents are used, the more robust they grow and the more difficult it is for rivals to copy them. The capabilities of an organization are significant organizational resources that help a company develop a competitive edge. It is necessary to create and maintain these talents dependent on the strategies and information systems of the companies for the company to gain long-term benefits in terms of its competitive position.

It is possible to consider capabilities to be the aggregate set of organizational skills or competencies the corporation possesses. Capabilities are complex phenomena that emerge as a direct consequence of organizational learning. Capabilities are never tangible (Prahalad and Hamel 1990). In many management and organizational studies, a variety of terms are used for capabilities, such as management process, roles, and skills. This is done in order to include all of these managerial requirements to reconfigure and transform organizations along with their resources and capabilities. The approach relates variation in management competencies to differences in business performance under situations of strategic change in an explicit manner (Helfat and Martin 2015). Therefore, skills are understood to be the capacity to organize and use resources in order to accomplish the objectives of the company. This suggests that disparities in performance are seldom brought about by differences in resources alone; rather, it is the use of resources that brings about differences in performance. Companies can better manage their operations and make better use of their resources when they have capabilities. Capabilities are characterized as complicated bundles of skills and accumulated knowledge (Day 1994).

The items that made up capabilities comprised the know-how of the firm's managers, the know-how of workers who were not in management positions, and the collective know-how in constructing and preserving partnerships with external parties. Castanias and Helfat (2001) provided more evidence for similar ideas, namely that the success of a company is substantially tied to the abilities, competence, and knowledge of its managers. The individual talents of a company's workers, in addition to the resources resulting from the myriad of transactions and routines that occur inside an organization, are referred to as the company's capabilities (Fahy 2002). According to Amlt and Schoemaker (1993), capabilities are defined as the ability of a company to deploy resources, often in combination, via organizational processes, to achieve a desired purpose. Those organizational procedures inside the company itself, which include the practice of managing reputational risk (Razak et al. 2016). Mauri and Michaels (1998) and the most current study on the topic, conducted by Galbreath and Galvin (2008), both hypothesized that their results demonstrated the primacy of firm effects, particularly the impact of capacity on firm performance.

2.3. Reputation Risk

According to Smith (2008), in general, a company's reputation is based on how stakeholders and people who are not affiliated with the company view the company's overall quality in terms of how it interacts with customers, employees, and vendors as well as how it manages its finances and fulfills its social responsibilities. It is very uncommon, and difficult to replicate, and there are no suitable alternatives, all of which contribute to the item's status as a strategic component of the company. As a result, a favorable association will be formed between it and the future success of the company. The existence of a cross-sectional link between reputation and financial success may be rationalized by many different possible advantages that come with having a good reputation (Roberts and Dowling 2002). Customers place a high value on affiliations and transactions with businesses that have a good reputation because of this. Reputation is prized for its own sake. When there is confusion over the fundamental quality of a company's products or services, reputation becomes crucial. It is difficult for competitive companies to swiftly produce quality demonstrations that would counteract the signaling advantages associated with having a strong reputation because of the same uncertainty.

Given the proliferation of social media and the rise of cybersecurity inside Industry 4.0, maintaining a positive corporate image is becoming ever more crucial. An emerging aspect of research is reputation risk, which keeps on expanding as more researchers and practitioners all over the world start to practice reputation risk holistically. This is due to the evolving nature of the risks, which correlates with the volatile and advanced tools of technology that relate to online and social media platforms in various business environments. Reputation risk research is growing as a result (Ben-Amar et al. 2014). Because of this, reputation risk is becoming an increasingly important problem in modern times since the success of a company is often dependent on its reputation (Boyd et al. 2010; Szyntar and Heijmeskamp 2020; Kunitsyna et al. 2018; Eckert 2017; Spanier 2015; Guo et al. 2020; Roberts and Dowling 2002; Etter et al. 2019). Accordingly, Peterson (2006) defines reputation risk as arising from a situation, occurrence, business practice, or event that tends to materially influence the perceived trust and confidence of the public or stakeholders in an institution. Reputation risk can be caused by a situation, occurrence, business practice, or event.

According to Stephen P. D'Arcy (2001), developments in technology, the quickening pace of business, globalization, rising levels of financial sophistication, and the unpredictability of the global economic environment all contribute to an increase in the overall number of risks as well as their level of complexity in today's world. This is especially true regarding the risk that pertains to the reputation of companies. According to Beasley et al. (2005), there is a favorable association between the knowledge and abilities of senior management team members, such as the chief risk officer, and the implementation and efficacy of a variety of risks, including reputation risk. By creating online communities that facilitate connection and interaction between users, national social media platforms such as Facebook, YouTube, Twitter, LinkedIn, and Instagram, among others, continue to draw billions of users. These web-enabled platforms provide novel opportunities for socialization and interaction among users, thereby transforming how individuals and groups share information such as personal data, news, opinions, and feedback regardless of whether the respective businesses are based solely or simply on the enhancement of the promotion of products and services, or the firm depends on its performance of online businesses (Raina and Rana 2019).

Both Barakat et al. (2018, 2019) stressed in their research that risk appetite, incorporated within the reputation risk factor, also indirectly boosted organizational performance (e.g., improved returns, profits, and growth. Because Hoyt et al. (2008) found a positive association between the practice of reputation risk and company performance, they stressed that there is an influence on firm performance of up to 17% due to the practice of reputation risk. The finding is also consistent with the findings of Liebenberg and Hoyt (2003) and Kleffner et al. (2003), who also emphasized the skills of senior management, such as the chief risk officer as risk champion, that can lead the firm toward the effectiveness of the reputation risk as part of specific risks that existed within the firm. The finding was made by Kleffner et al. and Liebenberg.

As a result, the following hypothesis has been proposed as a result of this research:

Thus, this research proposes the following hypothesis:

H1. *Capabilities have a positive relationship with firm performance;*

H2. *Capabilities have a positive relationship with reputation risk;*

H3. *Reputation risk has a positive relationship with firm performance.*

Great economic damage was caused as a result of the global financial crisis that occurred between 2008 and 2009. This demonstrates that the linkage of the relationship between capabilities and firm performance can be strengthened by effectively managing the various risks that the firm faces, even though this requires strict monitoring from regulators (Mikes and Kaplan 2013). Regardless, the preceding studies contribute to the growing theoretical void this study seeks to fill. Namely, reputation risk has never been regarded as a mediating variable between skills and the relationship between them and business performance.

H4. *The relationship between capabilities and firm performance will be mediated by the effectiveness of reputation risk.*

3. Methodology

This research applied a self-administered questionnaire to collect the data. Measurement of the variables adopted from Galbreath and Galvin (2008) involving 27 items in total. This research adopts the methodology of (Ghazali and Manab 2013), which used the same data sources from Bursa Malaysia but did not include PN17 or GN3 firms in its analysis. The PLS-SEM was used in this study to do data analysis. As the population size was derived at 928, the sample size for this study was determined through the sample size table established by Sekaran (2013), which is at 250. To achieve an adequate response rate, a total of 650 questionnaires were distributed. There were a total of 161 replies considered to be legitimate after the questionnaire was sent to 650 people. Out of a total of 928 firms that were listed in Malaysia, adequately, 24.7 percent responded to the survey.

4. Results and Discussion

As the companies were from a wide range of industry groups, variation in the samples shows the overall industries represented the population of publicly listed companies under Bursa Malaysia. The distribution showed that slightly more than half of the sample (55.6%) was from large-scale companies (more than 500 employees). As far as the length of operation is concerned, 8 percent of the companies have been in operation for more than 10 years. The demographic data of the respondents showed that they all come from diverse educational backgrounds. More than 72 percent of the respondents have more than 10 years of working experience, indicating they have many experiences in their respective departments of corporate affairs and communications and are capable and reliable of answering the survey questionnaire without bias which is important for the validity of this study.

4.1. Assessment of Reflected Measurement Model

All item loadings were greater than 0.50 and significant at the 0.01 level, indicating convergent validity at the indicator level (Hulland 1999). All average variance extracted (AVE) values were greater than 0.50, suggesting convergent validity at the construct level. A measurement model is considered to have satisfactory internal consistency reliability when the composite reliability (CR) of each construct exceeds the threshold value of 0.7 (Gefen et al. 2000). A CR number of more than 0.70 indicates the dependability is satisfactory.

Thus, the results indicate that the items used to represent the constructs have satisfactory internal consistency reliability. The results stated above are as follows in Table 1:

Table 1. AVE and Composite Reliability.

Latent Variable	Cronbach's Alpha	Composite Reliability	Average Variance Extracted
CAPABILITIES	0.814	0.867	0.525
REPUTATIONAL RISK	0.972	0.974	0.641

Based on Table 2, off-diagonal elements are lower than the square roots of AVE (bold on the diagonal). Hence, the result indicates an adequate discriminant validity for all the reflective constructs.

Table 2. Fornell-Larcker Criterion.

Latent Variable	Capabilities	
CAPABILITIES	**0.718**	
	0.303	**0.817**

As for Table 3, the HTMT criterion also indicates that the confidence interval does not show the value of 1 on any of the constructs, confirming discriminant validity.

Table 3. Heterotrait-Monotrait Ratio (HTMT).

Latent Variable	Capabilities	Reputational Risk
CAPABILITIES		
REPUTATIONAL RISK	0.560	

4.2. Assessment of Structural Model

Henseler et al. (2009) indicate that moderate or average R2 values are acceptable when the endogenous construct is explained by a few exogenous constructs. For this research, capabilities explained 9.2% of the variance with R2 = (0.092), considered moderate, and reputation risk (able to explain 53% of the variance in firm performance. The f2 value of 0.10 indicates capabilities have a small effect in producing the R2 for reputation risk. On the other hand, the f2 value of 0.65 indicates a large effect in producing the R2 for firm performance. The predictive relevance (Q2) has a value of greater than 0, which indicates that the model has a medium predictive relevance for the performance construct.

Results from the study indicate that capabilities are one of the significant predictors in explaining the relationships of resources towards the firm's performance. This supports the finding of Galbreath and Galvin (2008) concerning the significant impact of capabilities on any positive outcome for the organization. As such, H1 was supported. As shown in Table 4, the hypotheses relating capabilities to the reputation risk were unsupported as the structural path coefficient was in the negative range. Therefore, H2 was rejected. As for H3, the finding shows that reputation risk provides a significant impact on firm performance.

Table 4. Path coefficients, Observed t-statistics, and results for all hypothesized paths.

Hypothesis	Path Coefficient	t-Value
Capabilities -> Firm Performance	0.303	4.455 **
Capabilities -> Reputation Risk	−0.023	0.376
Reputation Risk -> Firm Performance	0.735	19.929 **

Note: ** t-values > 2.33 ($p < 0.01$) (one-tailed test).

Based on Table 5, as indicated by Preacher and Hayes (2008), the significant indirect effects 95% Boot CI: [LL = 0.123, UL = 0.319] did not straddle a 0 in between, indicating mediation. Thus, this testing concluded that reputation risk had a significant mediation effect between capabilities and firm performance relationships.

Table 5. Bootstrapped confidence interval calculation.

Indirect Effect	SE	t-Values	LL	UL
0.221	0.050	4.421 **	0.123	0.319

Note: ** t-values > 2.33 ($p < 0.01$) (one-tailed test).

5. The Final Thoughts

According to the findings of this research, the connection between capabilities and performance is favorably mediated by reputation risk as an element of particular holistic hazards within the context of the risk management framework of the company. According to the findings of the study, publicly traded companies in Malaysia need to pay more attention to the efficient allocation of intangible resources, such as capabilities, to guarantee that managing those elements, which undoubtedly exist in the company regardless of whether or not the company's employees like it, has been realized. This result is also supported by a number of previous studies, such as Galbreath and Galvin (2008), both of which indicate that human capital, specifically capabilities, which are part of the intangible resources domain, exerts a direct and indirect influence on firm performance through various factors such as business processes, etc. Therefore, investment and growth in conjunction with a robust domain of resources like capabilities are proposed in order to accomplish the goals of achieving steady and improved performance by the organization.

The findings of this research indicate that the hypothesis of a positive link between capabilities and reputation risk toward firm performance is confirmed. This is shown by the fact that the hypothesis was shown to be supported by the findings. On the other hand, there is no evidence to suggest that capabilities and reputation risk are positively correlated. The results demonstrated that capabilities, a kind of intangible element of the company, are embedded along with reputational assets, a type of domain inside the company's intangible resources (Galbreath and Galvin 2008).

As a result of this, there is a risk that redundant testing of the same piece may take place. The reputation of a company is considered to be the firm's asset within the wide category of intangible resources, and it coexists with the capabilities, which are also a part of the larger picture of intangible resources, and it is intimately connected with those skills (Razak et al. 2016). Previous empirical research contradicts the findings of this study by demonstrating a positive and substantial correlation between intangible resources and enterprise risk in a unidimensional way (including reputation risk). This link was shown to exist between the two variables (e.g., Wan Daud et al. 2011; Kimbrough and Componation 2009; Wanlapa and Saenchaiyathon 2014). Liu (2011) provided evidence that supported the hypothesis that a positive correlation exists between organizational culture and knowledge management and business risk, namely reputation risk, among other risks. According to Kimbrough and Componation (2009), it is realistic to predict that an organization's internal culture will play a substantial role in ERM deployment success. The research conducted by Kimbrough and Componation (2009) demonstrates that deploying ERM should demonstrate the desired cultural characteristics. These characteristics include cross-functional cooperation, open communication across a network, and trust in associates' competency and willingness to deal with risks that can affect the entire organization. Involving and testing these three constructs in a mediating manner shall bear positive results. It can lead to fruition steps initiated by the firm to ensure that the priority of specialized risks, such as reputation risk, has been closely monitored and scrutinized, and has been adopted continuously to ensure the firm's continued viability in the years to come. This study suggests an addition of various resources dimension be added in future research, as well as considering improving government policy concerning the implementation of a risks framework that relates to small to medium enterprises. It has limitless variation of dimension involves especially when the nature of companies are uniquely differs.

Author Contributions: Conceptualization, N.A.R. and M.Z.O.; methodology, N.A.R. and N.H.M.; software N.A.R., I.O.; validation, N.A.R., S.K. and N.H.M.; formal analysis, N.A.R. and M.Z.O.; investigation, F.A.M. and N.K.M.T.; resources, M.Z.O. and I.O., S.K.; data curation, N.A.R., N.H.M.; writing—original draft preparation, N.A.R.; writing—review and editing, N.A.R. and F.A.M.; visualization, S.K., I.O.; supervision, N.A.R.; project administration, M.Z.O.; funding acquisition, N.A.R., M.Z.O., S.K. All authors have read and agreed to the published version of the manuscript.

Funding: Fundamental Research Grant Scheme: FRGS/1/2021/SS01/UITM/02/9 by Ministry of Higher Education (MOHE) with RMC File No. 600-RMC/FRGS 5/3 (050/2021).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

Amlt, Raphael, and Paul J. H. Schoemaker. 1993. Strategic assets and organizational rent. *Strategic Management Journal* 14: 33–46.

Barakat, Ahmed, Simon Ashby, and Paul Fenn. 2018. The Reputational Effects of Analysts' Stock Recommendations and Credit Ratings: Evidence from Operational Risk Announcements in the Financial Industry. *International Review of Financial Analysis* 55: 1–22. [CrossRef]

Barakat, Ahmed, Simon Ashby, Paul Fenn, and Cormac Bryce. 2019. Operational Risk and Reputation in Financial Institutions: Does Media Tone Make a Difference? *Journal of Banking and Finance* 98: 1–24. [CrossRef]

Beasley, Mark, Richard Clune, and Dana R. Hermanson. 2005. Enterprise risk management: An empirical analysis of factors associated with the extent of implementation. *Journal of Accounting and Public Policy* 24: 521–31. [CrossRef]

Ben-Amar, Walid, Ameur Boujenoui, and Daniel Zéghal. 2014. The Relationship between Corporate Strategy and Enterprise Risk Management: Evidence from Canada. *Journal of Management and Strategy* 5: 1–17. [CrossRef]

Bhanot, Sandeep. 2011. *A Study on Impact of Social Media on Company Performance*. Navi Mumbai: College of Management Studies, pp. 1–14.

Blair, Margaret M., and Steven M. H. Wallman. 2001. *Unseen Wealth*. Boston: Brookings Institution Press.

Boyd, Brian K., Donald D. Bergh, and David J. Ketchen. 2010. Reconsidering the Reputation–Performance Relationship: A Resource-Based View. *Journal of Management* 36: 588–609. [CrossRef]

Brush, Candida G., and Pieter A. Vanderwerf. 1992. A Comparison of Methods and Sources for Obtaining Estimates of New Venture Performance. *Journal of Business Venturing* 7: 157–70. [CrossRef]

Carton, Robert B. 2005. Measuring Organizational Performance: An Exploratory Study. Ph.D. thesis, University of Georgia, Athens, GA, USA.

Carton, Robert B., and Charles W. Hofer. 2010. Organizational Financial Performance: Identifying and Testing Multiple Dimensions. *Strategic Management Journal* 11: 1–23.

Castanias, Richard P., and Constance E. Helfat. 2001. The managerial rents model: Theory and empirical analysis. *Journal of Management* 27: 661–78. [CrossRef]

Collis, David J. 1994. Research Note: How Valuable Are Organizational Capabilities? *Strategic Management Journal* 15: 143–52. [CrossRef]

Collis, David J., and Cynthia A. Montgomery. 1995. Competing on Resources: Strategy in the 1995s. *Harvard Business Review* 73: 118–28.

Collis, David J., and Cynthia A. Montgomery. 1998. Creating Corporate Advantage. *Harvard Business Review* 76: 70–83.

D'Arcy, Stephen P. 2001. Enterprise Risk Management. *Journal of Risk Management of Korea* 12: 1–24.

Day, George S. 1994. The Capabilities of Market-Drive Organizations. *Journal of Marketing* 58: 37–52. [CrossRef]

Demirbag, Mehmet, S. C. Lenny Koh, Ekrem Tatoglu, and Selim Zaim. 2006. TQM and Market Orientation's Impact on SMEs' Performance. *Industrial Management & Data Systems* 106: 1206–28.

Dess, Grogery G., and Peter S. Davis. 1984. Porter's (1980) Generic Strategies as Determinants of Strategic Group Membership and Organizational Performance. *Academy of Management Journal* 27: 467–88. Available online: http://amj.aom.org/content/27/3/467.short (accessed on 8 October 2022). [CrossRef]

Eckert, Christian. 2017. Corporate Reputation and Reputation Risk: Definition and Measurement from a (Risk) Management Perspective. *The Journal of Risk Finance* 18: 145–58. [CrossRef]

Erdil, Oya, Hakan Kitapci, and Beyza Timurlenk. 2010. Effects of Core Employees on Organizational Capabilities and Firm Performance. *Journal of Global Strategic Management* 1: 30–38. [CrossRef]

Etter, Michael, Davide Ravasi, and Elanor Colleoni. 2019. Social Media and the Formation of Organizational Reputation. *Academy of Management Review* 44: 28–52. [CrossRef]

Fahy, John, and Alan Smithee. 1999. Strategic Marketing and the Resource Based View of the Firm. *Academy of Marketing Science Review* 10: 1–21.

Fahy, John. 2002. A resource-based analysis of sustainable competitive advantage in a global environment. *International Business Review* 11: 57–77. [CrossRef]

Foss, Nicolai J. 1996. Capabilities and the Theory of the Firm. *Revue d'économie Industrielle* 77: 7–28. [CrossRef]

Foss, Nicolai J. 1998. The Resource-Based Perspective: An Assessment and Diagnosis of Problems. *Scandinavian Journal of Management* 14: 133–49. [CrossRef]

Foss, Nicolai J., and Thorbjørn Knudsen. 2003. The Resource-Based Tangle: Towards a Sustainable Explanation of Competitive Advantage. *Managerial and Decision Economics* 24: 291–307. [CrossRef]

Galbreath, Jeremy. 2005. Which resources matter the most to firm success? An exploratory study of resource-based theory. *Technovation* 25: 979–87. [CrossRef]

Galbreath, Jeremy, and Peter Galvin. 2008. Firm factors, industry structure and performance variation: New empirical evidence to a classic debate. *Journal of Business Research* 61: 109–17. [CrossRef]

Gatzert, Nadine. 2015. The Impact of Corporate Reputation and Reputation Damaging Events on Financial Performance: Empirical Evidence from the Literature. *European Management Journal* 33: 485–99. [CrossRef]

Gefen, David, Detmar W. Straub, and Marie-Claude Boudreau. 2000. Structural Equation Modeling Techniques and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems* 7: 1–78.

Ghazali, Zahiruddin, and Norlida Abdul Manab. 2013. The Influence of Malaysian Code of Corporate Governance (MCCG) to Enterprise Risk Management (ERM) and Value. 2000 (February). Available online: Wbiworldconpro.Com (accessed on 8 October 2022).

Grant, Robert M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109–22. [CrossRef]

Guo, Yujuan, Di Fan, and Xiao Zhang. 2020. Social Media–Based Customer Service and Firm Reputation. *International Journal of Operations & Production Management* 40: 575–601.

Hall, Richard. 1992. The Strategic Analysis of Intangible Resources. *Strategic Management Journal* 13: 135–44. [CrossRef]

Hall, Richard. 2002. Enterprise Resource Planning Systems and Organizational Change: Transforming Work Organization? *Strategic Change* 11: 263–70. [CrossRef]

HassabElnaby, Hassan R., Woosang Hwang, and Mark. A. Vonderembse. 2012. The Impact of ERP Implementation on Organizational Capabilities and Firm Performance. *Benchmarking: An International Journal* 19: 618–33. [CrossRef]

Heidinger, Dinah, and Nadine Gatzert. 2018. Awareness, Determinants, And Value Of Reputation Risk Management: Empirical Evidence From The Banking And Insurance Industry. *Journal of Banking & Finance* 91: 106–18.

Helfat, Constance E., and Jeffrey A. Martin. 2015. Dynamic Managerial Capabilities: Review and Assessment of Managerial Impact on Strategic Change. *Journal of Management* 41: 1281–312. [CrossRef]

Henseler, Jorg, Christian M. Ringle, and Rudolf R. Sinkovics. 2009. The Use of Partial Least Squares Path Modelling in International Marketing. *Advances in International Marketing* 20: 277–319.

Hoopes, David G., Tammy L. Madsen, and Gordon Walker. 2003. Guest Editors' Introduction to the Special Issue: Why Is There a Resource-Based View? Toward a Theory of Competitive Heterogeneity. *Strategic Management Journal* 24: 889–902. [CrossRef]

Hoyt, Robert E., Dudley L. Moore, and Andre P. Liebenberg. 2008. The Value of Enterprise Risk Management: Evidence from the U.S. Insurance Industry. Unpublished Paper. 1–22.

Hulland, John. 1999. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal* 20: 195–204. [CrossRef]

Jantunen, Ari. 2005. Knowledge-Processing Capabilities and Innovative Performance: An Empirical Study. *European Journal of Innovation Management* 8: 336–49. [CrossRef]

Kim, Eunjung, Tohyun Kim, and Mooweon Rhee. 2021. The Dynamics of Organizational Reputation and Status: Momentum and Volatility. *Management Decision* 59: 2442–55. [CrossRef]

Kimbrough, R. Lane, and Paul J. Componation. 2009. The Relationship Between Organizational Culture and Enterprise Risk Management. *Engineering Management Journal* 21: 18–26. [CrossRef]

Kleffner, Anne E., Ryan B. Lee, and Bill McGannon. 2003. The Effect of Corporate Governance on the Use of Enterprise Risk Management: Evidence From Canada. *Risk Management Insurance Review* 6: 53–73. [CrossRef]

Kunitsyna, Natalia, Igor Britchenko, and Igor Kunitsyn. 2018. Reputational Risks, Value of Losses and Financial Sustainability of Commercial Banks. *Entrepreneurship and Sustainability Issues* 5: 943–55. [CrossRef]

Liebenberg, Andre P., and Robert E. Hoyt. 2003. The Determinants of Enterprise Risk Management: Evidence From the Appointment of Chief Risk Officers. *Risk Management Insurance Review* 6: 37–52. [CrossRef]

Liu, Xin. 2011. A Holistic Perspective of Enterprise Risk Management. Ph.D. dissertation, College of Business, Washington State University, Pullman, WA, USA, August. Available online: https://s3.amazonaws.com/na-st01.ext.exlibrisgroup.com (accessed on 8 October 2022).

Magnus Boyd. n.d. Managing Risk to Reputation.Pdf. *Clinical Risk* 15: 221–23. [CrossRef]

Majlis Inovasi Negara. 2007. *National Innovation Model*. Malaysia: Majlis Inovasi Negara.

Mauri, Alfredo J., and Max P. Michaels. 1998. Firm and industry effects within strategic management: An empirical examination. *Strategic Management Journal* 19: 211–19. [CrossRef]

Mikes, Anette, and Robert Kaplan. 2013. *Managing Risks: Towards a Contingency Theory of Enterprise Risk Management*. Boston: Harvard Business School, In Available at SSRN 2311293.

Murphy, Gregory B., Jeff W. Trailer, and Robert C. Hill. 1996. Measuring Performance in Entrepreneurship Research. *Journal of Business Research* 36: 15–23. [CrossRef]

Peterson, Gary. 2006. Interactive Risk Management System and Method with Reputation Risk Management. U.S. Patents 2006,011,68,98, June 1.

Prahalad, C. K., and Gary Hamel. 1990. The Core Competence of the Corporation. *Harvard Business Review* 68: 79–91.

Preacher, Kristopher J., and Andrew F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 40: 879–91. [CrossRef]

Pretty, Deborah. 2018. Reputation Risk in the Cyber Age: The Impact on Shareholder Value. *Aon & Pentland Analytics*. Available online: https://www.aon.com/reputation-risk-cyber-social-media-pentland-analytics-aon/index.html (accessed on 8 October 2022).

Raina, Ashish, and Varinder Singh Rana. 2019. Popularity Of Online Food Ordering and Delivery Services-A Comparative Study Between Zomato, Swiggy and Uber Eats in Ludhiana. *International Journal of Management, Technology and Engineering* 9: 6080–88.

Rawley, Thomas, and Marvin Lipson. 1985. *Linking Corporate Return Measures to Stock Prices*. Ithaca: HOLT Planning Associates.

Ray, Gautam, Jay B. Barney, and Waleed A. Muhanna. 2004. Capabilities, Business Processes, and Competitive Advantage: Choosing the Dependent Variable in Empirical Tests of the Resource-Based View. *Strategic Management Journal* 25: 23–37. [CrossRef]

Razak, Noraznira Abd, Zuriah Abdul Rahman, and Halimahton Borhan. 2013. Impact of Enterprise Risk Management (ERM) on Firm Performance: The Case of Malaysian Listed Firms. In *Sustainable Development: New Multidisciplinary Approaches and Methodologies*. Edited by A. Ahmed. London: World Association for Sustainable Development (WASD), pp. 249–59.

Razak, Noraznira Abd, Zuriah Abdul Rahman, and Halimahton Borhan. 2016. Modeling Firm Resources-Enterprise Risk Management Relationships: An Empirical Finding using PLS-SEM. *World Journal of Entrepreneurship, Management and Sustainable Development* 12: 35–49. [CrossRef]

Roberts, Peter W., and Grahame R. Dowling. 2002. Corporate Reputation and Sustained Superior Financial Performance. *Strategic Management Journal* 23: 1077–93. [CrossRef]

Robinson, Kenneth C., and Patricia Phillips Mcdougall. 2001. Entry Barriers and New Venture Performance: A Comparison of Universal and Contingency Approaches. *Strategic Management Journal* 22: 659–85. [CrossRef]

Scott, Susan V., and Geoff Walsham. 2005. Reconceptualizing and Managing Reputation Risk in the Knowledge Economy: Toward Reputable Action. *Organization Science* 16: 308–22. [CrossRef]

Sekaran, Uma. 2013. *Research Methods for Business: A Skill-Building Approach*, 6th ed. Chichester: Wiley.

Smith, Alan D. 2008. Resource Based View of the Firm: Measures of Reputation among Health Service-Sector Businesses. *Health Marketing Quarterly* 25: 361–82. [CrossRef]

Spanier, Gideon. 2015. Reputation Risk in the Social Media Age. *Raconteur*. Available online: https://www.raconteur.net/reputational-risk-in-the-social-media-age/ (accessed on 8 October 2022).

Spanos, Yiannis E., and Spyros Lioukas. 2001. An examination into the causal logic of rent generation: Contrasting Porter's competitive strategy framework and the resource-based perspective. *Strategic Management Journal* 22: 907–34. [CrossRef]

Szyntar, Alicja, and Romee Heijmeskamp. 2020. The Power of Reputation: A Qualitative Study on the Impact of Influencers' Perceived Reputation towards Makeup Brands. Master's thesis, Lund University, Lund, Sweden.

Veh, A., M. Göbel, and R. Vogel. 2018. Corporate reputation in management research: A review of the literature and assessment of the concept. *Business Research* 12: 315–53. [CrossRef]

Venkatraman, Natarjan, and Vasudevan Ramanujam. 1986. Measurement of Business Performance in Strategy Research: A Comparison of Approaches. *Academy of Management Review* 11: 801–14. [CrossRef]

Venkatraman, Natarjan, and Vasudevan Ramanujam. 1987. Measurement of Business Economic Performance: An Examination of Method Convergence. *Journal of Management* 13: 109–22. [CrossRef]

Voskovskaya, Anzhela Sergeevna, Tatiana Anatolievna Karpova, Tatiana Anatolievna Tantsura, Anna Yurievna Shirokih, Olga Yevgenievna Lebedeva, and Kostyantyn Anatol'evich Lebedev. 2022. The Risk Management System as an Enhancement Factor for Investment Attractiveness of Russian Enterprises. *Risks* 10: 179. [CrossRef]

Walsh, Gianfranco, Mario Schaarschmidt, and Harald von Kortzfleisch. 2016. Employees' Company Reputation-Related Social Media Competence: Scale Development and Validation. *Journal of Interactive Marketing* 36: 46–59. [CrossRef]

Wanlapa, Thomya, and Krittapha Saenchaiyathon. 2014. The effects of organizational culture and enterprise risk management on organizational performance: A conceptual framework. Presented at the International Conference on Trends in Multidisciplinary Business and Economics Research, Bangkok, Thailand, March 27–28; pp. 27–28.

Wan Daud, Wan Norhayate, Hasnah Haron, and Nasir Daing Ibrahim. 2011. The Role of Quality Board of Directors in Enterprise Risk Management (ERM) Practices: Evidence from Binary Logistic Regression. *International Journal of Business and Management* 6: 205–12. [CrossRef]

Journal of
Risk and Financial Management

Article

Return Based Risk Measures for Non-Normally Distributed Returns: An Alternative Modelling Approach

Eyden Samunderu [1,*] and Yvonne T. Murahwa [2]

1 International School of Management (ISM), Otto Hahn Str., 19, 44227 Dortmund, Germany
2 Independent Researcher, 16 Cheshire Road, Mt Pleasant, Harare P.O. Box MP163, Zimbabwe; ymurahwa@gmail.com
* Correspondence: eyden.samunderu@ism.de; Tel.: +49-231-97-51-3968

Abstract: Developments in the world of finance have led the authors to assess the adequacy of using the normal distribution assumptions alone in measuring risk. Cushioning against risk has always created a plethora of complexities and challenges; hence, this paper attempts to analyse statistical properties of various risk measures in a not normal distribution and provide a financial blueprint on how to manage risk. It is assumed that using old assumptions of normality alone in a distribution is not as accurate, which has led to the use of models that do not give accurate risk measures. Our empirical design of study firstly examined an overview of the use of returns in measuring risk and an assessment of the current financial environment. As an alternative to conventional measures, our paper employs a mosaic of risk techniques in order to ascertain the fact that there is no one universal risk measure. The next step involved looking at the current risk proxy measures adopted, such as the Gaussian-based, value at risk (VaR) measure. Furthermore, the authors analysed multiple alternative approaches that do not take into account the normality assumption, such as other variations of VaR, as well as econometric models that can be used in risk measurement and forecasting. Value at risk (VaR) is a widely used measure of financial risk, which provides a way of quantifying and managing the risk of a portfolio. Arguably, VaR represents the most important tool for evaluating market risk as one of the several threats to the global financial system. Upon carrying out an extensive literature review, a data set was applied which was composed of three main asset classes: bonds, equities and hedge funds. The first part was to determine to what extent returns are not normally distributed. After testing the hypothesis, it was found that the majority of returns are not normally distributed but instead exhibit skewness and kurtosis greater or less than three. The study then applied various VaR methods to measure risk in order to determine the most efficient ones. Different timelines were used to carry out stressed value at risks, and it was seen that during periods of crisis, the volatility of asset returns was higher. The other steps that followed examined the relationship of the variables, correlation tests and time series analysis conducted and led to the forecasting of the returns. It was noted that these methods could not be used in isolation. We adopted the use of a mosaic of all the methods from the VaR measures, which included studying the behaviour and relation of assets with each other. Furthermore, we also examined the environment as a whole, then applied forecasting models to accurately value returns; this gave a much more accurate and relevant risk measure as compared to the initial assumption of normality.

Keywords: risk; bonds; equities; hedge funds; forecasting; GARCH; value at risk

Citation: Samunderu, Eyden, and Yvonne T. Murahwa. 2021. Return Based Risk Measures for Non-Normally Distributed Returns: An Alternative Modelling Approach. *Journal of Risk and Financial Management* 14: 540. https://doi.org/10.3390/jrfm14110540

Academic Editors: Robert Brooks and Adrian Cantemir Calin

Received: 30 June 2021
Accepted: 1 November 2021
Published: 10 November 2021

1. Introduction

Financial markets have always been prone to exogenous shocks, and as a result, the element of risk is the dominant factor when examining and observing global patterns of the financial world. In historical patterns of crises (for example, 2007–2008), global financial meltdown had a massive impact on how firms could construct their portfolio in an effort to mitigate risk impact. It has become prudent to analyse current risk measures and find better ways of measuring it to allow for more accurate risk management and decision making. Thus, the increased volatility of financial markets over the last decade has induced researchers, policy makers and practitioners to develop and design more sophisticated risk management tools. Value at risk (*VaR*) has become the most standard proxy used by financial analysts to measure and quantify risk impact. In mathematical terms, *VaR* is calculated as follows:

$$VaR_{\alpha} = \alpha * \sigma * W$$

where α reflects the selected confidence level, σ the standard deviation of the portfolio returns and W the initial portfolio (Jorion 2003).

"Risk measurement is the foundation of risk management and hence of vital importance in any financial institution and investor" (Goodfellow and Salm 2016, p. 80). The universe of risk measures is large, and so, there is no single measure of investment risk that is correct for all purposes or for all investors. This differs based on investor needs and requirements. Risk tends to be measured in relation to price changes, which can take various forms such as relative, absolute, or log price changes. This price change is what this paper bases the risk measures on.

Conditional value at risk (CVaR), which is also known as Mean Excess Loss, Mean Shortfall or Tail VaR, is an alternative to VaR and attempts to deal with its predecessor's shortcomings. It measures the losses beyond the VaR point and has better coherence with properties of a robust risk measure. It is also referred to as expected tail loss and satisfies the measures that defines risk. Therefore, it can be argued that is is an ideal alternative that overcomes the shortfalls of VaR because it represents the loss that is expected beyond the loss given by the VaR. It can be argued that CVaR is a coherent risk measure that inhibits the following properties, such as transition-equivariant, positively homogeneous, convex and monotonic (Pflug 2000).

VaR is such a prevalent method to estimate risk, partly due to the regulatory framework created by the Basel committee on Banking Supervision during the 1990s, which forces the banks to calculate risk adjusted measures of capital adequacy based on VaR for their portfolios. This requirement was put into place in order to mitigate banks from taking on too much financial risk (Basel Committee on Banking Supervision 2009).

A VaR estimate corresponds to a specific critical value of a portfolio's potential one-day profit and loss probability distribution. Given their function both as internal risk management tools and as potential regulatory measures of risk exposure, it is important to quantify the accuracy of an institution's VaR estimates. However, it has some undesirable mathematical characteristics, such as a lack of subadditivity and convexity.

VaR also tends to be popular due to its ease in use and implementation by the average investors. However, with studies emerging, it is being revealed that most data are not homoscedastic but exhibit heteroscedasticity in the variance. This implores the use of the autoregressive conditional heteroscedasticity model (ARCH) and the general autoregressive conditional heteroscedasticity model (GARCH) to capture the changing volatility over time, since they are conditional upon heteroscedasticity (Orhan and Köksal 2011).

This study initially looks at the normality/non-normality of most of the data of returns. If this is affirmed, incorporating non-normality in risk calculations should therefore provide a better way of understanding and measuring the downside risk that is in an asset. The rest of the paper assesses models that consider this and gives a glimpse into the many options available that can be used to assess risk in non-normal distributions.

2. Literature Review

Risk measurement has evolved over time, and it began with the introduction of probabilities. The discovery and estimation of probabilities paved the way towards the first step in risk quantification. Evidence from Marrison (2002) alludes to the fact that risk is unavoidable in the financial industry, and so the best way to deal with it is to manage how it is measured and how it can provide the greatest risk-adjusted return whilst minimizing the risk exposure of the asset. In their study, Lovreta and Pascual (2020) examined the impact of sovereign default risk by using a vector autogressive (VAR) framework and applying Grainger-causality tests to illustrate implications of parameter instability between banks and sovereign default risk. The findings revealed that structural dependence in the financial system extends between banks and sovereign default risk volatility. The findings made further revelations that structural changes are present in the short-run dynamic relations between the sovereign and banking sector default risk during the period under analysis.

A large strand of risk measurement has focused on risk aggregation concerning risk implied by an aggregate financial position (Francq and Zakoïan 2018). Risk aggregation with dependence uncertainty is commonly referred to as the sum of individual risks with known marginal distributions and unspecified dependence structure. In their study, Francq and Zakoïan (2018) investigated both the estimation of market and estimation risks in portfolios by observing individual returns that follow a semi parametric multivariate dynamic model, and the asset composition was time varying. They developed asymptotic theory, which is used to estimate conditional VaR, and proposed an alternative risk testing approach called Filtered Historical Simulation (FHS). They concluded that by neglecting the estimation risk, practitioners might believe risk is controlled at a given level.

The universe of risk measures is large, and so there is no single measure of investment risk that is correct for all purposes or for all investors. Thus, the vast number of financial risk measures in the literature can be broadly subsumed into two categories: 1. risk as the magnitude of deviations from a target (risk of first kind), and 2. risk as necessary capital with respect to necessary premium (risk of the second kind). This differs based on investor needs and requirements. Risk tends to be measured in relation to price changes, which can take various forms, such as relative, absolute or log price changes. This price change is what this paper bases the risk measures on. The analysis is of various asset returns across different classes.

Generally, there are three specific measures used to measure the risk of individual assets, namely standard deviation, beta and duration. The volatility of asset prices is determined by the standard deviation approach, beta tackles market risk and portfolio risk measures, whilst duration measures the sensitivity of debt security prices to changes in interest rates. As the topic suggests, this paper focuses on the risk measures that are determined by the returns of the assets, and so it first assesses standard deviation, its downfall and what alternatives can be used to make the risk measure more accurate and reliable. It is also important to note that whilst standard deviation may be the appropriate measure of risk for high-volume markets (for example, Germany, the UK and France) it is not the most reliable risk measurement in terms of volatility when considering low-volume markets. It is also imperative that extreme financial external shocks lead to sharp spike jumps in stock prices and subsequently, volatility patterns. Wang et al. (2020) adopted four Jump-GARCH models in order to forecast the jump diffusion volatility which is used as a risk factor. The authors (Wang et al. 2020) considered both linear and nonlinear effects and the VaR of financial institutions by using vector quantile regressions. Evidence from the investigation revealed three interesting findings, which showed that when observing the volatility process of bank stock prices, the jump diffusion GARCH model is better than the continuous diffusion GARCH model. The jump behaviour of bank stock was seen to be heterogeneous due the difference in the sensitivity of the abnormal information shock. The performance of the support vector regression was seen to be a better approach than parametric quantile regression and nonparametric quantile regression. These findings yet again show that there is no "one size fits all" risk measurement model; thus, observing and measuring risk on asset returns requires a much more bespoke approach. Tang and Su (2017) conducted an analysis of the impact of financial events on systematic risk by observing the Chinese banking system. The authors criticize the methodological use of Contingent Claims Analysis (CCA) as a theoretical model to measure systematic risk (or beta of an asset) due to its strict theoretical assumptions and single source of risk information, as well as the fact that it limits the asset value volatility of the macro sector to being stationary stochastic process. Instead, they proposed an alternative model by relaxing the assumptions of CCA, replacing pure diffusion with jump diffusion, and introducing a macro-jump CCA that acts as a proxy to predict early warning effects of financial events.

Liu et al. (2020) examined the estimation for conditional volatility models and proposed a model-averaging estimator for conditional volatility under a framework of zero conditional mean. The three most popular univariate conditional volatility models are the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986) and the exponential GARCH (or EGARCH) model of Nelson (1990). Using Monte Carlo experiments, Liu et al. (2020) showed that their approach, the model-averaging forecast, leads to a better forecast accuracy, unlike other commonly used methods.

The Gaussian distribution introduced the assumptions of a normal distribution, which inevitably ushered in the mean variance framework; commonly used measures of risk are modelled around the normal distribution. However, it has been seen that most distributions are in fact not normally distributed, and continuing to ignore this fact by maintaining the assumptions used in the current risk measures distorts outcomes and leaves the financial markets unprepared to deal with and manage the degree of risk impact.

On the contrary, most returns tend to be not normally distributed but instead have heavy or fat tails that tend to be skewed. As highlighted by the recent 2007/2008 financial crisis, there are serious deficiencies concerning the risk models used. Most of these stem from the assumptions used in inputting the models; therefore, there is a model risk that is prevalent and it is desirable to find one that has a lower model risk level. Risk management models will need to be improved, putting greater emphasis on stress tests, scenario analysis and the use of forecasting. Bhowmik and Wang (2020) conducted a comprehensive literature review in order to ascertain different risk models raging from GARCH family-based models and used stock market returns as a "barometer" and an "alarm for economic and financial activities". Therefore, in order to prevent market uncertainty and mitigate risk in the stock markets, it is fundamental that the volatility of the stock market index returns is effectively measured. This means that financial market volatility is mainly reflected in the deviation of the expected future value of assets, indicating the possibility that volatility represents the uncertainty of the future price of an asset. This uncertainty is usually characterized by variance or standard deviation. However, it is imperative to note that measuring stock market volatility is a complex process for researchers because volatility tends to cluster. In fact, volatility is seen to be a permanent behaviour of the stock market around the globe. The presence of the volatility in the stock price makes it possible to earn abnormal profits by risk seeking investors. Kumar and Biswal (2019) employed GARCH econometric models in their study, and the results confirmed the presence of volatility clustering and advantageous effects that affect the future of stock markets. When using GARCH family models to analyse and forecast return volatility, the selection of input variables for forecasting is paramount, as essential conditions will be given for the method to have a stationary solution and perfect matching (Nelson 1990).

Dixit and Agrawal (2019) also suggested that the P-GARCH model is the most suitable to predict and forecast stock market volatility for the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) of India. The authors also highlighted the fact that volatility in financial markets is reflected because of uncertainty in the price, return, unexpected events and non-constant variance, which can be measured through the GARCH models that will give insight for investment decisions.

Variance measures the average squared deviation from the expected value, but this measure does not really give the full picture or description of the risk. In addition, proponents of this approach, such as Bodie et al. (2001), suggest that even though the returns are not normally distributed, if returns of a larger portfolio are used, it begins to resemble a normal distribution as it grows. However, in the case of stock prices, these cannot be negative, and so, to say that the normal distribution will accurately represent this distribution is a stretch.

The 2007/2008 financial crisis revealed that most risk managers are moving away from the historical treatment of risk but instead are laying more emphasis on scenario analysis and stress testing (Crouchy and Mark 2014). Stress testing is a risk management tool used to evaluate the potential impact on portfolio values of unlikely, although plausible events or movements in a set of financial variables (Lopez 2005). They are designed to explore the tails of the distribution of losses beyond the threshold (typically, 99%) used in the VaR analysis. For now, it is worth noting that the reason for them moving more to this approach is that they want to focus on looking at the outcome of a given adverse stress/scenario on their company or on a portfolio.

Volatility as a risk measure is conditional on past information and, by definition, the distribution must be symmetric, and thus not capable of capturing any asymmetric aspect of risk. This makes this risk measure more useful during normal market conditions, but less helpful during crises when the market does not behave according to the normality assumptions (Billio et al. 2016). However, the relevance of volatility forecasting for financial risk management has drawn a significant level of interest in the literature. This volatility forecastibility tends to vary with horizon, and different horizons are relevant in different applications. The estimation of volatility is a key input for calculating VaR because it directly depends on the expected volatility, time horizon and confidence interval for the continuous returns under analysis (Lelasi et al. 2021).

The VaR framework can be defined as a risk-measuring framework developed by financial market professionals as a means of measuring and comparing risk inherent in different markets. It describes the tail loss that can occur over a given time period resulting from its exposure to market risk, at a given confidence level. "It is the loss in market value time of an asset over a given time period that is exceeded with a probability" (Baker and Filbeck 2015). In other words, at a given time, an investor wants to know what he may stand to lose at a set confidence level and ideally, this is what VaR aims to answer. Thus, VaR models were sanctioned for determining market risk capital requirements for large banks and international banking authorities through the 1996 Market Risk Amendment Basel Accord. Spurred on by these developments, VaR has become a standard measure of financial market risk used by financial and even non-financial firms. VaR forecasts are often used as a testing ground when fitting alternative models for representing the dynamic evolution of time series of financial returns.

The downside of using VaR as a risk management tool is that it is most accurate for the evaluation of very short periods, such as a one-day period, and becomes more inaccurate the longer the time period (Krause 2003). Another limitation with VaR is that it does not give any further details on the losses beyond this VaR point. This is a limiting factor because such extreme risk, which is unaccounted for, can even result in the failure of the business or portfolio. VaR assumes that historical correlations do not vary in times of market stress and that the historical correlations remain unchanged, which is an incorrect assumption that can be proven in the data analysis (Baker and Filbeck 2015). VaR usually underestimates ex post volatility. Several studies have tested the performance of VaR in volatility forecasts and tested its limits against ex-post benchmarks (Nieto and Ruiz 2016; Barns et al. 2017). These studies concluded that simpler volatility specifications produce better VaR forecasts. For example, the popular VaR methodology based on a normality assumption on the conditional distribution of returns often leads to the underestimation of financial losses, as fat tails are not accounted for.

Jorion (2003) reiterates how VaR is an incomplete measure and may need to be used in conjunction with other risk measures to make it more accurate regarding shortcomings highlighted in the above discussions. The prominent downside of using VaR measurements is that it more suitable and accurate for the evaluation of very short periods, such as one-day periods. It operates under the assumption that the distribution is normally distributed and so it becomes a more inaccurate measure as the period prolongs. Since it works under "normal market conditions", it can be deemed as misleading for measures with a longer time span, which do not exhibit normal conditions during abnormal market periods. Other factors such as volatility can also give inaccurate data as they are subject to human bias, further making this proxy measure inaccurate.

As a decision tool, VaR has certain limiting mathematical properties such as convexity and subadditivity, as well as the fact that it gives the point at which the value at risk is and fails to give further details beyond this VaR point. Properties of risk measures can be formulated in terms of preference structures induced by dominance relations (see Fishburn 1980). This is detrimental because such extreme risk, which is not accounted for, can have a severe impact on the business or portfolio failure.

Stress testing is the second most commonly used risk management tool together with VaR, which is preferably used when assessing the investment strategy of a portfolio, to determine portfolio risk. It can also be used in determining hedging strategies, which will reduce portfolio losses (Van Dyk 2016). Thus, stress tests conducted in the context of a risk model can provide a useful alternative or complement to the current ad hoc methods of stress testing (Alexander and Sheedy 2008).

Most areas that may have been neglected in stress testing, which may have resulted in the recent crisis, could include failure to include risks such as securitization risk in the stress testing process. According to Van Dyk (2016), the stress tests that were carried out were unable to recognize the risk inherent in structured products. These risks would differ with differing asset classes such as between equities and bonds.

2.1. Suggested Alternatives for Non-Normal Distributions—Unconditional Distributions

There are two ways that modelling methods can be classified: as unconditional distributions, which are time independent, and conditional distributions, which are time dependent. Those that are classified as unconditional are those used under the normality assumption, and they assume that distributions of returns are independent of each other and tend to be independent of their past data. These include different variations of VaR. The conditional distribution approach is a newer modelling method designed to cater for non-normal distributions and address the shortcomings of the normal distribution to accurately model returns and risk. This method disputes the idea that returns are identical and independent, and models used in this category include GARCH and stochastic volatility, which are time-dependent processes. A benefit of this approach is that the models account for volatility clustering, which is a frequently observed phenomenon among return series. Hansen and Lunde (2005) made a comparison of ARCH-type models in terms of their ability to describe variance. They tested different volatility models in order to ascertain a better and more robust description of financial time series. In total, 330 GARCH-type models were compared using DM–USD exchange rate data and daily IBM returns. Findings revealed that there was no evidence to support the notion that GARCH is inferior to other models, and the models that performed better were the ones that accommodated a leverage effect. This argument is consistent with Kumar and Biswal (2019). This view is also mirrored in the work of Kilai et al. (2018), who acknowledge the shortcomings of GARCH-normal because it underestimates risk.

Conditional value at risk (CVaR)/expected tail loss/expected shortfall (ES) are alternatives to VaR and try to deal with their predecessor's shortcomings. These measure the losses beyond the VaR and have better coherence with the aforementioned desired properties of a good risk measure, two of which are convexity and monotonicity, which are there to ensure a desirable optimum exists (Baker and Filbeck 2015). CVaR modelling is cast in terms of the regression quantile function—the inverse of the conditional distribution function. In principle, VaR and CVaR measure different properties of the distribution. VaR is a quantile and CVaR is a conditional tail expectation. The two values only coincide if the tail is cut off. However, the problem choice between VaR and CVaR, especially in financial risk management, has been a popular base for discussion. The reasons affecting choice are based on the differences in mathematical properties, the stability of statistical estimation, the simplicity of optimization procedures and acceptance by regulators, etc.

In their early work, Rockafellar and Uryasev (2000) focussed on minimizing CVaR, whereby they used a technique for portfolio optimization which calculated VaR and optimized CVaR simultaneously. This technique is fundamental for investment companies, brokerage firms, mutual funds, etc., and uses a combination of analytical or scenario-based methods. Several case studies showed that risk optimization with the CVaR performance function and constraints can be carried out for large portfolios and a large number of scenarios with relatively small computational resources. It is also possible to dynamically hedge CVaR with options, and this is because CVaR takes into account average loss exceeding VaR. CVaR is considered as a coherent risk measure (Acerbi and Tasche 2002). CVaR has superior mathematical properties compared to VaR. CVaR is so called a "coherent measure"; for example, the CVaR of a portfolio is a continuous and convex function with respect to positions in instruments, whereas the VaR may even be a discontinuous function (Sarykalin et al. 2014). The theory of coherence requires a measure of risk to satisfy four mathematical criteria: (1) translation or drift variance (linear), (2) homogeneity, (3) monotonicity and (4) subadditivity (Artzner et al. 1999; Sheppard 2013). Risk management conducted using CVaR functions can be performed efficiently, and it can be optimized and constrained with convex and linear programming methods. Overall, CVaR provides an adequate picture of risks reflected in extreme tails. This is a very important property if the extreme tail losses are correctly estimated. However, broadly speaking, problems of risk management with VaR and CVaR can be classified as falling under the heading of stochastic optimization.

By using this measure, it is possible to handle extreme events and to measure risk in a more precise and easier way. Instead of focusing on what is likely to happen, this is achieved through the examination of the conditional expectations, hence the name. Expected shortfall has been one of the methods proposed as an alternative to VaR. For instance, in a situation when imposing limits on traders, by using VaR, the firm will assume too much because it holds less capital to cover this risk. This then puts the company at risk (Danielsson et al. 2005).

As the economic financial meltdown of 2008–2009 demonstrated, VaR fares poorly during periods of market distress (Chen 2014). Basel 2.5 addressed some of the shortcomings of VaR and added stressed VaR (SVaR). SVaR is viewed as a response to the pro-cyclicality of classical VaR. This is "a forward-looking measure of portfolio risk that attempts to quantify extreme tail risk calculated over a long time horizon" (Berner 2012). Its aim is to provide a wider perspective of various inherent risks such as market and credit risk, as well as gap risks and jumps. This is because it analyses a time-period that experienced extremes and takes into account the sudden changes. In short, stressed VaR is used to obtain an idea of the possible losses likely to occur given worse market conditions. SVaR corrects various deficits of the ordinary VaR in times of market stress and it employs the Gaussian (normal) probability formalism in a completely different way than VaR and is designed to account for tail risk and collective behaviour (Billio et al. 2016). The two main properties of SVaR are "fat tail volatilities" that account for outlier events in the risk factors and stressed correlations between risk factors that account for collective market participant behaviour in stressed markets (Dash 2012).

Lichtner (2019) emphasized that what is fundamental to any SVaR model is the choice of the return-type model for each risk factor. In his study, he proposed a generalized return model. This is because of the sensitive nature of SVaR numbers to the chosen return-type model, and researchers have to make prudent choices on which return type to adopt. The findings reveal interesting, different SVaR dynamics for each return type such as absolute, relative, shifted relative, log, etc.

It is a regulatory requirement according to the Basel III framework that SVaR be calculated for banks and financial institutions according to the framework (European Banking Authority 2012). They state that the benefit of using it is that it is procyclical. Applying this to asset classes or portfolios will have positive benefits over the generic VaR. The European Banking Authority (2012) stated that historical data which covers a continuous 12 months must be used or if it is less, then it must cover the period under which the portfolio was under stress. There is no weighting of historical data when determining the relevant historical period or calibrating the stressed VaR model. This is because the weighting of data in a stressed period would not result in a true reflection of the potential stressed losses that could occur for an asset or portfolio. It is assumed that a stressed VaR can be exceptionally smaller than VaR, and so this is tested as the authors compare some VaR variations. In the Basel framework, it is also recommended that there should be ongoing monitoring of stressed VaR relative to VaR. The ratio between the two measures that is identified at the beginning of the monitoring period should be used as a reference to the continued monitoring. Should the ratio decrease significantly, it should signal the potential need for a review of the stress period. Should the ratio between the stressed VaR and VaR go below one, this should be taken into account as a warning signal that may warrant the review of the stressed period.

Marginal VaR (MVaR) is another variation of VaR and amplifies the notion of understanding an additional amount of risk for new investments. This allows portfolio managers to understand the impact of adding or deducting positions in their portfolios. Volatility measures the uncertainty of the return of an asset when it is considered in isolation. When this asset belongs to a portfolio, what is of critical importance is its contribution to the portfolio risk. The effect of the small changes in a part of the portfolio to the portfolio VaR is measured by marginal value at risk (MVaR).

Incremental value at risk (IVaR) is yet another approach adding onto the traditional VaR approach. It has become a standard tool for making portfolio-hedging decisions, in particular, hedging and speculating with options and reducing the risk in a risk return analysis (Mina 2002). In theory, IVaR is a metric that measures the contribution in terms of relative VaR of a position or a group of positions with respect to the total risk of a pre-existent portfolio. Its aim is to calculate the worst-case scenario most likely to occur for a portfolio in a particular given period. Investors can therefore be able to determine which investment to undertake based on how it affects the portfolio, choosing the one which has the least losses impact. More recently, Jain and Chakrabarty (2020) made further improvements upon the performance of a managed portfolio by proposing the use of MVaR in order to ascertain the desirability of assets for inclusion in a managed portfolio. They empirically conducted the study with S&P BSE 100 as the benchmark index with at least five different optimization problems. Thus, in order to capture the effect of the change in dollar exposure of assets on the overall risk portfolio, the use of VaR of the individual assets is not adequate, and therefore, MVaR provides a suitable fit. This means that MVaR is more amenable when it comes to identifying additional portfolio risk (Jain and Chakrabarty 2020).

Lower Partial Moment (LPM) is a measure of downside risk. This is calculated as the average of squared deviations below a given target return. It is also a supplement to the VaR LPM and applies specifically to the left tale of a distribution of the VaR. It can be the average across all negative deviations from the VaR or the variance of the negative deviations of a pre-defined VaR It is, therefore, a method that places heavier weight on larger deviations from the current VaR in comparison to the smaller deviations (Goodfellow and Salm 2016). An investor can select or set the risk-free rates, and by selecting the degree of the moment, one can specify the risk levels that suit the portfolio or risk needs. Goodfellow and Salm (2016) compared three different risk measures based on the same stock return data, the portfolio variance as in the seminal works of Markowitz (1952) and the VaR based on *t* copula. They concluded that normal assumption substantially underestimates the risk faced by an organization, and therefore, they discredited the use of risk measures based on normal assumption.

Multifactor and Proxy Models are another option. The difference between multifactor models and arbitrage pricing models is that the latter limit themselves to historic data, whilst the former expand the data and include other macro-economic factors in the mode such as ratios and market capitalization. Their assumptions are that market prices usually fluctuate for no specific reason. They also assume that the returns of a stock are related to the riskiness; if a stock has high returns over a period of time, it must be riskier than the one with lower returns.

Fama and French (1992) found that if markets are reasonably efficient in the long term, then the market capitalization and book to price ratios were good stand-ins or proxies for risk measures. These types of multifactor models will thus do better in comparison to conventional asset pricing models to explain the differences in returns. The Capital Asset Pricing Model (CAPM) is based on the idea that not all risks should affect asset prices. The CAPM gives us insights about what kind of risk is related to return.

2.2. Suggested Alternatives for Non-Normal Distributions—Conditional Distributions

Serial correlation in asset return time series analysis to determine volatility forecast—ACF PACF—falls under the conditional distribution aspect. The aim of time series analysis is to decompose data to find a trend or seasonal component to it. This is carried out to assess and predict, and serial correlation is one such methodology of time series analysis. This occurs when a given period has returns that are dependent on the previous time, also known as autocorrelation, and is used to describe the relationship between observations on the same variable over an independent period. A zero correlation means they are independent. If otherwise, then they do not evolve in a random process, but rather they are related to their prior values (Napper 2008).

Having asset returns that are serially correlated tends to distort an asset class's actual risk as it reduces risk estimates from a time series characteristic, because serial correlation improperly smoothens an asset class volatility (J.P. Morgan Asset Management 2009). If one has a positive autocorrelation, this will lead to a VaR that underestimates the actual volatility of the asset. If one has a negative autocorrelation, this can result in an overstated volatility. (Baker and Filbeck 2015).

The Copula (t and Gaussian distribution) is a more mathematical approach to risk measurement which is under the Integrated Risk Management (IRM) framework, measuring the error of normal assumption and therefore adjusting it to fit the non-normality aspects of a distribution (Embrecchts et al. 2001). It takes into account events such as shocks and, under this methodology, what is known as fatal shock models, which occur and result in the destruction of the component. Other types are non-fatal shock models, whereby there is a chance of surviving the shock. Shocks can result from natural catastrophes or underlying economic events.

This method allows for the construction of models that measure risk which go beyond the standard ones at the level of dependence. This can be used to stress test different products and portfolios for extreme correlations and for the measurement of dependence. This methodology also addresses the limitations in linear correlation. However, the linear correlation tends to be preferred because of its simplicity and assumes normality. The concept of Copulas is rather abstract, in that it does not capture the joint movements in extreme values. This is where it would be prudent to consider using the next method to take into account the joint movement aspects of the non-normal distribution. This can be achieved by looking at correlation from a portfolio context whilst taking into account the variations that may evolve given different periods and under differing market conditions.

Financial markets are dynamic and constantly changing, and so it is critical to obtain data that are accurate and relevant to the period. In their BIS article, Loretan and English (2000a) suggested that when determining the appropriate time interval to use, risk managers should consider the periods of relatively high or low volatility. This is because these periods will have relevant information on the true and underlying asset relationship. If volatility and correlation are not adequately assessed, it could cause problems for risk managers when they want to stress test or look at worst-case scenarios. The thinking behind this is that higher correlations between assets generally come with added increased volatility.

This can be performed by grouping a period of high volatility and assessing the historic correlation, and this can be applied to any future high volatility scenario projections. A time under normal conditions must also be assessed to see how the correlation of volatiles is under this. It would be good to assess the correlations in these periods and to determine whether or not they increase or decrease with increased volatility. Thus, the main challenge is how to model risk to accommodate the time-varying nature of asset return volatilities (Loretan and English 2000b).

The main reasons for forecasting can be classified as being for asset allocation, for risk management, and for taking bets on future volatility. When looking at risk management, risk is measured to find the potential losses that an asset or portfolio may incur in the future (Reider 2009). To ensure this is possible, there will be a need to have estimates of the future correlations of the assets, as was described above in looking at correlations in risk measures. Following the seminal work of Box and Jenkins (1976), Autoregressive Moving Average (ARMA) models have become the standard tool for modelling and forecasting univariate time series. The ARMA framework can also be extended in order to allow for the possibility of heteroscedasticity. It is also necessary to obtain estimates of the future volatilities of the assets and this is where the ARMA and generalized autoregressive heteroskedasticity (GARCH) family of models come into play. They come in various methods and hybrids, but their intention is to forecast volatility. Volatility is not the same as risk. When it is interpreted as uncertainty, it becomes a key input to many investment decisions and portfolio creations. Thus, a good forecast of the volatility of asset prices over the investment-holding period is a good starting point for assessing investment risk. However, as a concept, volatility is simple and intuitive. It measures variability or dispersion about central tendency. Thus, the greater the deviation, greater the volatility. Furthermore, volatility forecasts are sensitive to the specification of the volatility model, and it is paramount to establish a balance between capturing the salient features of the data and overfitting the data.

Value at risk forecasting with the ARMA models are ideal in times of increased volatility. As previously introduced, the autoregressive and moving average can be used in time series analysis, and stemming from this are the models, which this paper mostly focusses on in the risk forecasts for the asset classes. These are namely the ARMA model, which is a combination of the autoregressive model and the moving average, GARCH model and exponential generalized autoregressive conditional heteroscedasticity (EGARCH) models. These models are designed to deal with common financial data time series characteristics such as thick tails and can be used to forecast the returns. EGARCH models capture the most stylized features of stock volatility, mainly time series clustering, negative correlation returns, lognormality and under certain specifications (e.g., Bollerslev 1986).

While ARCH was developed to model the changing volatility of inflation series, the model and its later extensions were quickly adopted for use in modelling conditional volatility of financial returns. The GARCH technique is a more sensitive way to measure risk in a distribution. GARCH is useful for non-normal distributions to analyse an asset's risk as well as market risk.

It integrates the theory of dynamic volatilities and that is used to describe variance in terms of what is currently observable (Engle 2004). This method takes the weighted average of past squared errors to form a weighted variance. By using the weights, with the most weight going towards the most recent information, it gives more influence to current data.

GARCH was developed to tackle the belief and fact that returns tend to be unpredictable and that they have fat tails. The main advantage of using these GARCH models is that they captured both the heavy tails of a return series as well as capturing the volatility clustering (Mabrouk 2016). According to the model, where there is high volatility, it is likely to remain high, and where there is low volatility, it will be low, given the fact that the periods will be time limited. GARCH adds onto the ARCH model, where the process produces extremes that would naturally be expected from a normal distribution, since the extreme values during the high volatility period are larger than could be anticipated from a constant volatility process (Engle 2004).

In using the GARCH approach, the analysis uses the exponentially weighted moving average (EWMA), which solves the problem of weights, whereby the recent return has more influence on the variance than that from the previous month. EWMA as a measure of volatility uses a smoothing parameter called *Lambda,* which is there to ensure that the variance that is calculated is more biased towards recent data. The weights exponentially decline at a rate throughout time.

The use of EWMA in this model is more fitting because it incorporates external shocks better than equally weighted moving averages. These estimates are much more efficient than the standard simple moving averages because they provide a more realistic measure of current volatility. However, there are some downsides to the model, such as the use of EWMA. For example, it assumes that the trend will continue in the long term, and so the forecasts determined become increasingly inaccurate as the forecast horizon expands. However, for scenarios with medium volatility of volatility, there is very little penalty for using EWMA regardless of the volatility generating process. The robust structure of EWMA appears to contribute to its greater forecasting accuracy than flexible GARCH. Brooks (2014) shared the logic that GARCH models are designed to capture the volatility clustering effects in the returns. In particular, he mentioned the Exponential GARCH model (EGARCH), about which he says on the plus side, its application ensures that the conditional variance will always be positive.

The stochastic volatility model (SV) is a model that is similar to the GARCH model but differs in that it is parameter driven by latent shocks, whereas GARCH is observation driven (Diebold and Rudebusch 2012). This makes it easier to analyse. The logic behind this model is that it considers issues such as time deformation (the idea that calendar time does not have to move in tandem with financial markets). Calendar time is constant and predictable unlike economic time, which may slow down or speed up. SV model forecasts are noticeably more accurate than GARCH in scenarios with very high volatility of volatility and a stochastic volatility generating process.

Another variation is the realized volatility. This method differs from the GARCH and SV approach, which looked at conditional expectations of discrete time squared returns. Realised volatility looks at realizations and not expectations. This realized volatility facilitates superior risk measurement, which can translate into superior risk management. By using RV, the model is based on continuous time, evolving along the way to allow for more accurate volatility forecasts. It allows for jumps and can be used to improve GARCH models and SV models.

3. Methodology

Our research design attempted to answer the research questions proposed earlier in this study. In view of the research problem, the proposed study applies a quantitative research design, drawing the analysis from multiple risk model testing methods. Through the testing methods derived from the GARCH family, we ascertain that the traditional assumptions of normality in a distribution do not assist financial decision makers in determining the appropriate models that give accurate risk measures. The normal distribution assumption is based on forecasting the future price changes of assets and basing these forecasts on the belief that prices follow a random walk. A random walk is described as a movement of variables or numbers that do not follow a specific order, and two types of random walk theories can be identified. Firstly, one is based on single price assets and the other is the fundamental model of asset price dynamics. To give our research design more richness, we also employed measures of dispersion in order to ascertain the distance between values within the distribution (range, interquartile range). Overall, three major ways of measuring risk are discussed, namely the variance–covariance approach, historical data and Monte Carlo simulations:

Choice of Data sets

Our research study adopted an empirical approach by examining a data set that represents the following asset classes: equities (138), hedge funds (13) and bonds (+20). All the data used in this study were obtained from Bloomberg. There was a diverse range from different classifications, sectors, industry classes, asset types, geographic regions and so forth. However, from this initial data pool, there had to be a selection of only those assets that would exist over the selected test time period, as is explained in the time period cut off section that follows.

Backtesting VaR Models

Several backtests were conducted, but the discussion on these methods is by no means exhaustive, since in this context, it is impossible to go through the variety of different methods and their applications. VaR models are only useful if they predict future risks accurately. In order to evaluate the quality of the estimates, the models should always be backtested with appropriate methods. However, a thorough discussion on every backtest method is beyond the scope of this study.

Unconditional coverage was adopted to statistically examine whether the frequency of exceptions over the time interval is in line with the selected confidence level. This was carried out to determine when portfolio losses exceeded VaR estimates.

Conditional coverage was also conducted through the clustering of exceptions in order not only to examine the frequency of VaR violations but also the time when they occur. For this part, both Christoffersen's (1996) interval forecast test and the mixed Kupiec test by Haas (2001) were conducted. As discussed earlier, the choice of parameters in VaR calculations is not arbitrary whenever backtesting is involved. To construct a solid view on the validity of the model, a relatively low confidence level was used. According to Jorion (2003), a confidence level of 95% is well-suited to backtesting purposes. We argue that with this approach, it is possible to observe enough VaR violations within a time period. Since the software had a 95% confidence interval by default, the authors would have also opted to use levels of 90% and 99% and test each one individually. However, we do acknowledge that having more than one level of confidence in the backtesting process makes the testing more effective. The total number of trading days (observations) was 252, which was sufficient to produce some statistically significant backtests and was also in line with the Basel backtesting framework.

As discussed earlier, single backtests can never be enough to evaluate the goodness of a VaR model. If one test yields a decent outcome, the result should always be confirmed by another test (Haas 2001).

Various tests were conducted in order to ascertain the research question, "how to measure risk in a non-normal world". The data were clustered into generic groups based on certain characteristics, and a summary of each group was established whilst accommodating exogenous factors. The cut off point for the data was data existing from 1995, and any asset that had data after this period was filtered out. This was carried out to ensure homogeneity, which allowed for better comparisons. There was a diverse range from different classifications, sectors, industry classes, asset types and geographic regions. However, from this initial data pool, there was only a selection of those assets that existed over the selected test time, and this is explained in the time period section that follows. The returns were calculated normally as opposed to via the lognormal method, which is time additive and assumes normal distribution. The skew of a lognormal distribution is positive, unlike market returns that tend to be negatively skewed. This assumption first had to be proved in the normality tests. It can also be a problem when associating log returns with fat tail assumptions.

3.1. Time Period

The timeline used in this analysis was a 21-year period ranging from 1995 to 2016. This time encompassed various periods of stress, ranging from the Mexican peso crisis to the

EU debt crisis. These periods are very important, as they reflect the stress periods where one can observe the returns during this time and come up with the most accurate risk measures for these situations. As mentioned in the literature review sections, calculating risk is not only based on models but one has to take into account historic events and study trends, business cycles and other issues, so as to take a more comprehensive approach. The current period of volatility resulting from the COVID-19 effect has not been factored into the study, given that the data available are inadequate, as it is still being collected. This, however, may be investigated further in future studies to assess COVID-19's impact. However, Iuga and Mihalciuc (2020) analysed the impact of global shocks and highlighted the global effects of COVID-19 by developing two models, using a regression model and their study to analyse the time frame from 2001–2020 Q2. This is due to the fact that data are still unfolding, and we minimized our analysis by not putting so much focus on the effect of the COVID-19 pandemic. Nevertheless, we do acknowledge the adverse effect the pandemic has had on economies and financial markets at large.

3.2. Tests

The first test carried out was to test the non-normality (skewness, kurtosis) of the data in order to see if there was non-normality present in the returns. If this was the case, then the next step was to determine the type of non-normality in the observed returns. To assess the non-normality or normality of the data sets, the Jarque Bera test was used, which is based on the classical measures of skewness and kurtosis (third and fourth moments of a distribution). As these measures are based on moments of the data, this test has a zero breakdown value. This test is a goodness-of-fit test of whether sample data have skewness and kurtosis that matches a normal distribution. Once the distributions were determined, it was easier to select the most appropriate measure or set of measures based on the literature gathered on the various options that can be used to accurately measure risk for a given data set with a given distribution. For any distribution F with finite central moments $\mu_k (k \leq 3)$, the skewness is defined as follows:

$$\gamma_1(F) = \mu_3(F)/\mu_2(F)^{3/2}$$

For any distribution F with finite central moments $\mu_k (k \leq 4)$, the kurtosis is defined as follows:

$$\gamma_2(F) = \mu_4(F)/\mu_2(F)^2$$

The next step was to conduct risk measure tests. After carrying out the normality tests, the next step was to carry out VaR tests. The VaR measure concept condenses the entire distribution of the portfolio returns into a single number that investors have found useful and is easily interpreted as a measure of market risk. The VaR is essentially a p-percentage quantile of the conditional distribution of portfolio returns. We selected two assets from each asset class that represent different characteristics to make our study more comprehensive and inclusive of different assets. The first part examined the various ways that directly improve the standard VaR by using its hybrids. Thereafter, the analysis focused on the application of other models to conduct volatility forecasts through the use of modelling methods suitable for conditional distributions which are time dependent. The results were compared with those from the other variation of VaR, which is the CVaR. The theory stands to state that the CVaR focusses on downside or tail risk, and so this measure is likely to be higher than that of standard VaR. Risk managers have a plethora of volatility measures to choose from when calculating VaR measures. Time series models of volatility range from exponentially smoothed and simple autoregressive models, over single-shock GARCH models, to two-shock stochastic volatility models. The benchmark measure pivoted in Morgan Guaranty Trust Company (1996) RiskMetrics sets the conditional mean constant and specifies the variance as an exponential filter:

$$\sigma_t^2 = (1 - \lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2$$

In the GARCH model, the most widely used form is GARCH (1, 1) and it has some extensions. In a simple GARCH model, the squared volatility is assessed.

When applying the SVaR, according to the Basel III framework, this can be calculated on top of the VaR with a 99% confidence interval. Applying this, however, to more a practical scenario involves the selection of a stress period. For this analysis, the period adopted was for monthly returns for a non-stress period of January 2003 to December 2005 (36 months) and the stressed period was the financial crisis from August 2007 to December 2009 (29 months). In order to broaden the scope of the analysis, two sets of data were used on the equities over the same periods. One was the monthly returns, which was similar to what was applied to the bonds and hedge funds, and the other was the daily returns of the equities. This was carried out to observe whether there was much difference when SVaR and VaR are calculated using monthly versus daily returns. Another dimension to this analysis was applied to the bonds and the normal period: VaR and stressed VaR were compared between one asset that is normally distributed and one that is not normally distributed. The aim was to visualize the extent of the impact a distribution of a data set had on the results of the test conducted.

In practice, Beta is typically estimated using parametric estimators such as ordinary least squares (OLS) because OLS is the best linear unbiased estimator. Additionally, the monthly, quarterly or annual returns data are very often employed (see Kamara et al. 2018). This is because they are stable and likely distributed. However, the outliers—a very common problem in real asset returns—may seriously affect the performance of the parametric estimators compared with non-parametric estimators.

The test for autocorrelation was another critical test undertaken because it helps in seeking dependence in returns over time. This was achieved by plotting the ACF and PACF to observe whether any correlation in an asset's return existed. This will further help in the application of models such as the GARCH to better forecast volatility in the future. If one examines empirical time series, it is easy to observe that large fluctuations in asset prices are more often followed by large ones, while small fluctuations are more likely followed by small ones. This stylized fact is known as volatility clustering. In order to view the properties and determine whether there were more large fluctuations than pure random processes, we looked at the autocorrelations of the returns.

The autocorrelation function $C(x_t, x_{t+\tau})$ is defined as follows:

$$c(x_t, x_{t+\tau}) \equiv \langle (x_t - \langle x_t \rangle)(x_{t+\tau} - \langle x_{t+\tau} \rangle) \rangle$$

$$\sqrt{\langle x_{t^2} \rangle} - \langle x_t \rangle^2 \sqrt{\langle x_{t+\tau}^2 \rangle} - \langle x_{t+\tau} \rangle^2$$

where $\langle x \rangle$ denotes the expectation value of the variable x.

After this, autocorrelation for an asset with itself and from the portfolio perspective was carried out. The most commonly used correlation test is the Pearson's test, although it assumes normality. The hypothesis of this would be to see whether there is a relationship between the correlation of a portfolio of assets and volatility during the normal periods against that observed over the stress periods. The assumption is that asset classes have correlations that tend to increase during high volatility periods, unlike during normal market conditions. In a portfolio context, this could mean that the benefit of diversification may be nullified in high-stress conditions, leading to an understated risk.

The next step was to assess the time series of the data to observe how the series moved and thus determine if there was volatility clustering and the type of variance that exists in each asset. In financial markets, prices of stocks and commodities fluctuate over time, which then produce financial time series. These time series are in fact of great interest both to practitioners and theoreticians when making inferences and predictions. This was to determine if the volatility displayed was of a homoscedastic nature (same variance) or heteroskedastic (the opposite of homoskedasticity). This helped to determine whether the data behaved as is expected in financial markets, which is usually heteroskedastic.

A characteristic inherent in a robust model is that the financial data should be able to accurately show and capture that volatility clustering exists (Lok 2015) in the time series.

Regression analysis was adopted to predict future volatility. In the process of finding the ideal VaR forecaster, the first step was to calculate the AIC of each asset. The AIC is ideal to show the best model to use in predicting volatility. After completing this, the final step was to forecast the risk using the three models, namely the ARCH, GARCH and EGARCH models for each of the asset classes to see how well the model forecasts volatility for each given asset class.

In carrying out this process, other software packages such as R were considered, but due to the complex nature of the methods, the authors opted to use the Excel software package called NumXL to carry out the computations. This package was familiar as it is their main statistical analysis tool. The next step involved plotting the forecast for the data using the various parameters for each model. Using this package, the parameters were calculated and used to calibrate and find the most efficient parameters to apply to each model. Using this data, the forecasts were plotted over a 12-month time horizon to assess and obtain the most likely volatility movements.

4. Results

4.1. Jarque Bera Test—Non-Normality Test

Despite the remarkable qualities of the Shapiro–Wilk test, we adopted the Jarque Bera (JB) test, which is applied to various general models, such as nonlinear regressions and conditional heteroskedasticity models.

When using the JB test, the aim of the test is to test the null hypothesis. In this case, the null hypothesis was that it is normally distributed and this is represented by a *p*-value of greater than 0.05, which will lead to failure and to the rejection of the hypothesis. However, if it is less than 0.05, then it means that one rejects the null hypothesis, leading to the conclusion that it is not normally distributed. This was applied to the data below for all three assets classes using this formula:

$$JB = \frac{n}{6}\,(S^2 + \frac{K^2}{4}) \sim X^2_{V=2}$$

where

- S = the sample skew;
- K = the sample excess kurtosis;
- N = the number of non-missing values in the sample;
- JB = the test statistic; JB has an asymptotic chi-square distribution.

In our analysis, we first derived the asymptotic distribution of the sample skewness and kurtosis coefficients of the model's standardized residuals. After, we constructed an asymptotic X_2 test of normality based on the results.

4.1.1. Hedge Fund Indices

Upon carrying out the tests on the hedge fund asset class, a summary of the findings from the Jarque Bera test are shown in Appendix A. All 11 funds have returns that are not normally distributed. By assessing the skewness, it shows that the returns are mostly negatively skewed or close to zero. This confirms the previous hypothesis on the distribution of returns, which stated that most returns exhibit a negative skew and that they tend to not be normally distributed. The kurtosis of the distribution is positive, indicating that the data set has fat tails. One particular asset with a low skewness of 0.0002 is the Dow Jones Credit Suisse Long/Short Equity Hedge Fund index, which would make one assume the distribution is normal, but if one looks at the kurtosis, it is 4. This means that even though its skew is close to zero, it is leptokurtic, which indicates that there is a clustering of the distribution around the mean. This will result in the distribution having higher peaks and fat tails, since most of the returns are clustered around the mean. The HFRMI Index and the Dow Jones credit Suisse Dedicated Share Bias Hedge Funds have a kurtosis less than 3, of 0.83 and 1.53, respectively, indicating that the distributions are platykurtic, meaning that they have lighter tails.

4.1.2. Equity Indices

The represented equity indices in Appendix B were selected to cover various diversifications, such as different geographic locations, and were either price weighted or market weighted. There were also stocks that represent small capitalization, mid-capitalization as well as large capitalization. This selection of indices ensured that the selection was broader and covered a wider range of the indices and so it was more representative of the markets as a whole. In developing these classifications, the authors also applied the World Economic Situation and Prospects (WESP), which classifies all countries into three segments, which are developed economies, economies in transition and developing economies.

Twenty-three indices were used to carry out this test. The results show that (be it capital weighted, all index weighted or price weighted) the distributions are not normally distributed, as indicated by the N. The only anomaly that is observed regarding normality, with a Y representing a yes, was with the four Japanese Stock Indices: the TPX500 Index (high capitalization), TPXSM Index (small capitalized), TPXM400 Index (mid capitalized) and the TPX100 Index (high liquidity), t the null hypothesis is not rejected, implying that the distribution is normally distributed. This only occurs in 4 indices of 23 observations.

For the rest of the indices, due to a vast number, a large number of different observations are made, but overall, it can be observed that they do not have a normal distribution. The kurtosis is mixed, with some being leptokurtic and others being platykurtic; regarding the differences in tails, some are heavy, implying high tail risk, and some are light, implying less risk in comparison to the heavy tails. The skewness results also differ when looking at the ones classified as being not normally distributed. They all exhibit symmetric distribution (between minus 0.5 and 0.5), and some are moderately skewed (between 0.5 to 1 and minus 0.5 to 1). The other measure is highly skewed, and this is shown by the indices in the data set that have a skewness greater than plus or minus 1, such as the BGSMDC index. Out of these 19 indices, only 4 have a negative skew, whilst the rest are positively skewed. Generally, a negative skew implies long-left tails, which indicates a higher chance of extreme negative outcomes, and positive implies a lesser likelihood of poor outcomes.

4.1.3. Bond Indices

There were around 60 bond indices, but the cut off was made to only include those that cover the test period, since most of them were younger issued bonds. Appendix C has the summary of the results. The results from this data set indicate that all the indices are not normally distributed, and so this supports the argument. Here, kurtosis seems to be above 3 for almost all of the results, indicating the heavy presence of fat tails. Again, almost all of the skewness results indicate a strong negative skew. However, the JPM Global Aggregate Bond index is the only one that has a normal distribution. Its skewness and kurtosis are close to zero, indicating potential symmetry.

Since the initial tests proved the non-normality of majority of asset returns, the analysis was narrowed down to adequately analyse the three asset classes, and two from each asset class were selected. This was to ensure that there was diversity in the results, which were representative of the general population. Below in Table 1 is a summary of the chosen assets and the description that is used in the analysis.

The two indices selected to represent the hedge funds are the HFRI Fund of Funds Composite Index (HFRI) and the Dow Jones Credit Suisse Event Driven Hedge Fund Index (DJCS). They each look at a hedge fund where one is a composite whilst the other is specific (Hedge Fund Research Inc. 2017). The HFRI is a hedge fund that invests with multiple managers through funds or managed accounts. There is, therefore, a diversified portfolio, since the idea behind it is to lower risk significantly by gaining the advantage of investing with more than one investment manager.

Table 1. Summary of asset classes representatives.

ASSET CLASS	NAME	DESCRIPTION
Hedge Fund	HFRI Fund of Funds Composite Index	HFRI
Hedge Fund	Dow Jones Credit Suisse Event Driven Hedge Fund Index	DJCS
Equity Index	OMX Helsinki All-Share Index	HEX
Equity Index	The Botswana Domestic Companies Index	BGSMDC
Bond Index	JPMorgan Government Bond Index- Emerging Markets	GBI-EM
Bond Index	the Merrill Lynch US High Yield Master II Index	HOAO

Source: Authors.

Equity Index: In selecting the equity indices to use, the authors chose two based on variant geographic locations. The first one is an equity index from Finland called the OMX HELSINKI INDEX (HEX). This is an all-shares index which reflects the current status and changes in the stock market. The other index that is examined is the Botswana Domestic Companies Index (BGSMDC), and this is an index that tracks companies that are traded in Botswana. The intention behind these choices was to obtain diverse representation. In this case, the HEX represents a stock market in the developing world against one from a nation in the developed nations which are found on two different continents, namely Europe and Africa. The classification of developed and developing nations is given according to the World Economic Situations and Prospects (WESP), which classifies all countries of the world into three broad categories, namely developing economies, economies in transition and developed economies.

Bond Index: The Merrill Lynch US High Yield Master II Index (H0A0) is the index that represents the corporate side of bonds because it is normally used as a benchmark index for high yield corporate bonds. This differs from the other bond index, which is the JP Morgan Government Bond Index-Emerging Markets (GBI-EM). This covers comprehensive emerging market debt benchmarks that track local currency bonds issued by emerging market governments.

4.2. Summary Statistics on Asset Class

In the first part of the analysis, an overview of the three asset classes was carried out to test for normality using the Jarque Bera test. This information is applied on these asset classes to estimate model and forecast volatility and the software used in Excel aids in the calculations. In this section, the authors give an overview of initial observations, which are further proved and justified using other models as the analysis continues.

The first step is a summary of Table 2, below, which was generated in NUMXL for the three assets under study. Below is the summary.

Table 2. Summary statistics. Source: authors.

EQUITY INDEX			Test	*P*-Value			SIG?
	HEX	BGSMDC		HEX		BGSMDC	
Average:	0.60%	1.35%	White Noise?	1.00%	FALSE	0.00%	FALSE
Standard Deviation:	7.57%	4.29%	Normal Distributed?	0.00%	FALSE	0.00%	FALSE
Skew:	−0.31	1.9	ARCH Effect?	0.00%	TRUE	64.11%	FALSE
Excess Kurtosis:	2.11	9.6					

Table 2. *Cont.*

HEDGE FUND INDEX			Test	*P*-Value			SIG?
	HFRI	**DJCS**		**HFRI**		**DJCS**	
Average:	0.44%	0.68%	White Noise?	0.00%	FALSE	0.00%	FALSE
Standard Deviation:	1.65%	1.80%	Normal Distributed?	0.00%	FALSE	0.00%	FALSE
Skew:	−0.85	−2.32	ARCH Effect?	0.05%	TRUE	2.24×10^{-306}	FALSE
Excess Kurtosis:	4.7	11.65					
BOND INDEX			**Test**	***P*-Value**			**SIG?**
	GBI-EM	HOAO		GBI-EM		HOAO	
Average:	0.45%	0.62%	White Noise?	53.18%	TRUE	3.05×10^{-6}	FALSE
Standard Deviation:	1.64%	2.57%	Normal Distributed?	0.00%	TRUE	0	FALSE
Skew:	0.1	−1.65	ARCH Effect?	98.43%	TRUE	1.77×10^{-15}	TRUE
Excess Kurtosis:	0.68	12.01					

4.3. Standard Deviation

Regarding the equities, the HEX has a higher standard deviation compared to the BGSMDC of 7.57% and 4.29%, respectively, showing that is has higher volatility. For the hedge funds, they have a much lower standard deviation compared to equities of 1.65% and 1.80%, respectively, for the HFRI and the DJCS. On the bonds, the GBI-EM is lower than the HOAO, since the former is based on government issues that tend to have lower risk, whilst the latter has the corporate bonds, which carry more risks with them, such as credit and default risk. Thus, it is justified to be riskier. Overall, it can be seen that equities carry higher risks than the other asset classes, making it a riskier asset class characterized by high volatility.

4.4. Skew

The next result on the distribution of the asset returns is from the third central moment, which is skewness. Four out of six of the assets have a negative skew. The negative skew proves the assumptions, which states how the distributions of returns tend to be negatively skewed. The presence of skewness shows that the distributions are non-normal. However, the Government Bond index GBI-EM and the Botswana BGSMDC have positive skews of 0.1 and 1.9, respectively. The GBI-EM is normally distributed, and its skew is highly insignificant. For the BGSMDC, the positive skew needs to be investigated further. For the asset, it has a positive mean and so a combination of a positive mean and positive skew implies that investing in this asset not only results in positive expected returns but also positive surprises on the upside as well. Investments in this section are highly unstable and this is usually short term until the market catches on. This result is significant given that this is a stock index in a growing market with high growth potential.

4.5. Excess Kurtosis

The fourth central moment measure looks at the peakedness of a distribution. In Table 2, the NUMXL statistics output is calculated as excess kurtosis instead of kurtosis. Excess kurtosis is defined as kurtosis minus 3. A normal distribution should ideally have an excess kurtosis of 0. These tests also stand to prove otherwise that the returns are not only non-normal and skewed, but that they possess long tails to the left. The negative skew in the returns implies frequent small gains and a few extreme losses. If the figure is above 0, it is leptokurtic, and this signals the presence of fat tails in the data set. Fat tails are considered undesirable because they imply additional risk.

In the tests, the returns from the equity and hedge fund all have kurtosis greater than zero, implying that there are high levels of tail risk. However, on the bonds, a different result is seen. The H0A0 index that represents them is also similar to the equities and hedge funds, although it is unusually higher. This is because on the returns, it has one of the highest ranges of −18% to 11%. This index represents high yield corporate bonds and so with high reward comes high risk, hence the fat tails. Comparing it, however, to the government issue of tracking bonds, it has a kurtosis of 0.6 and was concluded to be normally distributed, and that is why its kurtosis is almost zero, reflecting the small risk exposure this asset has in comparison to its other peers that all showed non-normality of returns.

4.6. White Noise and ARCH Effect

White noise and the ARCH effect is further explored in the volatility forecasting section. However, as an overview, it can be seen that the ARCH effect gives details on the distribution's fat tails and excess kurtosis. All the assets with an excess kurtosis greater than 0 are assumed to be serially correlated and have fat tails, further affirming the non-normality of the returns. White noise is defined as an assumption that a series has zero mean and constant variance. When applying this concept to the Autoregressive and Moving Average models, these are there to fix any assumptions that go against the white noise assumption according to Katchova (2013). Ideally, for the data to be applied to the model, it must first have the white noise characteristic. Its relevance will be shown in the section on ACF and PACF.

4.7. VAR Tests

Given that the data sets exhibited non-normal distributions with the exception of the GBI-EM bond index, the next step was to take an example from each asset class to first calculate the VaR and then compare it with CVaR. This comparison sought to prove how VaR underestimates risk and determine if CVaR is a better method. The calculations for all the VAR and CVar were made at 95%, 99% and 99.9% confidence intervals. Below, Figure 1 shows the graphs plotted for the two equity indices under analysis.

Figure 1 illustrates that the conditional VaR has a higher risk measure, which indicates how these equities have a much higher downside risk impact when measuring risk. Thus, for equities, it would be ideal to use the CVaR to capture this very aspect.

Similar results of this asset class to the equities are illustrated in Figure 2. However, it can be seen that generally, the hedge fund measurements for both the VaR and CVaR are lower than those of the equities, showing that this asset class generally has lower risk in relation to equities.

The results of the analysis of the bonds are depicted in Figure 2. From previous tests, it was denoted that the sovereign bond index GBI-EM was normally distributed, and so one would expect that the ideal measure would be the standard VaR. This is because the distribution is assumed to have little tail risk and its risk forecasts can be based on the Gaussian assumptions. For the HOAO index, which is not normally distributed, due to its distribution type it tends to have thicker tails and so a high tail risk. Therefore, to cater to the downside risk inherent in the tails, this index would be best suited with the CVaR measure (See Figure 3).

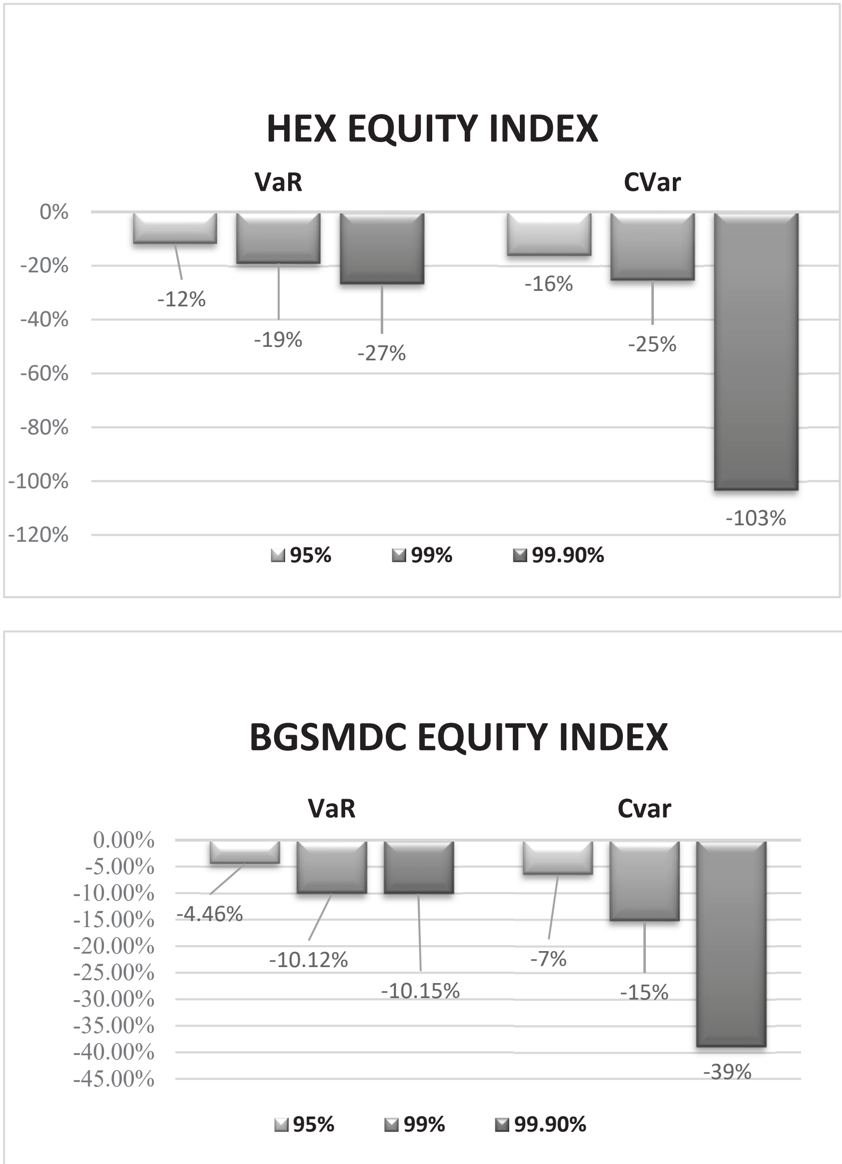


Figure 1. VaR versus CVaR for HEX and BGSMDC Equity Asset Index. Source: authors.

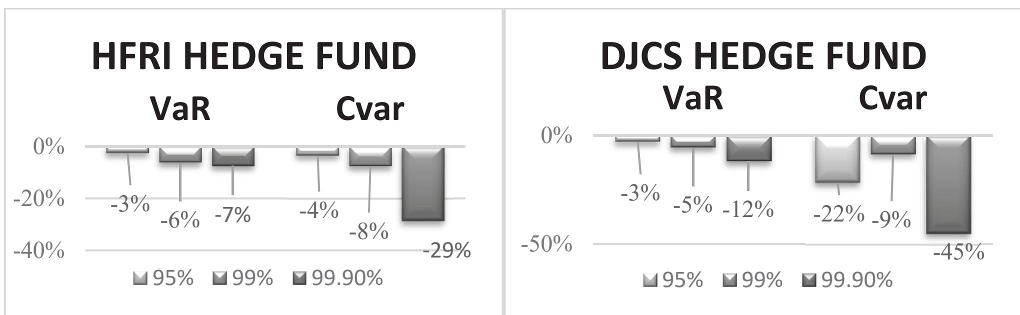


Figure 2. VaR vs. CVaR-HFRI and DJCS hedge fund.

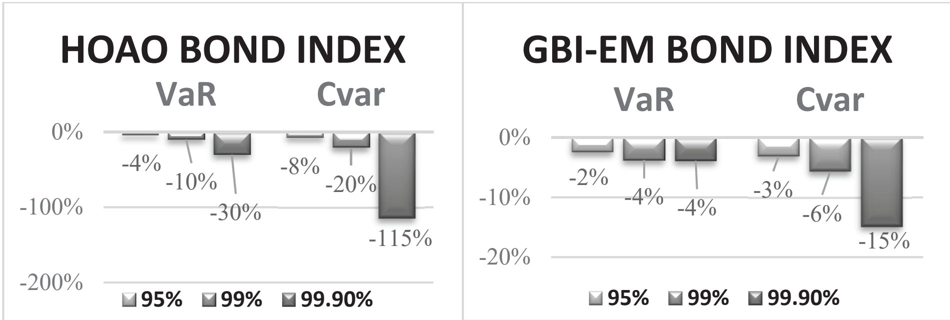


Figure 3. VaR and CVaR for HOAO and GBI-EM Bond Index Asset Class.

4.8. Stressed VaR versus VaR

In this case, the stressed VaR was calculated at a 99% confidence interval, as it is a regulatory requirement according to the Basel III framework in the case of banks and financial institutions. SVaR can be calculated by calculating the risk measures over a period where there was financial stress. These tests were then used as robustness checks. The stressed VaR must cover a period exceeding 12 months, and this was taken into account. The normal period was from January 2003 to December 2005 and the stressed period covered the financial crisis period from August 2007 to December 2009.

The results in Figure 4 above give a visual depiction of what is observed. The first analysis looks at the equities data. Because the data set also included the daily prices, the authors decided to calculate the normal and stressed VaR for both indices using the daily prices as well as the monthly prices to compare and see if there is an impact based on the time used in the calculation.

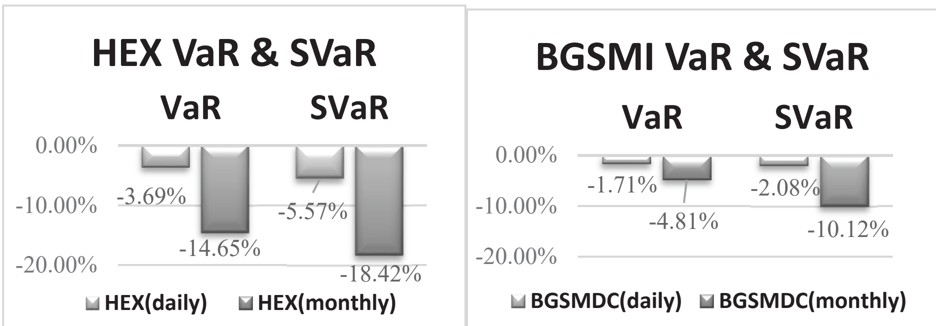


Figure 4. VaR versus SVaR equities. Source: authors.

Firstly, by comparing the VaR and SVaR for both scenarios, it can be observed that that the stressed VaR has a higher risk component in comparison to the normal period. The next step was to look at the different time horizons. When using the daily returns, the VaR and SVaR both show that the risk measure is lower in comparison to the value that is computed monthly. A reason for this could be that when data are observed daily, smaller incremental changes are shown as compared to data that are calculated over larger intervals, where higher deviations between the starting price and the next price will lead to higher return movements.

The next step was to look at the hedge fund data observed over a monthly period. This also displays similar results to the equities, showing that the stressed value at risk has a higher risk measure in comparison to the VaR in the normal period.

In this scenario, the authors compared the VaR observed in Figure 5 over the total time (1995–2016) to a normal period. The previous VaR readings were −7% and −12% for the HFRI and DJCS, respectively, whilst the normal time VaR in this section was −1% and −2%, respectively, showing that it is much lower than the 1995–2016 period.

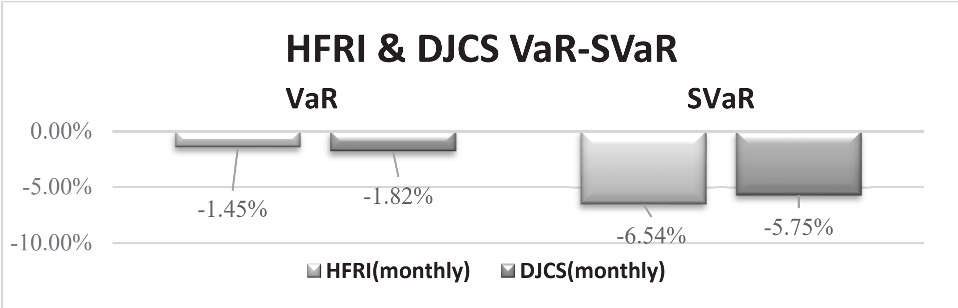


Figure 5. VaR versus SVaR hedge fund index.

The final step in this analysis was to look at the stressed VaR for the bonds in Figure 6. Firstly, it can be observed that the stressed VaR is higher than the normal period VaR, similar to the other asset classes' results. However, when looking at this date to make it more comprehensive, another angle is used to observe the data from Figure 6. Firstly, when observing the HOAO that is depicted in blue, it is seen that the SVaR is significantly higher than the VaR. However, when looking at the GBI-EM which is normally distributed, it can be seen that the readings for the VaR and SVaR are very similar, with the VaR being −3.57% whilst the SVaR is −3.08%, giving a negligible difference of 0.26%. This shows that in the case of a normal distribution without fat tails, the returns are generally uniform even in stressed periods.

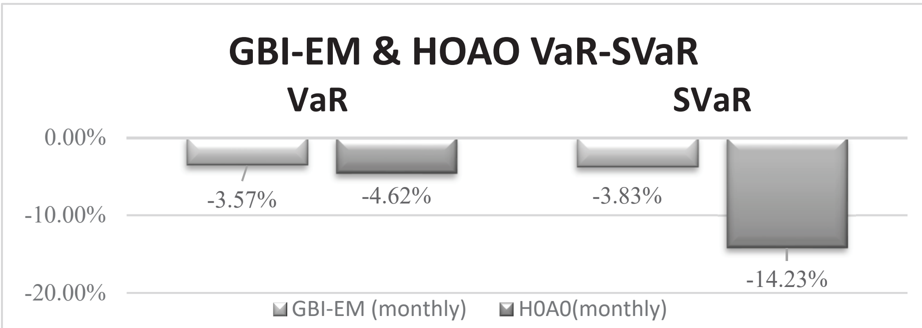


Figure 6. VaR versus SVaR bond index. Source: authors.

4.9. Volatility Analysis/Time Series Analysis

In this analysis, the aim is to calculate the weighted moving average (WMA) and the exponentially weighted moving average (EWMA), which is the volatility of the returns. The weighted moving average examines the stationary of the series. In this case, a 12-month equal weighted moving average is calculated, and the forecast horizon is the current period. By analysing the time series, it is possible to take note of outliers, seasonal or cyclical trends and other patterns, all of which will help in the prediction of future volatility in the asset returns. In addition, by looking at the time series, it is easier to identify the impact of stress periods described earlier in this paper. The WMA is calculated using the formula below:

$$wma_t^k = \frac{\sum_{i=1}^{k} x_{t-i} * w_i}{\sum_{i=1}^{k} w_i}$$

where:

- W_i = the weight of the i-th data point in the moving/rolling window.
- K = the size of the moving/rolling window.
- X_t = is the value of the time series at time t.

The EWMA is also calculated, and it shows the volatility over time of the asset and is calculated as follows:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)x_{t-1}^2$$

where:

X_t = the value of the time series value at time t.
Λ = the smoothing parameter (i.e., a non-negative constant between 0 and 1).

The HEX equity index in Figure 7 tracks stocks in an EU nation and it can be seen how volatile the returns were during the 2000 bubble and 2008 financial crisis. These markets were directly affected by the US economy situation, as their businesses and markets are interlinked. Generally, the trend, though volatile, is centred on the mean, signifying a stationary time series.

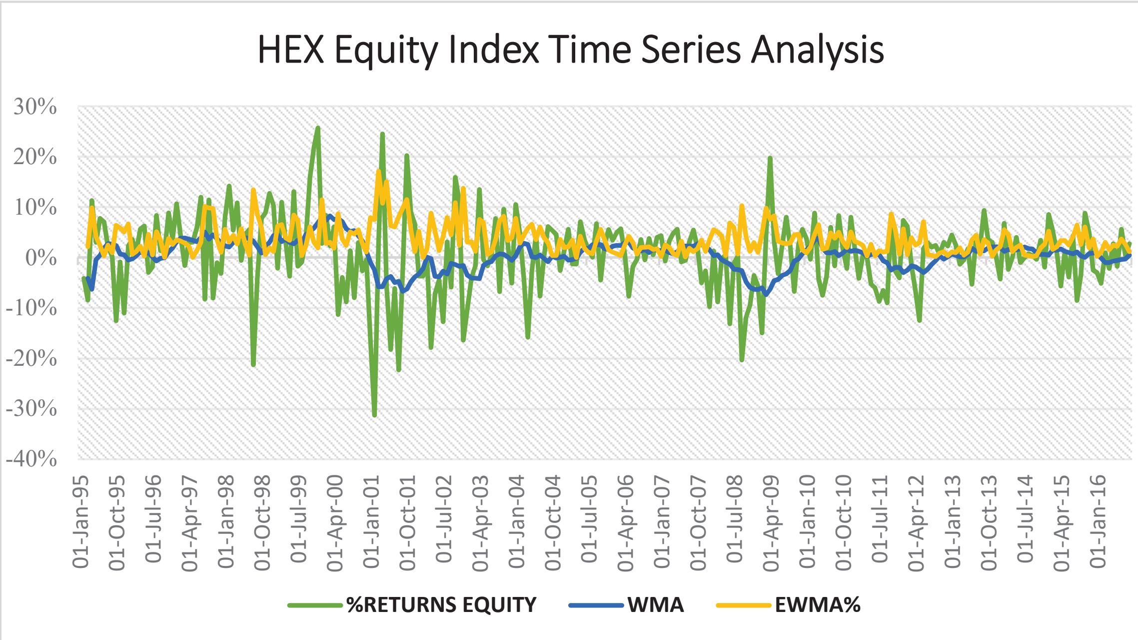


Figure 7. HEX monthly time series analysis, equity. Source: authors.

The time series in Figure 8 is different from the other equity returns, as it generally has a stable volatility, centring on the mean and fewer spikes during the 1997 Asian crisis, which could be because of the impact of Asian investors to the African nation being affected. It was not really affected by the tech bubble, implying that most of its investors were probably not in the tech industry. The WMA is generally more stable and smoothed out.

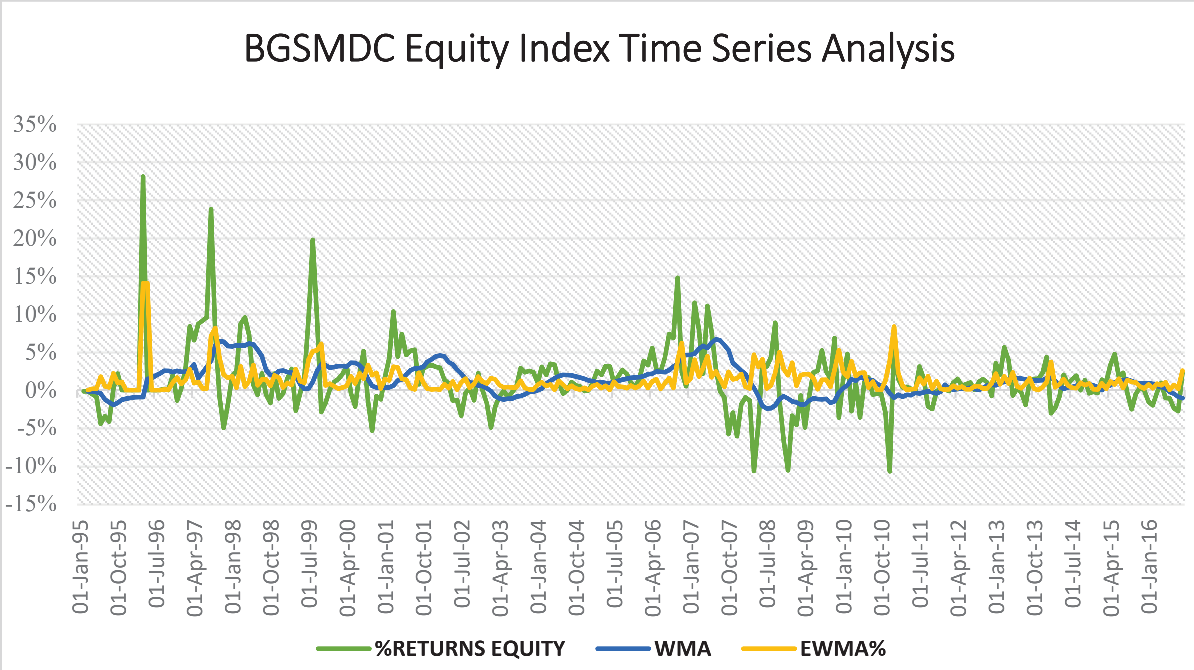


Figure 8. BGSMDC monthly time series analysis, equity. Source: authors.

Figure 9 reveals that the HFRI hedge fund has a unique characteristic, which is structured in a way so that it has many fund managers and so that it tends to be more diversified, and so volatility is around the mean and relatively stable. Clustering and huge sparks are shown in the crisis periods for the returns line.

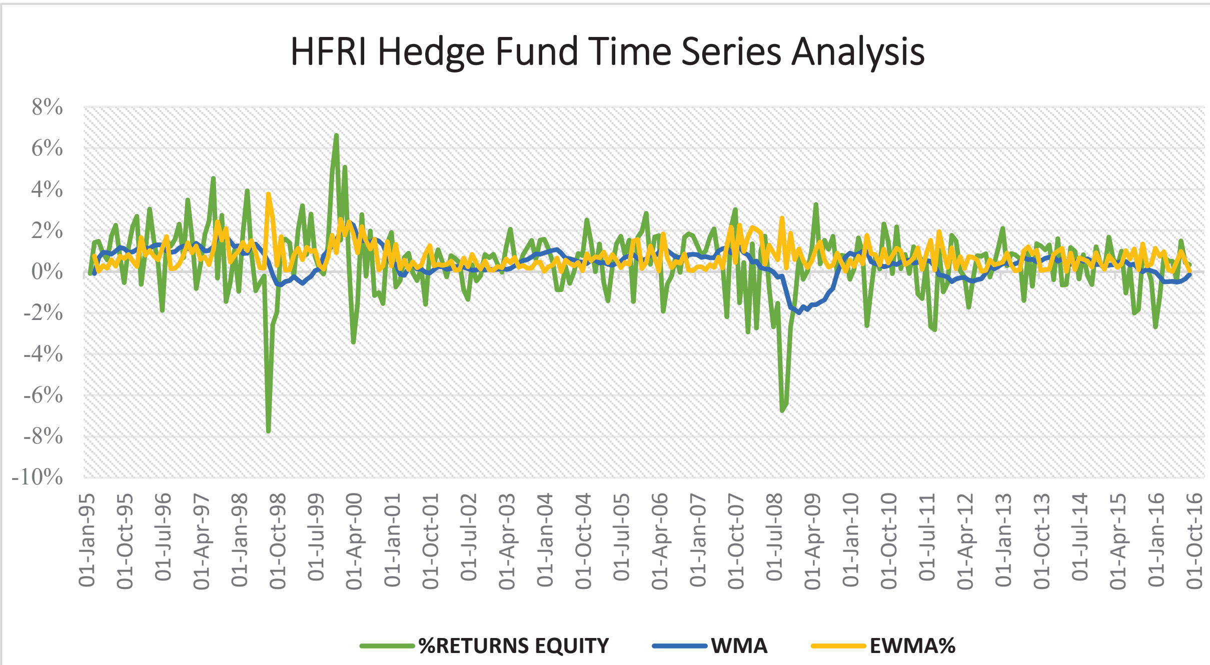


Figure 9. HFRI monthly time series analysis, hedge fund. Source: authors.

The DJCS in Figure 10 is also generally stable, since it is a hedge fund and experiences spikes during the stress period as it is based on liquid securities. This means that any change in the financial markets is reflected in the returns. It is noted that the WMA is relatively stable over time.

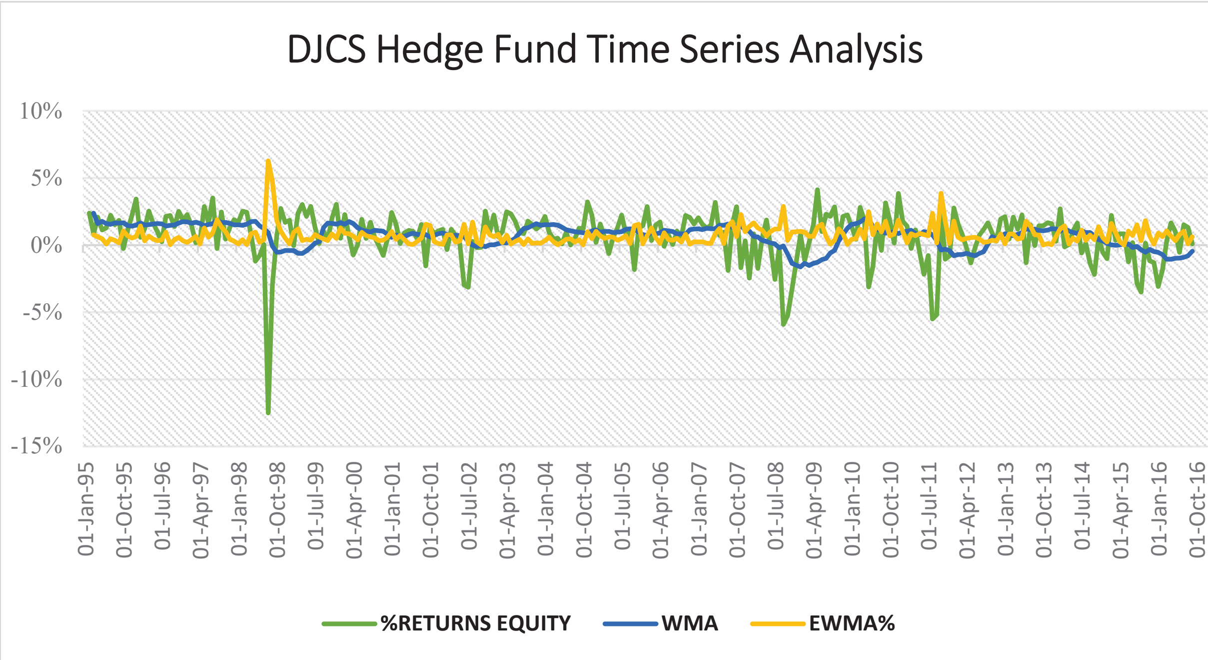


Figure 10. DJCS monthly time series analysis, hedge fund. Source: authors.

Figure 11 shows that the returns and the volatility are generally stable and only clustered around the stress period. There is little volatility because of the nature of bond pricing and its general predictability in comparison to equities. The notable high volatility is as a result of the overall financial markets being disrupted during periods of stress.

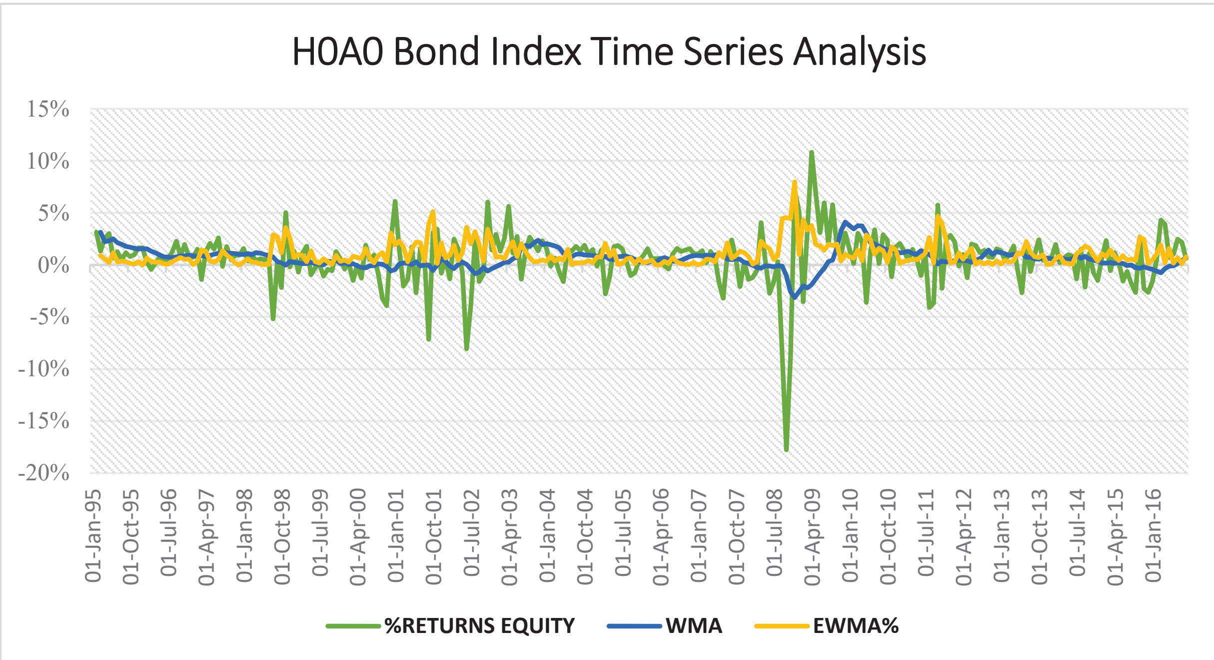


Figure 11. H0AO monthly time series analysis, bonds. Source: authors.

The GBI-EM in Figure 12 tracks the emerging markets' government bonds; by virtue of the debt being sovereign, it is generally stable, and the returns and volatility (EWMA) only spike in the stress period. The WMA is generally smooth and centred around the mean.

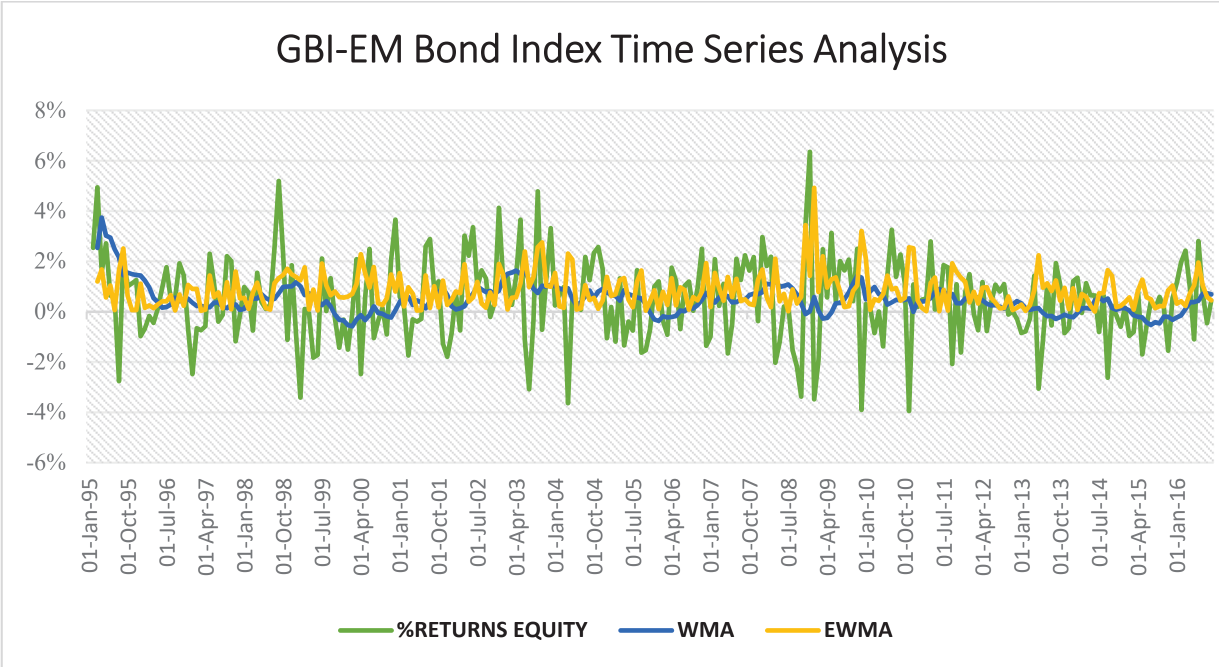


Figure 12. GBI-EM monthly time series analysis, bond index. Source: authors.

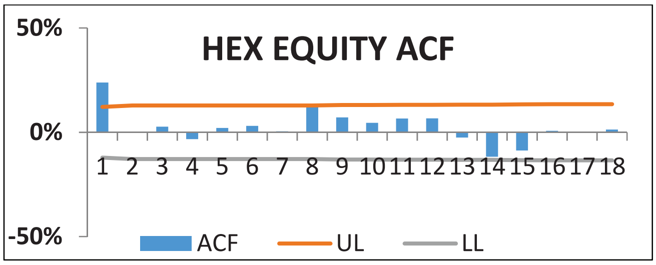
Overall, with the exception of the GBI-EM index (it seems to show a steady and constant volatility), most of the time series exhibit no constant volatility, which is shown by the separate periods of high volatility and low volatility. This is consistent with heteroskedasticity or conditional heteroskedasticity. This means that it is ideal to use more advanced forecasting financial modelling tools on it because it behaves in a way that is expected of financial markets that tend to exhibit periods of high volatility and low volatility.

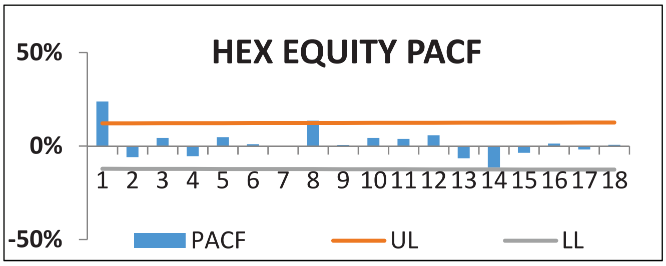
Through the application of time series analysis, it becomes necessary to pay attention to the order and so this is where the implementation of tests such as the serial correlation test is required. One such method was the time series analysis that was previously examined above. Now, the focus is on serial correlation, and this test is conducted with correlograms. This test is also critical in identifying the model order used in the tests such as ARMA models.

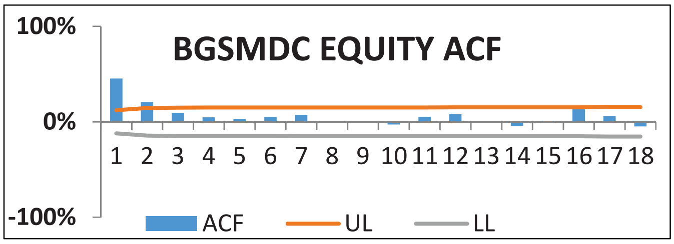
The ACF and PACF are used to detect the form of time series process. For it to be a suitable time series to apply forecasting models, it must be a stationery process, which is a desirable trait. This can be observed from the correlograms below (see Figure 13).

The log returns do not exhibit strong interdependency, though lag order 1 shows marginal significance. To shed more light on this, it is critical to assess whether the time series exhibits white noise, which implies no serial correlation. As shown in Table 2, the white noise test was conducted and it showed that both the HEX and BGSMDC tested false, implying that there is no white noise and so the series is serially correlated.

For the hedge fund plot (Figure 14), a similar pattern to that of the equities shows that the returns do not exhibit strong interdependency, though lag order 1 for the ACF although the PACF for HFRI has three significant lags. To assess this, Table 2 for the statistics summary, shows that the white noise test also showed that both assets tested false, which means there is also serial correlation of the returns of each asset.







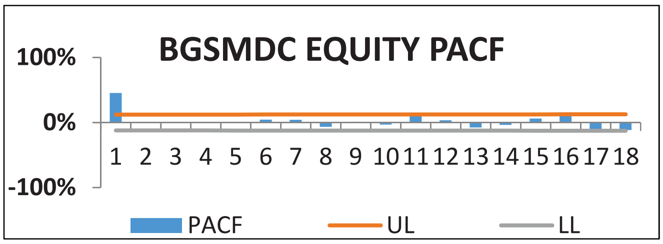


Figure 13. ACF and PACF plots for equities. Source: authors.

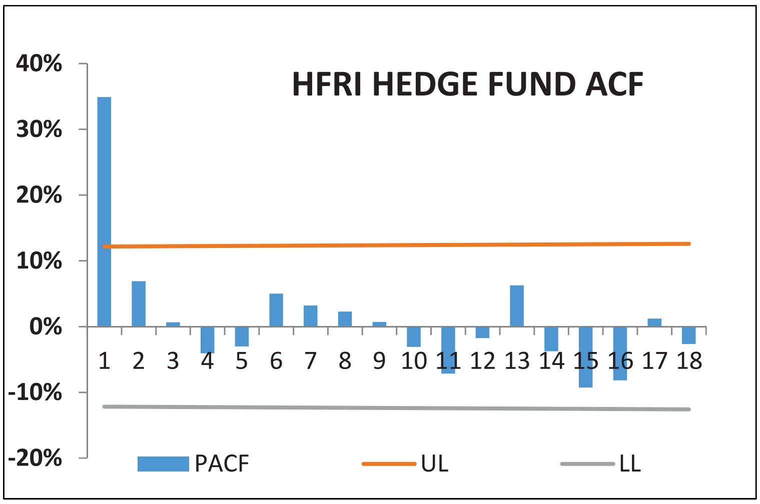


Figure 14. *Cont.*

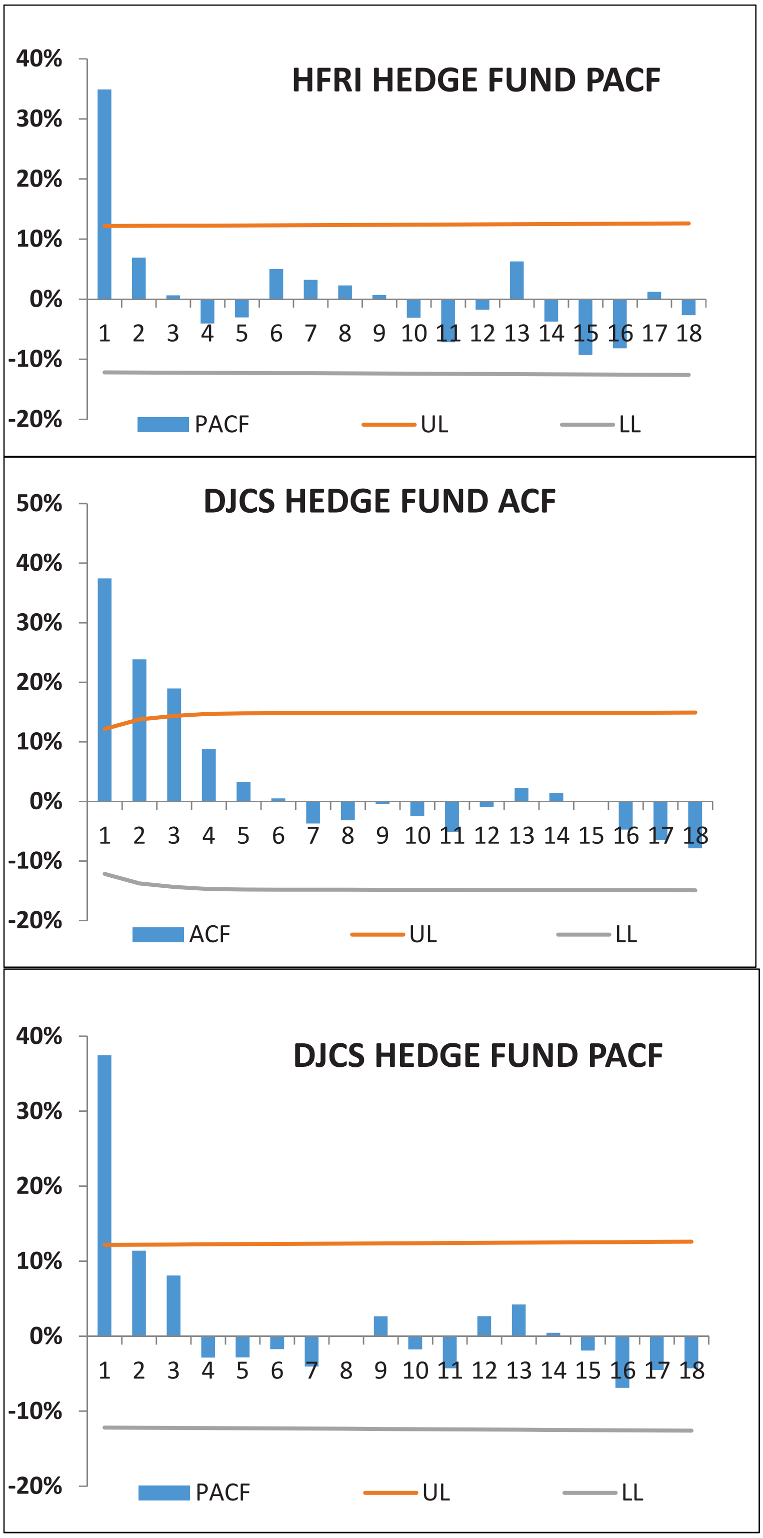


Figure 14. ACF and PACF plot for hedge funds. Source: authors.

Figure 15 illustrates the bond correlograms, and firstly, it shows that the corporate bond, H0AO, has one significant ACF lag, and for the PACF, it has four significant lags, which are lag 1, 2, 3 and 13. This result clearly shows that the asset has serial correlation. To affirm this Table 2, also shows false results for white noise, further confirming it is serially correlated.

However, when looking at the GBI-EM index, there a few significant lags for the ACF (7 and 13) and PACF (7 and 12). This asset was already affirmed to be normally distributed, and this further shows that the white noise test showed a true result, implying that the distribution is not serially correlated.

After looking at the correlation of the returns, it is also necessary to look at the squared values to assess whether there is correlation amongst them. This assists in decisions on the best model to use. Details on the arch effect are summarized in Table 2, and it shows that for bonds, the HFRI hedge fund and HEX equity, the ARCH effect is true, meaning that the squared values are correlated. This is the opposite for the BGSM and DJCS, which tested false in the tests.

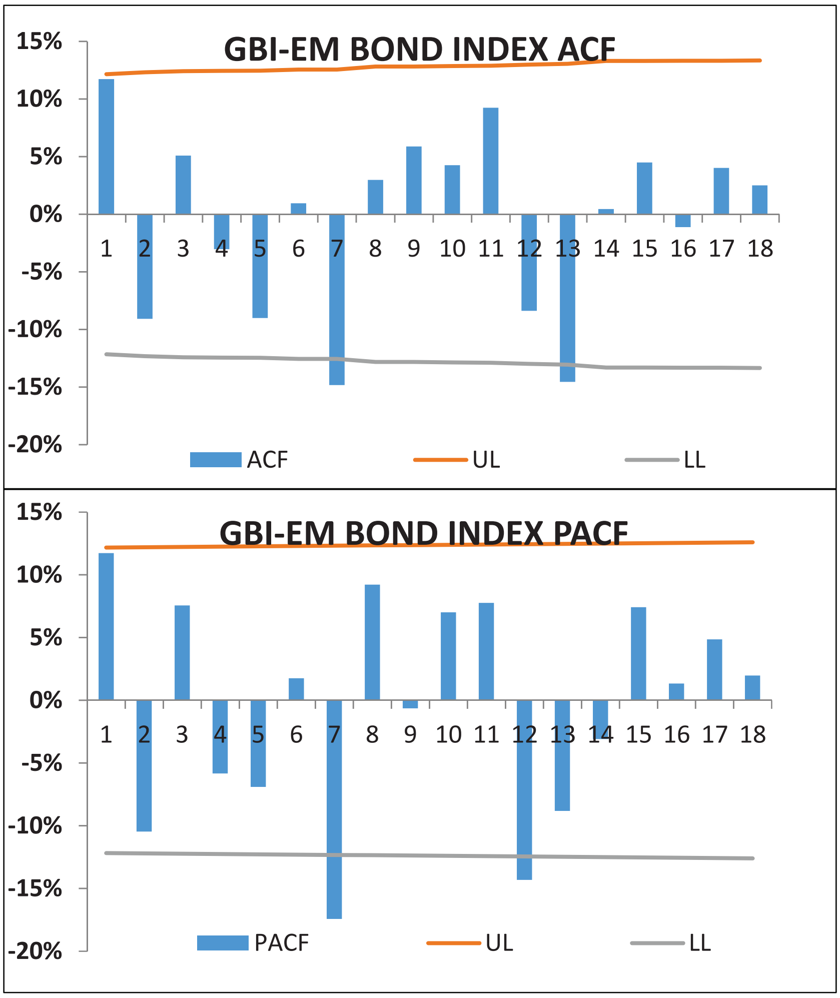


Figure 15. *Cont.*

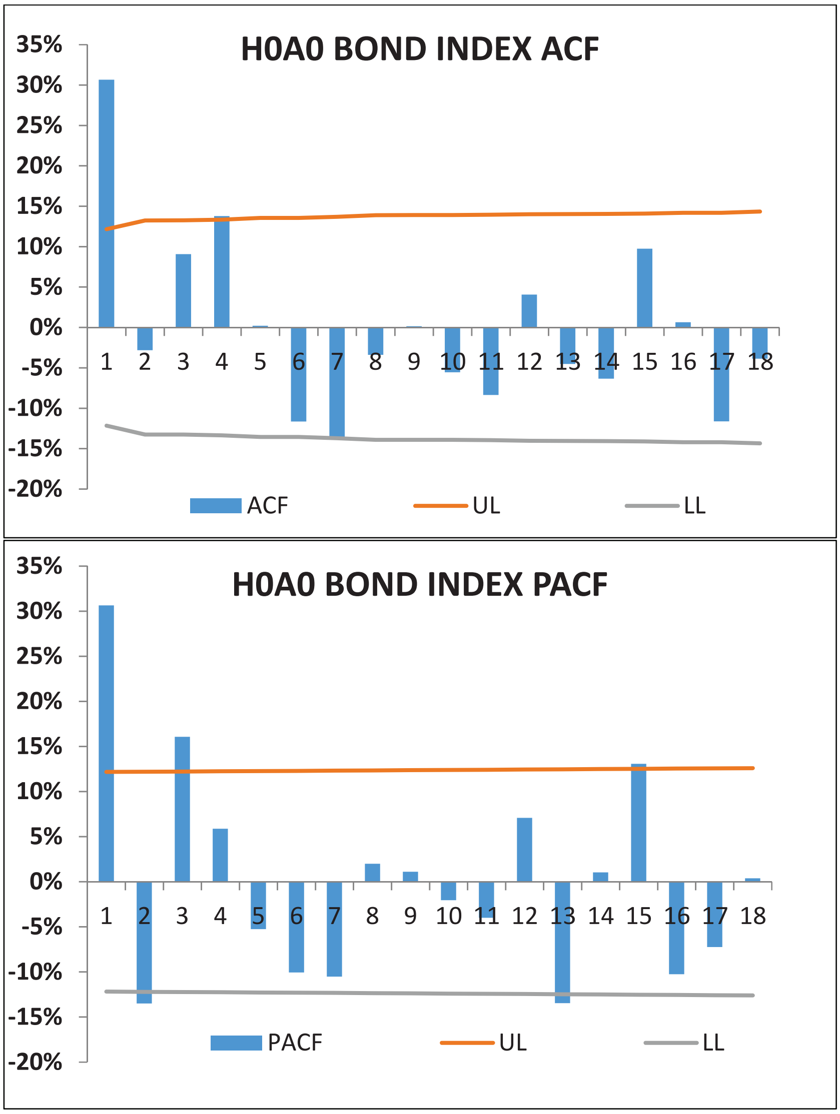


Figure 15. ACF and PACF plot for bonds. Source: authors.

4.10. Correlation Test—Portfolio

In this test, the authors carried out correlation tests between two time periods for each asset. For the purpose of uniformity, the same parameters applied in the stressed value at risk test are applied here as well.

By assessing Table 3, various pieces of information can be deduced, firstly by looking at the correlation of assets in a similar asset class. The equity indices BGSMI and HEX are positively correlated under normal conditions; they move together with a coefficient of 0.05 but, during the stress period, this relationship is reversed, and it can be seen that the correlation is now negative at -0.1. This means under normal market conditions they almost move together, though to a small significance, and soon after, they move away from each other.

The next class is the hedge funds, which both have a strong positive correlation that only strengthens during stress periods. In normal conditions, the correlation is 0.85, and it is 0.96 in a stressed scenario. For the bonds, it can be seen that the correlation coefficient is fairly stable and positive at 0.69 in normal conditions and 0.61 in the stressed period.

Table 3. Correlation matrix for all assets. Source: authors.

	CORRELATION 2003–2005							CORRELATION 2007–2009					
	BGMDC	**HEX**	**HFRI**	**DGCS**	**GBI-EM**	**H0A0**		**BGSMDC**	**HEX**	**HFRI**	**DGCS**	**GBI-EM**	**H0A0**
BGSMDC	1						**BGSMDC**	1					
HEX	0.05	1					**HEX**	−0.10	1				
HFRI	−0.10	0.47	1				**HFRI**	−0.22	0.64	1			
DGCS	−0.23	0.31	0.85	1			**DGCS**	−0.06	0.67	0.96	1		
GBI-EM	−0.14	−0.01	0.37	0.37	1		**GBI-EM**	−0.08	0.19	0.18	0.15	1	
H0A0	−0.34	0.31	0.39	0.29	0.69	1	**H0A0**	−0.22	0.52	0.59	0.55	0.61	1

By making combinations using all assets, it can be seen that the most drastic change in correlation was observed between the HEX-BGSMI equity indices and the HEX equity–GBI-EM bond, in which all the correlations shifted from being negatively correlated to positively correlated. However, for both index combinations, the initial coefficients were close to zero, and so they did not really have a strong relationship to start with. Even during the stress period, the coefficients remained close to zero, with the BGSMI-HEX moving from a positive 0.05 to a negative 0.1 and the HEX-GBI-EM moving from a negative 0.01 to a positive 0.19.

Overall, the results from this test are inconclusive, showing that the behaviour of assets does not rely on the class but on how they relate on an individual level with each other. In coming up with a portfolio, one would have to test all the assets, as was carried out in this section, and observe how each asset relates to the other so that the best risk measures can be applied to cater for periods when the normal volatility of the assets shifts in stress periods.

4.11. ARMA, GARCH, EGARCH Model Parameter

4.11.1. ARMA Parameters

The authors first used the ARMA model to forecast future risk. In coming up with the final forecast, there were a few assumptions made that should be noted. The order used in the simulation was an ARMA (1, 1). Using NUMXL, the first step was to come up with general parameters; after that, the author calibrated the parameters to find the optimal ones and all this conducted in the program.

μ is the long run mean, $\varphi 1$ is the first coefficient of the autoregressive (AR) component (first order correlation), θ,1 is the first coefficient of the moving average (MA) component and finally, σ is the standard deviation of the residuals (this measure looks at the dispersion of points and measures the accuracy of the variable being investigated).

For all six assets, these parameters were calculated to ascertain the final ARMA forecast over a period of 12 months, and the results are detailed in Table 4 below.

Table 4. ARMA parameters. Source: authors.

ARMA Model Parameters						
	Hedge Fund		**Equity**		**Bonds**	
Param	**HFRI**	**DJCS**	**HEX**	**BGSMDC**	**GBI-EM**	**HOAO**
μ	0.0043	0.0068	0.006	0.0135	0.0047	0.0088
φ1	0.5232	0.6602	−0.1408	0.4726	−0.3639	−0.0947
θ1	−0.1931	−0.3308	0.4	−0.0176	0.5243	0.4652
σ	0.0153	0.0165	0.0731	0.0382	0.0164	0.0241

Here, it can be observed that the standard deviation of the residuals is highest for the equities asset class. For the optimal coefficients for the autoregressive process and moving average process, the calibration process selected the ideal figures to input into the best ARMA model for each asset.

4.11.2. GARCH Parameters

In forecasting with GARCH, there were also some model parameters that were calculated and calibrated via the Excel spreadsheet. These are the ones input into the equation to come up with the volatility forecast. μ is the long run mean, $\alpha 0$ is the constant in the conditional volatility equation, $\alpha 1$ is the first coefficient of the ARCH component and $\beta 1$ is the first coefficient of the GARCH component. All of this is summarized in Table 5 below.

Table 5. GARCH parameters. Source: authors.

GARCH Model Parameters						
	Hedge Fund		Equity		Bonds	
Param	**HFRI**	**DJCS**	**HEX**	**BGSMDC**	**GBI-EM**	**HOAO**
μ	0.0043	0.0116	0.0095	0.0125	0.0038	0.0066
α_0	0	0.0001	0.0001	0.001	0.000	0.0001
α_1	0.1533	0.4621	0.1643	0.6781	0.0836	0.2496
β_1	0.8143	0.4417	0.8181	0	0.8204	0.6705

The GARCH model takes the ARMA process a step forward and factors in the conditional volatility element as well as the other GARCH components. The first coefficient of the ARCH component across all assets has a parameter close to zero.

4.11.3. EGARCH Parameters

Forecasting with the EGARCH model of the 1, 1 order has the parameters shown in Table 6 below.

Table 6. EGARCH parameters. Source: authors.

EGARCH Model Parameters						
	Hedge Fund		Equity		Bonds	
	HFRI	**DJCS**	**HEX**	**BGSMDC**	**GBI-EM**	**HOAO**
μ	0	0.01	0.01	0.01	0	0.01
$\alpha 0$	−0.45	−3.58	−0.53	−3.36	−7.9	−0.55
$\alpha 1$	0.28	0.15	0.35	−0.48	0.23	0.1
$\gamma 1$	0.209106	−3.04519	−0.10349	0.198544	1.066853	−1.99898
$\beta 1$	0.971952	0.584361	0.952573	0.568996	0.064752	0.941026

The μ is the long run mean; $\alpha 0$ is a constant in the conditional volatility equation; $\alpha 1$ is the first coefficient of the ARCH component, $\gamma 1$ is the first leverage coefficient and $\beta 1$ is the first coefficient of the ARMA component. The EGARCH has "the constant volatility embedded in the leverage coefficient which if included in a model of volatility it amplifies the volatility" (Engle and Siriwardane 2014).

4.11.4. ARMA, GARCH, EGARCH Model Forecasts

The models were used intend to capture the autocorrelation of squared returns and the volatility going back to the mean. These models took non-normal components into account, such as excess kurtosis, and factored them into the forecasts. The NumXL ARCH/GARCH modelling wizard automated the steps used to construct an ARCH model: guessing initial parameters, parameter validation, the goodness of fit testing, and residuals diagnosis. Therefore, by using these parameters, the final volatility forecasts were made using three models for the two equity assets, and Figure 16 shows the 12-month forecasts.

4.11.5. Equities Model Forecast

The first analysis is for the two equity assets, and Figure 16 gives the three forecasts made using all three models: ARMA, GARCH and EGARCH.

First, the HEX equity index and BGSMDC equity index and their 12-month volatility forecast based on the previously mentioned parameters are observed in Figure 16. It can be seen that the EGARCH has the lowest volatility forecast, followed by ARMA, and the highest is the GARCH model. Slight increases are anticipated in the first period, then they slowly taper off to become stable. This indicates that the nearer the forecast, the higher the level of accuracy. That is why the authors limited this to a 12-month forecast, as only current and relevant information affected the volatility forecast.

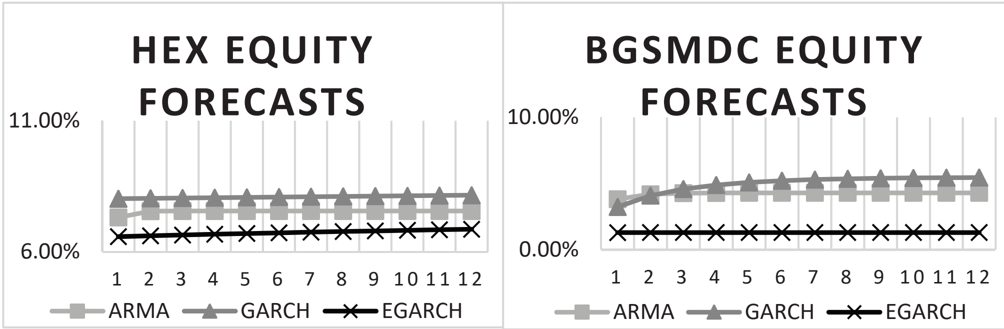


Figure 16. Equities model forecasts (12 months). Source: authors.

4.11.6. Hedge Fund Model Forecasts

For the hedge funds, the results are similar to the equity, with the EGARCH recording the lowest volatility relative to GARCH and ARMA. However, for the GARCH forecast, it is lower than the ARMA forecast for the HFRI (See Figure 17). This could be down to the difference in asset characteristics. The GARCH for both has a steady upward increase, whilst the ARMA indicates a steady, almost flat increase for the DJCS, which is different from the HFRI composite index, which increases up in the first period and then tapers off. This shows higher initial anticipated volatility for the composite and lower volatility for the event driven fund, which already anticipated the event shocks in its historic data.

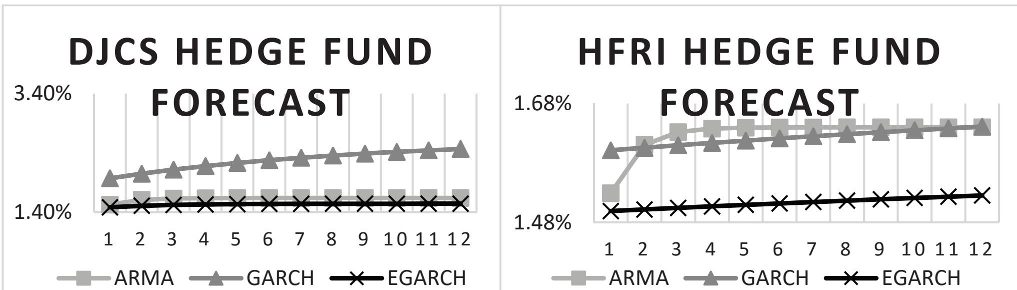


Figure 17. Hedge fund model forecasts (12 months). Source: authors.

4.11.7. Bond Model Forecasts

The bond forecasts have varying results from both the equity and hedge funds. It can be observed that the corporate bonds HOAO forecast has a higher volatility using the ARMA model but is similar to all other assets with the lowest EGARCH forecast. On the other hand, the sovereign GBI_EM bond forecast has higher volatility in the first period, falling sharply after that, with the lowest forecasts for the rest of the periods. In this situation, the GARCH has the highest volatility of all three (See Figure 18).

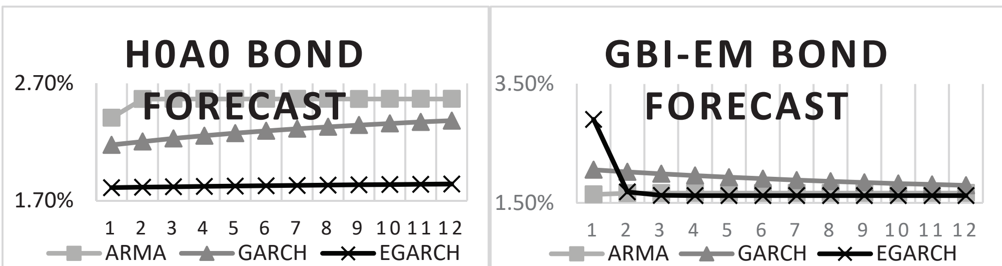


Figure 18. Bonds model forecasts (12 months). Source: authors.

After carrying out the tests for the three asset classes, there is need to adopt the best model to use. It cannot easily be interpreted by the forecast tables, and since the parameters used for all three assets are not similar, this is where the Akaike Information Criterion (AIC) test is considered. Akaike Information Criterion (AIC) is a method that is used to measure the goodness of fit of a model. Ideally, its lowest value is the best model to use. This can be used to determine which method between the three is ideal based on this test in order to obtain the best volatility forecast. Below is a table showing the results of the three tests for the six assets. In general, the AIC is defined as:

$$\text{AIC} = 2\text{k} - 2 \times \ln(\text{L})$$

where:

k is the number of model parameters.table
ln(L) is the log-likelihood function for the statistical model.

By looking overall at the above results, it is clear that the asset class with the lowest AIC is generally the hedge funds and bonds, whilst equity has the highest. This can therefore help to conclude that when applying these models, it is most suitable for hedge funds and bonds and not as much for the equities. However, it must be noted that the AIC figure for EGARCH of the BGSMDC is unavailable for comparison (See Table 7).

When analysing based on the model type, it can be seen that EGARCH is the most accurate model to use because for hedge funds and bonds, it has the lowest AIC for the three methods per asset. To analyse this further, the next step of assessing the volatility forecast for each asset critiques this discovery and seeks to validate it or otherwise.

Table 7. AIC model goodness of fit test for ARMA, GARCH and EGARCH. Source: authors.

AIC Model Goodness of Fit Test						
	Hedge Fund		**Equity**		**Bonds**	
	HFRI	**DJCS**	**HEX**	**BGSMDC**	**GBI-EM**	**H0A0**
ARMA	−1445.59	−1404.57	−614.82	−953.59	−1407.94	−1192.77
GARCH	−1457.21	−1391.03	−655.40	−938.08	−1404.11	−1276.66
EGARCH	−1459.71	−1422.17	−647.43		−1410.05	−1290.07

5. Discussion

5.1. Non-Normality Test Results

The normality tests indicated that for the hedge funds, all 11 were not normally distributed; for equities, 19 out of 23 assets were not normally distributed, with the exception of the Japanese indices that exhibited normality. When testing the bonds, it was also established that 9 out of the 10 data sets were not normally distributed, with the exception of the JP Morgan Government Bond Index-Emerging Markets (GBI-EM), which covers comprehensive emerging market debt benchmarks that track local currency bonds issued by emerging market governments.

The normality tests also yielded an analysis on various factors, and this was performed on the six selected assets. One was the standard deviation analysis, where it was discovered that hedge funds tend to have a lower standard deviation than equities, implying that equities are riskier than hedge funds. These are just generalized assumptions based on the initial analysis. For bonds, the level of risk differs based on the underlying asset. Corporate bonds have a higher standard deviation than sovereign bonds. The summary on the skewness was that most of the assets are negatively skewed with long tails, showing that they have a high downside risk. The results for the kurtosis showed that most of the distributions are leptokurtic. Another element in the summary was the ARCH effect, which gives details on the distribution's fat tails and excess kurtosis, and all of this was carried out to further affirm the non-normality of the returns. Once the argument on the type of distribution was resolved, the next step was to assess the risk measures relevant to this distribution.

VaR versus CVaR

After carrying out the analysis, it can be said that the CVaR is an extension of VaR that gives the total amount of loss given a loss event. If one is using VaR, according to Kidd (2012), the main question would be how often their portfolio could lose at least X amount of euros, whilst one using the CVaR method is concerned with answering the question of how much it could lose when its portfolio's losses exceed X amount of euros.

From the tests undertaken in the analysis, it can be seen that CVaR is a better measure for downside risk than VaR, as it has higher risk measures and is more sensitive to the shape of the tail loss the non-normal distributions. From an asset class perspective, it can be seen that the equities generally have a higher risk measure compared to the hedge funds. However, this is merely a generalization because what drives risk in an asset is not only its class but also what the asset is composed of. When drawing conclusions for this, it is critical to factor in other items such as the composition of the asset, its liquidity features, the industry, the geographic location and all other risk factors so that one can obtain an even more accurate measurement. This means combining this analysis asset by asset with the information in Figure 3, which gives information on the global risk landscape according to the world economic forum survey (World Economic Forum 2017).

5.2. VaR versus Stressed VaR

The results from the stressed VaR analysis show that when using different time horizons to compute the VaR and SVaR, it was observed that the returns are lower for the daily-observed data in comparison to the monthly observations. This shows that one needs to be aware of the reason for carrying out the tests so that they can use the correct data time horizons to obtain accurate results and not end up understating or overstating the risk measure. If it is for daily use, it would be ideal to use a shorter time to calculate, and if it is to observe an asset over a longer period, then the monthly time horizon could be adequate. All of this is situational, and so other factors should be taken into account together with this information.

Another interesting observation made was when looking at the bonds whereby the one asset was normally distributed and the other was non-normal. It was observed that the normally distributed GBI-EM had VaR and SVaR measures that were close together, unlike what was observed in the other asset classes. This therefore showed that the normal distribution, due to it not having the presence of fat tails, had less risk observed and so the returns were generally uniform, even in stressed time periods. This means that the standard VaR would be a good measure for this, but for all other data that is non-normal, it is critical to isolate the stress period and calculate its risks in comparison to a normal period. This makes SVaR a good measure to use in volatile periods because it captures the market risk inherent at that time. By knowing these measures, it can aid in improving the risk measures, in particular when forecasting data whereby the expectation is a downward one. This will enable the investor to assess if they have taken on excessive risk which may lead to their bankruptcy should another stress period occur. This then can be adjusted to suitable risk levels that do not threaten the loss of the assets.

5.3. Time Series Analysis

The time series analysis is useful in helping to identify the way security moves over time. By looking at the past trends, it can aid in anticipating the types of risks that affect that particular stock. For instance, one can refer to the World Economic Forum Global Risks Perception Survey 2016 indicated above, which identifies highly probable risks such as cyber-attacks and natural disasters. Using this information, one can look at the asset and what it's underlying asset is, and so determine the likelihood of adding more risk to it or reducing risk. For instance, if an index tracks agriculture companies that are based in an area where there is increasing amounts of natural disasters that occur and affect the industry, this means it is likely to affect the profitability and so lower returns will be forecasted. If an investor had calculated a VaR up to a certain point, it could now adjust that point based on anticipated risks it would have identified by analysing the time series.

It is also worth noting that when looking at a time series, it is essential to look at the underlying asset, as it can also give information on how the time series behaves. For instance, the HFRI hedge fund index, which is a fund that is managed by a diverse array of fund managers, is likely to exhibit a time series with a few extreme fluctuations, indicating that there is less volatility, probably resulting from the diversification aspect.

By using this information in conjunction with information obtained from the variations of VaR previously observed, one can decide whether the risk figure must be adjusted to accommodate the added risk based on the historic behaviour of the time series. The time series analysis also has a plot of the WMA; the authors are able to visually determine the stationarity of the time series. The EWMA plot shows the volatility of the asset over time, and it is easy to visually observe the periods of volatility clustering. The EWMA shows the periods with large returns that are clustered and very distinct from those observed periods of small returns, which may also be clustered. Not only does the volatility spike up during a crisis or stress period, but also it eventually drops back to approximately the same level of volatility as before the crisis, which is an indication of stationarity. This is a vital property which affirms the suitability of the use of advanced forecasting models in the time series volatility forecasting. A stationary time series is a time series that is mean reverting, meaning that it moves close to the mean.

5.4. Correlation Test—Single Asset

In time series analysis, it is said that a correlogram/analysis is one such tool that is necessary in identifying the model to use in the volatility forecasts. The correlogram analysis examines the time–spatial dependency within the sample data, and focuses on the empirical auto-covariance, auto-correlation and related statistical tests. It is key to the discovery of the interdependency of the observations.

Upon carrying out the tests, it can be seen that five of the six assets exhibit serial correlation. These results further affirm the general hypothesis that the returns of several asset classes tend to be serially correlated. For instance, it is commonly expected in hedge funds because they usually hold illiquid assets which can be hard to price and so for the fund managers to determine the prices, they rely on past information to derive the value. This therefore has an impact on the ideal method to use to forecast volatility in the next section.

The analysis after this discovery also looked at the squared values to assess if there is correlation amongst them. The results showed that even though the returns showed correlation, the ARCH effect showed that in both bonds, the HFRI hedge fund and HEX Equity, the ARCH effect is true, meaning that the squared values are correlated. This is the opposite for the BGSM and DJCS, which recorded false results in the tests. Overall, an assumption of correlation can be assumed and a GARCH model may be used.

5.5. Correlation Test-Portfolio Context

After carrying out all the tests and looking at the relationship between assets of a similar class or different combinations, it can be concluded that there is no particular pattern on asset classes and how they interact with each other in normal versus stressed periods. Therefore, it can be advised that for each portfolio, a portfolio manager needs to generate a similar correlation matrix to assess and see which assets are correlated, how they behave with each other during stress periods and how they are under normal periods. In the example used in the analysis, selecting the two hedge funds will a good example because whether or not the assets are in a normal or stressed period, they seem to maintain a correlation that is positive and close to each other. Therefore, this will allow an asset manager to maintain his portfolio in the same risk levels if that is his investment approach.

If there were a need to diversify and hedge, the portfolio manager would pick the assets with the lowest correlation or those that move in opposite directions, such as the two equity assets that switch correlations from being negatively correlated to being positively correlated during times of stress. Regardless of what the change in the correlation is, it is essential for the investor to know which direction volatility of the assets is likely to move. This will allow one to anticipate which changes may occur depending on how the assets relate to each other and how it affects the level of risk at a given time or when planning for future risk. Failure to correctly anticipate market risk may overstate the amount of diversification in a portfolio that can lead to an investor taking on excessive risk.

5.6. Volatility Forecast (ARMA, GARCH and EGARCH)

Beg and Anwar (2014) view the arrival of new information as a key driver of volatility in asset returns. According to the authors, the constant release of new information results in financial assets prices showing volatility, which in turn affects the financial risk analysis and risk management strategies. To consider this factor, models such as the Generalized Autoregressive Conditional Heteroscedasticity model can be used to analyse financial data.

Upon carrying out the studies, it can be said that the use of these volatility forecasts (ARMA, GARCH, EGARCH) is useful for non-normal distributions to analyse an asset's risk as well as market risk. This is because these models take into account the use of various parameters such as the use of a leverage coefficient in calculating volatility with the GARCH model. This can affect an asset's risk and incorporate them into the model. The type of model lags to consider can also be determined by looking at the ACF and PACF plots, as was carried out. These plots are useful in determining the best lag order to use for each individual asset class. In our case, most of the plots showed an ACF with one lag and then decay. Thus, for this study, the authors applied the order of 1, 1. However, in further use of these models, it is critical to look at other elements of the data, such as the stationarity, as all of this helps to determine whether the models are suitable for use with the chosen data set.

In the data analysis section, after the parameters were calculated and defined for all three models, the next step was to plot the 12-month forecasts using each model. It was generally observed that the GARCH model had the highest volatility of all three for each asset, whilst the EGARCH was the lowest. The main advantage of GARCH models is that they captured jointly heavy tails and volatility, which contribute to clustering (Jansky and Rippel 2011). This demonstrates a characteristic of the data sets being observed (except for the GBI-EM bond index, which is normally distributed and so its plot was very different from the other two, as it had a high EGARCH at period one which sharply declined.

Overall, when one examines the literature regarding the use of these models, it can be seen that they are vastly expanding to cater to different elements within the data set. Thus, it can be concluded that when carrying out a risk forecast, any of these models can be applied, but it all depends on the asset characteristics and what risk factors to consider.

5.7. AIC Test for Best Model

Akaike Information Criterion (AIC) was used to measure the goodness fit for each of the three models. Ideally, its lowest value is the best model to use, and in the test, it was seen that the EGARCH was seen as the best fit because it had the lowest AIC overall for all the assets examined (See Table 7).

Results from this test therefore give a guideline to the best method to use to forecast risk for these non-normally distributed assets (with the exception of the normally distributed JP Morgan Government Bond Index-Emerging Markets (GBI-EM)) out of all the methods. Hedge funds and bonds seemed best suited, with the lowest AIC when looking at asset classes. When looking at the best method to use overall, the EGARCH reigns superior, with the lowest AIC for the three methods per asset. Further tests to see which is better between GARCH and EGARCH could be carried out, as suggested by Malmsten (2004). He considers some of these tests to be symmetric models and asymmetric models. However, these go beyond the scope of this paper, but it is worth noting that various other tests can be used to determine the validity of model choices apart from the AIC, which is used in this study.

Results summary

VaR has become a popular method in measuring risk, and due to its simplicity, it became an overnight sensation; however, the main distribution assumption of normality led to it being discredited, leading to more investors moving away from it and focusing on its hybrids that take into account the fat tails. Although value at risk has been a measure that has been used in the recent past by most investors when measuring the risk exposure, it was inadequate when calculating how much would be lost, and this situation usually led to some investors underestimating risk measures such that when a crisis occurred, most of them suffered to the brink of collapse and bankruptcy. This led to the creation of other variations of this risk measure that specifically focus on the downside risks, and these included the conditional value at risk and stressed value at risk, amongst many. These were tested, and it was seen that the CVaR had a higher risk measure compared to the standard VaR, as it took into account the existence of heavy tails, and so this was a better measure of risk and would shield the asset holder from overexposure of their asset or portfolio which would lead to bankruptcy. Another measure that was considered was the stressed value at risk, and this measure showed that volatility differs during the stressed period and so if it is measured separately, it would give a better picture of how risky an asset could get in periods of high volatility.

The next step involved analysing the time series, and it was seen as a critical step in the research because it showed how the assets behave and can be used to see if any patterns are prevalent in the behaviour of the assets as they move over time. Observing a time series helped give information on issues such as the stationarity of the time series, which allow for the use of certain econometric models to establish volatility forecasts. After this, the test of correlation was performed for each asset to observe whether or not correlation of an asset's prices with one another exists. This was proven to be true, and so based on these data, it can be concluded that price dependence is a major theme occurring in the distribution of returns showing long periods of price dependence. It was also observed that there is the existence of price jumps, which are all key to obtaining a more realistic risk measure.

The next test was to find and compare volatility forecasting over the time horizon using three selected methods, namely ARCH, GARCH and its hybrid EGARCH. These methods were ideal to use because financial time series returns are skewed and fat tailed, and these models are geared to take this into account. They were able to capture the volatility clustering and took into account changes in trends, and all these were factored into the parameters which the authors calibrated using the NUMXL function in Excel to obtain the best combination of parameter values to use. The benefits of using these models are also that depending on the characteristics of the data being observed, there are hybrids to each method to take this into account. GARCH, for instance, has many of these, one of which is EGARCH, which was used here, and others include FIGARCH, HYGARCH and so forth. These model extensions are very relevant for the conditional volatility of financial returns.

However, in looking at the various methods, one must take into account and beware of the limits of statistics, as one cannot fully rely on this method for accurate measures. The data sets used in this paper aimed to offer a more robust representation of various assets and classifications of the market, such as sampling from a low developed economy, a composite index, a specialized index and so forth, but this does not mean all risks will be taken into account in the asset pricing or risk allocation in portfolios. There is still the risk of risks that are not quite well understood to either come in and drive returns up or down drastically, thus resulting in price jumps. Therefore, it can be observed that by having a combination of the risk measures which take into account other elements such as scenario analysis which involves looking at other macroeconomic variables, such as stress testing, which tests just how risky an asset can become, and forecasting with risk models, the ideal method for an investor would be to combine all elements to ensure they obtain the best risk measure possible.

6. Conclusions and Future Research

This paper made numerous attempts to elucidate the complexity of measuring risk. In order to obtain the best possible accurate risk measures, we recommend that researchers should incorporate historical data and adopt scenario analysis for a more robust outcome. This calls for observing asset behaviours and how they respond to external shocks and normal periods. This also means looking at an asset and how its individual returns are autocorrelated, and this will reveal how much of an impact historical data has on the future pricing of the asset. The findings also reveal several stylized facts about financial markets, such as fat tails in asset distributions, the absence of autocorrelations of asset returns, volatility clustering and aggregational normality.

Incorporating non-normality in risk calculations should provide a better way of understanding and measuring the downside risk that is in an asset. The models that have been analysed in this paper considered this, and so the results that are obtained reflect this aspect. All of the methods that were explored in the analysis section shed some light into the many options available that can be used to assess risk in non-normal distributions. One aspect identified was that if the methods are used together and not in isolation, they complement each other by addressing the shortcomings of the other to give the closest to accuracy in the measures. For instance, VaR can be used as the standard measure, and then adjusted based on the information obtained from the time series. However, it must be noted that there is no single risk measurement framework that can ensure portfolios and assets are immune to unexpected risk exposure, but by using a mosaic of methods, we can come close to this.

To summarize the study and tests conducted, the first step of testing the assumption that returns tend to be non-normal as opposed to the general assumption of normality was examined. The interpretation of findings from the tests clearly showed that the majority of the returns are not normally distributed, but instead are skewed and also possess excess kurtosis. Thus, it would be significant in the approach of assessing risk to first test the asset's historical data for normality. In some cases, one may find a distribution that is normal, such as the one found for the bonds index GBI-EM, in which case a basic risk test such as VaR could be adequate to some degree.

Finally, in order to obtain the most accurate possible risk measures, the best approach is not to just move away from the use of historic data. However, creating a balance yields better risk measures, whereby historic data are used to assess the behaviour of the asset and how it responds to shocks and normal periods by doing a time series analysis. This can also include looking at the asset and how its individual returns are correlated to each other (autocorrelation) to see how much of an impact previous data has on the future pricing of the asset. Since all models discussed have both merits and weaknesses, for the ones under non-normal assumption discussed in this paper, the risk measures that will be calculated will be closest to the level of accuracy in reflecting the market conditions.

7. Research Limitations

The ARCH and GARCH model family has been evolving since its introduction, such that a multitude of new varieties of these have emerged (Bollerslev et al. 2010). This poses a limitation on to how many varieties this study could have used due to the vastness of computations needed. This makes picking the best model to forecast volatility all types of financial data a bit cumbersome for this particular study. Another limitation faced was the availability of more recent data to adequately analyse the impact that COVID has had over the past year and more.

Another major limitation faced in providing universal data results was due to inadequate computing power to dissect all available asset classes. Instead, we picked three relatively common ones to apply. Even within these three asset classes, only two from each were used for volatility forecasting. For instance, the authors had 60 bond indices but only had to analyse two. It is hoped that in future studies, we can evaluate other asset classes such as forex, futures and other derivatives to apply the tests used in this study.

Future Research

Future research papers will be undertaken, implementing different asset classes as well as newer GARCH models for forecasting risk in the continued search for the most accurate risk measures. The authors are currently gathering information in the present COVID era to further study its impact on risk measures. It would be interesting to assess the short-term and long-term impacts the pandemic has had. This will allow the authors to gain a clearer picture on the future of similar events and find ways to circumvent negative financial crisis. Once the data are fully available, a follow up paper should be written to assess this new variable.

Another research area the authors wish to investigate are crypto currencies as an "asset class" more so than a digital currency. It would also be interesting to assess their price movements and behaviour and see how they affect the performance of other assets, as well as the type of risks involved with using these.

8. Patents

This section is not mandatory but may be added if there are patents resulting from the work reported in this manuscript.

Author Contributions: Conceptualization: E.S. & Y.T.M. methodology—Y.T.M. software, Y.T.M. validation, E.S. and Y.T.M. did formal analysis, Y.T.M. investigation, E.S. and Y.T.M. resources, E.S. and Y.T.M. data curation—E.S. and Y.T.M. writing—original draft preparation, E.S. and Y.T.M. writing—review and editing, E.S. visualization, E.S. supervision, E.S. project administration, N/A funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Jarque Bera Test On Hedge Fund Data 1995–2016.

	ASSET CLASS: HEDGE FUNDS					
	PERIOD: 1995–2015	**SKEWNESS**	**KURTOSIS**	**JARQUE BERA TEST**		
	FREQUENCY: MONTHLY DATA			***Test Stat***	***p-Value***	***Normality YES/NO (Y/N)***
1	**HFRIMI Index**	0.5535	0.830	21.053	2.68105×10^{-5}	N
2	**HFRI Fund of Funds Composite Index**	−0.6913	4.362	230.29	9.8651×10^{-51}	N
3	**HFRI Relative Value (Total) Index**	−2.7736	15.900	3119.6	0	N
4	**Dow Jones Credit Suisse Equity Market Neutral Hedge Fund Index**	−14.5592	227.511	578703	0	N
5	**Dow Jones Credit Suisse Convertible Arbitrage Hedge Fund Index**	−2.7917	18.642	4165.5	0	N
6	**Dow Jones Credit Suisse Fixed Income Arbitrage Hedge Fund Index**	−4.7834	37.000	16065	0	N
7	**Dow Jones Credit Suisse Event Driven Hedge Fund Index**	−2.1210	9.912	1278.7	2.1944×10^{-278}	N
8	**HFRI Merger Arbitrage Index**	−1.4211	5.903	472.21	2.8841×10^{-103}	N
9	**Dow Jones Credit Suisse Event Driven Distressed Hedge Fund Index**	−2.1875	12.024	1800.8	0	N
10	**Dow Jones Credit Suisse Dedicated Short Bias Hedge Fund Index**	0.7534	1.576	52.284	4.43371×10^{-12}	N
11	**Dow Jones Credit Suisse Long/Short Equity Hedge Fund Index**	0.0002	4.015	177.29	3.18098×10^{-39}	N

Note: N—not normal. Source: authors.

Appendix B

Table A2. Jarque Bera Test On Global Equity Indices Data 1995–2016.

	ASSET CLASS: HEDGE FUNDS PERIOD: 1995–2015 FREQUENCY: MONTHLY DATA	SKEWNESS	KURTOSIS	JARQUE BERA TEST		
				Test Stat	*p-Value*	*Normality YES/NO (Y/N)*
1	**HEX Index**	0.1334	2.0337	45.5759	0.0000000001	N
2	**OBXP Index**	−0.9527	2.7752	122.7647	0.0000000000	N
3	**WIG20 Index**	0.6914	4.3217	223.0550	0.0000000000	N
4	**SWIG80 Index**	0.5629	1.8088	49.1723	0.0000000000	N
5	**BUX Index**	−0.0829	3.2649	115.7762	0.0000000000	N
6	**XU100 Index**	1.3929	6.6046	556.6424	0.0000000000	N
7	**MOSENEW Index**	0.5202	2.2051	64.4031	0.0000000000	N
8	**BGSMDC Index**	2.4294	12.9725	2078.8516	0.0000000000	N
9	**SEMDEX Index**	0.0909	3.5010	133.1447	0.0000000000	N
10	**TPX500 Index**	−0.2149	0.4962	4.6678	0.0969191846	Y
11	**TPXSM Index**	−0.0077	−0.0732	0.0606	0.9701371267	Y
12	**TPXM400 Index**	−0.1909	0.4065	3.3701	0.1854371291	Y
13	**TPX100 Index**	−0.1852	0.4876	4.0620	0.1312014021	Y
14	**HSI Index**	−0.0250	2.2148	52.9616	0.0000000000	N
15	**HSCEI Index**	0.7712	3.8577	186.2729	0.0000000000	N
16	**SHCOMP Index**	0.2298	1.3361	21.5439	0.0000209803	N
17	**AS30 Index**	−0.7768	0.9155	35.0920	0.0000000240	N
18	**AS52 Index**	11.3502	127.9268	182,873.0544	0.0000000000	N
19	**NZSE Index**	−0.6398	1.1004	30.8580	0.0000001992	N
20	**CSEALL Index**	0.4701	1.1695	24.3918	0.0000050511	N
21	**SET Index**	0.1738	2.6434	77.0079	0.0000000000	N
22	**FBMKLCI Index**	0.6150	5.8711	389.8191	0.0000000000	N
23	**PCOMP Index**	0.1903	4.5279	223.6748	0.0000000000	N

Source: authors.

Appendix C

Table A3. Jarque Bera Test On Global Bond Indices Data 1995–2016.

	ASSET CLASS: HEDGE FUNDS					
	PERIOD: 1995–2015	**SKEWNESS**	**KURTOSIS**	**JARQUE BERA TEST**		
	FREQUENCY: MONTHLY DATA			*Test Stat*	*p-Value*	*Normality YES/NO (Y/N)*
Index						
DGOV	**TRR Index Val LOC U**	−2.87460108	23.259846	6314.8116	0	N
	PRR Index Val LOC U	−3.00174952	24.012297	6738.9563	0	N
	Market Value ex Cash LOC U	−2.4113742	17.219024	3517.2905	0	N
EMGB	**TRR Index Val LOC U**	−3.11385549	26.687749	8261.2239	0	N
	PRR Index Val LOC U	−3.26714759	27.654971	8882.4388	0	N
	Market Value ex Cash LOC U	−1.80246264	16.480611	3130.6662	0	N
H0A0	**TRR Index Val LOC U**	−1.25072768	9.9051886	1148.0704	5×10^{-250}	N
	PRR Index Val LOC U	−1.36111821	10.247232	1236.5797	3×10^{-269}	N
	Market Value ex Cash LOC U	−0.62749331	6.1434245	432.48321	1.22×10^{-94}	N
R0A0	**TRR Index Val LOC U**	−0.48191364	5.0908619	295.30421	7.51×10^{-65}	N
	PRR Index Val LOC U	−0.97506108	6.9261038	569.51279	2.1×10^{-124}	N
	Market Value ex Cash LOC U	1.032799838	6.9753815	582.14914	3.9×10^{-127}	N
R0O0	**TRR Index Val LOC U**	−14.8031114	232.9907	606,773.14	0	N
	PRR Index Val LOC U	−3.29260321	29.006072	9731.8887	0	N
R0U0	**TRR Index Val LOC U**	0.263632866	4.1566215	193.11063	1.17×10^{-42}	N
	PRR Index Val LOC U	−0.16858932	8.2990509	758.8673	1.6×10^{-165}	N
	Market Value ex Cash LOC U	1.350602438	10.142414	1211.8157	7.2×10^{-264}	N
R0M0	**TRR Index Val LOC U**	0.399379484	11.203383	1387.6919	4.6×10^{-302}	N
	PRR Index Val LOC U	0.048615775	10.789126	1280.5615	8.5×10^{-279}	N
	Market Value ex Cash LOC U	0.917248411	5.3501174	351.88049	3.89×10^{-77}	N

Table A3. *Cont.*

	ASSET CLASS: HEDGE FUNDS PERIOD: 1995–2015 FREQUENCY: MONTHLY DATA	SKEWNESS	KURTOSIS	JARQUE BERA TEST		
				Test Stat	*p-Value*	*Normality YES/NO (Y/N)*
Index						
R0C0	**TRR Index Val LOC U**	−1.65728372	27.476698	8425.5081	0	N
	PRR Index Val LOC U	−1.92002596	30.735521	10553.601	0	N
	Market Value ex Cash LOC U	0.99740086	3.4861319	177.45584	2.92×10^{-39}	N
R010	**TRR Index Val LOC U**	−3.12246366	16.642397	3475.6535	0	N
	PRR Index Val LOC U	−3.31780081	18.387309	4203.3677	0	N
	Market Value ex Cash LOC U	1.104372396	0.2637496	54.635463	1.37×10^{-12}	N

Source: authors.

References

Acerbi, Carlo, and Dirk Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26: 1487–503. [CrossRef]

Alexander, Carol, and Elizabeth Sheedy. 2008. Developing stress testing framework based on market risk models. *Journal of Banking and Finance* 32: 2220–36. [CrossRef]

Artzner, Philippe, Freddy Delbaen, Jean Marc Eber, and David Heath. 1999. Coherent measure of risk. *Mathematical Finance* 9: 203–28. [CrossRef]

Basel Committee on Banking Supervision. 2009. *Principles for Sound Stress Testing Practices and Supervision: January, 2009*. Basel: Bank for International Settlements (BIS).

Baker, Kent, and Greg Filbeck. 2015. *Investment Risk Management, Financial Markets and Investment Series*. Oxford: Oxford University Press.

Barns, Dennis, Gildas Blanchard, and Thorsten Lehnert. 2017. Volatility measures and value at risk. *International Journal of Forecasting* 33: 848–63.

Beg, Alam Rabiul, and Sajid Anwar. 2014. Detecting Volatility Persistence in GARCH Models in the Presence of the Leverage Effect. *Quantitative Finance* 14: 2205–13. [CrossRef]

Berner, Richard. 2012. Stress VaR and systematic risk indicators, private sector application of how to measure systematic interconnectedness. In *IMF Conference on Operationalizing Systematic Risk Monitoring*. New York: Morgan Stanley, May 28.

Bhowmik, Roni, and Shouyang Wang. 2020. Stock market volatility and return analysis: A systematic literature review. *Entropy* 22: 522. [CrossRef]

Billio, Monica, Lorenzo Frattarolo, and Pelizzon Loriana. 2016. Hedge Fund Tail Risk: An Investigation in Stressed Markets. *The Journal of Alternative Investments Spring* 18: 109–24. [CrossRef]

Bodie, Zvi, Alex Kane, and Allan Marcus. 2001. *Finance Investments*, 5th ed. London: McGraw-Hill.

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Bollerslev, Tim, Jeffrey Russell, and Mark Watson. 2010. *Volatility and Time Series Econometrics*. Oxford: Oxford University Press.

Box, George Pelham, and Gwilym Jenkins. 1976. *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day.

Brooks, C. 2014. *Introductory Econometrics for Finance*, 3rd ed. UK: University of Cambridge.

Chen, James Ming. 2014. Measuring market risk under Basel accords: VaR, stressed VaR and expected shortfall. *The IEB International Journal of Finance* 8: 184–201.

Christoffersen, Peter. 1996. Evaluating interval forecasts. *International Economic Review* 39: 841–62. [CrossRef]

Crouchy, Michel Dan Galal, and Robert Mark. 2014. *The Essentials of Risk Management*, 2nd ed. London: McGraw-Hill.

Iuga, Iulia Cristina, and Anastatsia Mihalciuc. 2020. Major Crises of the XXIst century and impact on economic growth. *Sustainability* 12: 9373. [CrossRef]

Danielsson, Jon, Bjørn Jorgensen, Gennady Samorodnitsky, Mandira Sarma, and Casper G. De Vries. 2005. *Subaddivity Re-Examined: The Case for Value at Risk*. New York: Cornell University Operations Research and Industrial Engineering.

Dash, Jan. 2012. Stressed Value at Risk. Paper presented at IEEE Conference on Computational Intelligence for Financial Engineering and Economics (CIFERr), Shenzhen, China, May 4–5.

Diebold, Francis Xavier, and Glen David Rudebusch. 2012. *Yield Curve Modelling and Forecasting*. The dynamic Nelson-Siegel approach. Princeton: Princeton University Press.

Dixit, Jitenda Kumar, and Vivek Agrawal. 2019. Foresight for stock market volatility: A study in the Indian perspective. *Foresight* 22: 1–13. [CrossRef]

Embrecchts, Paul, Filip Lindskog, and Alexander McNeil. 2001. Modelling Dependence with Copulas and Applications to Risk Management. September 10. Available online: https://people.math.ethz.ch/~{}embrecht/ftp/copchapter.pdf (accessed on 11 August 2017).

Engle, Robert F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50: 987–1007. [CrossRef]

Engle, Robert. 2004. Risk and Volatility: Econometric models and financial practice. *American Economic Review* 94: 405–20. [CrossRef]

Engle, Robert, and Emil Siriwardane. 2014. Structural GARCH: The Volatility-Leverage Connection. Available online: https://www.ecb.europa.eu/events/pdf/conferences/140623/Engle-and-Siriwardane_Structural-GARCH-the-volatility-leverage-connection.pdf?e24ef9f36b79ea29f0bb4f6e5a8fabf3 (accessed on 27 July 2017).

European Banking Authority. 2012. *EBA Guidelines on Stressed Value at Risk (Stressed VaR) EBA/GL/2012/2*. London: EBA, May 16.

Fama, Eugene, and Kenneth French. 1992. The cross-section of expected stock returns. *Journal of Finance* 57: 427–65. [CrossRef]

Fishburn, Peter Charles. 1980. Stochastic dominants and moments of distributions. *Mathematics of Operations Research* 5: 94–100. [CrossRef]

Francq, Christian, and Jean-Michel Zakoïan. 2018. Estimation risk for the VaR of portfolios driven by semi-parametric multivariate models. *Journal of Econometrics* 205: 381–401. [CrossRef]

Goodfellow, Christiane, and Christian Salm. 2016. Risky Risk Measures, A note on underestimating financial risk under the normal assumption. *Capernican Journal of Finance & Accounting* 5: 85–108.

Haas, Marcus. 2001. New methods in backtesting 2001. In *Financial Engineering*. Bonn: Research Center Ceasar.

Hansen, Peter Robert, and Asger Lunde. 2005. A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics* 20: 873–89. [CrossRef]

Hedge Fund Research Inc. 2017. HFRI Indices–Index Descriptions. Available online: https://www.hedgefundresearch.com/hfri-indices-index-descriptions (accessed on 22 August 2017).
J.P. Morgan Asset Management. 2009. Non-normality of Market Returns, a Framework for Asset Allocation Decision Making. Available online: https://jai.pm-research.com/content/12/3/8 (accessed on 14 June 2021).
Jain, Shrey, and Siddhartha Chakrabarty. 2020. Does Marginal VaR lead to improved performance of managed portfolios? A study of S&P BSE 100 and S&P BSE 200. *Asia Pacific Financial Markets* 27: 291–323. [CrossRef]
Jansky, Ivo, and Milan Rippel. 2011. *Value at Risk Forecasting with ARMA-GARCH Family of Models in Times of Increased Volatility*. IES Working Paper: 27/2011. Prague: Institute of Economic Studies, Faculty of Social Sciences, Charles University.
Jorion, Philippe. 2003. *Financial Risk Manager Handbook*, 2nd ed. Hoboken: GARP, Wiley & Sons.
Kamara, Avraham, Robert Korajczyk, Xiaoxia Lou, and Ronnie Sadka. 2018. Short-horizon beta or long-horizon alpha? *Journal of Portfolio Management* 45: 96–105. [CrossRef]
Katchova, Ani. 2013. Time Series ARIMA Models: Econometrics Academy. Available online: https://sites.google.com/site/econometricsacademy/econometrics-models/time-series-arima-models (accessed on 4 September 2017).
Kidd, D. 2012. Value at Risk and Conditional Value at Risk. Investment and Risk Performance. CFA Institute. Available online: http://www.cfapubs.org/doi/pdf/10.2469/irpn.v2012.n1.6 (accessed on 16 August 2020).
Kilai, Mutua, Anthony Gichuhi Waititu, and Anthony Wanjoya. 2018. Modelling Kenyan foreign exchange risk using asymmetry GARCH models and Extreme Value Theory approaches. *International Journal of Data Science and Analysis* 4: 38–45. [CrossRef]
Krause, Andreas. 2003. Exploring the limitations of Value at Risk: How good is it in practice? *Journal of Risk Finance* 4: 19–28. [CrossRef]
Kumar, Arya, and Saroj Kantar Biswal. 2019. Impulsive clustering and leverage effect of emerging stock market with special reference to Brazil, India, Indonesia and Pakistan. *Journal of Advanced Research in Dynamic Control System* 11: 33–37. [CrossRef]
Lelasi, Federica, Paolo Capelli, and Angeloantonio Russo. 2021. Forecasting volatility by integrating risk with environmental, social and governance risk. *Corporate Social Responsibility and Environmental Management* 28: 1483–95. [CrossRef]
Lichtner, Mark. 2019. How to choose the return model for market risk? Getting towards a right magnitude of stressed VaR. *Quantitative Finance* 19: 1391–407. [CrossRef]
Liu, Qingfeng, Qingsong Yao, and Guoqing Zhao. 2020. Model averaging estimation for conditional volatility models with an application to stock market volatility forecast. *Journal of Forecasting* 39: 841–63. [CrossRef]
Lok, Hsiao Yen. 2015. Different Methods of Back testing VaR and ES. Ph.D. dissertation, Actuarial Research Centre, Heriot Watt University, Edinburgh, UK. Available online: https://www.actuaries.org.uk/system/files/field/document/HY%20Lok%20PARTY%20Jan%202015.pdf (accessed on 15 February 2021).
Lopez, Jose. 2005. Stress tests: Useful complements to financial risk models. *FRBSF Economic Letter* 2005: 119–24.
Loretan, Mico, and William English. 2000a. *Evaluating "Correlation Breakdowns" during the Periods of Market Volatility*. Basel: Bank of International Settlements (BIS), Available online: http://www.bis.org/publ/confer08k.pdf (accessed on 18 September 2017).
Loretan, Mico, and William English. 2000b. III Special feature: Evaluating changes in correlations during periods of high market volatility: Bank for International Settlements (BIS). *Quarterly Review* 2000: 29–36.
Lovreta, Lidija, and Joaquin López Pascual. 2020. Structural breaks in the interaction between bank and sovereign default risk. *SERIEs* 2020. [CrossRef]
Mabrouk, Samir. 2016. Forecasting daily conditional volatility and h-step-ahead short and long Value-at-Risk accuracy: Evidence from financial data. *The Journal of Finance and Data Science* 2: 136–51. [CrossRef]
Malmsten, Hans. 2004. *Evaluating Exponential GARCH Models: SSE/EFI Working Paper Series in Economics and Finance, No. 564*. Stockholm: Stockholm School of Economics.
Markowitz, Harry. 1952. Portfolio selection. *Journal of Finance* 7: 71–91. [CrossRef]
Marrison, Chrsitopher. 2002. *The Fundamentals of Risk Measurement*. New York: Mc Graw Hill.
Mina, Jorge. 2002. Measuring bets with Relative Value at Risk. *Derivatives Week, Learning Curve* 2002: 14–5.
RiskMetrics Technical Document, 4th ed. New York: Available online: https://my.liuc.it/MatSup/2006/F85555/rmtd.pdf (accessed on 16 July 2021).
Napper, Kerry. 2008. Refining Risk Measures. MLC Investment Management. *White Paper—Serial Correlation* 1: 2–13.
Nelson, Daniel. 1990. ARCH models as diffusion approximations. *Journal of Econometrics* 45: 7–38. [CrossRef]
Nieto, Maria Rosa, and Esther Ruiz. 2016. Frontiers in VaR forecasting and back testing. *International Journal of Forecasting* 32: 475–501. [CrossRef]
Orhan, Mehmet, and Bülent Köksal. 2011. A comparison of GARCH models for VaR estimation. *ESWA* 2012: 2582–3592.
Pflug, Georg Ch. 2000. Some remarks on the Value at Risk and the Conditional Value at Risk. In *Probabilistic Constrained Optimization. Nonconvex Optimization and Its Applications 49*. Edited by Uryasev Stanislav. Boston: Springer. [CrossRef]
Reider, Rob. 2009. Volatility Forecasting 1: GARCH Models: October 2009. Available online: cims.nyu.edu/~{}almgren/timeseries/Vol_Forecast1.pdf (accessed on 10 June 2017).
Rockafellar, Ralph Tyrell, and Stanislav Uryasev. 2000. Optimization of conditional value at risk. *Journal of Risk* 2: 21–42. [CrossRef]
Sarykalin, Sergey, Gaia Serraino, and Stanislav Uryasev. 2014. Value at Risk vs. Conditional Value at Risk in Risk Management and Optimization. *INFORMS Tutorials in Operations Research* 2008: 270–94. [CrossRef]
Sheppard. 2013. *Financial Econometrics Notes*; UK: University of Oxford. Available online: http://www.kevinsheppard.com/images/b/bb/Financial_Econometrics_2013-2014.pdf (accessed on 10 June 2017).

Tang, Wang, and Fang Su. 2017. Analysis of the impact of extreme financial events on systematic risk. A case study of China's banking sector. *Economic Research Journal* 52: 17–33.

Van Dyk, Francois. 2016. Stress Testing 101: The method, recent developments, transparency concerns and some recommendations. *The SA Financial Markets Journal*. Available online: http://financialmarketsjournal.co.za/oldsite/16thedition/printedarticles/stresstesting.htm (accessed on 7 March 2017).

Wang, Zhouwei, Qicheng Zhao, Min Zhu, and Tao Pang. 2020. Jump Aggregation, Volatility Prediction, and Nonlinear Estimation of Banks' Sustainability Risk. *Sustainability* 12: 8849. [CrossRef]

World Economic Forum. 2017. *The Global Risks Report 2017*, 12th ed. Zurich: World Economic Forum.

Journal of
Risk and Financial Management

Article

Understanding Organisational Risks and Opportunities Associated with Implementing Australia's National Disability Insurance Scheme from the Nonprofit Service Provider Perspective—Findings from Quantitative Research

Hamin Hamin [1], David Rosenbaum [2,*] and Elizabeth More [3]

1 Magister Manajemen, Universitas Budi Luhur, Jakarta 12260, Indonesia
2 King Own Institute, Sydney, NSW 2000, Australia
3 Study Group, Sydney, NSW 2000, Australia
* Correspondence: david@koi.edu.au

Abstract: In this paper, we provide useful lessons from a quantitative analysis across several nonprofit organisations undergoing generational change due to the implementation of the Australian government's National Disability Insurance Scheme (NDIS). This paper contributes to the field in demonstrating the usefulness of the approach in revealing how change has to occur at both the micro and macro levels of the organisations involved, affecting both followers and transforming leadership, whilst simultaneously reinforcing the need to address the strategic and operational risks inherent in such transformational change. It represents a follow-up to an earlier published longitudinal qualitative research and provides further evidence on the key findings associated with the development of the NDIS Implementation Framework. The current paper considers the importance of the risk and opportunity conundrum associated with the implementation of the NDIS among Australian nonprofit service providers. This paper recognises that, as entities operating ostensibly outside the purely commercial realms of service design and delivery, nonprofit service providers are potentially handicapped by an historic lack of relevant and necessary market-based skills. The risks necessitate an accelerated programme of skill development and skill acquisition to enable the full range of opportunities to be realised. The change management processes, identified using the conceptual framework of readiness → implementation commitment → sustainability, as discussed in this paper, highlight the potential financial consequences which have substantial impacts on such nonprofit service providers. Organisations in these settings are challenged by ongoing financial sustainability issues where very small financial margins, resulting directly from the generational business model shift from a supply-driven system to a demand-driven system, may prove the difference between organisational survival and failure.

Keywords: transformational change; nonprofits; NDIS Implementation Framework; business models; sustainability; financial risk

Citation: Hamin, Hamin, David Rosenbaum, and Elizabeth More. 2022. Understanding Organisational Risks and Opportunities Associated with Implementing Australia's National Disability Insurance Scheme from the Nonprofit Service Provider Perspective—Findings from Quantitative Research. *Journal of Risk and Financial Management* 15: 614. https://doi.org/10.3390/jrfm15120614

Academic Editors: George Halkos and Robert Hudson

Received: 26 October 2022
Accepted: 14 December 2022
Published: 16 December 2022
Corrected: 17 February 2023

1. Introduction

The National Disability Insurance Scheme (NDIS) represents a generational shift in the way in which services are delivered to people with a disability. Prior to its initial roll-out, disability services were more securely funded within a supply-driven process, meaning that governments at the Federal and State levels funded service delivery organisations directly to provide a wide range of standardised services to people with a disability. The bulk of these services were provided by nonprofit service providers. The NDIS sought to restructure this process and introduce a demand-driven system whereby 'funding packages' were developed by the Federal Government. These represented a maximum annual dollar value for packages based on the complexity of medical needs of people with a disability.

With their packages confirmed, these individuals would source their required services from registered nonprofit as well as commercial service providers, meaning that funding would now flow from the government to people with a disability, who would then source their services from the market, resulting in what has become recognised as a more financially risky demand-driven system for providers. In this way, services could become more bespoke, responding to the needs of their clients.

"The instrument of collaborative inquiry ... is an approach in which the scholar, aware of the problems that managers and operators daily experience in the field, sets up a collaborative study of these needs through engagement and active involvement of organization members and collaborative interaction with management" (Shani 2021, p. xiii). Whilst ours was a very broad canvas of collaborative learning, since we were focused on a particular nonprofit sector involved in the change management of introducing the NDIS, across several organisations, our participant organisations were involved in interviews for an earlier qualitative research (Rosenbaum and More 2022) and surveys for this quantitative research, as well as some follow-up meetings and presentations to participant organisation boards. Our aim for the qualitative research was to develop a pragmatic framework to enhance the practice of managing the major change wrought by the NDIS to diverse large and small organisations in the sector, to offer a pragmatic tool for managing such changes for both organisations and policymakers, and to add to the body of change management literature. Our aim for this quantitative research is to assess the efficacy of the National Disability Insurance Implementation Framework (NDISIF), developed in the initial qualitative study. As reflected in capturing our intent with the research: "The focus of collaborative inquiry is to generate the practical knowledge that enables the organization to make relevant changes and to contribute actionable knowledge to the social science of organizational change and development." (Shani 2021, p. 7). In doing so, we seek to aid nonprofit service delivery organisations in successfully implementing the NDIS and, in doing so, mitigate the risks of failure, while visioning and acting upon the organisational opportunities offered by the NDIS, and, in the bigger picture context, ensuring the long-term success of the NDIS for the sake of people with disabilities. This is crucial given the ongoing escalation of cost blowouts with the system.

The change wrought by the implementation of the NDIS brings with it many strategic and operational risks for employees, involving multiple dimensions of stress, uncertainty, and diverse responses of various kinds. It tests the very change capability, adaptability, organizational learning, sustainability, risk management, and flexibility of organisations and their leadership. Organisational change is impacted by context and environment, and dynamic change in a reciprocal way, that may impact the very sustainability, nature, structure, and processes of the organisation. This is as true for the nonprofit sector as for the commercial sector, although variables, constraints, and opportunities may differ.

The NDIS is a generational shift in, and an altered model of, the way services to people with disabilities are developed, provided, and funded. Its focus is on a person-centred approach in which services are reflective of individuals who remain active participants and decision makers in their own lives (Green and Mears 2014). Whilst it is not unique in a global context, it represents a major paradigm shift in Australia. It is a prime example of a force for good, given that it reflects the rights of people with a disability to live free from abuse, exploitation, and violence. These are in keeping with Australia's commitment to the United Nations' Convention on the Rights of Person with Disabilities (UN 2006), and reflected in the Vision Statement in the recent Australia's Disability Strategy 2021–2031 as " ... an inclusive Australian society that ensures people with disability can fulfill their potential, as equal members of the community" (Commonwealth of Australia 2021, p. 2).

At the heart of the creation and implementation of the NDIS is also the context of the United Nations' Sustainability Goals, particularly those of Goal 10—Reduce Inequalities, and Goal 16—Stand Up for Human Rights. One clear challenge to fully achieving both Disability and Sustainability goals is the need for a sustainable funding model which adequately addresses the associated funding risks, which have been prophesized in the early days of the system and are now gathering pace as the costs continue to escalate. The NDIS is not alone in this challenge as Gilchrist and Perks (2022, p. ii) state: “Indeed, many not-for-profit social services organisations are challenged in terms of sustainability by what might be termed a malevolent cycle, where poor quality impacts staffing, in turn impacting service capacity, thereby reducing income while, at the same time, infrastructure and other elements must continue to be paid for”. As Gilchrist and Perks (2022, p. 5) puts it: “The survival of the Not-for-profit and/or charitable corporation is a secondary issue to the sustainability of services given the risk borne by people with disability who need reliable, appropriate quality services and supports in order to live their lives”.

The introduction of the NDIS in Australia in 2016, which is yet to be fully implemented, shifts the focus of service design, delivery, and financial support from a supply-driven business model to a demand-driven business model. The former is structured around service providers developing programmes and being funded by Federal and State Governments. The demand-driven approach requires service providers to develop and deliver programmes based on the requirements and demands of service users, with the Federal Government providing ”packaged” funding directly to service users who then pay service providers for using services that meet their specific needs (Rosenbaum and More 2022). These ”needs” are reflective of their specific disabilities and their aspirations for living a fulfilled life (Taylor et al. 2020). The origins of the NDIS sought, in part, to address the ”contracting culture” inherent in the steady shift from the state to the market, which flowed from the impact of neo-liberalism in Australia (Onyx et al. 2016). However, the political repercussions of ongoing funding for the NDIS have not been fully successful. The Federal Government has sought to increasingly contract a range of operational activities that underpin the NDIS to the private sector, moving away from the person-centric focus and negatively impacting those the NDIS originally sought to benefit.

The NDIS offers a new challenge and associated risks as well as opportunities to clients and organisations, representing a mindset and generational shift for those with disabilities being supported with services that enable them to lead fruitful and meaningful lives (Meltzer and Davy 2019). The focus of change is the movement from a supply-driven approach to a demand-driven one, with market mechanisms becoming the focus of both service design and service delivery (Foster et al. 2021). The former model relies on service providers developing programmes and being historically funded by Federal and State governments, with service providers reacting to pooled funding available through different centralised funding sources and their services not necessarily reflecting the requirements of service users. Under the NDIS, we find a demand-driven approach requiring service providers to develop and deliver programmes grounded in the demands and requirements of service users, with funding moving from direct funding of the providers to direct funding of the users through a mechanism of assessed annual support funding packages. Service users then have choice in a way not previously offered, now using service providers who can best meet their personal requirements. The landscape is changed to a more competitive and riskier marketplace, with service providers now competing for clients, against other nonprofit and for-profit organisations (Green et al. 2018). This change in focus means that service providers are now having to consider competitive market pressures, the way they structure themselves, and the skills that are now required, which are well-beyond the somewhat fewer complex parameters of service design and delivery (Rosenbaum and More 2021).

Nonprofit service providers must now grasp the new reality of both opportunities and challenges. Decision making must acknowledge the wide-ranging strategic implications that place organisational sustainability at its core, now needing to respond to service user requirements as they are indirectly funded by central governments through the NDIS funding packages paid directly to NDIS “clients” who then decide where their funding packages will be spent. This represents a shift in the approach by leadership in these organisations, where changes to organisational culture results from a change in the prevailing business models. This challenges many in this sector who have often differentiated themselves from commercial organisations by focusing on the needs of their clients rather than, in part, on commercial outcomes that may now be required as a direct result of this shift.

This paper considers from a quantitative perspective the efficacy of the NDIS Implementation Framework (“NDISIF”) (Rosenbaum and More 2022) derived from the earlier qualitative research, which identified the relevant considerations necessary for the successful implementation of the NDIS amongst nonprofit service providers. It was derived from an extensive interviewing process and considered a range of Perspectives and Influencers in a three-stage framework of Readiness → Implementation → Sustainability. Here, we report a quantitative research methodology that investigates, through surveying, the accuracy and validity of connections identified in the NDISIF. In doing so, this paper is based on a mixed methods approach to research (Denscombe 2008), with the original framework being developed and reported using a qualitative approach, whilst this follow-up paper describes a useful quantitative approach seeking to assess the validity of the key elements of the NDISIF. Both types of analyses and reporting are necessary.

As further explored in the Section 4.2 of this paper, the aim of this research is to identify the extent to which the lessons learned from the implementation of the NDIS in the nonprofit sector can be applied to the development of sector-specific change management approaches. As mentioned earlier, a mixed methods research methodology was applied in this study, combining elements of qualitative (derived through open-ended semi-structured interviews and available corporate data) and quantitative analyses (derived through detailed mathematical-based analysis of larger population-sized questionnaires). From a quantitative perspective, the detailed organisation-wide questionnaires enabled the researchers to test the Framework’s validity and potentially support several research outcomes drawn from the analysis of the interview data. The use of a quantitative methodology in this manner may also enable “ … generalizations to be made beyond the boundaries of the situation under study … ” (Easterby-Smith et al. 1997, p. 75).

2. Research Question

This research seeks to validate the findings from an earlier qualitative research from which the NDIS Implementation Framework was developed. This Framework is reproduced below in Figure 1.

The elements contained within the above Framework are explained further in Table 1, which supports the key research question relevant to this current quantitative research, namely, what is the extent to which the NDISIF can be substantiated through rigorous quantitative analysis? The Conceptual Framework identified in Figure 2, including the resulting hypotheses, supports the response to this research question.

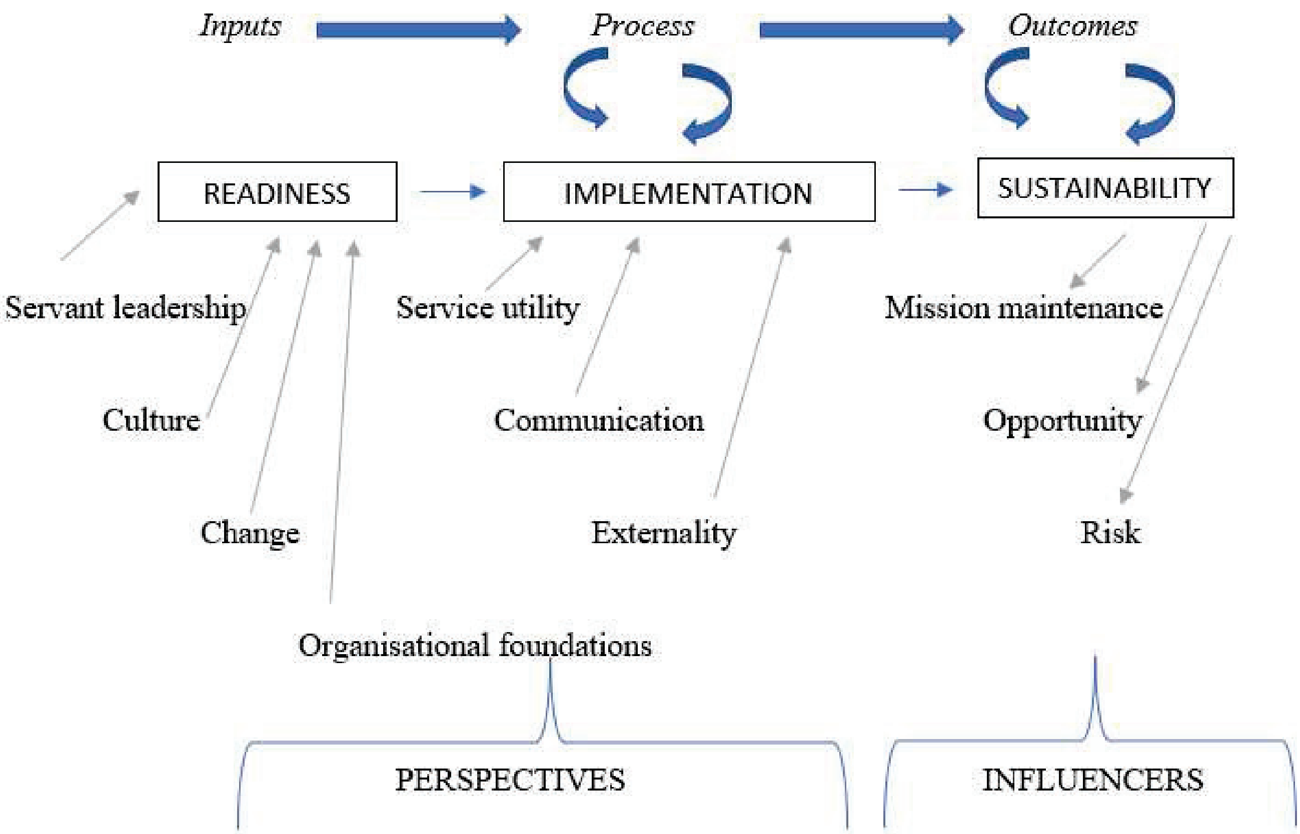


Figure 1. The NDISIF (Implementation Framework) (Rosenbaum and More 2022, p. 427).

Table 1. Success factors identified from the development of the NDIS Implementation Framework resulting from the qualitative research (Rosenbaum and More 2022).

	Readiness
Leadership	Servant leadership has been identified as the most relevant leadership style for nonprofit disability service organisations implementing the NDIS. The characteristics of empowerment, humility, authenticity, interpersonal acceptance, provision of direction and stewardship, as identified in the latest research into this leadership style, support the approaches necessary in this sector (Van Dierendonck 2011).
Culture	Cultural adaptability (Corritore et al. 2020) and a strong focus on organisational trust (Page et al. 2019) that underpins a supportive, inclusive, empowering, and accountable culture appear as fundamental requirements in these organisations.
Change management	Use and application of approaches to managing change must be adaptable where the change process must be organisationally aligned and reflective of wide-ranging nonprofit attributes. An appropriate approach is the reconsidered Lewin 3-step model of change (Lewin 1947), as discussed in specific nonprofit research into change management (Rosenbaum et al. 2018).
Organisational foundations	The absence of a range of restricting forces, which must be addressed either before or during the change process, requires a review of the organisational structures (Waddell et al. 2019) and the role of organisational human resource functions (El-Dirani et al. 2019), and addressing issues associated with what has become known as the 'head office syndrome' (Bouquet et al. 2016).
	Implementation/Commitment
Service utility	The ability to provide service design and delivery in a clear manner in a contested marketplace whilst maintaining advocacy as an important element of staff engagement in a changing internal and external environment (Kimberlin 2010).

Table 1. *Cont.*

Implementation/Commitment	
Communication	Wide-ranging elements of internal communications, including coordinated top-down messaging; consistency in change communication; focused customer choice communication; addressing organisational silos linked to both service design and delivery; and the use of carefully crafted language. Additionally, communication must be authentic and sincere in order to strengthen an emotional connection and, therefore, trust between service provider and service user (Frei and Morriss 2020).
Externality	Reliance on effective and efficient interactions with the NDIA, which is the Federal Government Agency tasked with the rollout of the NDIS as recognised by the federal Government Joint Standing Committee on the NDIS in its 2019 Report (Andrews 2019). This reinforces the advantages resulting from a well-considered external networking approach to the implementation of the NDIS at the organsiational level.
Sustainability	
Mission	The ongoing maintenance of the organisational mission must be prominent in order to ensure staff acceptance of the necessary changes required to make the NDIS implementation successful (Rosenbaum et al. 2017), accepting that any apparent conflict between a values-based mission and the commercial realities of a demand-driven NDIS market place is adequately addressed from the perspective of client well-being (Dawson and Daniel 2010). This goes to the heart of organisational identity and its maintenance during all phases of the changes deemed necessary to successfully implement the NDIS (Venus et al. 2019)
Risk and Opportunity	The mind shift related to seeing the NDIS as an organisational and market opportunity, rather than purely a risk which requires mitigation. Such an approach supports staff in embracing the necessary changes required to successfully implement the NDIS in an uncertain and far risker context.

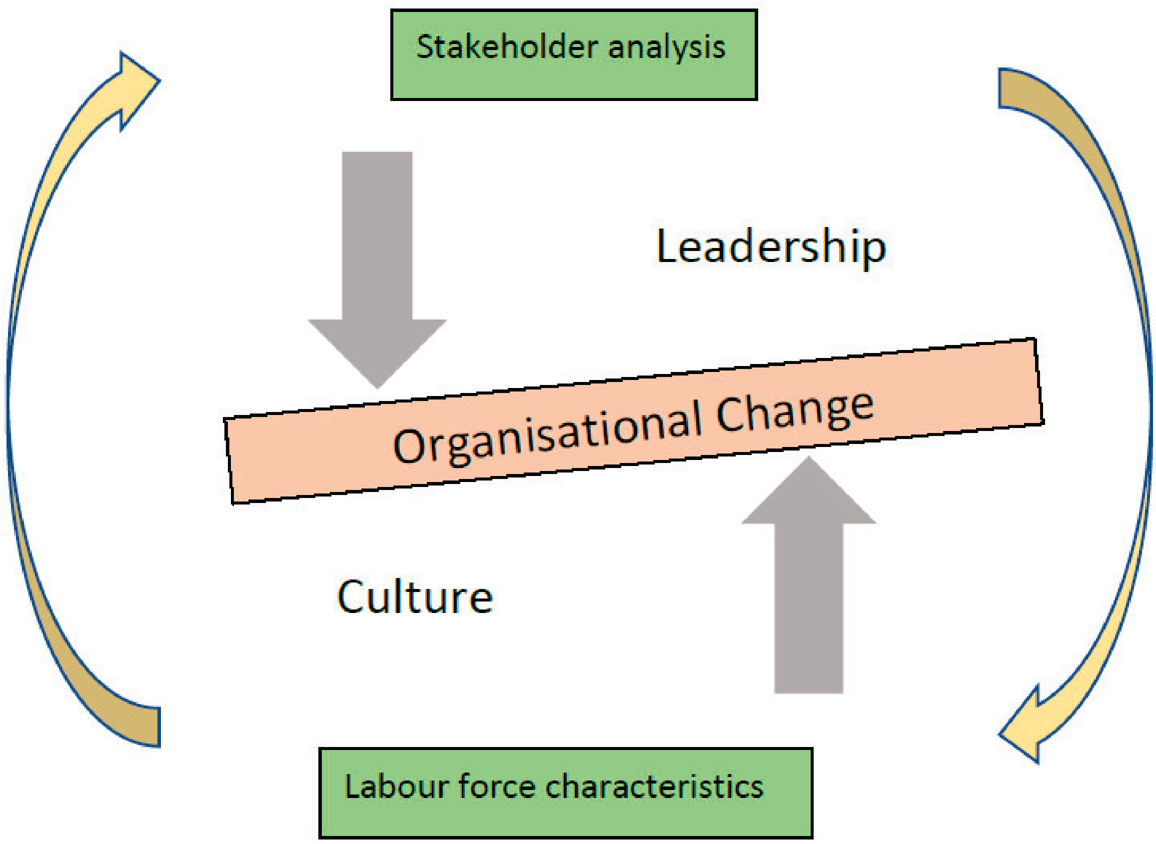


Figure 2. Key considerations in successful organisational change for nonprofit disability service providers.

3. Theoretical Context of Change

As outlined in our earlier published paper on the original qualitative aspect of our research (Rosenbaum and More 2021), the change framework for our work remains the same. Refocusing an organisation to reflect a major generational shift from its current to a future desired state though organisational change has many dimensions, including leadership styles, followership, context, nature of the change—evolutionary, change capability, culture, trust, resourcing, timing, productivity, and financial risk. Unsurprisingly, numerous models of change abound from classic to more contemporary ones (Burke 2017; Rothwell et al. 2021).

For the application of the theory, a key challenge is how to explain the way to implement organisational change that facilitates a movement from taken-for-granted mindsets, processes, and actions to new systems and being able to institutionalise such new thinking and approaches. This is exacerbated by an increasingly volatile, uncertain, complex, ambiguous, and diverse (VUCAD) environment and the need for almost constant change. This is further compounded as a direct result of our recent challenges of changes in a pandemic period that has wrought its own menu of rapid and prolonged changes.

There have been diverse approaches to understanding and explaining organisational change across numerous theoretical and disciplinary bases, from individual and group psychological and communication perspectives to those dealing with organisation-wide change. Change types also provide a different approach—incremental to radical, small- vs. large-scale change, and proactive vs. reactive. These have been well-explained some time ago by Stace and Dunphy (2001, p. 97).

Earlier approaches to change were mired in a static linear perspective, but successful change nowadays recognises the real dynamism, richness, and complexity of major change as is the NDIS, not only during implementation but also in the future to ensure the embedding of such change in the long term without individuals and organisations falling back into outmoded thoughts and practices—the unfreezing of the old state, change to the new state, and refreezing the new state for permanency (Lewin 1951).

3.1. Considering the Management of Change as a Sector-Specific Challenge

Organisational change management remains of substantial academic and practitioner interest (Rosenbaum et al. 2018), and, to some extent, it is supported by an ongoing dialogue as to the specifics of its execution, be that at an industry-wide level (Kätelhön et al. 2019) or at an organisation-specific level (Beniflah and Veloz 2021). From this perspective, organisational change can be viewed across a spectrum of macro-considerations through to micro-considerations. One approach sees organisational change in terms of the application of models that potentially apply to all organisations (Smith et al. 2020), thereby largely ignoring the contextual specifics that may both support and hinder successful implementation. Such an approach may, to some extent, force the processes of change through the arteries of these models, where the lifeblood of successful change develops from a focused understanding of how change unfolds. Once an understanding of this is confirmed, these models underpin an approach of adaptation. However, what may be adapting is not the model to the organisation and its context but, rather, the organisation to the model. This ignores context and reinforces a falsity that organisations in and of themselves change. However, this one-dimensional approach may prove to be inaccurate as organisations tend not to change successfully unless their people change their beliefs and adapt to the changing processes and circumstances. Therefore, the focus of change must first rest on, amongst other things, changing people's views and attitudes (Rosenbaum et al. 2018), understanding resilience (Parker and Ameen 2018), and developing employees' skill characteristics (Stouten et al. 2018) to improve change outcomes.

Following on from the potential shortcomings identified above in this one-dimensional view of change, especially with regard to the focus on individuals, is a different understanding of how change management may need to be handled. Here, we seek to include the notion of a sector-specific approach which becomes more apparent when considering the nonprofit sector. The reason this becomes relevant is the range of characteristics that uniquely define this sector and its workforce (Rosenbaum et al. 2017), and understanding that, in any change management approach, if we fail to change the way individuals both view the change and function within it, it is likely that such change will fail, or, at the very least, result in delayed outcomes.

It is at this point that this research recognises the importance of change being considered at the level of each organisation. This suggests that unique industrial and sectoral characteristics may benefit from a framework approach to organisational change management, rather than a model approach. The latter tends to be viewed from a procedural perspective containing somewhat standardised elements, as distinct to the former approach which emphasises a wide range of contextual characteristics that must be integrated with the realities of managing change. By identifying the key Perspectives and Influencers within the broad 3-stage process of Change Readiness, Change Implementation, and Organisational Sustainability, the NDISIF provides a roadmap for how nonprofit disability service providers may implement the NDIS within their organisations (Rosenbaum and More 2022).

3.2. Key Considerations in Successful Change Management

From the development of the NDISIF, our research identifies a range of organisational processes and structural considerations that are necessary to maximise the change outcomes for implementing the NDIS, especially within nonprofit service delivery organisations. One key element relates to organisational flexibility, where any procedural approach to change must be balanced with appropriate leadership and cultural characteristics to ensure that the context of the setting is a key consideration. This recognises the need to create a strategic change guide that reflects the risk appetite and the variability associated with the planning and execution phases, leading to the institutionalisation of the implemented change (Rosenbaum and More 2022). Such an approach considers the comprehensive method undertaken by Lewin (Lewin 1947), which accounts for the integrated steps of action research, group dynamics, and force field analysis (Rosenbaum et al. 2018).

From a change management perspective, the involvement of stakeholder analysis appears pivotal to understanding the organisational context which enables change to be understood and structured (Vargas et al. 2019). This reflects the organisational uniqueness of the settings and further focuses attention on the human element of change—the organisational actors who need to both plan for and drive change to lead to long-term sustainability. This awareness of stakeholders also focuses attention to outside of the organisation and relates to the interactions necessary to develop and maintain the networks necessary to support change. These networks are important as they ground change in a broader context by linking not only other service providers together, but also reinforcing the necessary relationships with external government agencies that determine the necessity for change. This supports the broader issue of advocacy to further strengthen staff support during change complexities in an often highly emotional setting, given the nature of the clients these organisations deal with on a day-to-day basis.

A further element to this implementation challenge is leadership understanding of the work characteristics associated with such change. The NDIS implementations challenge the historical understanding of client service delivery. Effectively, this shifts the focus from the person who has a disability and is availing themselves of a particular service to one where a more focused commercial arrangement evolves and the person with a disability becomes the client of the service delivery organisation, and the services need to be delivered in the context of a customer relationship (Rosenbaum and More 2021). This subtle, but important, shift in the relationship, has challenged many service delivery personnel, and understanding the human side of this shift becomes an important element for consideration. Addressing this changing environment is considered fundamental to the ongoing mind shift needed to guarantee success in the implementation. The focus on this rebalancing is highlighted in Figure 2 with the fulcrum of change evidencing leadership and culture as having essential roles in ensuring effective change outcomes. Additionally, the impact of stakeholders, both internal and external, and the skill characteristics of staff, along with the interplay between these, reflect a level of both diversity and complexity that underpin successful organisational change in this sector.

This is diagrammatically represented in Figure 2 above.

3.3. The Australian National Disability Insurance Scheme as the Research Setting

As identified earlier in this paper, the primary research setting is the implementation of Australia's National Disability Insurance Scheme. We have focused our analysis on several disability service providers operating in the nonprofit sector, where nonprofit service providers compete with for-profit service providers. Our interest in nonprofit service providers stems from an understanding that these organisations face the challenges of managing transformational change, as outlined above. These challenges bring into sharp focus the extra dimension of traditional mission/margin conflicts within these institutions, which are less prevalent amongst commercial providers entering the market later and not having to address the pre-implementation challenges faced by their nonprofit competitors.

4. Research Context: Australian Nonprofit Disability Service Sector

This research embraces a quantitative approach, motivated by a need to understand the past of nonprofit disability service delivery in this complex environment and to inform the future design and implementation aspects critical to the ongoing success of such a large generational shift in the NDIS social initiative, especially considering its escalating costs.

Our NDISIF (Rosenbaum and More 2022) is premised on the key factors of organisational readiness for change, the organisational implementation strategies that support the implementation, and the sustainability challenges (see Figure 3: Conceptual Framework) that must be addressed to make the change successful. These factors appear in Table 1 above, including the elements that support them.

Hypotheses

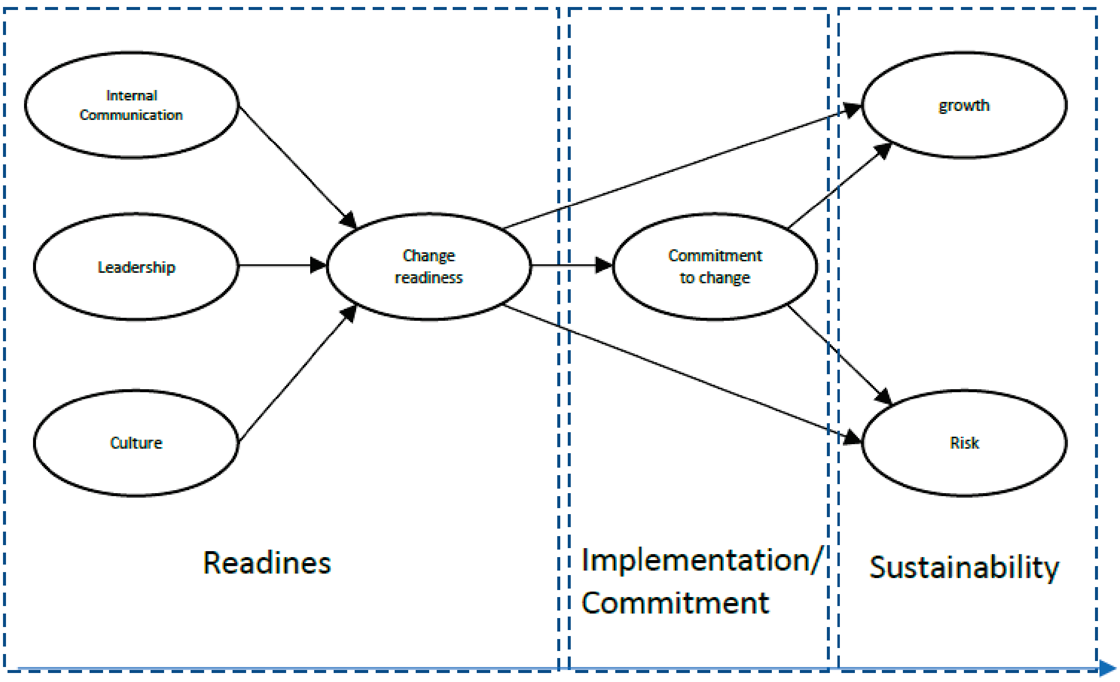


Figure 3. Conceptual framework developed and applied in this research stemming from the NDIS Implementation Framework developed in the earlier qualitative research, as reported in Rosenbaum and More (2022).

4.1. Hypotheses and Associated Methodology

4.1.1. Readiness

H1a: *Change readiness is significantly influenced by internal communication.*

H1b: *Change readiness is significantly influenced by leadership.*

H1c: *Change readiness is significantly influenced by culture.*

H2a: *Change readiness has a direct positive impact on growth.*

H2b: *Change readiness has a direct negative impact on risk.*

H3a: *Change readiness has a positive impact on growth mediated by commitment to change.*

H3b: *Change readiness has a negative impact on risk mediated by commitment to change*.

H4a: *Commitment to change has a significant mediating effect between change readiness and growth.*

H4b: *Commitment to change has a significant mediating effect between change readiness and risk.*

4.1.2. The Mediating Effect of Change Readiness between Change Drivers of Unfreezing and Refreezing (Sustainability)

H1a: *Change readiness has a significant mediating effect between internal communication and growth.*

H1b: *Change readiness has a significant mediating effect between leadership and growth.*

H1c: *Change readiness has a significant mediating effect between culture and growth.*

H2a: *Change readiness has a significant mediating effect between internal communication and risk.*

H2b: *Change readiness has a significant mediating effect between leadership and risk.*

H2c: *Change readiness has a significant mediating effect between culture and risk.*

4.1.3. The Mediating Effect of Commitment to Change between Readiness to Change (Unfreezing) and Sustainability (Refreezing)

Growth

H3a: *Change readiness and commitment to change has a significant mediating effect between internal communication and growth.*

H3b: *Change readiness and commitment to change has a significant mediating effect between leadership and growth.*

H3c: *Change readiness and commitment to change has a significant mediating effect between culture and growth.*

Risk

H4a: *Change readiness and commitment to change has a significant mediating effect between internal communication and risk.*

H4b: *Change readiness and commitment to change has a significant mediating effect between leadership and risk.*

H4c: *Change readiness and commitment to change has a significant mediating effect between culture and risk.*

4.2. Methodology

Research problems and research objectives were initially explored through qualitative data collection, with the conceptual framework and the hypotheses being also developed using qualitative research and previous original research. The study reported here follows this former exploration with quantitative data that are collected from a larger sample so that the results can be better concluded for the population of interest. The purpose of collecting the quantitative data is to examine the variables with a larger sample and then perform a deeper exploration of some NDIS cases known during the qualitative phase. In particular, this quantitative study determines whether there are different mediating effects of the implementation constructs (e.g., commitment and risk) between readiness and sustainability in the NDIS Implementation Framework (NDISIF).

4.2.1. A Three-Stage Model Comparison

The main objective of this quantitative study is to examine whether the hypotheses regarding planned organisational change management developed in the qualitative study demonstrates differences in the drivers of change among the three stages of pre-, during, and post-implementation. This quantitative study employs a two-step analysis, namely empirical estimation and three-stage model comparison. First, this quantitative study investigates how the "moving" variables, such as "change readiness" and "commitment to change", intervene in the relationships between the unfreezing and refreezing variables. It provides findings from an empirical relationship testing of how change drivers of unfreeze factors, such as internal communication, leadership, and culture, have significant effects on the institutionalisation of organisational changes of both sustainability and risk. Second, we deal with a three-stage model comparison. The three-stage model refers to the testing of the planned organisational change management model in three different circumstances, namely pre-, during, and post-change. In this second step, a comparative analysis of the empirical testing of planned organisational change management models in three different circumstances is presented.

4.2.2. Sample and Procedure

This study applied partially mixed sequential qualitative—quantitative methods (Leech and Onwuegbuzie 2009). The qualitative study was conducted prior to the quantitative empirical study. The qualitative phase of the study produced a conceptual framework of planned organisational change management, as presented in the previous sections. The quantitative phase of the study recruited additional participants to those in the qualitative phase.

The questionnaire for this study was designed to expose participants to the three stages mentioned above in managing transformational change attributed to the NDIS implementation. The same set of questions was purposely designed to investigate whether there were differences in transformational changes in three different scenarios. In total, 135 employees from nonprofit service organisations completed the online questionnaire. The response rate was approximately 73%. After data cleaning and validation of excluding nonsensical responses on items, the total number of respondents who completed a set of questions pre-, during, and post-NDIS implementation totalled to 68%.

Nearly two-thirds of the sample were male (63.3%; female: 36.7%), the average age was 43.4 years (SD = 9.84), and the mean tenure was 17.85 years (SD = 11.25). Most of the sample worked in a non-managerial position (90.8%). More than half (56.2%) held a predominantly operational position, while 43.8% of the sample held a predominantly support position (administrative, IT, HR, or finance tasks supporting the operational processes).

4.2.3. Operationalisation Definition and Measurement

To analyse the data, this study employed partial least square (PLS) implemented in the SmartPLS program. Prior to conducting model testing, the process of turning abstract concepts of planned organisational change management into measurable observations was developed based on the qualitative study. Table 2 shows the change driver latent variables of unfreezing as specified by the three dimensions of internal communication, leadership, and culture. The mediating variables of moving are specified by the dimensions of change readiness and commitment to change. Lastly, the refreezing variables are implied by sustainability and risk.

Table 2. Operational indicators of latent dimensions.

Construct	Dimension	Definition	Indicator
Unfreezing	Internal communication	Contains elements of strong top-down but coordinated communication pathways; consistency regarding the change messaging; communication that underpins a strong understanding of the role that customer choice plays in both programme design and service delivery; recognising the need to minimise the deleterious impacts of organisational silos in the delivery of integrated services; and the effective use of wide-ranging brainstorming sessions which involve extensive cross-sections of the organisation in order to consistently address implementation challenges. Dimensions include the nature and style of language, ability to positively react to bottom-up communications, use of relevant communication media, and acknowledging authenticity and sincerity in messaging.	Intcom1: I have considered leaving this sector as a direct result of the move to a commercial model for service delivery. Intcom2: I have reluctantly shifted my approach towards the commercial realities of service delivery in the post-NDIS environment.
	Leadership	Focuses on the application of Servant Leadership as the appropriate focus for nonprofit disability service organisations, where leadership focuses on followers and the ability of leaders to enable followers to fully realise their own potential.	Lead1: Team-based award is a vital emotional support mechanism. Lead2: I feel that my views are considered in decision-making.
	Culture	Culture is represented by a range of attributes which highlight the internal organisational conflict between the purposeful mindset of service provision, based on mission, with the pragmatic reality of commercialism, as represented by the shift from a supply-driven to a demand-driven business model. This is further defined by the existence of organisational sub-cultures evidenced in internal operational silos that have historically existed within many disability service organisations. This understanding of cultural attributes is further supported by both change optimism as well as change pessimism within these organisations.	Culture1: I believe commercial practices are appropriate for nonprofit disability service providers in dealing with the NDIS implementation. Culture2: Different aspects of our organisational culture helps us to overcome difficulties.
	Change readiness	Change readiness identifies the extent to which the organisation can effectively introduce the necessary changes. It contains two key elements. On the one hand is the extent to which human resources within the organisation have the necessary personal skills to accept and deal with change, whilst on the other hand, it is the organisational infrastructure that supports staff as they work with the change. This latter element highlights the importance of training and upskilling staff so that the challenges of change, at the human level, can be addressed.	Ready1: People who work here feel confident that the organisation can get people invested in implementing this change. Ready2: People who work here feel confident that they can keep track of progress in implementing this change. Ready3: People who work here want to implement this change.

Table 2. *Cont.*

Construct	Dimension	Definition	Indicator
Moving	Commitment to change (Willingness to change)	Change willingness tends to be a direct product of change readiness as it is the result of the application of change skills that can then be applied to the practice of change. Willingness to change is a product of numerous factors, including an individual's understanding of change, the skills that the individual has and/or has developed to cope with change, and the attitude of the individual with regard to change. The latter can also be the product of a common vision which focuses attention on the need for change and, in the case of nonprofit disability organisations, the belief that clients of the organisation will benefit from the proposed changes.	Will1: I intend to put effort into achieving the goals of the change. Will2: I am supportive of the change. Will3: People who work here are motivated to implement this change.
Refreezing	Sustainability of growth	Sustainability refers to the extent to which the organisation, having undertaken the implementation and is progressively moving through the process, is able to maintain the changes over a longer period, especially when the initial change focus, which can coincide with heightened internal activity, tends to wind down. A key issue for nonprofit disability service organisations is the ability to have an ongoing focus on the original mission of the organisation to provide services to people with disabilities, whilst keeping an eye on the commercial realities that are necessary in a changing demand-driven environment.	Sustain1: I feel that my DSO has maintained its focus on its original mission and values. Sustain2: I feel that any changes to our mission and/or values are consistent with my DSO's focus on its clients. Sustain3: I agree with the need for my DSO to be 'profitable'. Sustain4: I accept the changing focus of my organisation regarding commercial outcomes that are in line with client outcomes.
	Risk	Risk, in the context of our nonprofit disability service organisations, is the process of understanding and managing the opportunities that invariably arise with the movement from a supply-driven to a demand-driven business model. Evaluating and mitigating risk to conceive and deal with opportunities are the foundations for ongoing sustainability by these organisations. This moves these organisations from a traditional conservative approach to one that is more reactive to options that may provide potential for growth, in terms of expanded service offerings and, therefore, financial positioning.	Riskop1: Lived experience within the organisation is considered necessary for our DSO's success. Riskop2: I believe that our DSO needs to be bold when it comes to service design in order to remain competitive.

Unless otherwise stated, items of indicators are measured on 5-point Likert scales from 1 = *strongly disagree* to 5 = *strongly agree*.

The initial results of the qualitative study found many operational indicators to measure the dimensions of the conceptual model. After conducting a principal component analysis to reduce the items of the indicators into a set of interpretable factors for confirmatory factor analysis (CFA), an item-level analysis was used to examine the unidimensional aspects of the latent variables. The indicators of each of the latent dimensions that were used in the questionnaire are operationalised by the items, as shown in Table 2.

4.2.4. Strategy of Data Analysis

This study uses two steps of fit measurement evaluation, which includes the measurement model, and structural model evaluation (Chin 2010). The first step of the measurement model evaluation focuses on the validity and reliability of the indicators that are used for each latent variable, as shown in the conceptual model. The objective of the measurement model is to justify whether there are relationships between the latent variables and its measures/indictors. Since, in this study, the conceptual model consists of reflective measurement models for all latent variables, composite reliability (rho A), convergent validity (AVE), indicator reliability (factor loading), and discriminant validity (HTMT) were evaluated (Benitez et al. 2020). The second step of structural model evaluation focuses on the assessment of the theoretical model (Akter et al. 2011). To examine the fit of the theoretical model, predictive relevance (Q2) and GoF index were used.

4.2.5. Assessment of the Reflective Measurement Model

The quality of the reflective measurement model was assessed using the measurement model evaluation criteria of Cronbach's alpha, rho A, average variance extracted (AVE), weight factor, and loading factor. For the first criterion, the composite reliability was assessed using Dijkstra–Henseler's rho A to estimate the correlation between the latent variable and the construct scores. The composite reliability checks the amount of random error contained in the construct scores, which is expected to be limited. The values of Dijkstra–Henseler's rho A for internal communication, change readiness, commitment to change, and sustainability as shown in Table 3, and are larger than 0.707. These values can be considered as reasonable, indicating reliable construct scores. However, the values of Dijkstra–Henseler's rho A for other latent variables, such as leadership, culture, and risk, are lower than the standard criterion of 0.707.

Convergent validity represents the extent to which the indicators' variance is explained by a latent variable. It indicates that the indicators belong to one latent variable measure, or the same construct. The average variance extracted (AVE) was used to evaluate convergent validity. Table 3 shows that all values of AVE of the latent variables are greater than 0.5, indicating that there is empirical evidence for the convergent validity of all latent variables. It means that more than half of the variance is explained by the latent variables (Bagozzi and Yi 1988).

The next assessment is indicator reliability, which measures the amount of variance presented in a latent variable in terms of the contribution of each indicator. The loadings, also called factor loadings, are a good measurement of this matter (Henseler et al. 2014). In this study, the standard estimate of factor loading is 0.707 or higher, indicating that more than 50% of the indicator variance is explained by the corresponding latent variable (Benitez et al. 2020). Table 3 shows that the factor loadings are all significant with a *p*-value less than 0.001, with estimates ranging from 0.780 to 0.919. The factor loading estimates are greater than 0.707, suggesting that the measures are reliable.

The last measurement model assessment is discriminant validity, which measures differentiation between the different aspects measured by the latent variables. It shows the degree to which a measure of construct diverges from (or has no correlation with) another measure, which underlying construct is conceptually unrelated to it. The Heterotrait-Monotrait Ratio of Correlations (HTMT) is used to provide evidence for discriminant validity. Table 4 shows that almost all the HTMT estimates are under 0.85, except $\text{HTMT}_{\text{leadership-culture}}$ and $\text{HTMT}_{\text{culture—commitment to change}}$ which are under 0.9. Discriminant validity is verified through cross loadings (Benitez et al. 2020) for pre-, during, and post-NDIS implementation, as presented in Appendix B.

Table 3. Measurement model evaluation.

Construct	Indicator	Pre-NDIS Implementation				During NDIS Implementation				Post-NDIS Implementation			
		rho A	AVE	Weight	Loading	rho A	AVE	Weight	Loading	rho A	AVE	Weight	Loading
Internal communication		0.729	0.76			0.737	0.71			0.79	0.761		
	Intcom1			0.567 ***	0.888 ***			0.735 ***	0.924 ***			0.688 ***	0.927 ***
	Intcom2			0.561 ***	0.886 ***			0.427 ***	0.752 ***			0.445 ***	0.814 ***
Leadership		0.665	0.7			0.656	0.719			0.646	0.697		
	lead1			0.591 ***	0.872 ***			0.496 ***	0.797 ***			0.469 ***	0.759 ***
	lead2			0.565 ***	0.858 ***			0.675 ***	0.896 ***			0.713 ***	0.904 ***
Culture		0.632	0.74			0.657	0.708			0.708	0.741		
	culture1			0.479 ***	0.764 ***			0.477 ***	0.775 ***			0.482 ***	0.812 ***
	culture2			0.705 ***	0.899 ***			0.698 ***	0.902 ***			0.671 ***	0.908 ***
Change readiness		0.897	0.83			0.858	0.771			0.898	0.826		
	ready1			0.357 ***	0.897 ***			0.416 ***	0.91 ***			0.375 ***	0.923 ***
	ready2			0.354 ***	0.934 ***			0.354 ***	0.885 ***			0.335 ***	0.917 ***
	ready3			0.389 ***	0.897 ***			0.368 ***	0.836 ***			0.391 ***	0.886 ***
Commitment to change		0.752	0.78			0.79	0.697			0.858	0.779		
	will1			0.387 ***	0.829 ***			0.354 ***	0.806 ***			0.367 ***	0.913 ***
	will2			0.434 ***	0.869 ***			0.407 ***	0.874 ***			0.372 ***	0.885 ***
	will3			0.403 ***	0.749 ***			0.436 ***	0.824 ***			0.395 ***	0.849 ***
Growth		0.905	0.78			0.906	0.759			0.922	0.777		
	sustain1			0.281 ***	0.869 ***			0.273 ***	0.864 ***			0.3 ***	0.88 ***
	sustain2			0.294 ***	0.895 ***			0.324 ***	0.911 ***			0.327 ***	0.928 ***
	sustain3			0.251 ***	0.835 ***			0.238 ***	0.842 ***			0.219 ***	0.816 ***
	sustain4			0.313 ***	0.905 ***			0.31 ***	0.865 ***			0.282 ***	0.899 ***
Risk		0.786	0.74			0.896	0.751			0.647	0.735		
	riskop1			0.458 ***	0.831 ***			0.389 ***	0.781 ***			0.549 ***	0.84 ***
	riskop2			0.67 ***	0.925 ***			0.737 ***	0.944 ***			0.616 ***	0.875 ***

4.2.6. Assessment of the Structural Model

The second step of fit measurement evaluation is the assessment of the structural model. Since this study compared the role of transformational change drivers in those three different implementation stages, the Partial Least Squares Multi-Group Analysis (PLS MGA) was employed to analyse the differences in the transformational changes. To perform PLS MGA, measurement invariance must be assessed to confirm that the measurement models specify measures of the same attribute under different conditions (Henseler et al. 2014). The measurement invariance test is meant to ensure that the construct measures are invariant across the groups (Steenkamp and Baumgartner 1998; Sarstedt and Ringle 2011). Before discussing the fit measurement testing results of the structural model, the measurement invariance was examined.

Table 4. Overall construct correlation matrix (HTMT).

	Change Readiness	Culture	Internal Communication	Leadership	Risk	Growth	Commitment to Change
Change readiness	1						
Culture	0.742	1					
Internal communication	0.332	0.525	1				
Leadership	0.791	0.89	0.38	1			
Risk	0.274	0.524	0.319	0.326	1		
Growth	0.549	0.841	0.578	0.595	0.499	1	
Commitment to change	0.819	0.869	0.576	0.733	0.568	0.783	1

Note: Construct correlation matrix for pre-, during, and post-NDIS implementation is presented in Appendix B.

4.2.7. Test for Measurement Invariance

To assess measurement invariance, this study used the measurement invariance of composite models (MICOM) procedure developed by Henseler et al. (2014). Appendix A presents the results of three comparisons' MICOM test of compositional invariance and composite equality estimates. The results of the measurement invariance test confirmed that, generally, the multigroup comparison test results corresponded very closely since the compositional invariance and full measurement invariance were established. The justification of compositional invariance was supported by the parametric test that yielded, in all cases, higher t-values than the permutation test (all *p*-values were insignificant). It was also shown by the fact that all "original correlations" were greater than/equal to the 5% quantile. The establishment of full invariance is justified by the composite equality test, as shown in Appendix A, in which most of the mean differences and all variance differences fall between the 2.5% and 97.5% boundaries. The measurement invariance test discovers that, in respect of all three structural model relations, the three path coefficients are equal across the three stages (pre-, during, and post-NDIS implementation). Since the measurement invariance test using the MICOM procedure is achieved, the group comparisons can be proceeded with Multigroup Analysis (MGA).

4.2.8. Test for Structural Model

The last step of a fit measurement evaluation is a structural model assessment, which evaluates, with respect to the estimates and hypothesis tests, the causal relations between the exogenous and endogenous variables. The results of overall fit of the estimated model, such as path coefficient estimates, effect sizes (f^2), and coefficient of determination (R^2), meet the minimum model fit, as shown in Table 5. The overall fit of the estimated model was evaluated using a bootstrap-based test of the overall model fit and the SRMR. The purpose of this evaluation is to measure of approximate fit to obtain an empirical model for the proposed theory. Table 5 contains the value of the complete model SRMR, which is below the recommended threshold value of 0.080 (Henseler et al. 2014; Hu et al. 1992). However, the values of the SRMR for pre- and post-NDIS implementation are slightly greater than the threshold value due to the small sample size (N) and low degree of freedom (df) (Baron and Kenny 1986). Hu and Bentler (1999) advised that the model should be neglected if the SRMR is greater than 0.1.

Table 5. Structural model evaluation.

	Path Coefficient			f-Square			
	P1	P2	P3	Complete	P1	P2	P3
Internal communication -> Change readiness	−0.092 ns	0.208 ***	−0.007 ns	0.18	0.011	0.089	0.232
Leadership -> Change readiness	0.289 **	0.442 ***	0.485 ***	0.005	0.089	0.25	0
Culture -> Change readiness	0.372 ***	0.279 ***	0.331 ***	0.109	0.128	0.096	0.157
Change readiness -> Commitment to change	0.567 ***	0.727 ***	0.782 ***	0.003	0.475	1.122	0.006
Change readiness -> Growth	−0.113 ns	0.29 ***	0.082 ns	0.376	0.017	0.074	0.36
Commitment to change -> Growth	0.754 ***	0.444 ***	0.662 ***	0.941	0.748	0.174	1.579
Change readiness -> Risk	−0.143 ns	−0.156 ns	−0.164 ns	0.185	0.018	0.013	0.316
Commitment to change -> Risk	0.531 ***	0.45 ***	0.653 ***	0.014	0.244	0.109	0.015
				R Square			
				Complete	P1	P2	P3
Change readiness				0.427	0.308	0.557	0.507
Risk				0.194	0.217	0.124	0.286
Growth				0.452	0.484	0.468	0.529
Commitment to change				0.485	0.322	0.529	0.612
Overall fit of the estimated model				Value			
SRMR				0.072	0.099	0.069	0.085
d_ULS				0.883	1.685	0.824	1.229
d_G				0.475	0.719	0.589	1.114

The second evaluation of the structural model is path coefficients and their significance levels. The path coefficient estimates for the hypothesised relationships range from −0.164 to 0.754. Most of these coefficient estimates are significant at a 5% significance level. The next structural model evaluation is to examine the effect sizes of the relationships between the constructs. This study used f^2 values to measure the magnitude of an effect independent of the sample size. The effect sizes range from weak to large, with the relationship between Leadership and Change Readiness having the weakest effect size. The final evaluation of the structural model is R-square which assesses goodness of fit in regression analysis. The R-square value gives the share of variance explained in a dependent construct. An evaluation of R-square values should be judged relative to studies that investigate the same dependent variable (Benitez et al. 2020). In this study, the R-square values range from 0.124 to 0.612, which are considered to be acceptable values since this study of transformational change in non-profit organisational setting is in its initial stages.

Findings

The purpose of this quantitative study is to examine empirically the moderating effects of commitment to change on the relationships between change readiness (ready to change) and organisational sustainability in three scenarios of transformational changes—pre-, during, and post-NDIS implementation. Table 6a presents the path coefficients of the direct relationships between change readiness and sustainability, and the indirect relationships between change readiness and sustainability moderated by commitment to change. The results of the partial least squares (PLS) analysis show the significant relationships between change readiness, commitment to change, and sustainability at different levels of effects.

Table 6. a. Three-stage model comparisons. b. Regression coefficient comparison.

			a			
	Pre-		During		Post-NDIS Implementation	
	Path Coeff (STDEV)	*p*-Value	Path Coeff (STDEV)	*p*-Value	Path Coeff (STDEV	*p*-Value
Internal communication -> **Change readiness**	−0.092 (0.097)	0.342	0.208 (0.062)	0.001	−0.007 (0.08)	0.927
Leadership -> **Change readiness**	0.289 (0.105)	0.006	0.442 (0.071)	0.000	0.485 (0.083)	0.000
Culture -> **Change readiness**	0.372 (0.109)	0.001	0.279 (0.07)	0.000	0.331 (0.091)	0.000
Change readiness -> Risk	−0.143 (0.114)	0.21	−0.156 (0.126)	0.217	−0.164 (0.136)	0.227
Change readiness -> Growth	−0.113 (0.073)	0.122	0.29 (0.084)	0.001	0.082 (0.112)	0.465
Change readiness -> Commitment to change	0.567 (0.101)	0.000	0.727 (0.047)	0.000	0.782 (0.046)	0.000
Commitment to change -> Risk	0.531 (0.159)	0.001	0.45 (0.119)	0.000	0.653 (0.146)	0.000
Commitment to change -> Growth	0.754 (0.062)	0.000	0.444 (0.081)	0.000	0.662 (0.101)	0.000
			b			
	Pre- vs. During		During vs. Post-		Pre- vs. Post-	
	Path Coeff Diff	*p*-Value	Path Coeff Diff	*p*-Value	Path Coeff Diff	*p*-Value
Internal communication -> Change readiness	0.301	0.006	0.216	0.034	0.085	0.506
Leadership -> Change readiness	0.153	0.23	−0.043	0.696	0.195	0.145
Culture -> Change readiness	−0.093	0.479	−0.052	0.655	−0.041	0.773
Change readiness -> Risk	−0.013	0.936	0.008	0.983	−0.021	0.925
Change readiness -> Growth	0.403	0.001	0.208	0.142	0.195	0.144
Change readiness -> Commitment to change	0.16	0.114	−0.055	0.39	0.215	0.027
Commitment to change -> Risk	−0.082	0.671	−0.203	0.282	0.122	0.585
Commitment to change -> Growth	−0.31	0.003	−0.218	0.095	−0.092	0.435

4.3. The Effects of Change Readiness on Sustainability

Internal communication, both in terms of quantity and design, has been determined to be a key ingredient by employees in order to support organisational readiness to change during the implementation of the NDIS. The quantitative analysis supports the view that, during the implementation of the NDIS, internal communication, along with culture and leadership, supports the ability of employees to support organisational growth, leading to organisational sustainability, whilst having no effect on risk (see Table 6a). As it is presented in the difference testing of the regression coefficients, the role of internal communication is significantly different during the implementation of NDIS from the pre-NDIS implementation (see Table 6b, path coefficient difference = 0.301, p-value < 0.01) and post-NDIS implementation phases (path coefficient difference = 0.216, p-value < 0.05)). This means that internal communication plays a significant role largely during the implementation of NDIS. It can, therefore, be determined that internal communication, as described above, assists employees to accept the necessary organisational changes driving the shift from the previous supply-driven model to the newly created demand-driven model for service-provision design and delivery. In this manner, it becomes an important feature in overall change readiness considerations, both from a timing and an execution perspective.

In all three phases, employee attitude towards leadership and culture has a significant influence on the creation of organisational sustainability (growth and risk) mediated by change readiness and commitment to change.

During the implementation of the NDIS, employees must be ready to change, as reflected in Table 6a. Change readiness has both a direct and an indirect effect on the sustainability of organisational growth. Based on Table 6a, the direct effect of change readiness on organisational growth and sustainability is significant in the scenario of "during" NDIS implementation (standardised coefficient = 0.29, p-value < 0.001). However, change readiness does not directly affect organisational risk in relation to sustainability in the three scenarios of transformation change.

4.4. Commitment as a Change Moderator

In routine/regular activity, which occurs in the pre-and post-NDIS implementation phases, change readiness alone is not enough to impact organisational sustainability unless employees also have adequate commitment to the change process and recognise the need for change. Our research reflects the interdependencies associated with change readiness and change commitment. Commitment to change is needed in all situations to drive employees to make the necessary organisational changes that can lead to sustainability, both from a growth and a risk perspective (see Table 6a), recognising that the latter also reflects opportunities associated with the organisational view of a future in a post-NDIS environment. However, during the implementation of the NDIS, change commitment has a lower significant role compared to the other two phases of NDIS implementation (see Table 6a,b the regression of coefficient of implementation is lower than those of the other two phases with p-value < 0.001). Our research suggests that there may be higher levels of change readiness and change commitment after the implementation phase of the NDIS than before. This points to the realities of the implementation of the NDIS moving employees to a state of acceptance that could support any ongoing post-implementation changes that may be required. Accordingly, the fact that employee commitment grows as the implementation progresses reinforces the view that the pre-NDIS implementation and the implementation phases play an important role leading to the full implementation.

4.5. Sustainability: The Outcome of Change

Our research suggests that, in all stages, organisational sustainability (both risk and growth) is affected by the levels of change readiness and change commitment. The path coefficients between commitment to change and sustainability (both risk and growth) range from 0.44 to 0.754, with p-value < 0.001. However, the influence of change commitment on the growth factors of sustainability, after the implementation of the NDIS, is lower than that before the implementation (path coefficient of 0.754 compared to 0.662), meaning a timely focus on change commitment is an important overall ingredient. Although the path coefficient difference between the pre- and post-NDIS implementation phases is not significant (−0.92, *p*-value = 0.435), commitment to change has a significant contribution effect on organisational growth and sustainability (see Table 5, R-squares are around 0.5).

5. Discussion

Continuous change in most organisations seems to be the norm, especially in healthcare, although not all are as radical a change as is the introduction and implementation of the NDIS. It appears to have proven even more complex and challenging since its first steps in 2016 and, as it continues to grow and escalate in costs, it is imperative to increase our understanding of such change and its management for success. This represents ongoing challenges for organisational leadership in this sector as the NDIS reflects societal expectations which must be interpreted and managed at an organisational level, ensuring that service design and delivery can meet the expectations of those whom the NDIS was designed to service. Other challenges associated with the NDIS highlight possible linkages with socio-economic issues (Cortese et al. 2021), which, whilst focused on a policy level, will impact service delivery organisations over time as service design and delivery could, to some extent, be partially impacted by Australia's geography and population locations (Wiesel et al. 2017).

Both positive and negative views of major change management theories and models abound. Most recently, Chowthi-Williams and Davis (2022, p. 1) claim that "readiness for change could provide the energy, motivation, and engagement for successful change management. . . . " urging that leaders must deal with change management inhibitors and focus more on their people and energy." Lailla (2022, p. 404) found that "organizational change was related to changes in strategy, culture, employee attitudes, organizational structure, technology, communication leadership, and employee development affecting employee performance."

But, here, rather than taking on an a priori model and testing it, we have developed a pragmatic framework from the change management experience of a variety of nonprofit organisations dealing with the challenges wrought by implementing the NDIS. In doing so, some of the views expressed above, coming after our work, nevertheless come to the fore in our research.

Consequently, there were many lessons learnt during this research on the move from a supply-driven to a demand-driven approach, which has challenged the structures, skills, processes, and mindsets that are long-embedded and need to change in a competitive environment where client service delivery is turned on its head. The sustainability of the participating organisations emerges as a key challenge in the generational shift, leading us to explore the efficacy of the framework for successful NDIS implementation in the nonprofit sector. This is a broad change management approach in terms of its characteristics, instead of a narrow prescriptive model that would inhibit useful individual organizational characteristics and idiosyncrasies that are necessary in innovation and resilience.

Across change readiness, implementation, and sustainability, we teased out critical components for consideration, leading to the creation of the NDISIF for the sector. Stakeholder analysis, both internal and external, is important, as are other components, such as leadership and culture. Based on a qualitative study of seven organisations using grounded theory and framework analysis, for the quantitative study to consolidate and teste the findings of the original work, five more organisations were added.

The lessons learnt are in the success factors across readiness, implementation/commitment, and finally sustainability-leadership, culture, change management, organizational foundations, service utility, communication, externality, mission, risk, and opportunity. These are found to be crucial in both the qualitative and quantitative research studies.

6. Conclusions

6.1. Implications for Theory

First, the research undertaken and reported in this study adds to the body of change management work in the nonprofit sector, which is often ignored in favour of the for-profit sector. Moreover, the mixed methods approach we used is different from many other studies, ending in the creation of a novel framework, which is grounded in the initial qualitative work, through interviews primarily, and then re-examined and tested for validity in the quantitative work through surveys.

6.2. Implications for Practice

Given the ongoing critiques of the NDIS and the government body overseeing its operations, the National Disability Insurance Agency (NDIA), since its pilot trials in 2016, it remains important in 2022 and beyond to provide the sort of framework that our research has produced, the NDISIF, to assist the ongoing NDIS implementation challenges in the nonprofit sector. Moreover, whilst the original figure of those to be serviced by the NDIS was given as 45,000, this has now grown to be 560,000 and continues to grow, as does the ballooning cost and risks of the scheme, a challenge for whatever government is in office. Furthermore, with the current challenges of rising inflation, competition for workforce talent, and other economic and political challenges, it is imperative that research provides the much-needed support for disability organisations in the nonprofit sector to succeed, or we will fail both the provider organisations and their clients. The ramifications for nonprofit NDIS service delivery organisations are substantial. Managerial challenges for leadership in these organisations are embodied in sustainability issues stemming directly from an inability to mitigate the substantial risks associated with implementation, whilst on the other hand, failure to identify and act, in a timely fashion, on the equally substantial opportunities that present themselves as a direct result of the changes to service design and delivery that the NDIS relies on. Accordingly, the implementation of the NDIS will result in "winners" and "losers" both at the organisational level as well as amongst service users (Green and Mears 2014).

6.3. Future Research

We are hoping that the NDISIF framework, in providing guidance and an approach necessary, we believe, for success, can also be a useful framework for other similar social systemic change management, including both large-scale change projects as in, for example, aged care services, housing, and education, as well as other smaller-scale change projects within organisations.

We also hope to refine our methodological approach to see how useful it is in understanding some of the current "wicked problems" confronting society, especially so we can enhance it for future research in the nonprofit sector and publication in a proposed future Handbook in Nonprofit Change Management.

6.4. Limitations

A key limitation is that we will need to broaden the focus area of NDIS implementation as it may not always be a basis for the change management framework we have devised when the framework is tested against other change management approaches, contexts, and locations, given its characteristics may not cross over well into other challenging areas. The addition of new organisations only in the quantitative part of the research may also have had unintended consequences, although we feel this strategy enhanced the richness of the data for the study. The issue of the usefulness of the NDISIF in the for-profit sector is also yet to be explored more fully.

Author Contributions: All authors contributed to all sections of the paper. Additionally, All authors have read and agreed to the published version of the manuscript.

Funding: This research was in part funded by three industry partners within the Australian non-rpofit sector including Koorana Child and Family Services Inc. (Sydney, Australia); Junction Works Ltd. (Melbourne, Australia), and Wayside Chapel (Sydney, Australia).

Data Availability Statement: Data comprises responses from online survey. These are held on file by the researchers.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Measurement Invariance of Composite Models (MICOM)

	Compositional Invariance				Composite Equality									
	Original Correlation	Correlation Permutation Mean	5.00%	Permutation *p*-Values	Mean—Original Difference	Mean—Permutation Mean Difference	2.50%	97.50%	Permutation *p*-Values	Variance—Original Difference	Variance—Permutation Mean Difference	2.50%	97.50%	Permutation *p*-Values
Pre vs. During														
Change readiness	0.999	0.999	0.997	0.474	−0.362	0	−0.27	0.267	0.002	0.191	0	−0.39	0.387	0.358
Culture	1	0.994	0.978	0.984	−0.054	0.006	−0.249	0.267	0.699	0.012	−0.003	−0.38	0.368	0.957
Internal communication	0.989	0.965	0.873	0.549	0	0.001	−0.253	0.253	0.992	−0.081	0	−0.29	0.272	0.578
Leadership	0.994	0.997	0.99	0.147	−0.051	0.004	−0.259	0.252	0.717	−0.126	0	−0.38	0.381	0.55
Risk	0.998	0.991	0.966	0.632	−0.169	0.004	−0.261	0.265	0.196	0.108	−0.004	−0.43	0.452	0.691
Sustainability	1	0.999	0.998	0.808	0.005	0.001	−0.253	0.261	0.959	−0.027	0.003	−0.39	0.391	0.899
Commitment to change	0.999	0.999	0.996	0.532	−0.216	0.008	−0.237	0.253	0.092	−0.102	0.001	−0.34	0.346	0.555
During vs. Post														
Change readiness	1	0.999	0.998	0.474	0.26	−0.001	−0.24	0.235	0.043	−0.053	0.002	−0.34	0.317	0.381
Culture	1	0.996	0.983	0.856	−0.086	0.001	−0.223	0.243	0.25	−0.025	0.005	−0.32	0.345	0.457
Internal communication	1	0.985	0.944	0.88	0.047	0	−0.227	0.228	0.362	0.002	−0.009	−0.24	0.209	0.486
Leadership	0.999	0.997	0.99	0.64	0.083	0.001	−0.223	0.226	0.267	0.05	0.003	−0.32	0.339	0.408
Risk	0.99	0.992	0.968	0.259	0.182	0.004	−0.22	0.23	0.096	−0.01	0.003	−0.26	0.283	0.496
Sustainability	1	0.999	0.998	0.516	−0.16	0.002	−0.219	0.237	0.11	0.11	−0.002	−0.36	0.367	0.305
Commitment to change	1	0.999	0.998	0.629	0.23	0.003	−0.234	0.236	0.055	−0.274	0.003	−0.36	0.364	0.097
Pre vs. Post														
Change readiness	1	1	0.999	0.826	−0.112	0.003	−0.234	0.234	0.214	0.142	−0.008	−0.32	0.33	0.224
Culture	1	0.994	0.977	0.878	−0.138	−0.002	−0.237	0.225	0.159	−0.013	0.005	−0.27	0.274	0.456
Internal communication	0.994	0.951	0.794	0.679	0.047	0.002	−0.237	0.238	0.378	−0.078	0.002	−0.25	0.259	0.283
Leadership	0.99	0.996	0.985	0.126	0.042	0.002	−0.235	0.239	0.393	−0.083	−0.005	−0.34	0.32	0.361
Risk	0.997	0.993	0.972	0.559	0.003	0.003	−0.226	0.225	0.508	0.121	−0.004	−0.43	0.395	0.345
Sustainability	0.999	0.999	0.997	0.355	−0.156	0.001	−0.249	0.236	0.142	0.08	0.007	−0.33	0.347	0.363
Commitment to change	1	0.999	0.997	0.66	0.031	0.002	−0.247	0.229	0.426	−0.373	0.005	−0.35	0.362	0.032

Appendix B. Construct correlation matrix of Pre-, During, and Post-NDIS Implementation

	Pre-NDIS Implementation							**During NDIS Implementation**							**Post-NDIS Implementation**						
	1	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1: Change readiness	1							1							1						
2: Culture	0.655	1						0.853	1						0.739	1					
3: Internal communication	0.103	0.579	1					0.514	0.487	1					0.384	0.523	1				
4: Leadership	0.607	0.808	0.191	1				0.911	1.055	0.344	1				0.876	0.768	0.637	1			
5: Risk	0.194	0.532	0.234	0.264	1			0.193	0.342	0.292	0.165	1			0.447	0.816	0.472	0.646	1		
6: Sustainability	0.346	0.843	0.54	0.5	0.465	1		0.691	0.758	0.64	0.678	0.388	1		0.656	0.961	0.551	0.601	0.743	1	
7: Commitment to change	0.691	1	0.469	0.588	0.602	0.833	1	0.876	0.855	0.657	0.792	0.435	0.769	1	0.883	0.798	0.589	0.795	0.704	0.812	1

References

Akter, Sharhriar, John D'ambra, and Pradeep Ray. 2011. An evaluation of PLS based complex models: The roles of power analysis, predictive relevance and GoF index. Paper presented at 17th Americas Conference on Information Systems (AMCIS 2011), Detroit, MI, USA, August 4–8; Available online: https://aisel.aisnet.org/amcis2011_submissions/151 (accessed on 15 March 2022).

Andrews, Kevin. 2019. *NDIS Planning Interim Report by Joint Standing Committee on the NDIS*. Commonwealth of Australia. Canberra: Joint Standing Committee on the National Disability Insurance Scheme.

Bagozzi, Richard, and Youjae Yi. 1988. On the evaluation of structural equation models. *Journal of the Academy of Marketing Science* 16: 74–94. [CrossRef]

Baron, Reuben, and David Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51: 1173. [CrossRef] [PubMed]

Beniflah, Jake, and Julie Veloz. 2021. Building a multicultural organisation: A conceptual model for organisational change in the 21st century. *Journal of Cultural Marketing Strategy* 5: 153–68.

Benitez, Jose, Jorge Henseler, Ana Castillo, and Florian Schuberth. 2020. How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management* 57: 103168.

Bouquet, Cyril, Julian Birkinshaw, and Jean-Louis Barsoux. 2016. Fighting the headquarters knows best syndrome. *MIT Sloan Management Review* 57: 59.

Burke, Warner. 2017. *Organization Change: Theory and Practice*. London: Sage Publications.

Chin, Wynne. 2010. How to write up and report PLS analyses. In *Handbook of Partial Least Squares*. Berlin: Springer.

Chowthi-Williams, Annette, and Geraldine Davis. 2022. *Successful Change Management in Health Care: Being Emotionally and Cognitively Ready*. Oxon: Routledge.

Commonwealth of Australia. 2021. *Summary of Australia's Disability Strategy 2021–2031*. Canberra: Commonwealth of Australia.

Corritore, Matthew, Amir Goldberg, and Sameer Srivastava. 2020. The New Analytics of Culture. *Harvard Business Review* 98: 77–83.

Cortese, Corinne, Freda Truscott, Mona Nikidehaghani, and Sandra Chapple. 2021. Hard-to-reach: The NDIS, disability, and socio-economic disadvantage. *Disability & Society* 36: 883–903.

Dawson, Patrick, and Lisa Daniel. 2010. Understanding social innovation: A provisional framework. *International Journal of Technology Management* 51: 9–21. [CrossRef]

Denscombe, Martyn. 2008. Communities of practice: A research paradigm for the mixed methods approach. *Journal of MIxed Methods Research* 2: 270–83. [CrossRef]

Easterby-Smith, Mark, Richard Thorpe, and Andy Lowe. 1997. *Management Research: An Introduction*. London: Sage.

El-Dirani, Ali, Mohammad Houssein, and Hussin Hejase. 2019. An Exploratory Study of the Role of Human Resources Management in the Process of Change. *Open Journal of Business and Management* 8: 156–74. [CrossRef]

Foster, Michele, Eloise Hummell, Karen Fisher, Samantha Borg, Catherine Needham, and Alyssa Venning. 2021. Organisations adapting to dual aspirations of individualisation and collaboration in the National Disability Insurance Scheme (NDIS) market. *Australian Journal of Public Administration* 81: 127–44. [CrossRef]

Frei, Frances, and Anne Morriss. 2020. Begin with Trust—The first step to becoming a genuinely empowering leader. *Harvard Business Review* 98: 112–21.

Gilchrist, David, and Ben Perks. 2022. The Challenge of Sustainability: Queensland's Not-for-profit Sector and the Impact of Growing Financial Pressure. Available online: https://api.research-repository.uwa.edu.au/ws/portalfiles/portal/173099048/2022_FINAL_Queensland_NFPs_Cost_Increases.pdf (accessed on 15 March 2022).

Green, Jenny, and Jane Mears. 2014. The implementation of the NDIS: Who wins, who loses? *Cosmopolitan Civil Societies: An Interdisciplinary Journal* 6: 25–39. [CrossRef]

Green, Celia, Eleanor Malbon, Gemma Carey, Helen Dickinson, and Daniel Reeders. 2018. *Competition and Collaboration between Service Providers in the NDIS*. Sydney: Centre for Social Impact, UNSW Sydney.

Henseler, Jorg, Theo Dijkstra, Marko Sarstedt, Christian Ringle, Adamantios Diamantopoulos, Detmar Straub, David Ketchen Jr., Joseph Hair, Tomas Hult, and Roger Calantone. 2014. Common beliefs and reality about PLS: Comments on Rönkkö and Evermann' (2013). *Organizational Research Methods* 17: 182–209. [CrossRef]

Hu, Li-tze, and Peter Bentler. 1999. Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6: 1–55. [CrossRef]

Hu, Li-tze, Peter Bentler, and Yutaka Kano. 1992. Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin* 112: 351. [CrossRef] [PubMed]

Kätelhön, Arne, Raoul Meys, Sarah Deutz, Sangwon Suh, and Andre Bardow. 2019. Climate change mitigation potential of carbon capture and utilization in the chemical industry. *Proceedings of the National Academy of Sciences* 116: 11187–94. [CrossRef]

Kimberlin, Sara. 2010. Advocacy by nonprofits: Roles and practices of core advocacy organizations and direct service agencies. *Journal of Policy Practice* 9: 164–82. [CrossRef]

Lailla, Nor. 2022. Organizational Change On Employee Performance. *Procedia of Social Sciences and Humanities* 3: 404–10. [CrossRef]

Leech, Nancy, and Anthony Onwuegbuzie. 2009. A typology of mixed methods research designs. *Quality & Quantity* 43: 265–75.

Lewin, Kurt. 1947. Frontiers in group dynamics: Channels of group life; social planning and action research. *Human Relations* 1: 143–53. [CrossRef]

Lewin, Kurt. 1951. *Field Theory in Social Science: Selected theoretical Papers*. Edited by Dorwin Cartwright. New York: Harper & Brothers.

Meltzer, Ariella, and Luara Davy. 2019. Opportunities to enhance relational wellbeing through the National Disability Insurance Scheme: Implications from research on relationships and a content analysis of NDIS documentation. *Australian Journal of Public Administration* 78: 250–64. [CrossRef]

Onyx, Jenny, Liz Cham, and Bronwen Dalton. 2016. *Current Trends in Australian Nonprofit Policy*. Paper Presented at the Nonprofit Policy Forum. Berlin: De Gruyter.

Page, Lesley, Sheila Boysen, and Terry Arya. 2019. Creating a culture that thrives. *Organization Development Review* 51: 28–35.

Parker, Hamieda, and Khadija Ameen. 2018. The role of resilience capabilities in shaping how firms respond to disruptions. *Journal of Business Research* 88: 535–41. [CrossRef]

Rosenbaum, David, and Elizabeth More. 2021. Risk and Opportunity—The Leadership Challenge in a World of Uncertainty—Learnings from Research into the Implementation of the Australian National Disability Insurance Scheme. *Journal of Risk and Financial Management* 14: 383. [CrossRef]

Rosenbaum, David, and Elizabeth More. 2022. Towards a Strategic Change Management Framework for the Nonprofit Sector: The Roll-out of Australia's National Disability Insurance Scheme. In *Effective Implementation of Transformation Strategies: How to Navigate the Strategy and Change Interface Successfully*. Edited by Shelley Kirkham, Danielle Tucker, Kate Hughes, Angelina Zubac and Ofer Zwikael. London: Palgrave Macmillan, pp. 415–50.

Rosenbaum, David, Elizabeth More, and Peter Steane. 2017. A longitudinal qualitative case study of change in nonprofits: Suggesting a new approach to the management of change. *Journal of Management & Organization* 23: 74–91.

Rosenbaum, David, Elizabeth More, and Peter Steane. 2018. Planned organisational change management: Forward to the past? An exploratory literature review. *Journal of Organizational Change Management* 31: 286–303. [CrossRef]

Rothwell, William, Sohel Imroz, and Behnam Bakhshandeh. 2021. *Organization Development Interventions: Executing Effective Organizational Change*. New York: Taylor and Francis.

Sarstedt, Marko, and Christian Ringle. 2011. Multigroup Analysis in Partial Least Squares (PLS) Path Modeling: Alternative Methods and Empirical Results. *Measurement and Research Methods in International Marketing Advances. International Marketing* 22: 195–218.

Shani, Abraham. 2021. *Collaborative Inquiry for Organization Development and Change*. Cheltenham: Edward Elgar Publishing.

Smith, Aaron, James Skinner, and Daniel Read. 2020. *Philosophies of Organizational Change: Perspectives, Models and Theories for Managing Change*. Cheltenham: Edward Elgar Publishing.

Stace, Doug, and Dexter Dunphy. 2001. *Beyond the Boundaries: Leading and Re-Creating the Successful Enterprise*, 2nd ed. New York: McGraw-Hill.

Steenkamp, Jan-Benedict, and Hans Baumgartner. 1998. Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research* 25: 78–90. [CrossRef]

Stouten, Jeroen, Demise Rousseau, and David De Cremer. 2018. Successful organizational change: Integrating the management practice and scholarly literatures. *Academy of Management Annals* 12: 752–88. [CrossRef]

Taylor, Rachel, Nuttaneeya Torugsa, and Anthony Arundel. 2020. Social innovation in disability nonprofits: An abductive study of capabilities for social change. *Nonprofit and Voluntary Sector Quarterly* 49: 399–423. [CrossRef]

UN. 2006. Convention on the Rights of Persons with Disabilities. Available online: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html (accessed on 15 September 2022).

Van Dierendonck, Dirk. 2011. Servant leadership: A review and synthesis. *Journal of Management* 37: 1228–61. [CrossRef]

Vargas, Valerie, Rebecca Lawthom, Alicia Prowse, Sally Randles, and Konstantinos Tzoulas. 2019. Sustainable development stakeholder networks for organisational change in higher education institutions: A case study from the UK. *Journal of Cleaner Production* 208: 470–78. [CrossRef]

Venus, Merlijn, Daan Stam, and Daan Van Knippenberg. 2019. Visions of change as visions of continuity. *Academy of Management Journal* 62: 667–90. [CrossRef]

Waddell, Dianne, Andrew Creed, Thomas Cummings, and Christopher Worley. 2019. *Organisational Change: Development and Transformation*. South Melbourne: Cengage Learning.

Wiesel, Ilan, Carolyn Whitzman, Christine Bigby, and Brendan Gleeson. 2017. *How Will the NDIS Change Australian Ciities?* Parkville: Melbourne Sustainable Society Institute.

Journal of
Risk and Financial Management

Article

Benchmarking—A Way of Finding Risk Factors in Business Performance

Jarmila Horváthová, Martina Mokrišová * and Mária Vrábliková

Faculty of Management, University of Prešov, Konštantínova 16, 080 01 Prešov, Slovakia; jarmila.horvathova@unipo.sk (J.H.); maria.vrablikova@smail.unipo.sk (M.V.)
* Correspondence: martina.mokrisova@unipo.sk

Abstract: The purpose of this study was to emphasize that the Data Envelopment Analysis (DEA) method is an important benchmarking tool which provides necessary information for improving business performance. To fulfil the abovementioned goal, we used a sample of 48 Slovak companies involved in the field of heat supply. As their position in the economic and social environment of the country is essential, considerable attention should be paid to improving their performance. In addition to the DEA method, we applied the Best Value Method (BVM). We found that DEA is a highly important benchmarking tool, as it provides benchmarks for units that have problems with performance and helps us to reveal risk performance factors. The DEA method also allows us to determine target values of indicators. The originality of this paper is in its comparison of the results of the BVM and the DEA methods.

Keywords: benchmarking; best value; business; data envelopment analysis (DEA); performance; risk factors

Citation: Horváthová, Jarmila, Martina Mokrišová, and Mária Vrábliková. 2021. Benchmarking—A Way of Finding Risk Factors in Business Performance. *Journal of Risk and Financial Management* 14: 221. https://doi.org/10.3390/jrfm14050221

Academic Editor: George Halkos

Received: 9 April 2021
Accepted: 11 May 2021
Published: 14 May 2021

1. Introduction

In the current turbulent period, it is necessary for every company to monitor market changes, not only in the areas of marketing and innovation, but also in the field of business performance and the application of financial indicators in its measurement. Businesses are looking for different ways to increase their performance and achieve a competitive advantage over their competitors. Therefore, they use different methods, techniques and indicators. One of the methods than can be applied in this area is benchmarking. Benchmarking finds the best practices for knowledge and know-how by learning from other companies, competitors and industry leaders to gain a competitive advantage in global competition (Tian and Ketsaraporn 2013).

Benchmarking is part of the concept of quality management. It is a method of analyzing and comparing practices and experiences in various areas of business operations (Kaczmarska 2010). This method was first used in the 1980s by the Xerox Corporation to improve its competitive position (Demjanová 2006). According to Veber (2000), benchmarking is based on two principles. The first is the principle of the Chinese general Sun-c (500 BC): "If you know the enemy and know yourself, you need not fear the result of a hundred battles." The second principle is being the best of the best.

The purpose of benchmarking is to be inspired by the best competitors (best practices). This, however, does not imply imitation. This is also confirmed by one of the definitions of benchmarking by R. C. Camp as "finding best practices in business that lead to excellent results" (Kiseľáková and Šofranková 2014). Benchmarking is currently one of the most widely used management tools, and it is applied in order to increase business performance (Bogetoft 2012). Benchmarking includes the benchmarking of products and services, business processes and performance measures (Maleyeff 2003).

Based on the above, we set the research problem and aim of the paper. The research problem is the following: Which benchmarking method is able to identify performance risk factors? The aim of the paper considers how to improve businesses' performances by applying selected benchmarking methods.

The next part of the text is structured as follows: The first section outlines the theoretical basis of benchmarking, performance benchmarking, performance benchmarking tools and performance risk factors. The second section describes the research sample and the methodology. The research problem and research questions are formulated in this part of the paper. When addressing the research problem, we made use of selected financial indicators, the BVM, the DEA and the Spearman's rank correlation coefficient. The third section includes results and a discussion of the results achieved. This section lists scores and rankings of companies obtained by the BVM and the DEA. The strength and association between rankings are determined by Spearman's rank correlation coefficient. A special part of this section is devoted to the calculation of target values—benchmarks for improving businesses' performances. The final part of the paper is the conclusion, which provides recommendations for improving business performance and the benefits of the DEA in terms of performance improvement. The process of the research is illustrated in Figure 1.

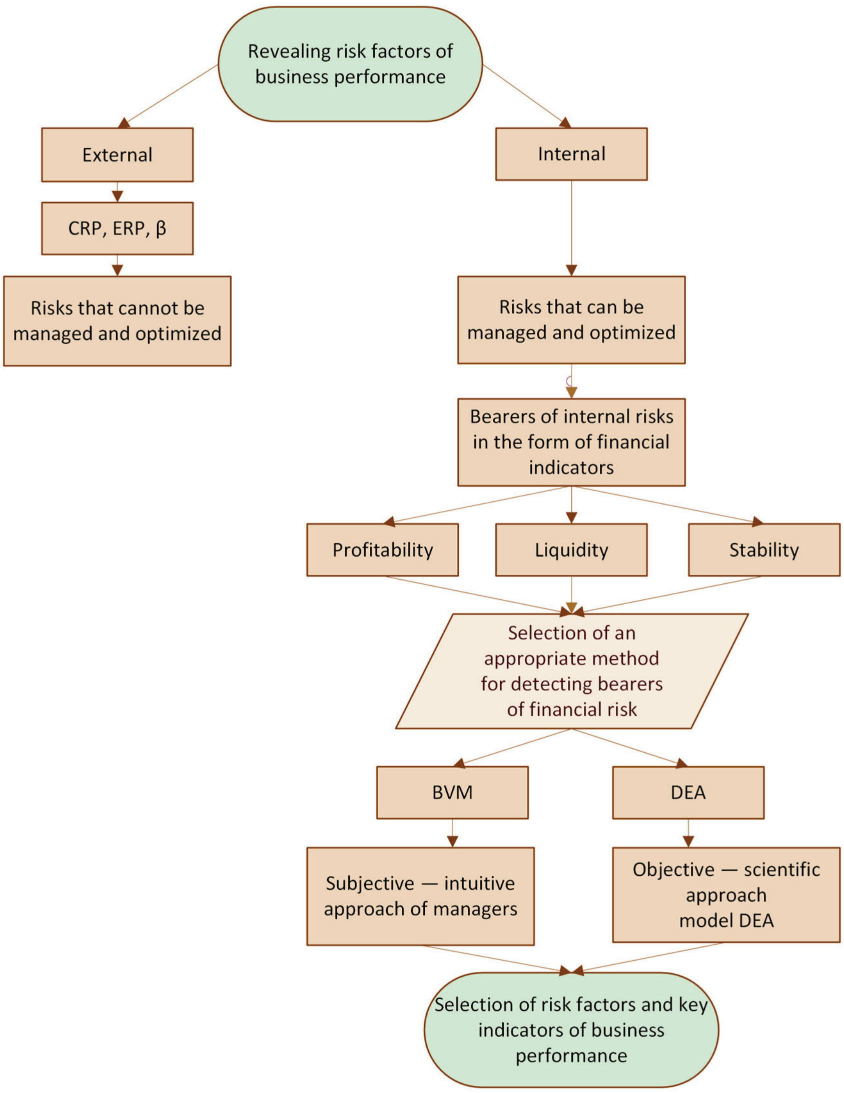


Figure 1. Flowchart of the research. Legend: CRP—Country Risk Premium, ERP—Equity Risk Premium, β—systematic risk.

2. Literature Review

Benchmarking studies can be divided according to the type of benchmarking into: process, functional and performance benchmarking studies (Nenadál et al. 2011). A similar classification is given by (Fong et al. 1998 in Anand and Kodali 2008), who divide benchmarking according to its content into process, functional, performance and strategic benchmarking. As part of our research, we focused on solving the issue of performance benchmarking, with the aim of increasing the performance of businesses.

Performance benchmarking is used by an organization to determine its market position compared to the performance characteristics of other organizations. To maintain ethical rules in the commercial environment, it is appropriate to use a third party for this type of benchmarking. It is the most common type and will probably be the starting point for any procedures aiming at the overall operation of an organization (Pogranová et al. 2011).

Performance benchmarking is a combination of two methodologies: benchmarking and performance management. According to Camp (1995, p. 16), the aim of benchmarking performance measures is to set and validate goals for several vital performance measures which guide an organization. Benchmarking is imperative for performance (Krishnamoorthy and DLima 2014). Many organizations include benchmarking as a component of their performance management system (Maleyeff 2003).

Benchmarking is also an important tool for identifying key performance indicators (Ho et al. 2000). Well-known key performance indicators are related to financial analysis (Bogetoft 2012). If we, during a performance measurement, find out which key financial indicators are lagging behind our competitors, it is advantageous for us to improve their values while preventing other indicators from creating an imbalance. Conversely, if we find out in which aspects we are significantly better than our competitors, and this indicator does not have a negative impact on other financial objectives, we can present this as a competitive advantage. Benchmarking is not only a measurement tool but a process of identifying gaps in performance, while the elimination of gaps and continuous improvement can bring the company closer to its competitors (Zairi and Leonard 1996).

When comparing businesses' performances, it is possible to apply different indicators (Štefko et al. 2012). Bogetoft (2012) states ROA, gross margin, debt ratio and price/book (stock capitalization/book values) are important performance indicators. According to Štefko et al. (2021), these measures include selected indicators of liquidity, profitability, capital structure and activity. According to Wood and McConney (2018), important performance indicators include, for example, return on assets (ROA), return on equity (ROE), return on capital employed (ROCE) and net interest margin (NIM). Additionally, according to Bărbută-Misu et al. (2019), the most used performance measures include ROA and ROE, but they mention also liquidity, leverage, productivity, solvency and assets turnover. One of the modern benchmarking models of enterprise performance evaluation is the Creditworthy Model (CWM), which is the most suitable model for the comparison of financial performance between two or more enterprises (Kisel'áková et al. 2018).

The selection of indicators must be based on the definition of performance. Business performance is understood by most authors as the ability of a business to value the investment invested in its business activities in the best way (Johnson and Kevan 2000). Therefore, the most commonly used performance measurement indicators are financial indicators based on the primary objective of business—to maximize profits. We talk about financial indicators (profitability, liquidity, indebtedness, activity and market value), but new indicators are being added, such as EVA (economic value added), MVA (market value added), CVA (cash value added), WACC (weighted average costs of capital), RONA (return on net assets) and the like (Neumaierová and Neumaier 2002; Ittner et al. 2003; Frost 2005; Šulák and Vacík 2005; Petřík 2009; Kislingerová et al. 2011).

Accounting measures of performance have been the traditional mainstay of quantitative approaches to organizational performance measurement. However, over the past two decades, a great deal of attention has been paid to the development and use of non-financial measures (customer loyalty, customer satisfaction, capacity utilization and prevention of heat losses, renewal of production facilities, employee satisfaction, employee qualification, water protection, waste reduction, employee environmental awareness) of performance (Neely 2002; Otley 1999).

When solving the given problem, various methods from the area of performance benchmarking were applied in individual studies. For example, in the gas industry, the following methods were applied: Agrell and Bogetoft (2007) estimated the efficiency of gas distribution in Germany by applying a method of the Ordinary least squares, the Stochastic frontier analysis and the DEA; Rossi (2001) estimated the efficiency of the gas distribution sector in Argentina, and he used the Stochastic frontier and Malmquist index; Farsi et al. (2007) estimated the economic efficiency of the gas distribution sector in Switzerland by applying SFA; Erbetta and Rappuoli (2003) estimated the optimal scale and technical efficiency of the Italian gas distribution industry with the use of the DEA (Goncharuk 2008).

Significant methods that can be applied in the field of evaluation, and improvements in business performance include multidimensional benchmarking analyses based on the measurement of a number of criteria. These methods include, e.g., network charts, AHP maturity index and Z charts (Vochozka et al. 2017). Multidimensional methods also include the BVM method. It is possible to mention the following studies of authors who applied the BVM method in the field of performance benchmarking. Magd and Curry (2003) focused on the application of this method in the public sector. The Best Practice methodology was also applied in an article by Asrofah et al. (2010), in terms of increasing the efficiency of the manufacturing industry. The interconnection of DEA and BVM methods was realized in an article by the authors Yang et al. (2016), who, using the given methods, searched for winners in the field of supply and provided for their needs. A more detailed description of the BVM method is given in the section Data and Methodology.

An effort in the application of benchmarking methods in business performance improvement has been to develop benchmarking models that are able to perform multidimensional measurement (Vochozka et al. 2017). Benchmarking can be seen as a process of determining valid measurements for comparing the performance of comparable units in order to determine their relative positions and thus set the standard for highest quality. In this regard, the DEA method can be considered as a multidimensional benchmarking tool. This quantitative analysis method for evaluating the relative efficiency of a set of comparable, homogenous decision-making units (DMUs) has been often used as a practical benchmarking tool in management (An et al. 2021). The foundations of the DEA method were laid by Farrell (1957). At that time, this researcher thought methods such as measuring labor productivity or capital productivity were restrictive, as they did not cover the possibility of combining multiple input measurements, and were thus insufficient for expressing the overall efficiency of the company. These shortcomings have resulted in a more relevant approach applicable to any company, leading to a broader concept of productivity and a more general concept of efficiency. He was inspired by the work of Vilfred Paret, who, in his welfare theory, considered, in addition to increases and decreases, their importance, and assigned them some weights—this principle is now known as the Pareto criterion. Farrell is considered to be the father of the DEA, as DEA is based on the principles of linear programming. The basic task of DEA models is to compare DMUs within a group of units with the same or similar units. The definition of DMU is relatively flexible. DMU is the entity responsible for converting inputs to outputs, the effectiveness of which needs to be evaluated (Cooper et al. 2007; Cooper et al. 2011; Hatami-Marbini 2019; Ruiz and Sirvent 2019).

Klieštik (2009) explains that input-oriented approaches measure efficiency based on input variables and, in order to improve efficiency, a company should reduce the volume of inputs. Opposite these are output-oriented approaches that recommend increasing outputs to improve efficiency. Another criterion for the breakdown of DEA models is the return on scale. Constant returns to scale (CRS) occur when a 1% increase in input yields a 1% increase in output (CCR models). Variable returns to scale (VRS) occur when output increases by 1% or output changes by less than or more than 1% (BCC models).

The disadvantage of CCR and BCC models, the need to choose whether we are input or output oriented, has been overcome by the SBM model (Slack-Based Measure) by Tone (2001). The model works well for complicated tasks where it is not possible to determine whether a DMU should minimize inputs or maximize outputs; therefore, it is necessary to combine these requirements (Zimková 2015).

Measuring business efficiency using DEA models also has various advantages and disadvantages. The advantages, according to Majorová (2007) and Klieštik (2009), include working with multiple inputs and outputs at the same time, not requiring the normality of data distribution and quick identification of inefficient companies (companies below the efficiency limit), a comparison of such companies with efficient companies or objective measuring.

Another advantage of the DEA method is that it provides a more scientific basis for setting goals, and thus allows inefficient firms to find the easiest way to improve their performance. In 2008, Goncharuk applied three DEA models, as well as other performance benchmarking tools, to increase efficiency and effectiveness in the gas industry. The DEA method was used as a benchmarking method by Ruiz and Sirvent (2019), who applied it to improve business performance. Using the DEA for benchmarking ensures an evaluation in terms of targets that are attainable. Determining benchmarking information through closest efficient targets is one of the relevant topics in the recent DEA literature (Aparicio et al. 2014).

The above was confirmed also by Cooper et al. 2004, in Shewell and Migiro 2016, who state that studies of benchmarking practices using DEA have shown inefficiencies in some of the most profitable firms; therefore, DEA has been found to provide a better vehicle for establishing benchmarks than using profitability as a criterion. The DEA method was also used as a benchmarking method by Deville (2009), who compared branches and regional banks of a large French banking group. He applied DEA in the area of operational performance and, as a comparative variable, used the DEA score.

The benefits of the DEA method are also described in the work of Díaz et al. (2004). According to these authors, the DEA method can be used to assess and compare quantitative efficiencies and the weighting of any performance indicator, permitting managers to obtain a well-defined performance ranking.

The disadvantages of the DEA method include inefficiency deviations, which may be due to statistical distortion and a non-parametric approach that make it difficult to test hypotheses about inefficiency and the structure of the production function. DEA is non-statistical method; therefore, it does not yield estimates that can be easily validated with conventional statistical procedures (Banker 1990; Färe et al. 2001). Other disadvantages of the method concern its sensitivity to outliers (Coelli et al. 1998) and the fact that the method deals with relative efficiency (Farantos 2015). Another disadvantage of the DEA method is the sensitivity of its results to the selection of inputs and outputs, so their relative importance needs to be analyzed prior to the calculation. However, there is no way to test the appropriateness of inputs and outputs. Additionally, the number of DMUs on the frontier tends to increase with the number of inputs and outputs entering the model (Berg 2010; Zbranek 2013).

3. Data and Methodology

The research aimed at improving performance with the use of benchmarking methods was carried out on a sample of 48 Slovak companies in the field of heat supply. Companies in this sector use local district heating systems. Their sources and distribution of heat were built together with the development of urban agglomerations. Systems of these companies enable the efficient use of various energy sources produced in the city, including renewables, waste heat, etc. These systems are integrators of energy infrastructure, which can effectively link production and consumption and allow the storage of energy (in the form of heat) in times of surplus (Janiš 2018). The structure of these systems is provided by the climate and segmentation of the territory, historical development, demographic conditions, regional structure, the nature of residential, commercial and industrial construction, economic activity and the availability of fuel sources for heating (Antimonopoly Office of the Slovak Republic 2013). The analyzed industry is important from an economic as well as a social point of view, and plays an important role in the daily life of society and consumers. In this industry, a larger number of companies go bankrupt every year when compared to other Slovak industries. These companies have an important position in many Slovak districts. They are not subsidiaries of the parent company, but they are independent entities. They do not have interconnected management and each company represents a separate unit. In the area of fixed indicators and legislation, these companies have regulated heat prices. They are used to protect risk groups of the population from existential problems. Low energy prices have a positive effect on inflation, as well as on business development. However, on the other hand, heat management companies, in many cases, cannot adjust their variable and fixed costs to the level of regulated prices. Other important regulated indicators are indicators of environmental policy and environmental protection—emission limits for pollutants in the air, emission limits for pollutants in the water, wastewater limits or limits for waste reduction. The analyzed companies are not able to pay the costs and meet limits; therefore, they go bankrupt. However, they could be beneficial for the state and its people, as they provide alternative options for heat production and heat supply. Since they have an important position in many Slovak districts, in this paper, we focused on the evaluation, comparison and improvement of the performance and competitiveness of these companies. The data from the financial statements for the year 2016 were obtained from the Slovak analytical agency CRIF—Slovak Credit Bureau, s.r.o (CRIF 2016). The comparison of performance and subsequently competitiveness of these companies was realized with the use of two benchmarking methods—the BVM and the DEA.

In our research, we focused on the following 9 financial indicators which have a significant impact on the basic financial objectives of these companies. These indicators are the following: current ratio (CL) (1), average collection period (ACP) (2), creditors payment period (CPP) (3), return on assets (ROA) (4), return on equity (ROE) (5), return on sales (ROS) (6), equity ratio (ER) (7), interest coverage (IC) (8) and cost ratio (CR) (9).

Table 1 lists descriptive statistics for the analyzed businesses. The median of the current ratio indicates that half of the analyzed sample of businesses achieves a value of liquidity higher than 0.93, which can be considered appropriate in relation to the characteristics of the industry. The analyzed businesses have high creditors payment periods, the median of which is 183 days. They also achieve good results in profitability indicators in terms of both median and average. The capital structure of these companies is, on average, 30%: 70% in favor of debt, which may be the reason for the lower stability of these companies. From the point of view of interest coverage, the analyzed businesses are able to pay interest. The median of the cost ratio is 0.66.

Table 1. Descriptive statistics for the analyzed businesses.

Variable	Descriptive Statistics					
	Valid N	**Mean**	**Median**	**Minimum**	**Maximum**	**Std. Dev.**
Current ratio	48	2.5220	0.9337	0.032532	55.580	8.4011
Average collection period	48	61.6805	34.8444	5.028613	651.676	97.9389
Creditors payment period	48	297.3943	182.3338	2.176053	2941.961	473.0657
Return on assets	48	0.2332	0.1384	0.110630	2.071	0.3052
Return on equity	48	0.7502	0.5733	0.125611	6.227	0.9378
Return on sales	48	0.6160	0.2982	0.018377	7.215	1.4150
Equity ratio	48	0.3008	0.1814	0.054834	0.733	0.2190
Interest coverage	48	111.7940	5.1498	0.000000	2242.891	384.1357
Cost ratio	48	0.6862	0.6594	−0.251537	0.976	0.2139

As we stated in the Introduction, the research problem is as follows: Which benchmarking method is able not only to evaluate but also to improve business performance? In line with the research problem, we asked these research questions:

RQ1: What performance do individual companies achieve based on the BVM?
RQ2: What performance do individual companies achieve based on the results of an input-oriented DEA?
RQ3: What is the strength and direction of the association between the rankings achieved by the BVM and the DEA?

To analyze differences among businesses in space, we used a multidimensional map of objects, which uses an output of multidimensional scaling (MDS). Several authors (Rhee et al. 2009; Zema et al. 2020) used MDS as benchmarking technique. This map was also applied by Lukáčová et al. (2020) when analyzing and applying indicators in the field of tax harmonization and competitiveness. MDS allows us to test whether and how certain criteria by which one can distinguish among different objects of interest are mirrored in corresponding empirical differences of these objects (Borg and Groenen 1997). To express how well a data are represented by an MDS map, we applied Kruskal's *Stress*, which is the most widely used goodness-of-fit statistic. *Stress* is calculated according to Formula (1) (Kruskal 1964):

$$Stress = \sqrt{\frac{\sum_{k=1}^{m}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{k=1}^{m} d_{ij}^2}} \tag{1}$$

where $\hat{d}_{ij}$ expresses the predicted distance between objects i and j, and d_{ij} is the actual distance between objects i and j.

The smaller the value of the *Stress*, the more the calculated and entered object coordinates fit. According to Kruskal (1964) *Stress* around 0.20 means insufficient overlap, while 0.10 is sufficient, 0.05 is good, 0.025 is excellent and 0.00 is a perfect fit. When applying MDS, an important task is to determine the total number of required dimensions. The goal is to keep the number of dimensions as small as possible (usually, we choose 2-dimensional; maximum is 3-dimensional space). The number of dimensions is chosen based on the lowest possible value of the *Stress* criterion.

To deal with the research problem, we applied two methods. Using the BVM method, we selected the companies that achieve the best values within the individual indicators. Subsequently, the sum of the final evaluation of individual indicators was processed to the final ranking of companies. The DEA method worked with an already created model, which processed all input values of indicators and created a ranking of companies in the field of performance, according to the achieved values of indicators.

In order to answer the research question 1 (RQ1), we used the benchmarking tool which we adopted from the product benchmarking methodology (Kovaľová and Nogová 2016; Vrábliková and Loučanová 2017). This method is used to determine innovative intention of a product, and consists of the following steps:

- Selection of subjects for comparison (48 companies operating in the heating industry);
- Selection of evaluation criteria (9 indicators);
- Determining the weight of the criteria by means of a paired comparison under the questionnaire survey (48 respondents—financial managers of surveyed companies). The weight of the criteria is determined as follows:
    - Average significance: $Øv_i = 100/\text{number of criteria}$;
    - Significance coefficient: k_i = frequency of occurrence (based on pairwise comparison);
    - Average significance coefficient: $Øk_i = \sum k_i/\text{number of criteria}$;
    - Conversion using deviation: $D = (k_i - Øk_i) \times d$;
    - Deviation: $d = \sum v_i / \sum k_i$;
    - Real significance: $w = Øv_i + D$.

- the last phase is the BVM, which is based on the knowledge of the fair values of the benchmarking criteria and the real significance (w), resulting from the previous benchmarking steps. The tendency of the criterion (t) may increase or decrease depending on whether we want to maximize or minimize the criterion value. In order to identify the global benchmark, we proceed as follows:
  - The actual values of the criteria (x);
  - Transformed criterion values (a): if the criterion's tendency increases: a = actual value/highest value; if the criterion's tendency decreases: a = lowest value/actual value;
  - Point values for individual criteria (b): b = a × w;
  - Total score for the company (B): B = ∑b
  - Ranking of businesses according to their score.

The aim of the second step (RQ2) was to measure the performance of companies with the application of the input-oriented DEA CCR model. To solve this problem, we chose a model that was built as a dual task of linear programming, which uses the same data as the multiplicative model but reduces the number of model constraints (Kočišová 2012). Klieštik (2009) also considers it more advantageous and practical to work with a model that is a dual-model to the primary CCR model. In this case, the dual model will have $(m + s)$ constraints and $(n + m + s + 1)$ variables. Let us denote λj, $j = 1, 2, \ldots, n$, as dual variables belonging to the first set of constraints of the model, θ_o as scalar dual variable assigned to the next constraint and s_k^+, $k = 1, 2, \ldots, s$, and s_i^-, $i = 1, 2, \ldots, m$, as dual variables assigned to the lower limits for weights of outputs and inputs (slacks). A dual input-oriented CCR model can be written as follows (2):

$$
\begin{aligned}
&\text{Minimize } \theta_o - \varepsilon\left(\textstyle\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{s} s_k^+\right)\\
&\text{s. t. } \textstyle\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_o x_{io},\\
&\quad \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{ko},\\
&\quad \lambda_j \geq 0, s_i^- \geq 0, s_k^+ \geq 0.
\end{aligned} \tag{2}
$$

Complementary variables are introduced in this model s_i^- and s_k^+ (3):

$$
\begin{aligned}
s_i^- &= \theta_o x_{io} - \textstyle\sum_{j=1}^{n} x_{ij}\lambda_j,\\
s_k^+ &= \textstyle\sum_{j=1}^{n} y_{kj}\lambda_j - y_{ko}
\end{aligned} \tag{3}
$$

Slacks indicate how far the unit (DMU_o) is from the efficiency frontier. The variable s_i^- expresses excessive input consumption and s_k^+ expresses the lack of outputs. The unit that is effective in the CCR model has a value of the objective function $\theta_o = 1$, and values s_i^- and s_k^+ (slacks) equal to zero in every optimal solution.

The DEA model also provides information on how the DMU has to change its behavior to become efficient. The DMU, which is a projection of an inefficient unit on the efficiency frontier, is called a peer-unit. The inputs and outputs of this unit are target values of inputs x'_{io} and target values of outputs y'_{ko}, which can be calculated in two ways using optimal values of variables θ_o^*, λ_j^*, s_i^{*-}, s_k^{*+} and the following Formulas (4) and (5):

$$x'_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j^*,\ i = 1,\ 2,\ \ldots,\ m, y'_{ko} = \sum_{j=1}^{n} y_{kj}\lambda_j^*,\ k = 1,\ 2,\ \ldots,\ s. \tag{4}$$

$$x'_{io} = \theta_o^* x_{io} - s_i^{*-},\ i = 1,\ 2,\ \ldots,\ m, y'_{ko} = y_{ko} + s_k^{*+},\ k = 1,\ 2,\ \ldots,\ s. \tag{5}$$

The aim of the final step (RQ3) was to measure the strength and direction of association between the results achieved in the case of the BVM and the input-oriented DEA model.

We used Spearman's rank correlation coefficient to determine the correlation between the rankings of companies. We calculated the Spearman's correlation coefficient according to relationship (6) (Štiglic 2009), as there were several identical orders in the database of companies:

$$p = \frac{\frac{1}{6}\left(n^3-n\right)-\sum_1^n d_i^2-T_x-T_y}{\sqrt{\left[\frac{1}{6}\left(n^3-n-2T_x\right)\right]*\left[\frac{1}{6}\left(n^3-n-2T_y\right)\right]}},$$
$$T_x = \sum_1^{S_x}\left(t_{i(x)}^3 - t_{i(x)}\right),$$
$$T_y = \sum_1^{S_y}\left(t_{i(y)}^3 - t_{i(y)}\right). \tag{6}$$

where T_x, T_y, for each repeated value in the group X, Y is the number of repetitions deducted from the third power of the repetition in question, and the results are summed as S_x, S_y, number of repetitions in the group X (Y), $t_{i(x)}$, $t_{i(y)}$ represents the number of repetition of the value X_j, Y_j in the group X (Y).

For larger sets ($n > 30$), the probability distribution can be approximated by a t-distribution with $(n-2)$ degrees of freedom. Then, the test statistic for the Spearman coefficient has the form (7):

$$t = p\sqrt{\frac{(n-2)}{(1-p^2)}} \tag{7}$$

We use this random variable even if the groups X and Y feature several repeated values. The null hypothesis H_0, which states that there is no statistically significant relationship at significance level $\alpha = 0.05$, is rejected (we accept alternative hypothesis H_1), assuming that $|t| \geq t_{1-\frac{\alpha}{2}}$ (for one-sided test); $|t| \geq t_{1-\alpha}$ (for two-sided test), $t_{1-\frac{\alpha}{2}}$; $(t_{1-\alpha})$ is the critical value of Student's t-distribution, with $(n-2)$ degrees of freedom. We chose the correlation coefficient based on the results of the normality tests and a two-dimensional point diagram with an ellipse of 95% constant probability density. This verification, as well as the calculations of test statistics and Spearman's coefficient, were performed using software Statistica.

4. Results and Discussion

The starting point for assessing the performance of the analyzed sample of businesses was the calculation of nine financial indicators listed in Data and Methodology.

As the next step, we performed an initial analysis of businesses in space with the use of an MDS map (Figure 2). We were looking for an optimal MDS solution for analyzed sample of businesses with the use of software Statistica. We applied three dimensions, since MDS with four and five dimensions provided the same results. The number of dimensions was chosen based on Kruskal's *Stress*. Value of *Stress* for the three-dimensional model achieved 0.003. Therefore, we can state that the fit of objects in the constructed MDS map is perfect. On the left side of Figure 2, we can see a cluster of businesses. The distances between these businesses and the values of their indicators are highly correlated. We can say that the distances represent the values of indicators well in a linear sense. The cluster marked in Figure 2 with a red circle is created by these businesses: 19, 20, 21, 22, 23, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 38, 40, 41, 44, 45, 46, 48. They occur in the area of the negative values of dimension one and dimension two, but dimension three ranges from negative to positive values. These businesses have low liquidity, high indebtedness and are undercapitalized. According to the DEA, they achieved an efficiency score of 0.4–0.6. Business 48 is located near businesses 46, 40, 28, 35, 31 and 32; these businesses have a high creditors payment period, low interest coverage, insufficient liquidity and are undercapitalized. Business 43, which has a negative value of dimension three, is highly undercapitalized. Business 30 stands out from the cluster. It has a positive value in dimension one and negative values in dimensions two and three; the only indicator which achieved negative values is ROS. Business 13 has positive values in all dimensions, and all indicators achieve very good results. Business five also has positive values in all dimensions. It achieves a high average collection period and creditors payment period, but they are also balanced. Business two has negative values in dimensions one and two; it achieves a high creditors payment period and is undercapitalized.

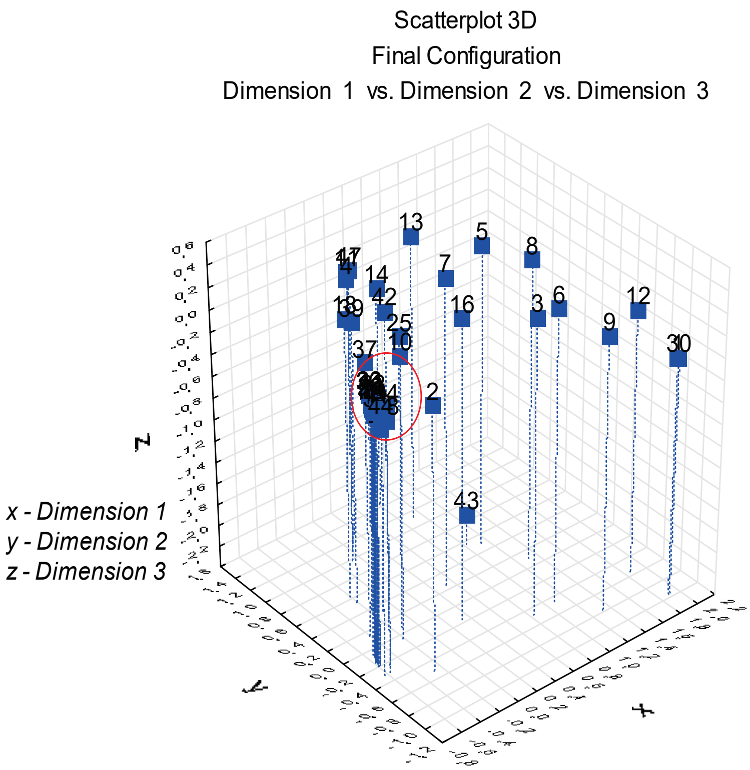


Figure 2. MDS map.

Depending on the tendency of the indicator (increasing or decreasing), we applied the BVM, which represents the final stage of benchmarking. We first found local benchmarks (LB 1–9) for each criterion and then the global benchmark (GB), which is DMU1. DMU1 is also a local benchmark, as it reaches the best values in ROA (4) and interest coverage (8). Overall, DMU2 finished in second place, and is a local benchmark in profitability indicators ROE (5) and ROS (6). DMU44 achieved the best values in terms of current liquidity (1) and the creditors payment period (3), and ended up in fourth place overall. The score (B) of all 48 companies is shown in Table 2.

Table 2. Score and ranking of DMUs in the research sample, according to the BVM.

DMU	Score (B)	Ranking	DMU	Score (B)	Ranking	DMU	Score (B)	Ranking
DMU1	42.27	1	DMU17	7.56	23	DMU33	2.95	39
DMU2	36.12	2	DMU18	3.07	36	DMU34	4.04	30
DMU3	15.20	5	DMU19	4.72	29	DMU35	9.89	16
DMU4	−45.14	48	DMU20	9.07	18	DMU36	2.92	41
DMU5	27.30	3	DMU21	2.96	38	DMU37	2.65	45
DMU6	13.62	7	DMU22	3.37	33	DMU38	8.12	21
DMU7	14.28	6	DMU23	2.80	42	DMU39	2.33	47
DMU8	11.45	11	DMU24	7.56	22	DMU40	5.34	28
DMU9	11.00	14	DMU25	3.14	35	DMU41	2.66	44
DMU10	9.78	17	DMU26	3.95	31	DMU42	5.37	27
DMU11	5.93	26	DMU27	3.06	37	DMU43	10.65	15
DMU12	11.57	10	DMU28	8.60	20	DMU44	19.99	4
DMU13	11.19	13	DMU29	3.34	34	DMU45	6.36	25
DMU14	12.67	8	DMU30	3.78	32	DMU46	2.53	46
DMU15	6.43	24	DMU31	11.39	12	DMU47	2.69	43
DMU16	8.75	19	DMU32	2.95	40	DMU48	12.60	9

LB4 = DMU1, LB5 = DMU2, LB6 = DMU2, LB7 = DMU35, LB9 = DMU6, LB1 = DMU44, LB2 = DMU5, LB3 = DMU44, LB8 = DMU1, GB = DMU1.

The results of input-oriented DEA (RQ2) shows that 12 companies achieved an efficiency score of "1" (DMU1, DMU2, DMU3, DMU6, DMU13, DMU19, DMU27, DMU30, DMU34, DMU43, DMU44 and DMU47). We can say that these companies use inputs efficiently, since they achieve zero slacks. The least efficient business is DMU36, which achieved a score of "0.23648". This company achieves very low liquidity, at the level of 0.07, a high creditors payment period of 2941 days, an equity ratio of 12% and is highly undercapitalized. The overall score and ranking of the companies' performances are shown in Table 3. The DEA model was processed in DEAFrontier (Zhu 2019).

Table 3. Score and ranking of DMUs in the research sample, according to the DEA.

DMU	Score (B)	Ranking	DMU	Score (B)	Ranking	DMU	Score (B)	Ranking
DMU1	1.00000	1	DMU17	0.27254	47	DMU33	0.62739	23
DMU2	1.00000	1	DMU18	0.32642	44	DMU34	1.00000	1
DMU3	1.00000	1	DMU19	1.00000	1	DMU35	0.36533	43
DMU4	0.29331	46	DMU20	0.39885	38	DMU36	0.23648	48
DMU5	0.71299	21	DMU21	0.42563	36	DMU37	0.62626	24
DMU6	1.00000	1	DMU22	0.56887	26	DMU38	0.39143	39
DMU7	0.69771	22	DMU23	0.79442	20	DMU39	0.56164	27
DMU8	0.79822	19	DMU24	0.45843	32	DMU40	0.43853	33
DMU9	0.88252	15	DMU25	0.31385	45	DMU41	0.81895	18
DMU10	0.50705	28	DMU26	0.39032	40	DMU42	0.91712	11
DMU11	0.84913	16	DMU27	1.00000	1	DMU43	1.00000	1
DMU12	0.90023	14	DMU28	0.41295	37	DMU44	1.00000	1
DMU13	1.00000	19	DMU29	0.45948	31	DMU45	0.43431	35
DMU14	0.83223	17	DMU30	1.00000	1	DMU46	0.48972	29
DMU15	0.47912	30	DMU31	0.36818	42	DMU47	1.00000	1
DMU16	0.61270	25	DMU32	0.37122	41	DMU48	0.43446	34

The DEA method allowed us to calculate target values-benchmarks for analyzed businesses, which were not efficient according to the DEA model. Since we used an input-oriented DEA model, we were able to calculate target values for inputs. Figure 3 shows current and target values of indicators ACP and CPP for inefficient businesses. The red line in the figure represents 90 days. Values of ACP and CPP above this line can be evaluated negatively. A majority of the businesses achieved current values of CPP above 90 days, and current values of ACP below 90 days. When calculating goal values, we can see a significant reduction in ACP and CPP. A majority of target values of these indicators are below 90 days.

We also calculated the target values of indicator CR for inefficient businesses with the use of the DEA. Figure 4 shows the comparison of current and target values of this indicator. The red line represents the value above which this indicator can be evaluated negatively. Despite the fact that all current values of CR are below the red line, businesses need to reduce them to target values in order to be efficient.

The third task of our research was to determine the strength and association between the abovementioned rankings (RQ3), where the first variable was the order of individual enterprises determined by the BVM and the second variable was the order of companies determined by the DEA.

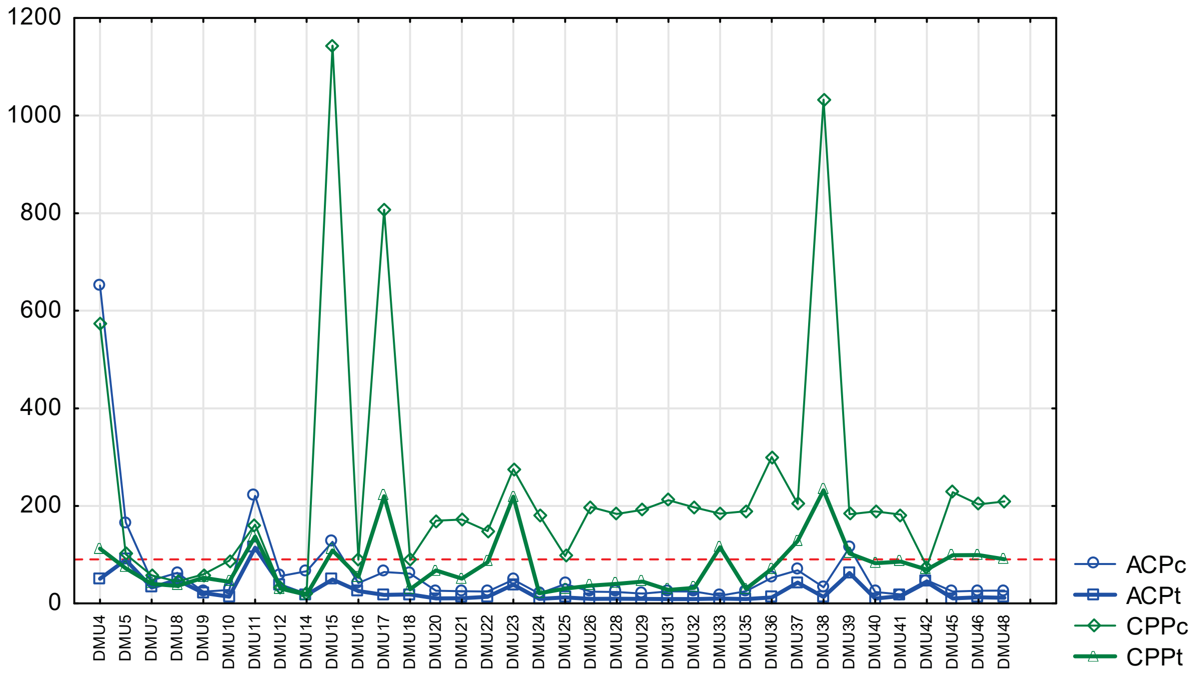


Figure 3. Comparison of current and target values of ACP and CPP (days). Legend: ACP_c—current Average collection period, ACP_t—target Average collection period, CPP_c—current Creditors payment period, CPP_t—target Creditors payment period.

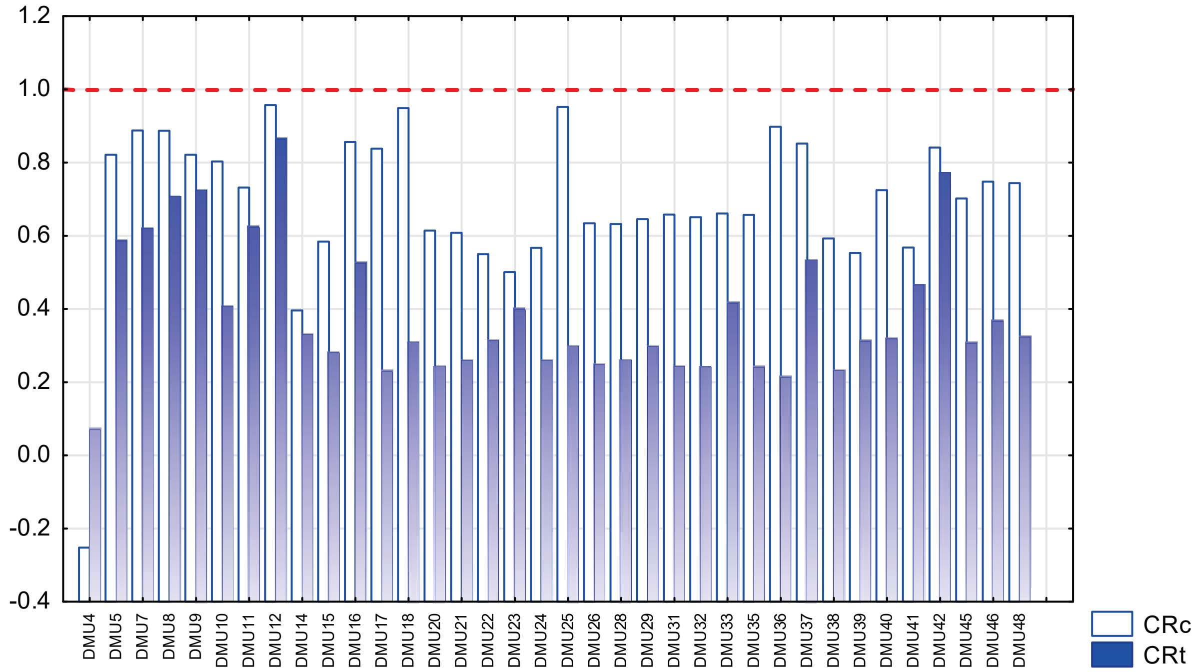


Figure 4. Comparison of current and target values of CR. Legend: CR_c—current Cost ratio, CR_t—target Cost ratio.

The Spearman's coefficient is 0.314965, i.e., 31.50%, which is a medium positive dependence. Table 4 shows the value of test statistic t ($n - 2$). The quantile of Student's distribution found in the statistical tables for reliability of 1-α and degrees of freedom ($n - 2$) is $t_{0.95}$ (45) = 1.671 (statistical tables). After determining the right-hand critical domain W $\leq$ 1.671, ∞), we found out that the test variable is located in the critical domain; therefore, we reject the null hypothesis and accept the alternative hypothesis of H_1. We conclude that there is statistically significant relationship between rankings achieved by the BVM and the DEA.

Table 4. Spearman's Rank Order Correlation.

Pair of Variables	Valid N	Spearman R	t ($n-2$)	p-Value
BVM ranking and DEA ranking	48	0.314965	2.250750	0.029224

Spearman's rank order correlations; marked correlations are significant at $p < 0.05000$.

The financial ratios used in this paper have an impact on the overall performance of the company. However, when applying benchmarking based on the BVM, weights are determined subjectively, according to a financial managers' preferences (pairwise comparison). With DEA, the measurement is more objective and accurate; however, it does not have to take into account a managers' preferences when making financial decisions. In order to determine the performance of a company in business environment, it is appropriate to combine these forms effectively.

5. Conclusions

In order to remain in a stable financial situation, a company must not only monitor the development of key financial indicators over time or in the future, but also analyze the status of the indicators in the industry with regard to its direct competition, and to be inspired by competition when solving financial problems (look for "best practices").

This research focused on measuring the financial performance of 48 Slovak companies in the field of heat supply. We calculated nine key financial indicators of analyzed companies and identified three research questions: determining the financial performance of the company through BVM, determining the financial performance of the company through DEA and analyzing the strength and direction of the association between the results achieved.

Benchmarking based on the BVM is one of the more subjective methods of measuring performance, as it uses a paired comparison method to determine the weights of indicators. This has to be carried out through a questionnaire survey. The advantage of using benchmarking is the motivation of financial managers to prioritize financial decision making. Financial managers evaluated the indicators as expected. The most important indicators are profitability indicators ROA, ROE and ROS, which were confirmed by several other authors (Wood and McConney 2018; Bărbută-Misu et al. 2019; Bogetoft 2012). These are important financial indicators which determine internal risk factors, such as business risk. The least preferred indicator is interest coverage. DMU1 became the benchmark, DMU2 came second, DMU5 was third, DMU44 was fourth and DMU3 was fifth. Compared to the BVM, the DEA is a benchmarking tool that offers more objective results, while inputs to the DEA model can be indicators expressing the preferences of managers. In our case, the input-oriented DEA CCR model was applied. Within the sample, there were 12 businesses that used their inputs efficiently (including DMU1, DMU2, DMU3 and DMU44).

Based on the Spearman's rank correlation coefficient, there is a 31.50% dependence among the companies' rankings. It is a positive dependence, and it is at a given level of significance. The value of the correlation coefficient may be lower due to the different nature of the methods and the fact that the BVM is a more subjective method than the DEA.

The originality of the paper is in its the use of a combination of DEA and BVM methods to improve performance. Both these methods are multidimensional ones. Both are also benchmarking methods, while each comes from a different set of benchmarking techniques which are used to improve performance. The contribution to the literature is that DEA can be considered a benchmarking tool and can be mentioned in the literature as a benchmarking technique for increasing business performance.

Some limitations of the research may be the smaller sample of companies or the subjective and benevolent attitude to managers when completing questionnaires. In our future research, we will improve these shortcomings.

For business practice, we recommend combining these methods, depending on whether the goal is to clarify priorities in terms of preference for financial ratios or to accurately determine financial performance. At the same time, DEA can be an important learning tool for managers to understand the importance of their decisions and their impact on business performance. It can motivate and stimulate them to focus on those indicators that are significant in terms of the financial situation of an enterprise. It can also be a tool to implement the need to identify key business performance indicators and performance risk factors.

Based on the obtained results, we can also generalize some recommendations for comparison, which could be useful for any company:

- To make a comparison with competitors as a multi-criteria comparison and not only on the basis of one criterion (for each indicator, the benchmark was also achieved by another company);
- To convert values of indicators into common unit;
- To involve managers in the selection of appropriate indicators and take into account their view on the significance of the applied indicators;
- To apply at least two evaluation methods when evaluating enterprises and compare their results;
- To monitor and optimize the profitability of the company, which has been confirmed as an important indicator of business performance.

Author Contributions: All authors contributed to all aspects of this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Financial statements in the form of balance sheets and profit and loss statements were obtained from the agency CRIF - Slovak Credit Bureau, s.r.o., which deals with the collection and processing of financial statements of Slovak companies according to individual SK NACE. Information were provided by a third party, which is focused on data collection and cooperation with academic institutions and supports them in obtaining the necessary data for their research activities. These financial statements have been prepared by the company by mutual agreement and according to the requirements of the authors.

Acknowledgments: This paper was prepared within the grant scheme VEGA No. 1/0741/20 (The application of variant methods in detecting symptoms of possible bankruptcy of Slovak businesses in order to ensure their sustainable development) and VEGA No. 1/0090/19 (Research into the financial situation and financial literacy of low-income population groups in context of socio-economic problems of Eastern Slovakia).

Conflicts of Interest: The authors declare no conflict of interest.

References

Agrell, Per Joakim, and Peter Bogetoft. 2007. *Development of Benchmarking Models for German Electricity and Gas Distribution*. Sundsvall: Sumicsid. [CrossRef]

An, Qingxian, Xiangyang Tao, and Beibei Xiong. 2021. Benchmarking with data envelopment analysis: An agency perspective. *Omega* 101: 102235. [CrossRef]

Anand, G., and Rambabu Kodali. 2008. Benchmarking the benchmarking models. *Benchmarking: An International Journal* 15: 257–91. [CrossRef]

Antimonopoly Office of the Slovak Republic. 2013. Functioning and Problems in the Heat Management Sector in the Slovak Republic with a Focus on Local Central Heat Supply Systems from the Perspective of the Antimonopoly Office of the Slovak Republic. Available online: http://www.antimon.gov.sk/data/att/365.pdf (accessed on 21 June 2020).

Aparicio, Juan, Jose Juan Lopez-Espin, Raul Martinez-Moreno, and Jesús Tadeo Pastor. 2014. Benchmarking in Data Envelopment Analysis: An Approach Based on Genetic Algorithms and Parallel Programming. *Advances in Operations Research* 431749. [CrossRef]

Asrofah, Tutie, Suhaiza Zailani, and Yudi Fernando. 2010. Best practices for the effectiveness of benchmarking in the Indonesian manufacturing companies. *Benchmarking: An International Journal* 17: 115–43. [CrossRef]

Banker, Rajiv D. 1990. *Stochastic Data Envelopment Analysis, Working Paper*. Minnesota: Carlton School of Management.

Bărbută-Misu, Nicoleta, Mara Madaleno, and Vasile Ilie. 2019. Analysis of risk factors affecting firms' financial performance—Support for managerial decision-making. *Sustainability* 11: 4838. [CrossRef]

Berg, Sanford. 2010. *Water Utility Benchmarking: Measurement, Methodology, and Performance Incentives*. London: IWA Publishing.

Bogetoft, Peter. 2012. *Performance Benchmarking: Measuring and Managing Performance*. New York: Springer. [CrossRef]

Borg, Ingwer, and Patrick J. F. Groenen. 1997. *Modern Multidimensional Scaling. Theory and Application*. New York: Springer.

Camp, Robert C. 1995. *Business Process Benchmarking*. Milwaukee: ASQC Quality Press.

Coelli, Tim, D. S. Prasada Rao, and George G. Batteese. 1998. *An Introduction to Efficiency and Productivity Analysis*. London: Kluwer Academic Publishers.

Cooper, William, Lawrence M. Seiford, and Joe Zhu. 2004. *Handbook on Data Envelopment Analysis*. Boston: Kluwer Academic.

Cooper, William W., Lawrence M. Seiford, and Kaoru Tone. 2007. *Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA—Solver Software*. New York: Springer.

Cooper, William, Lawrence M. Seiford, and Joe Zhu. 2011. *Data Envelopment Analysis: History, Models and Interpretations*. New York: Springer.

CRIF. 2016. *Financial Statements of Businesses*. Bratislava: Slovak Credit Bureau.

Demjanová, Lucia. 2006. Benchmarking as a Way to Change and Innovation in Slovakia. *Management in Theory and Practice* 2: 44–50. Available online: http://casopisy.euke.sk/mtp/clanky/2--2006/demjanova.pdf (accessed on 3 September 2020). (In Slovak).

Deville, Aude. 2009. Branch Banking Network Assessment Using DEA: A Benchmarking Analysis—A Note. *Management Accounting Research* 20: 252–61. [CrossRef]

Díaz, Juan Antonio Rodríguez, Emilio Camacho Poyato, and Rafael López Luque. 2004. Applying benchmarking and data envelopment analysis (DEA) techniques to irrigation districts in Spain. *Irrigation and Drainage* 53: 135–43. [CrossRef]

Erbetta, Fabrizio, and Luca Rappuoli. 2003. *Estimating Optimal Scale and Technical Efficiency in the Italian Gas Distribution Industry, HERMES Working Paper, No. 6*. Vercelli: University of Eastern Piedmont.

Farantos, Georgios I. 2015. The Data Envelopment Analysis Method and the influence of a phenomenon in organizational Efficiency: A literature review and the Data Envelopment Contrast Analysis new application. *Journal of Data Envelopment Analysis and Decision Science* 15: 101–17. [CrossRef]

Färe, Rolf, Shawna Grosskopf, James L. Kirkley, and Dale Squires. 2001. Data Envelopment Analysis (DEA): A Framework for Assessing Capacity in Fisheries When Data are Limited. In *Microbehavior and Macroresults, Paper presented at the Tenth Biennial Conference of the International Institute of Fisheries Economics and Trade, Corvallis, Oregon, USA, July 10–14*. Edited by Ann L. Shriver and Richard S. Johnston. Corvallis: International Institute of Fisheries Economics and Trade.

Farrell, Michael James. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A.* 120: 253–90. [CrossRef]

Farsi, Mehdi, Massimo Filippini, and Michael Kuenzle. 2007. Cost efficiency in the Swiss gas distribution sector. *Energy Economics* 29: 64–78. [CrossRef]

Fong, Sik Wah, Eddie W. L. Cheng, and Danny C. K. Ho. 1998. Benchmarking: A general reading for management practitioners. *Management Decision* 36: 407–18. [CrossRef]

Frost, William. 2005. *ABCs of Activity Based Management—Crushing Competition Throught Performance Improvement*. Bloomington: iUniverse.

Goncharuk, Anatoliy G. 2008. Performance benchmarking in gas distribution industry. *Benchmarking: An International Journal* 15: 548–59. [CrossRef]

Hatami-Marbini, Adel. 2019. Benchmarking with network DEA in a fuzzy environment. *Rairo—Operations Research* 53: 687–703. [CrossRef]

Ho, Daniel C. W., Edwin H. W. Chan, Nicole Y. Wong, and Man-Wai Chan. 2000. Significant metrics for facilities management benchmarking in the Asia Pacific region. *Facilities* 18: 545–56. [CrossRef]

Ittner, Christopher D., David F. Larcker, and Taylor Randall. 2003. Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations & Society* 28: 715–41. [CrossRef]

Janiš, Stanislav. 2018. Slovak heating companies are preparing for the fourth generation of heating. In *Slovak Energy Annual 2018*. Bratislava: RE-PUBLIC, pp. 76–79.

Johnson, Gerry, and Scholes Kevan. 2000. *Ways to a Successful Business*. Praha: Computer Press.

Kaczmarska, Bozena. 2010. The Data Envelopment Analysis method in benchmarking of technological incubators. *Operations Research and Decisions* 20: 79–95.

Kiseľáková, Dana, and Beáta Šofranková. 2014. *Modern Financial Management of Business Entities*. Prešov: Bookman.

Kiseľáková, Dana, Beáta Šofranková, Veronika Čabinová, and Janka Šoltésová. 2018. Analysis of enterprise performance and competitiveness to streamline managerial decisions. *Polish Journal of Management Studies* 17: 101–11. [CrossRef]

Kislingerová, Eva, Jindřich Soukup, Petr Boukal, Patrik Sieber, Jan Heřman, Ivan Nový, Josef Krause, Hana Scholleová, Jiří Fotr, Miroslav Špaček, and et al. 2011. *New Economy. New Opportunities?* Praha: C. H. Beck.

Klieštik, Tomáš. 2009. Quantifying the Efficiency of Transport Companies' Activities Using Data Envelopment Analysis. *E + M Ekonomie a Management* 9: 133–45. Available online: http://www.ekonomie-management.cz/download/1331826707_1add/13_kliestik.pdf (accessed on 25 June 2020). (In Slovak).

Kočišová, Kristína. 2012. Application of DEA models in the analysis of technical efficiency of bank branches. *Economic Journal* 60: 169–86. Available online: https://www.sav.sk/journals/uploads/0622125302%2012%20Ko%C4%8Di%C5%A1ov%C3%A1%20RS.pdf (accessed on 5 May 2020). (In Slovak).

Kovaľová, Marcela, and Zuzana Nogová. 2016. Creative approach to the innovations based on the product benchmarking results. In *Innovation Management, Entrepreneurship and Corporate Sustainability, Paper presented at the 4th International Conference, Prague, Czechia, May 26–27*. Edited by Martin Lukeš, Jan Misař, Jan Mareš, Ondřej Dvouletý, Miroslav Špaček and Ivana Svobodová. Prague: Oeconomica, pp. 338–48.

Krishnamoorthy, Bala, and Christine DLima. 2014. Benchmarking as a measure of competitiveness. *International Journal of Process Management and Benchmarking* 4: 342–59. [CrossRef]

Kruskal, Joseph B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27. [CrossRef]

Lukáčová, Marta, Jaroslav Korečko, Sylvia Jenčová, and Mária Jusková. 2020. Analysis of selected indicators of tax competition and tax harmonization in the EU. *Entrepreneurship and Sustainability Issues* 8: 123–37. [CrossRef]

Magd, Hesham, and Adrienne Curry. 2003. Benchmarking: Achieving best value in public-sector organisations. *Benchmarking: An International Journal* 10: 261–86. [CrossRef]

Majorová, Matina. 2007. *Reference Materials for Course Optimal Programming II—Part DEA—Data Envelopment Analysis*. Nitra: Faculty of Economics and Management, Slovak University of Agriculture. (In Slovak)

Maleyeff, John. 2003. Benchmarking performance indices: pitfalls and solutions. *Benchmarking: An International Journal* 10: 9–28. [CrossRef]

Neely, Andy. 2002. *Business Performance Measurement. Theory and Practice*. Cambridge: Cambridge University Press—The Edinburgh Building.

Nenadál, Jaroslav, David Vykydal, and Petra Halfarová. 2011. *Benchmarking—Myths and Reality*. Praha: Managemet Press, ISBN 978-80-7261-224-6. (In Czech)

Neumaierová, Inka, and Ivan Neumaier. 2002. *Performance and Market Value of the Company*. Prague: Grada Publishing. (In Czech)

Otley, David. 1999. Performance management: A framework for management control systems research. *Management Accounting Research* 10: 363–82. [CrossRef]

Petřík, Tomáš. 2009. *Economic and Financial Management of the Company*. Prague: Grada Publishing. (In Czech)

Pogranová, Zuzana, Radovan Savov, Drahoslav Lančarič, and Lucia Lenčéšová. 2011. *Benchmarking—A Tool for Improving Business Management*. Nitra: Slovak University of Agriculture in Nitra, ISBN 978-80-552-0672-1. (In Slovak)

Rhee, Cheul, Junghoon Moon, G. Lawrence Sanders, and Youngchan Choe. 2009. Benchmarking Technique in E-Business Using the Galileo Method. Paper presented at the 42nd Hawaii International Conference on System Sciences, Waikoloa, HI, USA, January 5–8; pp. 1–10. [CrossRef]

Rossi, Martin A. 2001. Technical change and efficiency measures: The post-privatisation in the gas distribution sector in Argentina. *Energy Economics* 23: 295–304. [CrossRef]

Ruiz, José L., and Inmaculada Sirvent. 2019. Performance evaluation through DEA benchmarking adjusted to goals. *Omega—International Journal of Management Science* 87: 150–57. [CrossRef]

Shewell, Patricia, and Stephen Migiro. 2016. Data envelopment analysis in performance measurement: A critical analysis of the literature. *Problems and Perspectives in Management* 14: 705–13. [CrossRef]

Štefko, Róbert, Beata Slusarczyk, Sebastian Kot, and Cezary Kolmasiak. 2012. Transformation on steel products distribution in Poland and Slovakia. *Metalurgija* 51: 133–36.

Štefko, Róbert, Petra Vašaničová, Sylvia Jenčová, and Aneta Pachura. 2021. Management and Economic Sustainability of the Slovak Industrial Companies with Medium Energy Intensity. *Energies* 14: 267. [CrossRef]

Štiglic, Martin. 2009. Nonparametric Statistical Methods and Their Economic Applications. Available online: http://web.ics.upjs.sk/svoc2009/prace/3/Stiglic.pdf (accessed on 10 August 2019). (In Slovak).

Šulák, Milan, and Emil Vacík. 2005. *Measuring the Performance of Companies*. Pilsen: University of West Bohemia. (In Czech)

Tian, Zhilong, and Suttapong Ketsaraporn. 2013. Performance benchmarking for building best practice in business competitiveness and case study. *International Journal of Networking and Virtual Organisation* 12: 40–55. [CrossRef]

Tone, Kaoru. 2001. A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130: 498–509. [CrossRef]

Veber, Jaromír. 2000. *Management: The Basics, Prosperity, Globalization*. Praha: Management Press. (In Czech)

Vochozka, Marek, Jiří Jelínek, Ján Váchal, Jarmila Straková, and Vojtěch Stehel. 2017. *The Use of Neural Networks in Comprehensive Evaluation of the Company*. Praha: C.H.Beck, ISBN 978-80-7400-642-5. (In Czech)

Vrábliková, Mária, and Erika Loučanová. 2017. Benchmarking Application in Selected Wood Processing Enterprise. *Economic Spectrum* 12: 1–6. Available online: https://www.sjf.tuke.sk/umpadi/taipvpp/2017/index.files/prispevky/25_Loucanova_Nosalova_Parobek_Dopico_AObaly.pdf (accessed on 20 June 2020). (In Slovak).

Wood, Anthony, and Shanise McConney. 2018. The impact of risk factors on the financial performance of the commercial banking sector in Barbados. *Journal of Governance and Regulation* 7: 76–93. [CrossRef]

Yang, Jyh-Bin, Han-Hsiang Wang, Wei-Chih Wang, and Shin-Min Ma. 2016. Using data envelopment analysis to support best-value contractor selection. *Journal of Civil Engineering and Management* 22: 199–209. [CrossRef]

Zairi, Mohamed, and Paul Leonard. 1996. The link between benchmarking and performance measurement. In *Practical Benchmarking: The Complete Guide*. Dordrecht: Springer, pp. 81–94.

Zbranek, Peter. 2013. Data Envelopment Analysis as a tool for evaluation of employees' performance. *Acta Oeconomica et Informatica* 16: 1–10. [CrossRef]

Zema, Demetrio Antonio, Pasquale Filianoti, Daniela D'Agostino, Antonino Labate, Manuel Esteban Lucas-Borja, Angelo Nicotra, and Santo Marcello Zimbone. 2020. Analyzing the Performances of Water User Associations to Increase the Irrigation Sustainability: An Application of Multivariate Statistics to a Case Study in Italy. *Sustainability* 12: 6327. [CrossRef]

Zhu, Joe. 2019. *DEA Frontier Software*. Worcester: Foisie Business School, Worcester Polytechnic Institute.

Zimková, Emília. 2015. *Retail Production Process in Commercial Banking and Its Evaluation*. Bratislava: Wolters Kluwer. (In Slovak)

Journal of
Risk and Financial Management

MDPI

Article

Does Risk Disclosure Matter for Trade Credit?

Issal Haj-Salem [1] and Khaled Hussainey [2,*]

[1] Department of Accounting, Institute of Higher Commercial Studies of Carthage, University of Carthage, Carthage 1054, Tunisia; issal.hajsalem@gmail.com
[2] Accounting and Financial Management Group, Faculty of Business and Law, University of Portsmouth, Portsmouth PO1 3DE, UK
* Correspondence: Khaled.Hussainey@port.ac.uk; Tel.: +44-239-2844-715

Abstract: In this paper, we examine the impact of risk disclosure practices on trade credit. We hypothesize that risk information could reduce information opacity that arises between companies and their suppliers. We collected annual reports for Tunisian listed companies for the period 2008–2013. This gives us 146 firm-year observations. We find that risk disclosure has a positive impact on the level of trade credit. Our paper offers a new empirical evidence on the role of risk disclosure in reducing information asymmetry and increase companies' access to short-term external funds. Our study provides managerial implications for firms, suppliers, and regulatory authorities.

Keywords: risk disclosure; trade credit; content analysis; Tunisian listed companies

Citation: Haj-Salem, Issal, and Khaled Hussainey. 2021. Does Risk Disclosure Matter for Trade Credit? *Journal of Risk and Financial Management* 14: 133. https://doi.org/10.3390/jrfm14030133

Academic Editor: George Halkos

Received: 16 February 2021
Accepted: 17 March 2021
Published: 20 March 2021

1. Introduction

Access to finance is a challenge for companies worldwide to preserve and enhance their growth. This was also considered a major problem for emerging companies (Brown et al. 2011) and particularly for African ones (Fowowe 2017). Many African countries have witnessed many financial reforms (e.g., financial regulatory framework, monetary policy framework, supervisory framework, etc.) over the last three decades to overcome a long period of weak performance (Banya and Biekpe 2017). Debt financing and trade credit used to be the two main sources of financing. Nevertheless, the first refers to loans provided by financial institutions and the latter refers to loans provided by suppliers for a short period when firms purchase their products (Xu et al. 2020). Researchers believe that using trade credit may help firms to access bank loans since it is considered a good signal (Andrieu et al. 2017). Moreover, firms are likely to use trade credit rather than bank credit during a tight money period (Xu et al. 2020). In this context, in Tunisia, the Arab spring undoubtedly mitigates bank credit availability for Tunisian firms. Many reforms took place to improve the financial soundness indicators and financial stability. The restrictive monetary policy conducted by the Tunisian Central Bank (BCT) since 2017 with five increases in the interest rate had a considerable effect on the evolution of loans to the economy which severely decreased to reach +4.9% in August 2019 against +5.7% in July and + 10.9% in August 2018. Furthermore, to protect banks against excessive risk-taking, the BCT, through circular n° 2018-10 of 1 November 2018, instituted a new credit/deposit ratio. This was a restriction likely to rationalize the granting of loans since Tunisian banks must not exceed the limit of 120% for this latter. Consequently, a clear slowdown in the pace of credit evolution has been established. Accordingly, firms are likely to search for other financing sources. For that, the trade-credit was considered one of the most important short-term financing sources.

Although the important role of trade credit as one of the most important sources of the short-term financing, the association between corporate risk disclosure and trade credit remains largely unexplored. Recently, there are only a few researchers, such as (Xu et al. 2020) and Ceustermans et al. (2017) that have started to examine, respectively, the impact of the readability of annual reports on trade credit; and the impact of corporate voluntary disclosure on trade credit for small private firms.

The objective of our paper is to examine the impact of risk disclosure on trade credit. We aim to answer to the following research question "Does risk disclosure matter for trade credit? To achieve our research objective, we measured levels of risk disclosure for a sample of Tunisian listed firms during 2008–2013 and manually collected accounting information from corporate annual reports. We find that risk disclosure determines the trade credit provided by suppliers to their customers. Indeed, risk disclosure mitigates information asymmetry which leads to a positive association with the obtained trade credit. Moreover, risk disclosure sends signals to suppliers helping them to assess adequately the firm risk situation and to avoid any misinterpretation or to imagine the worst situation. This helps suppliers to extend more trade credit. Our results are robust after using, on the one hand, a log transformation and, on the other hand, another measure of trade credit.

Our paper offers interesting contributions to disclosure literature. We focus on the Tunisian context, as an example of an emerging market during a period of crisis. Moreover, we contribute to the economic consequences of risk disclosure literature by examining its impact on trade credit, which has received a considerable attention since the global financial crisis. To the best of our knowledge, no study, to date, has examined the impact of risk disclosure practices on trade credit.

The rest of the paper is organized as follows. In Section 2, we review the literature and develop our research hypothesis. In Section 3, we discuss the research method (sample, research model, and variables' measurement). In Section 4, we present and discuss our findings. Section 5 concludes the paper.

2. Literature Review and Hypothesis Development

2.1. Motivations of Trade Credit

Trade credit appears when firms have differed payment of sales or purchases. Overall, the payment period is about 30 days and it could be extended according to the arrangement between customers and suppliers. Trade credit is omnipresent in different industries and economies (Islam and Wheatley 2020). It was considered the second most important source of short-term financing behind bank financing as well for developed countries or developing countries (Fisman 2001). Moreover, prior researchers state that trade credit is most used in developing countries, where credit from financial institutions is limited (Giannetti et al. 2011). Many empirical studies find that trade credit could be a substitute for bank loans particularly in the early stage and then it could become its complement (Huyghebaert et al. 2007).

The motivations of trade credit could be classified into financial, operational/transaction, and commercial ones. Firstly, according to the financing theory market imperfections lead to limited credit accessibility for several firms (Schwartz 1974). Also, suppliers are likely to extend credit to their customers at lower costs than their financial institutions due to information asymmetry for the latter (Petersen and Rajan 1997). Secondly, according to commercial theories extending trade credit may build a long-term relationship and reinforce the loyalty of customers to the firms' products (Cheng and Pike 2003), or even expand their markets by attracting new customers (Van Horen 2005). Moreover, the trade credit could be used as a quality guarantee to allow customers verifying products' quality before the pay for these products. In addition, it could be used by suppliers as a signal for the quality of their goods (Long et al. 1993; Smith 1987).

2.2. Risk Disclosure Literature

Corporate risk disclosure (CRD) is one of the hottest research areas as it offers managerial and policy/regulatory implications. The first studies on risk disclosure have been conducted to understand the attitudes of stakeholders and the relevance of such practice for their decision-making (e.g., Deumes 2008; Solomon et al. 2000) and its usefulness to investors since it alleviates the information asymmetry (Miihkinen 2013). Furthermore, many studies were conducted to investigate the determinants of risk disclosure rather than its relevance. Those studies consider corporate governance mechanisms and firm characteristics as potential drivers for corporate risk disclosure. Through prior studies, we find that studies in developing countries (Al-Maghzom et al. 2016; Alzead and Hussainey 2017) are scarce compared to developed ones (Abraham and Cox 2007; Beretta and Bozzolan 2004; Elshandidy et al. 2013; Elshandidy and Shrives 2016; Elzahar and Hussainey 2012; Linsley et al. 2006; Miihkinen 2012, 2013). Prior empirical studies (Al-Maghzom et al. 2016; Alkurdi et al. 2019; Allini et al. 2015; Alzead and Hussainey 2017; Bufarwa et al. 2020; Elshandidy and Neri 2015; Haj-Salem et al. 2019; Nahar et al. 2020) have been mainly focused on the determinants of risk disclosure such as corporate governance attributes (board size, independent directors, CEO duality, ownership structure, audit quality, etc.) and firm-specific characteristics (firm size, growth, industry type, etc.). Moreover, the economic consequences of the risk disclosure practices start being investigated (Haj-Salem et al. 2020). In our paper, we are wondering whether the risk disclosure may enhance or hinders suppliers to extend trade credit as it is one of the most important sources of short-term financing. No study has investigated this issue, to the best of the authors' knowledge. Recently, few researchers start investigating the association between trade credit and disclosure in general and particularly annual report readability (Xu et al. 2020) or voluntary disclosure (Ceustermans et al. 2017). For instance, Ceustermans et al. (2017) found that voluntary disclosure is positively related to the level of trade credit for small and private companies. They argued, that providing more financial information mitigates traditional information asymmetry between firms and their creditors. Consequently, trade creditors are likely to react louder than banks to voluntary disclosure since they do not ask usually for private information from their customers to keep their trust relationship. In the same vein, the results of Xu et al. (2020) indicate that firms with more readable annual reports are likely to benefit from more trade credit.

Since empirical research linking risk disclosure with trade credit has not been yet undertaken despite the importance to understand this association, our study aims to fill this void in the literature and we expect that risk disclosure practices influence creditors' behaviors.

2.3. Hypothesis Development

The agency theory shows that, there is information asymmetry between principals and agent (Fama and Jensen 1983; Jensen and Meckling 1976). Accordingly, the agent (the firm) tries to undertake several ways to mitigate the information asymmetry with the principals (stakeholders). For that, prior studies considered risk disclosure as a mean to alleviate this information opacity with stakeholders, since it enables the latter to estimate the firm risk profile. Moreover, overall companies disclose risks and the way to manage them. This information will be relevant for different stakeholders (investors, creditors, etc.) before making financial decisions (Haj-Salem et al. 2020). Accordingly, risk disclosure could have a positive impact the level of trade credit. Moreover, based on signaling theory, the risk disclosure could be considered as a signal for suppliers and tend to mitigate the information asymmetry. This signal may enhance the company's reputation and its market valuation (Al-Maghzom et al. 2016). Indeed, a higher information asymmetry is likely to increase the risk of customers' payment default. Consequently, suppliers may hinder granting trade credit (Ceustermans et al. 2017). Furthermore, suppliers rely more on available information rather than asking for sensitive information from their customers since this may destroy their trust relationship (Arruñada 2010). Accordingly, to overcome any misinterpretation suppliers need more risk information to better assess the firm situation, otherwise, they will consider it as withholding the worst possible information (Spence 1973) which will refrain from lending decisions. Prior researchers indicate that companies using more trade credit are more likely to provide financial statements (Yohn and Allee 2009). Moreover, others predict a positive association between disclosure and a firm's dependence on external finance (Diamond and Verrecchia 1991). Recently, Ceustermans et al. (2017) find that voluntary disclosure mitigates information asymmetry and consequently there is a positive association between disclosure in financial statements and the ability to receive more trade credit. Furthermore, the empirical findings of Xu et al. (2020) indicate that suppliers are likely to grant more trade credit to customers with more readable annual reports. We therefore hypothesize that:

Risk disclosure positively affects the level of trade credit.

3. Research Method

3.1. Our Sample

Our sampling process started by collecting annual reports of non-financial listed companies from the BVMT (Tunisian Stock Exchange) for the period 2008–2013.

In line with prior research (Haj-Salem et al. 2020), we excluded financial companies because of their specific requirements and regulation for risk disclosure. Our initial sample involves 28 Tunisian listed firms which represent approximately 90 percent of all non-financial Tunisian listed companies. We manually collect our data from corporate annual reports and CMF (Financial Market Council). We focused on annual reports since there were considered as the primary mean of information for stakeholders (Marston and Shrives 1991; Miihkinen 2013).

We checked the existence of outliers to avoid bias in our findings. This is in line with prior literature. For that, using a Student's t-test helps us to remove all observations with a t greater than 2 in absolute value (Fox 2015). This gives us a final sample of 146 firm-years observations.

We choose Tunisia as a context for our study for many reasons. Tunisia is an emerging market, and recently there are several calls for more research on CRD in developing countries (Khlif and Hussainey 2016). Moreover, during constraint's periods, firms are more vulnerable to financing, and the dependence on bank financing and trade credit increases (Lin and Qiao 2020). Furthermore, firms usually suffer from bank credit-granting during the crisis and tend to use more trade credit (Chen et al. 2019). For that, suppliers were considered as viable financing sources during the global financial crisis (Bazzana et al. 2019) and play an important role in the survival of Small to Medium Enterprise (SMEs) during the crisis (Pattnaik et al. 2020). Accordingly, it is interesting to investigate the association between risk disclosure and trade credit in the Tunisian context during 2008–2013, since it was a period characterized by economic and political instability. It is important to highlight that the year 2014 was characterized by a political transition in Tunisia, which was the legislative and presidential elections that were considered an official beginning for a democratic transition. This new political system may undoubtedly impact the socio-economic situation of the country. For that, our sample period ends on 2013. Table 1 provides a description of our sample by industry.

Table 1. Sample description.

Industry	Firms Numbers
Telecommunications	2
Consumer services	5
Health	2
Consumer goods	8
Industrials	7
Basic Material	3
Oil and Gas	1
TOTAL	28
Original sample	168
(−) Missing reports	13
(−) Outliers	9
Final sample	146

3.2. Measuring Trade Credit

The dependent variable in our study is trade credit (TC). According to prior works (e.g., Deloof and La Rocca 2015) there are several measures that could be used. For instance, the accounts payable over total assets are used to take into account trade credit demand. Moreover, in order to reflect the supply of trade credit the accounts receivables over total assets is more suitable. In addition, we could measure the net trade credit by the differences of the two latter proxies. Furthermore, other researchers (e.g., Chen et al. 2019) used the sum of customer deposits, accounts payable and notes payable, divided by the cost of goods sold. For our study, and following prior research (e.g., Ceustermans et al. 2017; Chen et al. 2017; Islam and Wheatley 2020; Petersen and Rajan 1997; Xu et al. 2020), we measure the trade credit as the ratio of accounts payable to total assets as shown on the balance sheet.

3.3. Measuring the Risk Disclosure

We use the content analysis method to measure risk disclosure levels in Tunisian annual reports. We refer to the definition of risk disclosure as advanced by (Linsley et al. 2006) as "any opportunity or prospect, or of any hazard, danger, harm, threat or exposure, that has already impacted upon the company or may impact upon the company in the future or of the management of any such opportunity, prospect, hazard, harm, or threat or exposure". Following prior research (Al-Shammari 2014; Amran et al. 2008; Beretta and Bozzolan 2004) we use "sentence" as the unit of our textual analysis. We also use a disclosure index adapted to the Tunisian context, developed by Haj-Salem et al. (2020). This index is composed of 48 items and it was inspired by Linsley et al. (2006) and Moumen et al. (2015) ("Appendix A").

We use the stability, reproducibility, and accuracy tests to check the reliability of our risk disclosure measure Krippendorff (2004). For that, five annual reports were independently coded by the coder at different times. Moreover, another coder used the same index to code the same annual reports. After comparing results using the "ReCal" online statistical software the results present a satisfactory level of inter-coder reliability.

3.4. Our Control Variables

Our regression model includes a number of control variables. Consistent with prior literature, we control for firm age measured by calculating the natural logarithm of the number of years since the incorporation of the company (Xu et al. 2020). There is no consensus about the impact of firm age on trade credit. Some researchers find that there is a negative association between trade credit and firm age (Chen et al. 2017). Nevertheless, others believe that the use of trade credit is more important in older firms (Petersen and Rajan 1997). In our study, we expect a negative impact of firm age on trade credit since younger firms are likely to have difficulty, comparing to older ones, obtaining bank loans. Accordingly, they rely more on trade credit as a source of short-term finance (Deloof and La Rocca 2015).

We control for firm size measured by calculating the natural logarithm of total assets. Larger firms are considered more creditworthy. They are more likely to have access easily to trade credit thanks to their negotiation power with suppliers (Ceustermans et al. 2017).

We also control for profitability measured by market to book value. This is consistent with prior works on trade credit (Xu et al. 2020). We expect a negative coefficient since firms with higher profitability may easier have more financing from other sources because they have less default risk than less profitable firms. Accordingly, they will use less trade credit (Chen et al. 2017; García-Teruel and Martínez-Solano 2007; Petersen and Rajan 1997).

We control for solvency. We use equity divided by total assets as a measure for solvency (Ceustermans et al. 2017). We also control for liquidity measured by the calculation of the ratio of current assets to current liabilities. Overall, firms with liquidity shortages are likely to use more trade credit to overcome such situations (Boissay and Gropp 2007). Moreover, according to financing and pecking order theory, more solvent and liquid firms tend to rely less on trade credit due to their higher costs compared to bank loans (Ceustermans et al. 2017). Accordingly, we expect a negative Effect of both solvency and liquidity on trade credit.

We control for firm growth. This was measured by the difference between profit current year and profit of previous year; divided by profit of previous year (Lin and Qiao 2020). We believe that the more it is the more is the need to finance new investments (Lin and Qiao 2020; Van den Bogaerd and Aerts 2014). Accordingly, they rely on an external source of finance particularly trade credit. We expect to have a positive coefficient.

Following prior research, firms with an inventory ratio will undoubtedly have more purchases and hence, they will use more trade credit (Elliehausen and Wolken 1993). We therefore control for inventory ratio. The inventory ratio is measured by the ratio of inventory to total assets. We expect to have a positive association between trade credit and inventory ratio.

Finally, we included accounts receivable as a control variable. It is measured by the ratio of accounts receivable to total assets. We expect a positive coefficient with trade credit since companies are likely to compensate accounts receivable with accounts payables (Paul and Wilson 2007).

3.5. Research Model

$$\text{TC i,t} = \alpha 0 + \beta 1 \text{ CRDi,t} + \beta 2 \text{ AGEi,t} + \beta 3 \text{ SIZEi,t} + \beta 4 \text{ TQi,t} + \beta 5 \text{ SOLVEi,t} + \\ \beta 6 \text{ LIQUIi,t} + \beta 7\text{GROWTH i,t} + \beta 8\text{INVEN i,t} + \beta 9\text{AR i,t} + \epsilon$$

where:

α = the intercept.
$\beta 1, \ldots, \beta 9$ = Regression coefficients.
ε = Error term.

The following Table 2 presents variables' description:

Table 2. Variables' description.

Variable	Symbol	Description
Trade credit	TC	Accounts payable/total assets
Corporate Risk disclosure	CRD	Risk disclosure index as described in "Appendix A"
Firm age	AGE	Natural logarithm of the number of years since the incorporation of the company
Firm size	SIZE	Natural logarithm of total assets at the end of the year
Profitability	TQ	(Market value + total debts)/total assets
Inventory	INVEN	Inventory/total assets
Solvency	SOLVE	Equity/total assets
Liquidity	LIQUI	Current assets/current liabilities
Growth	GROWTH	(Profit current year (N)- profit of (N-1))/profit (N-1)
Accounts receivable	AR	Accounts receivable/total assets

4. Empirical Analysis and Robustness Check

4.1. Descriptive Analysis

Table 3 reports our descriptive analysis. On average, during 2008–2013, the accounts payable-to-total asset ratio (TC) is about 0.138. This is consistent with prior research (Ceustermans et al. 2017). The CRD is on average about 5.424. Accordingly, there is a low level of risk disclosure for Tunisian-listed companies. This is in line with previous studies on risk disclosure in emerging markets particularly in Tunisia (e.g., Haj-Salem et al. 2020). The average age of the companies in our sample is about 44 years (the natural logarithm of age is on average 3.790). Hence, overall, our sample is characterized by old firms. The mean value of logged total assets (SIZE) is 7.96. We can consider that the companies in our sample are medium-sized with approximately a total asset of 2864.07 TND. Furthermore, the market value of the companies (TQ) generates a mean of 1.660 and a mean of inventory (INVEN) ratio of 20.94%. Regarding, solvency (SOLVE) and liquidity (LIQUI) ratios they report a mean of 47.69% and 217%, respectively. Moreover, the growth ratio (GROW) is about −13.72%. Hence, Tunisian firms during 2008–2013 are facing a period of crisis that negatively influence firms' profitability and growth. Finally, accounts receivables make up an average of 17.810% of total assets.

Table 3. Descriptive analysis.

Variable	Obs	Mean	Std. Dev.	Min	Max
TC	146	0.1382692	0.1033652	0.0019	0.399
CRD	146	5.424658	3.062196	1	13
AGE	146	3.790805	0.4049906	2.5649	4.5643
SIZE	146	7.965512	0.4891449	7.0494	9.2242
TQ	146	1.660255	0.9608301	0.3574	5.8751
INVEN	140	0.2094964	0.1964394	0.0012	0.9219
SOLVE	146	0.4769548	0.379287	−1.2082	0.993
LIQUI	146	2.170023	1.751522	0.3241	9.3967
GROW	146	−0.1372239	6.637974	−42.1083	34.5245
AR	146	0.1781027	0.1522464	0	0.5895

Note: This table presents the descriptive statistics for our variables. All variables are defined in Table 2.

4.2. Correlation Analysis

Table 4 reports the Spearman correlation. The multicollinearity issue is checked using this matrix. No value is greater than 0.8 which confirms the non-existence of multicollinearity between variables. Furthermore, we addressed also multicollinearity by examining the variance inflation factors (VIF) which provides a VIF that does not exceed 10. This is presented in Table 5. Hence, the multicollinearity assumption was not breached.

Table 4. Correlations matrix.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) TC	1.00									
(2) CRD	0.05	1.00								
(3) AGE	0.09	−0.08	1.00							
(4) SIZE	0.01	0.17 *	0.21 *	1.00						
(5) TQ	0.06	−0.27 *	0.34 *	0.02	1.00					
(6) INVEN	0.03	−0.11	0.00	−0.22 *	−0.08	1.00				
(7) SOLVE	−0.20 *	−0.24 *	0.17 *	−0.12	−0.01	−0.09	1.00			
(8) LIQUI	−0.34 *	−0.15	0.01	−0.25 *	0.03	0.17	0.57 *	1.00		
(9) GROWTH	−0.04	0.06	0.10	0.20 *	0.01	0.00	0.06	0.06	1.00	
(10) AR	0.24 *	0.02	−0.09	−0.32 *	−0.01	−0.11	−0.05	−0.04	−0.17 *	1.00

Note: This table demonstrates the correlation matrix among variables utilized in the study's model. All variables are defined in Table 2.
* demonstrate the levels of significance at 0.10.

Table 5. Variance inflation factors (VIF).

Variable	VIF	1/VIF
SOLVE	1.98	0.504221
LIQUI	1.93	0.519378
SIZE	1.63	0.614319
TQ	1.33	0.750454
AGE	1.33	0.754157
INVEN	1.25	0.802816
CRD	1.22	0.821464
AR	1.12	0.892728
GROW	1.12	0.894476
Mean VIF	1.43	

All variables are defined in Table 2.

4.3. Regression Results

We conducted a panel data approach. Before starting, we checked for omitted variables. This was done through the Ramsey test. We found an F (3, 127) = 0.85 and Prob > F = 0.4665. Hence, no omitted variables are revealed for our model. First, we generated the specification panel test to judge the homogeneity or heterogeneity among individuals. The Chow test showed a prob > F = 0.000. Accordingly, the null hypothesis of homogeneity is rejected. Secondly, we check for the individual effect. For that, we generate respectively, a fixed-effect model, and a random-effect model. The comparison between the two models, by the Hausman test, we found that the random effect is the most suitable model since chi2(8) = 13.73 and Prob > chi2 = 0.0890. Hence, the interpretations are according to the findings of the random effect regression. Moreover, we check for homoscedasticity. The Breusch and Pagan Lagrangian multiplier test for random effects was generated. Consequently, the findings of this test report that chibar2(01) = 125.31 and Prob > chibar2 = 0.0000. This indicates, that there is a heteroscedasticity problem. This issue has been corrected to avoid unbiased results following the method of White (1980). Furthermore, the autocorrelation between residuals was checked following the test of the Wooldridge (2002). According to the F statistics displayed through this test, the null hypothesis of autocorrelation is rejected (F (1, 25) = 1.118; Prob > F = 0.3005).

Table 6 reports the findings of the random effect model. The results show a positive and significant association between CRD and trade credit. This suggests that firms with higher risk disclosure receive more trade financing from suppliers. The results are consistent with H1. The results could be explained according to signaling and agency theory. Indeed, the CRD decrease information asymmetry. Consequently, suppliers may avoid customers' risk payment default, better assess the firm wealth, and are likely to provide more trade credit (Ceustermans et al. 2017). Moreover, retaining information may send a bad signal for suppliers and they will consider it as withholding the worst possible information which may refrain from lending decisions (Spence 1973). Our findings are consistent with previous studies (Ceustermans et al. 2017; Xu et al. 2020).

Table 6. Results.

TC	Coef.	Std. Err.	z	P > z	[95% Conf.	Interval]
CRD	0.002461	0.0013731	1.79	0.073 *	−0.0002303	0.0051523
AGE	0.0295114	0.0146549	2.01	0.044 **	0.0007883	0.0582346
SIZE	0.0250873	0.0108888	2.30	0.021 **	0.0037457	0.0464289
TQ	−0.0078299	0.0056521	−1.39	0.166	−0.0189078	0.0032479
INVEN	0.0771459	0.0221102	3.49	0.000 ***	0.0338108	0.120481
SOLVE	0.0533021	0.0163949	3.25	0.001 ***	0.0211687	0.0854355
LIQUI	−0.0218254	0.003589	−6.08	0.000 ***	−0.0288596	−0.0147912
GROW	−0.0002238	0.0007146	−0.31	0.754	−0.0016244	0.0011768
AR	0.1756127	0.0317851	5.52	0.000 ***	0.113315	0.2379104
_cons	−0.203238	0.094263	−2.16	0.031	−0.3879901	−0.0184858

Note: This table illustrates the findings of the random-effect model after dealing with the heteroscedasticity problem. ***, **, and * demonstrate the levels of significance at: 0.01, 0.05, and 0.10, respectively.

Looking at the coefficients on the control variables, we find that respectively, age, size, inventory, solvency, and accounts receivable are positively and significantly associated with trade credit. Indeed, larger firms are more likely to have access easily to trade credit (Ceustermans et al. 2017). This is in line with our expectations. However, we expected that there is a negative association between solvency and trade credit since, according to financing and pecking order theory, more solvent and liquid firms tend to rely less on trade credit. Additionally, we find a negative and significant association between trade credit and liquidity, which suggests that firms with a lower liquidity ratio use more trade credit. No association was revealed between profitability, growth, and trade credit.

4.4. Robustness Check

It is crucial to check for the robustness of the empirical findings to ensure their validity. According to previous studies (Cooke 1998), the transformation allows solving several issues such as non-normal distribution, non-linearity between dependent and independent variables, etc. Hence a log transformation for the dependent variable is performed. Table 7 provides the findings of the random effect regression and those after, both, log transformation and using another measure for trade credit. The similarity of the results allows us to conclude about their validity.

Table 7. Robustness check.

Robustness Check			
Different Measure for the Dependent Variable	**Log Transformation**	**Results**	
0.014 ***	0.020 *	0.002 *	CRD
(4.46)	(1.82)	(1.79)	
0.091 ***	0.266 **	0.030 **	AGE
(2.99)	(2.39)	(2.01)	
−0.023	0.040	0.025 **	SIZE
(−0.74)	(0.38)	(2.30)	
−0.017	−0.047	−0.008	TQ
(−1.63)	(−1.31)	(−1.39)	
−0.012	0.803 ***	0.077 ***	INVEN
(−0.23)	(3.94)	(3.49)	
0.212 ***	0.436 ***	0.053 ***	SOLV
(9.86)	(3.78)	(3.25)	
−0.007	−0.250 ***	−0.022 ***	LIQUI
(−0.85)	(−7.46)	(−6.08)	
0.000	−0.001	−0.000	GROW
(0.10)	(−0.16)	(−0.31)	
0.229 ***	1.688 ***	0.176 ***	AR
(2.92)	(6.45)	(5.52)	
−0.028	−3.694 ***	−0.203 **	_cons
(−0.10)	(−4.00)	(−2.16)	
140	140	140	N

***, **, and * demonstrate the levels of significance at: 0.01, 0.05, and 0.10, respectively.

5. Conclusions

We investigate the impact of risk disclosure on trade credit for Tunisian listed companies during 2008–2013. We find that risk disclosure determines the trade credit provided by suppliers to their customers. Hence, risk disclosure mitigates information asymmetry and this leads to a positive association with the obtained trade credit. Moreover, risk disclosure sends signals to suppliers helping them to assess adequately the firm risk situation and to avoid any misinterpretation or to imagine the worst situation. This helps suppliers to extend more trade credit. Our results are robust after using on the one hand a log transformation and on the other hand another measure of trade credit.

We choose Tunisia, as an example of the emerging economies, to undertake our analysis and hence we are filling an important research gap in the literature. In addition, examining the impact of risk disclosure on trade credit is a novel feature of our paper.

Our findings have several implications for different stakeholders. First, our study showed that the risk information in annual reports is important for a firm since it helps for having more trade credit which is an important source of financing. Accordingly, managers and the board of directors have to pay attention to the communication of more risk information when preparing annual reports to beneficiate from more trade credit.

Second, this study may be relevant to regulators to emphasize the importance of risk disclosures since it will have a positive impact on economic growth. Finally, we may consider the risk disclosure relevant to suppliers who are likely to grant more trade credit to their customers.

Our study is subject to a number of limitations. The first limitation is the relatively small sample. Nevertheless, this is due to the unavailability of annual reports in addition to the manual data collection. However, our sample is representative since it is composed most of non-financial-listed companies. The second limitation is inherent to the subjectivity of manual content analysis. However, we followed previous literature to assess the reliability and validity of our scores. Further research could investigate this association for other research contexts. Further research could also investigate the impact of governance mechanisms on the disclosure–trade credit relationship.

Author Contributions: Conceptualization, I.H.-S. and K.H.; methodology, K.H.; software, I.H.-S.; validation, I.H.-S., K.H.; formal analysis, I.H.-S.; investigation, I.H.-S.; resources, I.H.-S.; data curation, I.H.-S.; writing—original draft preparation, I.H.-S.; writing—review and editing, K.H.; visualization, I.H.-S.; supervision, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions.

Acknowledgments: The authors would like to thank the editor and three anonymous referees for helpful comments on this research.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Risk disclosure level index adopted from Haj-Salem et al. (2020). *: Items added after reading annual reports.

Operations Risk
Customer satisfaction
Product development
Efficiency and performance
Sourcing
Stock obsolescence and shrinkage
Product and service failure
Environmental
Health and safety
Brand name erosion
Out of stock *
Quality of products *
Suppliers *
Employees *
Empowerment Risk
Leadership and management
Outsourcing
Performance incentives
Change readiness
Communications

*Control **
Information Processing And Technology Risk
Integrity
Access
Availability
Infrastructure
Integrity Risk
Management and employee fraud
Illegal acts
Reputation
*Litigation risk **
Strategic Risk
Environmental scan
Industry
Business portfolio
Competitors
Pricing
Valuation
Planning
Life cycle
Performance measurement
Regulatory
Sovereign and political
Financial risk
Interest rate
Exchange rate
Commodity
Liquidity
Credit
*Solvency ratio **
*Profitability ratio **
*Management ratio **
*Working capital fun **
*Downside risk **

References

Abraham, Santhosh, and Paul Cox. 2007. Analysing the determinants of narrative risk information in UK FTSE 100 annual reports. *The British Accounting Review* 39: 227–48. [CrossRef]

Alkurdi, Amneh, Khaled Hussainey, Yasean Tahat, and Mohammad Aladwan. 2019. The impact of corporate governance on risk disclosure: Jordanian evidence. *Academy of Accounting and Financial Studies Journal* 23: 1–16.

Allini, Alessandra, Francesca Manes Rossi, and Khaled Hussainey. 2015. The board's role in risk disclosure: An exploratory study of Italian listed state-owned enterprises. *Public Money & Management* 36: 113–20. [CrossRef]

Al-Maghzom, Abdullah, Khaled Hussainey, and Doaa A. Aly. 2016. Corporate governance and risk disclosure: Evidence from Saudi Arabia. *Corporate Ownership and Control* 13: 145–166. [CrossRef]

Al-Shammari, Bader. 2014. Kuwait corporate characteristics and level of risk disclosure: A content analysis approach. *Journal of Contemporary Issues in Business Research* 3: 128–53.

Alzead, Ramzi, and Khaled Hussainey. 2017. Risk disclosure practice in Saudi non-financial listed companies. *Corporate Ownership and Control* 14: 262–75. [CrossRef]

Amran, Azlan, Abdul Manaf Rosli Bin, and Bin Che Haat Mohd Hassan. 2008. Risk reporting. *Managerial Auditing Journal* 24: 39–57. [CrossRef]

Andrieu, Guillaume, Raffaele Staglianò, and Peter van der Zwan. 2017. Bank debt and trade credit for SMEs in Europe: Firm-, industry-, and country-level determinants. *Small Business Economics* 51: 245–64. [CrossRef]

Arruñada, Benito. 2010. Mandatory accounting disclosure by small private companies. *European Journal of Law and Economics* 32: 377–413. [CrossRef]

Banya, Roland Mwesigwa, and Nicholas Biekpe. 2017. Bank competition and economic growth. *Journal of Economic Studies* 44: 245–65. [CrossRef]

Bazzana, Flavio, Giacomo De Laurentis, Raoul Pisani, and Renata Trinca Colonel. 2019. Can domestic trade credit insurance contracts be effective collateral for banks? A quantitative study of the Italian market. *The European Journal of Finance* 26: 1239–52. [CrossRef]

Beretta, Sergio, and Saverio Bozzolan. 2004. A framework for the analysis of firm risk communication. *The International Journal of Accounting* 39: 265–88. [CrossRef]

Boissay, Frédéric, and Reint Gropp. 2007. *Trade Credit Defaults and Liquidity Provision by Firms*. ECB Working Paper No. 753. Frankfurt: European Central Bank.

Brown, Philip, Wendy Beekes, and Peter Verhoeven. 2011. Corporate governance, accounting and finance: A review. *Accounting & Finance* 51: 96–172. [CrossRef]

Bufarwa, Idris M., Ahmed A. Elamer, Collins G. Ntim, and Aws AlHares. 2020. Gender diversity, corporate governance and financial risk disclosure in the UK. *International Journal of Law and Management* 62: 521–38. [CrossRef]

Ceustermans, Stefanie, Diane Breesch, and Joël Branson. 2017. Voluntary Disclosure of Sales and the Extent of Trade Credit in Small Private Companies. *Accounting in Europe* 14: 388–406. [CrossRef]

Chen, Deqiu, Ming Liu, Tao Ma, and Xiumin Martin. 2017. Accounting quality and trade credit. *Accounting Horizons* 31: 69–83. [CrossRef]

Chen, Shenglan, Hui Ma, and Qiang Wu. 2019. Bank credit and trade credit: Evidence from natural experiments. *Journal of Banking & Finance* 108: 105616. [CrossRef]

Cheng, Nam Sang, and Richard Pike. 2003. The trade credit decision: Evidence of UK firms. *Managerial and Decision Economics* 24: 419–38. [CrossRef]

Cooke, Terry E. 1998. Regression analysis in accounting disclosure studies. *Accounting and Business Research* 28: 209–24. [CrossRef]

Deloof, Marc, and Maurizio La Rocca. 2015. Local financial development and the trade credit policy of Italian SMEs. *Small Business Economics* 44: 905–24. [CrossRef]

Deumes, R. 2008. Corporate Risk Reporting: A Content Analysis of Narrative Risk Disclosures in Prospectuses. *Journal of Business Communication* 45: 120–57. [CrossRef]

Diamond, Douglas W., and Robert E. Verrecchia. 1991. Disclosure, liquidity, and the cost of capital. *The journal of Finance* 46: 1325–59. [CrossRef]

Elliehausen, Gregory E., and John D. Wolken. 1993. The demand for trade credit: An investigation of motives for trade credit use by small businesses. *Federal Reserve Bulletin* 79: 929.

Elshandidy, Tamer, and Lorenzo Neri. 2015. Corporate Governance, Risk Disclosure Practices, and Market Liquidity: Comparative Evidence from the UK and Italy. *Corporate Governance: An International Review* 23: 331–56. [CrossRef]

Elshandidy, Tamer, and Philip J. Shrives. 2016. Environmental Incentives for and Usefulness of Textual Risk Reporting: Evidence from Germany. *The International Journal of Accounting* 51: 464–86. [CrossRef]

Elshandidy, Tamer, Ian Fraser, and Khaled Hussainey. 2013. Aggregated, voluntary, and mandatory risk disclosure incentives: Evidence from UK FTSE all-share companies. *International Review of Financial Analysis* 30: 320–33. [CrossRef]

Elzahar, Hany, and Khaled Hussainey. 2012. Determinants of narrative risk disclosures in UK interim reports. *The Journal of Risk Finance* 13: 133–47. [CrossRef]

Fama, Eugene F., and Michael C. Jensen. 1983. Agency problems and residual claims. *The Journal of Law and Economics* 26: 327–49. [CrossRef]

Fisman, Raymond. 2001. Trade credit and productive efficiency in developing countries. *World Development* 29: 311–21. [CrossRef]

Fowowe, Babajide. 2017. Access to finance and firm performance: Evidence from African countries. *Review of Development Finance* 7: 6–17. [CrossRef]

Fox, John. 2015. *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks: Sage Publications.

García-Teruel, Pedro Juan, and Pedro Martínez-Solano. 2007. Effects of working capital management on SME profitability. *International Journal of Managerial Finance* 3: 164–77. [CrossRef]

Giannetti, Mariassunta, Mike Burkart, and Tore Ellingsen. 2011. What You Sell Is What You Lend? Explaining Trade Credit Contracts. *Review of Financial Studies* 24: 1261–98. [CrossRef]

Haj-Salem, Issal, Salma Damak Ayadi, and Khaled Hussainey. 2019. Corporate governance and risk disclosure quality: Tunisian evidence. *Journal of Accounting in Emerging Economies* 9: 567–602. [CrossRef]

Haj-Salem, Issal, Salma Damak Ayadi, and Khaled Hussainey. 2020. The joint effect of corporate risk disclosure and corporate governance on firm value. *International Journal of Disclosure and Governance* 17: 123–40. [CrossRef]

Huyghebaert, Nancy, Linda Van de Gucht, and Cynthia Van Hulle. 2007. The Choice between Bank Debt and Trace Credit in Business Start-ups. *Small Business Economics* 29: 435–52. [CrossRef]

Islam, Mohammad Nazrul, and Clark M. Wheatley. 2020. Impact of Climate Risk on Firms' Use of Trade Credit: International Evidence. *The International Trade Journal* 35: 40–59. [CrossRef]

White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838. [CrossRef]

Jensen, Michael C., and William H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3: 305–60. [CrossRef]

Khlif, Hichem, and Khaled Hussainey. 2016. The association between risk disclosure and firm characteristics: A meta-analysis. *Journal of Risk Research* 19: 181–211. [CrossRef]

Krippendorff, Klaus. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research* 30: 411–33. [CrossRef]

Lin, Qiang, and Beibei Qiao. 2020. The relationship between trade credit and bank loans under economic fluctuations-based on the perspective of the supply chain. *Applied Economics* 53: 688–702. [CrossRef]

Linsley, Philip M., Phillip J. Shrives, and Mandy Crumpton. 2006. Risk disclosure: An exploratory study of UK and Canadian banks. *Journal of Banking Regulation* 7: 268–82. [CrossRef]

Long, Michael S., Ileen B. Malitz, and S. Abraham Ravid. 1993. Trade credit, quality guarantees, and product marketability. *Financial Management* 22: 117–127. [CrossRef]

Marston, Claire L., and Philip J. Shrives. 1991. The use of disclosure indices in accounting research: A review article. *The British Accounting Review* 23: 195–210. [CrossRef]

Miihkinen, Antti. 2012. What Drives Quality of Firm Risk Disclosure? *The International Journal of Accounting* 47: 437–68. [CrossRef]

Miihkinen, Antti. 2013. The usefulness of firm risk disclosures under different firm riskiness, investor-interest, and market conditions: New evidence from Finland. *Advances in Accounting* 29: 312–31. [CrossRef]

Moumen, Néjia, Hakim Ben Othman, and Khaled Hussainey. 2015. The value relevance of risk disclosure in annual reports: Evidence from MENA emerging markets. *Research in International Business and Finance* 34: 177–204. [CrossRef]

Nahar, Shamsun, Mohammad Istiaq Azim, and Md Moazzem Hossain. 2020. Risk disclosure and risk governance characteristics: Evidence from a developing economy. *International Journal of Accounting & Information Management* 28: 577–605. [CrossRef]

Pattnaik, Debidutta, Mohammad Kabir Hassan, Satish Kumar, and Justin Paul. 2020. Trade credit research before and after the global financial crisis of 2008—A bibliometric overview. *Research in International Business and Finance* 54: 101287. [CrossRef]

Paul, Salima, and N. Wilson. 2007. The determinants of trade credit demand: Survey evidence and empirical analysis. *Journal of Accounting, Business and Management* 14: 96–116.

Petersen, Mitchell A., and Raghuram G. Rajan. 1997. Trade credit: Theories and evidence. *The Review of Financial Studies* 10: 661–91. [CrossRef]

Schwartz, Robert A. 1974. An economic model of trade credit. *Journal of Financial and Quantitative Analysis*, 643–57. [CrossRef]

Smith, Janet Kiholm. 1987. Trade credit and informational asymmetry. *The Journal of Finance* 42: 863–72. [CrossRef]

Solomon, Jill F., Aris Solomon, Simon D. Norton, and Nathan L. Joseph. 2000. A Conceptual Framework for Corporate Risk Disclosure Emerging from the Agenda for Corporate Governance Reform. *The British Accounting Review* 32: 447–78. [CrossRef]

Spence, Michael. 1973. Job Market Signalling. *The Quarterly Journal of Economics* 87: 355–74. [CrossRef]

Van den Bogaerd, Machteld, and Walter Aerts. 2014. Media reputation of a firm and extent of trade credit supply. *Corporate Reputation Review* 17: 28–45. [CrossRef]

Van Horen, Neeltje. 2005. Trade Credit as a Competitiveness Tool: Evidence from Developing Countries. Available online: https://mpra.ub.uni-muenchen.de/2792/1/MPRA_paper_2792.pdf (accessed on 1 January 2021).

Wooldridge, Jeffrey. 2002. Econometrics of cross section and panel data. Cambridge, MA: MIT Press, 2002a). "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition and Stratification". *Portuguese Economic Journal* 1: 117–39. [CrossRef]

Xu, Hongkang, Trung H. Pham, and Mai Dao. 2020. Annual report readability and trade credit. *Review of Accounting and Finance* 19: 363–85. [CrossRef]

Yohn, Teri Lombardi, and Kristian D. Allee. 2009. The Demand for Financial Statements in an Unregulated Environment: An Examination of the Production and Use of Financial Statements by Privately Held Small Businesses. *The Accounting Review* 84: 1–25. [CrossRef]

Journal of
Risk and Financial Management

Review

Energy Crisis Risk Mitigation through Nuclear Power and RES as Alternative Solutions towards Self-Sufficiency

George Halkos * and Argyro Zisiadou

Laboratory of Operations Research, Department of Economics, University of Thessaly, 38333 Volos, Greece
* Correspondence: halkos@econ.uth.gr

Abstract: This paper reviews the case of nuclear energy. Currently, the worldworld is facing one of the greatest energy crises due to the Russo-Ukrainian war. This conflict has lead to limited sources of gas, causing a dramatic decrease in energy supply, leading to emerging energy crisis risks. This is one on the main purposes of reviewing nuclear energy as a possible energy alternative in the future. Apart from presenting the basis of nuclear energy and nuclear reactors, we attempt to compare this source of electricity with other renewable energy forms, such as solar, wind and hydroelectric power. Furthermore, we illustrate the benefits and drawbacks that have been observed regarding nuclear power as well as its contribution to economic growth and the impact it has had on the environment. It has been said that, with the use of nuclear power, air pollution will be reduced because of the elimination of greenhouse gases. However, nuclear power, apart from the final product, generates waste that in this case is radioactive, meaning that the management and disposal techniques are of the utmost importance. Of course, unfortunate events that involved nuclear power do exist and are unfortunately engraved in our memories. Both the nuclear accidents, such as Three Mile Island, Chernobyl and Fukushima, and nuclear weapons usage by military forces, the well-known atomic bombing of Hiroshima and Nagasaki, bring great controversy regarding the adaptation of nuclear power. As is presented in the paper, since the beginning of the new millennium the scheme of energy production and electricity production appears to have changed drastically. By using available data reported by BR, we illustrated that the production of energy and electricity has increased over the last 22 years (2000–2021) due to excessive demand; however, what is more important to mention is the share of both electricity and energy derived from renewable forms such as solar, wind and hydroelectric power. It is shown that more and more countries adopt those sources of energy than did in previous decades. It is crucial to note that it is not the science that causes catastrophic events, but rather the errors of humans.

Keywords: energy crisis risks; nuclear energy; renewable energy; greenhouse gases emissions; nuclear accidents; nuclear weapons; nuclear waste management

JEL Classification: Q01; Q40; Q42; Q52; Q53

Citation: Halkos, George, and Argyro Zisiadou. 2023. Energy Crisis Risk Mitigation through Nuclear Power and RES as Alternative Solutions towards Self-Sufficiency. *Journal of Risk and Financial Management* 16: 45. https://doi.org/10.3390/jrfm16010045

Academic Editor: Thanasis Stengos

Received: 12 December 2022
Revised: 6 January 2023
Accepted: 10 January 2023
Published: 11 January 2023

1. Introduction

In present times, humanity is facing one of the most difficult periods in modern history. Over the last years, the quality of the environment has changed dramatically. Halkos and Zisiadou (2019) mention that nature acts independently, and human beings cannot manipulate the results of this independency. Nevertheless, this statement refers to the natural phenomena that occur on Earth, such as earthquakes, volcanic eruptions, tsunamis, etc., so even if we exclude human activity from the scheme, it is logical that the Earth will continuously transform. However, the impact of human activity should always be taken into consideration because it causes significant, and sometimes disastrous, alterations. It has been said that technology can either be a blessing or a curse depending on the way it is used (Halkos and Zisiadou 2020a).

Since the advent of the Industrial Revolution, there has been a sharp increase in the demand for electricity, and as a result the massive production of electricity by non-renewable sources. Because of the use of fossil fuels as main production raw materials, leading to enormous greenhouse gases emissions, environmental quality has been degraded. Greenhouse gases have caused the increase in the temperature of the terrestrial atmosphere, which is also known as global warming, and combined with air pollution caused due to increased industrialization, the Earth is facing the phenomenon of climate change. The situation is getting worse year by year due to the fact that there is a global annual increase in energy demand. In order to generate these amounts of energy needed, we should either increase the production of electricity using fossil fuels, which will then increase the greenhouse gases emissions and air pollution in general (Barros and Managi 2016), or we should promote more sustainable forms of energy. Solar, wind, hydroelectric and nuclear power can generate electricity while at the same time reducing greenhouse gases and carbon dioxide emissions. Renewable sources of energy (solar power, wind power and hydroelectric power) characterized by unpredictable power production (Basit et al. 2020) do not totally fulfill the criteria of sufficient productivity (Mastrocinque et al. 2020; Mourmouris and Potolias 2013) due to the fact that they are affected by the acts of nature. Nuclear power, on the other hand, may be a suitable option for power production replacement.

Another problem that the worldEarth is currently facing is the energy crisis that has resulted from the Russo-Ukrainian war. Following Alam et al. (2022) and Prohorovs's (2022) work regarding the Russo-Ukrainian war, we aim to underline possible alternatives that can be used to address the global energy crisis. Researchers emphasize that the existing war has pushed the global economy to a crucial point (Alam et al. 2022; Prohorovs 2022), where increased inflation rates observed globally as well as the instability of commodity markets and the increased uncertainty indicate a not so prosperous economic future. The limited fossil fuels that are available in nature, together with the energy embargo of Russia have led to a sharp increase in prices, with consequences to the global economy through inflation. Prohorovs (2022) examined the consequences for European countries' businesses and economies and highlighted that most countries need to expedite their development of sufficiency with regard to enerty resources. This appears to be a powerful factor that brought nuclear power into discussion as an attempt to ensure that there is going to be sufficient energy worldfor the Earth, inthe following years, and at the same time provide sustainable solutions for the next generations. What is important to examine, however, is whether nuclear energy appears to be trustable by nations around the worldworld as a source of energy production. For that reason, we are going to investigate, using reported data, whether nuclear energy has been adopted by more countries over the last two decades. Moreover, based on the same dataset, we are going to address the questions regarding renewable forms of energy and electricity production. A main question raised about renewables is if they are assumed to be a possible extension of energy production. Do nations trust and adopt those sources in order to cover a part of the needed energy and electricity production? Do they increase their share of global production?

The structure of the paper is as follows following: Section 2.1 provides a literature review on nuclear energy and specifically the reactors, while Section 2.2 presents the other renewable energy forms. Section 2.3 underlines the advantages and disadvantages of nuclear power, while Section 2.4 draws attention to the economic growth and environmental impact that have resulted from it. The following section provides information regarding nuclear waste management and disposal. Sections 2.6 and 2.7 review the most known nuclear accidents and military attacks using nuclear weapons. Section 3 illustrates the current conditions regarding the global energy mix and electricity since 2000. Finally, the last section concludes with significant statements regarding the topics analyzed above.

2. Literature Review

2.1. Nuclear Energy

Over the last several years, more and more nuclear power reactors have been constructed in an attempt to increase the nuclear energy produced globally. Many reasons are hidden behind these decisions, but energy sufficiency and climate change appear to be the most important. The first nuclear power plants in the world started operating in the 1950s. Since then, their expansion has been noticeable. Deutch et al. (2003), underline that during this time, 44 new nuclear power plants were under construction in 12 different countries, with most of the constructions being in China, India, Korea and Russia. As for the USA, at that point there was no plan for additional nuclear power plant construction. A fewew years later, Ozcan et al. (2016) showed that new power plants were being constructed. At that period of time, the number of reactors under construction increased to 59, with four more countries entering into their construction. Akyuz (2017), emphasizes that, to date, 31 countries around the world have nuclear power plants, with a total of 447 nuclear reactors which annually produce about 17% of the world's electricity (da Mata et al. 2017). Elliott (2016) highlights that the net nuclear power capacity equals to 392 gigawatts (GW).

Topal-Namli and Namli (2014) provide information regarding the percentage of nuclear power of a portion of total energy for some countries. More specifically, 75% of the electricity used in France derives from nuclear power, while other european countries such as Belgium, the Czech Republic, Hungary, Slovakia, Slovenia, Sweden, Switzerland and Ukraine depend on nuclear power for up to 1/3 of their total consumption. Additionally, more than 30% of the electricity used in South Korea, Bulgaria and Finland derives from nuclear power. High dependency on nuclear power is also reported in USA, United Kingdom, Spain and Russia, with nuclear power consumption reaching almost the 20% of the total energy expenditures. Japan, which is known for its nuclear facilities, derives than 25% of its annual electricity supply from nuclear power. Lastly, countries such as Italy and Denmark that do not have any nuclear power plants satisfy approximately 10% of their electricity demand through nuclear power.

The great concern that has been raised regarding the phenomenon of climate change in recent years has forced scientists and policy makers around the world to explore new forms of energy production. What makes nuclear power a great candidate is the fact that, unlike oil and gas, the raw materials that are used in nuclear power plants, such as uranium, are available in large quantities in nature, which can ensure adequate energy production in the present, but also in the long-run, maintaining the viability of this form of energy (Jewell 2011; NEA 2008; Macfarlane and Miller 2007). Moreover, as da da Mata et al. (2017) mention, nuclear energy tends to be more competitive compared to other forms of energy, due to the fact that it produces a large amount of electric power, with low operational costs and minimum polluting gas emissions, which mainly occur during material and waste transport, while at the same time providing great financial returns. Additionally, Brook et al. (2014), indicate that nuclear fission has the lowest environmental impact, especially because of limited gas emissions, a statement which classifies nuclear energy among the cleanest sources of power. This statement, together with the increasing concern regarding climate change, has forced governments and policy makers to rethink the possibility of launching national nuclear power programs (Adamantiades and Kessides 2009; Nuttall 2005; Sauga 2008).

Reactors

Vaillancourt et al. (2008) released a thorough analysis regarding the reactors that are currently in use worldwide. Based on their analysis, there are five different types of reactors available, from which four types of reactors produce electricity with nuclear fission using uranium as a fuel, while only one of the reactor types produces electricity with nuclear fusion using tritium, which is produced from lithium, as a fuel. More specifically, the five available reactor types are: (a) Light water reactors (LWR), (b) Advanced light water reactors (A-LWR), (c) Pressurized heavy water reactors (PHWR), (d) High-temperature gas-cooled reactors, and (e) Fusion power reactors.

More analytically, these five types of reactors are divided into those categories based on the technology used to produce electricity, as well as the primary fuel they are using during the main procedure. Starting from the first type of reactor, the Light water reactor (LWR), it is important to mention that it is a second-generation reactor and based on the NEA (Nuclear Energy Agency) (NEA 2005), 90% of the total reactors in use globally belong to this specific type of reactor. The two subtypes of Light water reactors are the pressurized water reactors (PWR) and the boiling water reactors (BWR). The OECD (Organization for Economic Co-operation and Development) (OECD 2001) mentions that the development of the Light water reactors took place in the USA and FSU (Former Soviet Union). Vaillancourt et al. (2008) explain that LWR tend to be replaced by newer technologies and more advanced reactors when their useful life cycle expires.

The second most common type of reactor worldwide is the Advanced light water reactor (A-LWR), which is described as an upgrade of the LWR. These third-generation reactors (Hamacher and Sheffield 2004) appear to be the upgraded version of LWR in terms of lower investment and operational costs, a higher availability factor, and a longer lifetime, and as mentioned above they are part of the LWR replacement process. It is interesting to note that 80% of the under-construction reactors globally are of this type (NEA 2005). On the contrary, a not so selected reactor is the Pressurized heavy water reactor. As EIA (2006) illustrates, the PHWR is designed by Atomic Energy of Canada Ltd., and includes amongst others the CANDU (Canada Deuterium Uranium) reactor. However, Vaillancourt et al. mention that only 5% of the reactors in use globally belong to this category, and only constitute 15% of the under-construction reactors. Moreover, apart from Canada, which is their country of origin, five more countries have adopted these types of reactors, including Argentina, China, India, Romania and South Korea. India and Romania are currently constructing more of the PHWR (EIA 2006).

The next two types of reactors, the High-temperature gas-cooled reactor and the Fusion power reactor, are under development, with the expectation that they will be released around 2030 and 2050, respectively. More specifically, the High-temperature gas-cooled reactor is characterized as the fourth-generation reactor, is developed in United Kingdom, and its main difference, compared to the previous three reactors, is the fact that it uses helium as a coolant element instead of water, and is planned for construction in South Africa (OECD 2001). Last is the Fusion power reactor, which is an ITER project, and it is also currently under development. Researchers support the belief that this type of reactor will be available in 2050 based on estimates (Fiore 2006; Hamacher and Sheffield 2004; Tokimatsu et al. 2003). Based on Fiore (2006), the factor that establishes the Fusion power reactor as the best possible choice is its difference with regard to the type of fuel used. As already mentioned, the first four reactors use uranium as a fuel, while the fifth reactor will use lithium, an element for which reserves are reported to be unlimited and, indeed, will be available for a thousand years (Vaillancourt et al. 2008; Fiore 2006). What is crucial to mention is that there are more types of reactors such as the Small Modular Reactor (SMR), Gen-IV reactors and others; however, the five reactors analyzed above by Vaillancourt et al. (2008) are assumed to be the most common globally. Reviews regarding the Gen-IV reactors and the status of their operations has been conducted by many researchers focused on the nuclear energy (Lorusso et al. 2018; Kim 2013; Murty and Charit 2008; Driscoll and Hejzlar 2005). Similarly, regarding the operation and progress of SMR and whether this type of reactor can be the cheapest and safest choice of nuclear energy production as a result the possible future of energy globally has been discussed by numerous scientists (Kumar et al. 2021; Rowinski et al. 2015; Cooper 2014; Liu and Fan 2014; Vujić et al. 2012; Wang et al. 2011).

2.2. Renewable Energy Sources

Renewable energy forms have gained great public attention over the last years, possibly due to the opposition of institutions to nuclear energy and with regard to their efforts in reducing GHG emissions (Hossain et al. 2022; Immonen and Kopsakangas-Savolainen 2022;

Zhao et al. 2022; Brook et al. 2014). As has already been mentioned, there is a great concern in the circles of environmental and social scientists due to the sharp increase of reported Greenhouse Gases (GHG) emissions and the adverse outcomes leading to climate change. In an attempt to reduce GHG, a great majority of countries are trying to adopt alternate forms of energy. As Brook et al. (2014) underline in their paper, institutions and public opinion appear to be against nuclear energy production because of the belief that this form of energy is assumed to be unsustainable, uneconomic and unsafe, while it increases the risk of nuclear weapons production and use. Renewable energy appears to be one of the main options countries have, with some nations announcing that their main purpose is to substitute, if not all, at least the majority of their electricity production by using renewables (Brook et al. 2014). Another reason that made people turn their focus to renewables is the realization that there is a limited availability of fossil fuels, meaning that in the future they will not be able satisfy the required energy demands. Thus, if we exclude the energy produced by fossil fuels and/or nuclear energy, the alternatives we have are (a) solar power, (b) wind power, (c) hydroelectric power, (d) geothermal power, and (e) biomass. Regarding the sustainability of renewable energy, it is crucial to mention that a source can be assumed as sustainable only if it can provide a great amount of energy in the long-run, is capable of fulfilling daily demands, without depriving future generations, while at the same time being reliable, safe and environmentally friendly, and of course economically feasible. Therefore, the origin of renewable energy types (sunlight, wind, water etc.) sets limitations on the availability of energy production due to weather phenomena, leading to the belief that they may not be able to be characterized as sustainable (Brook et al. 2014). However, what is important to mention is that due to the development of technology, more and more solutions have been found in order to solve the problems that are occurring. In other words, scientists try to advance the technology with regard to the energy storage. Alkhalidi et al. (2022), Kebede et al. (2022), and Aneke and Wang (2016) have made a thorough analysis of all available energy storage technologies that are currently available around the world and may be assumed as a solution to the renewable source projects by improving the efficiency of energy production as well as the energy storage applications.

As Aneke and Wang (2016) mention, there are two different types of energy storage technologies, the electrical and the thermal, with subtypes such as mechanical energy storage (MES), chemical energy storage (CES), electrochemical energy storage, superconducting magnetic energy storage (SMES), cryogenic energy storage, sensible heat storage, latent heat storage, thermochemical heat storage, and possibly even more subtypes of energy storage technologies. The importance of energy storage technologies has also been described by Rahman et al. (2020), who also reported the challenges attached to the stability and reliability of renewables with regard to sufficient energy supplies. Through an overview of energy storage technologies, Rahman et al. (2020) have concluded that energy storage systems may provide a solution to the uncertainty regarding the sufficiency of energy produced and stored by renewable sources. Having these solutions in mind, more nations will be able to adopt the renewable forms of energy production, and hopefully in the future the environmental impact of energy production will be reduced.

2.2.1. Solar Power

One of the most common forms of renewable energy is solar. It is based on solar radiation, which is captured through photovoltaic cells (PV), which, by heating water, produce electricity. As da Mata et al. (2017) underline in their paper, the photovoltaic flat plates can be installed either on buildings or special structures. In recent years there has been a significant increase in the number of sales of photovoltaic modules, which therefore led to a decrease in their price. Although there was a significant price drop, the average cost per photovoltaic unit is still high, sometimes even estimated at 10 times the cost of more common sources of electricity. This drawback may be a limitation to the expansion of the solar power industry.

2.2.2. Wind Power

Another form of renewable energy that has gained increasing attention in recent years is wind power. According to da Mata et al. (2017) there has been a recorded annual increase of about 20% over the last 5 years. Brook et al. (2014) emphasize that fact that wind power is one of the most ancient types of power generation. Over the centuries, wind provided great benefits to humanity for everyday activities such as wheat grinding, wood sawing and water pumping. Moreover, for many centuries, travelling around the world was only possible due to wind used in sailing, making transportation possible. One of the main drawbacks of wind power, however, is the limited availability. Electricity, or power in general, cannot be produced when there is no wind. In previous centuries, in there was no wind for a day or two, transportation was likely to be delayed. However, in current times, if there is no wind, the electricity already produced may not be able to cover the daily power demands.

2.2.3. Hydroelectric Power

Another renewable form of energy da Mata et al. (2017) illustrate is hydroelectric power, which is also believed to be an ancient type of energy production. The generation of hydroelectric power can be achieved through the use of the force of water, usually taking the form of rivers and dams. In contrast to solar and wind power, which are not fully reliable due to the possible lack of sunlight and/or wind, hydroelectric power appears to be a more stable source of energy. da Mata et al. (2017) underline the fact that by 2015, 16% of global electricity was provided by hydroelectric power, with some countries depending mainly on it as their main energy source. Characteristic examples are Norway, with 99% of the country's electric consumption produced by hydroelectric power, while Canada, Switzerland and Sweden have usage percentages of 58%, 55% and 45%, respectively.

Obviously, this type of energy does have its own disadvantages. The first and most serious disadvantage is the fact that in order to have a sufficient, continuous water supply, there is a need to construct dams, which has been shown to cause significant environmental damage to the surrounding ecosystem. The other noticeable disadvantage is a possible drought occurrence, which will decrease the volume of the available water, a phenomenon which is currently more visible due to climate change and global warming (da Mata et al. 2017).

2.2.4. Geothermal Power

Geothermal power, which is recognized as a renewable energy resource (Rybach 2003) is the energy created by the heat of the Earth, as Kagel et al. (2005) have mentioned. Kagel et al. (2005) analyzed the geothermal energy produced by United States since 1960 and pointed out that due to the fact that during geothermal energy production no fossil fuels are used, there are no harmful air emissions that will burden the environment. Moreover, they clarify that although sulfur dioxide is not directly emitted during the energy production through geothermal power plants, the hydrogen sulfide is transformed into sulfur dioxide and sulfuric acid when released to the atmosphere. Rybach (2003), following the Kyoto Protocol suggestions, emphasizes that although all different types of energy production may have an impact on the environment, the technology used during the energy production is what affects the level of encumbrance. As for geothermal power production, based on figures presented by Rybach (2003), coal, oil, and gas have higher GHG emissions, while hydroelectric, solar and wind have lower GHG emissions. These indications characterize geothermal as an attractive alternative solution for GHG emissions reduction. Based on Kagel et al. (2005), due to the fact that the Earth's interior temperature is expected to remain high for billions of years, geothermal energy can be assumed to be a sustainable alternative for a long time.

Regarding the main disadvantages of geothermal energy production, Bagher et al. (2014) underline that this source of energy cannot be available everywhere in the world, while at the same time the power plant sites are not close to the distribution areas. Moreover, when analyzing the investment cost, the steam power plant cost is relatively high, and it is not certain whether a depreciation of such an investment will be achieved.

2.2.5. Biomass

Another renewable energy source alternative is biomass, which is assumed to be a great source of energy that can contribute to the increased energy demands of the modern society (Bridgwater 2006). Compared to other renewable energy sources, biomass is the one that produces the most emissions during energy generation that burdens the environment with gases (HCI, SO_2, HF, HCN, NOx and CO), particulates (PM 2.5, PM 10), as well as toxins that, if not handled properly, may become hazardous (Song and Hall 2020). A great disadvantage of biomass as described by Vassilev et al. (2015) is the high investment cost required for a biomass power plant installation, thus increasing the risk of non-depreciation, as is the case with geothermal energy.

2.3. Advantages & Disadvantages of Nuclear Energy

Living in the real world, none of the decisions or actions human beings choose is perfect. The same idea is observed in terms of nuclear energy as well. When examining the option of adopting a new process, we have to thoroughly comprehend all different aspects, positive and negative, that may appear, so as to weight them properly on our final decision. For that reason, before we reach any conclusions about nuclear energy, we need to evaluate both the advantages and disadvantages that exist. As has already been mentioned above, 90% of the in-use reactors belong to the LWR type. Based on Schiermeier et al. (2008), this specific type of reactor can produce electricity with a cost varying between $0.025 and $0.07 USD per kilowatt-hour, with the fluctuation appearing due to the design and/or requirements of each reactor. The equivalent cost per kilowatt-hour when using natural gas is approximately $0.05 USD. The low cost of production is mainly affected by the low cost of raw material (Temurçin and Aliağaoğlu 2003). A similar comparison between nuclear energy and production cost was also made by Akyuz (2017) and Aras (2013). That factor highlighted the fact that nuclear energy can sometimes be an cheaper solution compared to the known and excessively used methods.

Following the shame concept, other researchers have estimated the cost per MWh both for nuclear energy generation and renewable energy generation. More specifically, Korkmaz and Önöz (2022) provide evidence that the average cost system cost regarding the nuclear energy scenario equals to $64 USD per MWh (giving a $0.064 USD per KWh), while the renewable energy scenario provides an average system cost of $62.3 USD per MWh (giving a $0.0623 USD per KWh). As they mention, the reported costs are the levelized costs of electricity generation (LCOE). Barthelmie and Pryor (2021) reported the LCOE of wind energy generation and underlined that regarding the onshore wind en-ergy LCOE is almost $40 USD per MWh in U.S.A and Europe (giving a $0.04 USD per KWh) and almost $60 USD per MWh in Asia (giving a $0.06 USD per KWh) while the offshore LCOE is higher compared to the onshore LCOE and ranges between $106-$171 USD per MWh in specific countries (giving a $0.106-$0.171 USD per MWh). The results provides by Korkmaz and Önöz (2022) and Barthelmie and Pryor (2021) are comparable with those provides by Schiermeier et al. (2008). Hansen (2019) followed a similar ap-proach in reporting LCOE of many different electricity generation sources, however, the results are presented in €. As Hansen (2019) mentions, onshore wind power has the lowest LCOE, which equals to €34 per MWh (€0.034 per KWh), followed by coal power plants with LCOE equal to €39 per MWh (€0.039 per KWh), followed by large-scale hydro-power with a LCOE equal to €40 per MWh (€0.04 per KWh) and small-scale hydro-power with a LCOE equal to €45 per MWh (€0.045 per KWh). Combined Circle Gas Turbines provide a LCOE equal to €58 per MWh (€0.058) while the offshore wind power gives a LCOE equal to €66 per MWh (€0.066 per KWh) and nuclear power generations provides a LCOE equal to €69 per MWh (€0.069 per KWh). Similarly, large-scale photovoltaics (solar power) report a LCOE equal to €72 per MWh (€0.072 per KWh) while small-scale photovoltaics report a LCOE equal to

€104 per MWh (€0.104 per KWh). Finally, biomass power plants provide a LCOE equal to €88 per MWh (€0.088 per KWh).

If we compare nuclear energy to other renewable energy forms such as solar power and/or wind power, it is proven that nuclear energy is assumed to be a sustainable source due to the fact that it fulfills all required criteria. Renewable energy forms depend on meteorological conditions, as we have already stated. Those conditions may reduce or even eliminate electricity production using solar, wind or hydroelectric power. Nuclear power, on the other hand, is not affected by similar conditions, leading the supporters of the NNPs to emphasize that sustainability is one of the main advantages of nuclear energy (Coskun and Tanriover 2016). Researchers (Akyuz 2017; Kok and Benli 2017; Kurt 2014; Aras 2013) who have analyzed Turkey's case regarding its dependence in oil and natural gas imports conclude that the country is heavily dependent of those imports, and even a slight decrease on those imports may have a huge negative impact to the electrical situation of the entire nation. If establishing nuclear power plants in Turkey, they firmly believe that the risk of electrical insufficiency is drastically reduced. Moreover, apart from the risk of sufficient supply, Topal-Namli and Namli (2014) state that Turkey will also be able to reduce energy expenditures.

Another great advantage supporters of NPPs underline is the longevity of the main equipment of the nuclear power plants, the reactors. More specifically, the initial reactor designs were guaranteeing a 40-year operational lifespan of the reactors used; however, the evolution of technology extends their operational ability up to 80 years (Ozcan et al. 2016). This is the strongest and most alluring benefit that nuclear energy has to offer. Contrary to fossil fuels, which are the main producers of greenhouse gas emissions, causing global warming and climate change, nuclear power plants do not emit greenhouse gases during the process of electricity production (Akyuz 2017; Ilhan et al. 2010; Sirin 2010; Erdogdu 2007).

Although nuclear energy and nuclear power plants have numerous benefits, both for the economy and the environment that have already been mentioned, their opponents express their firm dissatisfaction based on their significant drawbacks (Akyuz 2017). The initial statement that is raised by Gunay and Iseri (2017) is related to the insufficient and ineffective regulations that many countries have regarding nuclear energy. More specifically, many countries around the world, such as Turkey, do not have sufficient and laws and regulations, nor highly-trained operators, that are required to observe and guarantee nuclear safety. This factor significantly increases the risk of a potential industrial accident. In order to tackle this possibility, the governments of these countries should establish a comprehensive nuclear energy plan in an attempt to reduce current uncertainty (Topal-Namli and Namli 2014).

Despite the fact that Schiermeier et al. (2008), underline the low cost of production compared to fossil fuels, Akyuz (2017), states that nuclear energy can be assumed to be a very costly energy source due to the excessive investment cost. This is a critical reason for Turkey's NPP cancellations. Another factor that increases the operational costs of NPPs is nuclear waste management, which requires substantial standard techniques and equipment (Ozcan et al. 2016). Regarding waste management and disposal, this will be thoroughly described in a following Section (see Section 2.5), Cohen (1983) and Sirin (2010), emphasize the great danger that lies beneath waste management and disposal. Nuclear power plants create toxic, radioactive and extremely hazardous waste that remains hazardous for sometimes even hundreds of thousands of years. Cohen (1983) specifically mentions that 99% of their contaminants remain on materials even after 600 years. Even if we dispose of them in the burying zones, there is an unacceptable level of risk both for people and the environment.

Regarding spent fuel and its recycling process, apart from being costly, it increases the possibility of nuclear theft risk. Due to the fact that nuclear weapons can be created from nuclear energy fuel residuals, the proliferation danger is increased. (Sokolski 2010). Moreover, nuclear power plants can produce plutonium for military purposes (Akcay 2009), and although most countries of the world have signed the Non-Proliferation Treaty (see

Section 2.7.3), no one can ensure that they will not create nuclear bombs which will disturb worldwide peace.

History shows that the next three drawbacks sometimes may be connected with each other. Topal-Namli and Namli (2014) state that nuclear power plant locations are chosen based on their surroundings, due to the fact that those facilities have need immediate access to water for cooling purposes. Therefore, seaside, riverside and lakeside areas are the perfect location for the construction of a nuclear power plant. On the other hand, based on recorded events, nuclear power plants are vulnerable to natural phenomena, such as earthquakes and tsunamis. The most well-known case is the Fukushima Daiichi Nuclear Power Plant case (see Section 2.6.3), a facility which was built close to the coast in order to have access to water for cooling purposes, and a tsunami caused the cooling system shutdown, leading to the greatest nuclear disaster of the new millennium. The last disadvantage mentioned by Akyuz (2015) is the risk of nuclear accidents in general. In spite of not being a frequent phenomenon, with only 33 accidents and incidents reported to date, the disastrous impact of such events is devastating. The worst case in history was the Chernobyl disaster (see Section 2.6.2).

2.4. Economic Growth and Environmental Impact

Another aspect that authorities have to weigh for a possible expansion of nuclear energy is the economic growth that will accrue to each nation along with the range of environmental impacts. An in-depth study by Apergis and Payne (2010) examined the linkage between nuclear energy consumption and economic growth for a period of 26 years (1980–2005) and included 16 countries. The factor that triggered the researches to examine this linkage was the highly volatile price of oil and gas, including the fact that a great number of countries depend on the imports of such products, as well as the excessive greenhouse gases emissions produced from fossil fuels that have a negative impact on the environment. As stated by Fiore (2006) as well as Toth and Rogner (2006), nuclear energy may be be highly significant when used in regions where energy demand growth is rapid if we consider that the scientific society has announced that fossil fuel reserves will not be sufficient to cover the global energy demand sometime in the near future.

One of the most well-known environmental hypotheses that provides a linkage between economic growth and environmental pollution is the Environmental Kuznets Curve (EKC) hypothesis, which graphically is presented as an inverted U-shaped relationship (Iwata Hiroki and Samreth 2010). Based on this theory, during to the initial stages of economic growth and increasing incomes, environmental pollution and natural resource degradation is observed, while after a certain point of increasing income, an additional unit of income leads to a reduction in environmental pollution. A number of scientists have focused their research on CO_2 emissions and energy consumption, mainly nuclear and renewable energy. Irfan and Shaw (2017) state that renewable energy leads to carbon emissions reduction, while non-renewable energy leads to an increase in it. Apergis and Payne (2010) underline that renewable energy consumption does not have a decreasing impact on carbon emissions, a similar finding to the Menyah and Wolde-Rufael (2010) study. Forsberg (2009) and Heal (2009) emphasize that this reduction may not have been achieved due to the lack of adequate storage technology for renewable energy; however, in the long-run renewable energy technologies appear to improve environmental quality. Regarding the environmental impact that s connected to the nuclear energy production, it is crucial to mention that uranium mining may increase the risk of environmental danger. The good news, however, is that due to the new technology in nuclear reactors, the quantity of uranium needed for nuclear energy production will sharply decrease. In that way, the volume of waste that needs to be managed and isolated, as well as radioactivity hazards, will be drastically reduced (Brook et al. 2014).

Franco et al. (2017) undertook a study in India for the period 1901–2011 and their findings indicate that although urbanization leads to economic growth and improvements in quality of life, carbon dioxide emissions are significantly increased due to the higher levels of energy consumption. Many researchers (Shafiei and Salim 2014; Jebli Mehdi Ben and Ozturk 2016), when examining the CO_2 emissions in OECD countries, separate the energy variable into two subcategories, renewable and non-renewable, indicating that non-renewable energy increases CO_2 emissions, while renewable energy decreases CO_2 emissions in these countries. Lastly, Al-Mulali et al. (2015) examine the Vietnam case and observe that although non-renewable energy consumption increases CO_2 emissions, renewable energy consumption has no effect on carbon dioxide emissions. On the other hand, Al-Mulali et al. (2016), examine the Kenya case (1980–2012) and find that renewable energy consumption reduces carbon dioxide emissions, while non-renewable energy consumption and urbanization increase carbon dioxide emissions, and as a result, environmental pollution.

Paramati et al. (2017), on their econometric analysis over the period 1990 to 2012, regarding developing countries, state that renewable energy consumption can positively affect economic growth, as well as the quality of the environment. Menyah and Wolde-Rufael (2010), underline the fact that the economic impact of the global warming phenomenon can be incredibly costly and lead to a 25% decline of global GDP (Gross Domestic Product). If we take into consideration the energy crisis many countries will face due to the fact that energy imports will be declined (Hedenus et al. 2010), we can imagine the great decrease that will be reported in the global GDP. It is crucial to understand that the phenomenon of global warming is not only an environmental disaster, but also a huge threat to the global economy (Adamantiades and Kessides 2009; DeCanio 2009; Reddy and Assenza 2009).

Omri et al. (2015) conducted an in-depth study in order to examine the causality between nuclear and renewable energy consumption and economic growth. They proposed four hypotheses regarding causality, the growth hypothesis, the conservation hypothesis, the neutrality hypothesis and the feedback hypothesis. The growth hypothesis indicates that there is a unidirectional causal relationship running from energy consumption to economic growth. The conservation hypothesis indicates that it is assumed to be a unidirectional causality running from economic growth to energy consumption. The neutrality hypothesis indicates that there is no causality between energy consumption and economic growth, and the feedback hypothesis posits that there is a bi-directional causality between energy consumption and economic growth. Regarding nuclear energy, the growth hypothesis is accepted for Belgium and Spain, while the conservation hypothesis is accepted for Bulgaria, Canada, The Netherlands and Sweden. The neutrality hypothesis is accepted for Finland, Hungary, India, Japan, Switzerland and the U.K. Finally, the feedback hypothesis is accepted for Argentina, Brazil, France, Pakistan, and the United States (Omri et al. 2015). Similarly, regarding renewable energy, the growth hypothesis is accepted for Hungary, India, Japan, The Netherlands and Sweden, while the conservation hypothesis is accepted for Argentina, Spain, and Switzerland. The neutrality hypothesis is accepted for Brazil and Finland. Finally, the feedback hypothesis is accepted for Belgium, Bulgaria, Canada, France, Pakistan and the United States (Omri et al. 2015).

2.5. Waste Management & Disposal

As we have already mentioned in Section 2.1, nuclear energy production, like any other type of production, apart from the final product, also produces waste. However, not all waste is the same. In the case of nuclear waste, we should always have in mind the statement that Corkhill and Hyatt (2018) made regarding the high radioactivity of fission products and actinides. In other words, the most crucial process after power generation is the procedure of waste management and disposal. These radioactive materials need to be treated and stored properly so that they will not have a hazardous effect on humans and the environment (Corkhill and Hyatt 2018; Keeney and Winterfeldt 1994). As is well-known, the level of radioactivity declines over the years. Nevertheless, this is a lengthy process that in some cases may exceed hundreds of thousands of years. The two terms used in order to describe the period of time that is needed for nuclear waste to reduce its radioactivity are half-life and hazardous life, with half-life being the period of time that is needed to decay half of the element's radioactivity, while hazardous life refers to the period of time that is needed to decay to a thousandth or millionth of an element's original amount of radioactivity, and this amount is usually is 10 or 20 half-lives.

For instance, the half-life of Tritium is 12 years, while its hazardous life is equal to 120–240 years. Similarly, the half-life of Strontium-90 equals 28 years, while its hazardous life equals to 280–560 years. A great example is Nickel-25, which has a half-life equal to 76,000 years, and a hazardous life equal to 760,000–1,520,000 years. Having this a priori information in mind, it is essential to mention that is not only important how the waste is treated after nuclear power production, but it is also crucial to find suitable methods and storage for these radioactive elements in the long-run, and of course avoid the corrosion of the storage material before the end of the hazardous life, otherwise the radioactivity will be spread to the environment. This is the main reason why nuclear waste management is significantly more expensive compared to nuclear energy production.

The four main nuclear waste types are the following: Spent Nuclear Fuel (SNF), High-Level Waste (HLW), Intermediate-Level Waste (ILW) and Low-Level Waste (LLW). Each type of waste has a different process of management and disposal. Before we move forward to the waste management and disposal processes, it is important to clarify the differences between these types of waste. The nuclear fuel used in a reactor in order to produce electricity is uranium oxide. Since uranium oxide has undergone fission, it is removed from the reactor and is called Spent Nuclear Fuel (Corkhill and Hyatt 2018) and is assumed to be the most controversial type of nuclear waste (Ramana 2018; Flynn et al. 1995), due to the fact that it continuously transforms and includes high levels of radioactivity. If a person was exposed to such level of radioactivity, they would absorb a lethal dose in less than a minute. The most common method of SNF management and disposal is to cool the fuel for several years (2–5 years) under water in special engineered cooling ponds and then stored in containers and transferred to long-term storage locations. Moving forward, nuclear waste that includes concentrations of radionuclides, which increase radiogenic self-heating, are known as high-level waste, and contain substantial levels of radioactivity. Similar to the high-level waste, the intermediate-level wastes contain significant levels of radioactivity, but it is lower compared to high-level waste, and still requires caution with regard to the waste management and disposal process. Last are the low-level wastes which cannot generate heat, have low levels of radioactivity, and do not require any special consideration with regard to the management process (Corkhill and Hyatt 2018). Below, we will analyze all possible methods that can be used in the process of waste management as well as waste disposal.

Waste immobilization is a waste management technique that is divided into two types based on the materials used. The purpose of waste immobilization is to ensure that the radioactive materials will be handled, transported and stored safely, minimizing the potential risks to human health and the environmental. In other words, the material should remain stable during the transport and storage, as well as remain stable under required temperature ranges and radiation fields, and finally, durable and not easily dissolvable under conditions of long-term storage and disposal. Vitrification using borosilicate glass is an attractive immobilization technique, especially for high-level wastes. Not only can borosilicate glass be used during this process, but other kinds of glass, such as aluminophosphate glass, are commonly used in Russia (Corkhill and Hyatt 2018; Stefanovsky et al. 2004; Yim and Linga Murty 2000; Weber et al. 1997; Lutze 1988). The second waste immobilization technique is cementation. In contrast to borosilicate glass, cementation is preferable for intermediate-level waste. In this process, the radioactive waste is simply surrounded or encapsulated by a wet cement paste which gradually turns into a hard cement block. A great advantage of this process is the fact that cement is a low-cost material which can be simply processed and handled. (Corkhill and Hyatt 2018; Yim and Linga Murty 2000).

Another option for waste management that has been proposed was to remove nuclear waste from the Earth and to store it in space. A more suitable option is assumed to be the removal of nuclear waste from the surface of the Earth. The reason behind this is that the upper layers of the surface are exposed to environmental risks, such as climate change, tectonic plate movements, and of course human intrusion. Thus, geological disposal of high-level waste is proposed, basically by burying the wastes several hundreds of meters or more below ground into special containers. This concept is suitable for the long-term storage of nuclear waste (Corkhill and Hyatt 2018). In order to successfully dispose of nuclear waste using this proposed method, it is crucial to follow the proposed principle so as to maintain safety. This concept is known as the multi-barrier concept and indicates that, first of all, specially engineered multi-layer barriers will contain the nuclear waste until the moment that the majority of the radioactivity will be decayed. Secondly, the host geology will successfully isolate the barrier from the biosphere so as to reduce the likelihood of human intrusion and, finally, the location of the storage facility will be several hundreds of meters below ground and will ensure long transport pathways to decay any significant migration of radionuclides from the waste to the biosphere (Corkhill and Hyatt 2018). Based on Hench et al. (1986), an engineered waste package is a seven-layer container consisting of (a) alkali borosilicate glass or titanate based polyphase serving as a host matrix for the high-level waste, (b) a metal canister, usually stainless steel surrounding the glass compartment, (c) a metallic overpack such as mild steel ductile iron, pure titanium, titanium alloy, or nickel alloy, (d) a sleeve which is used in order to assure the clearance for the package, (e) the backfill of a material contained between the other engineered waste package components and the host rock, (f) a buffer, which is a material used to facilitate the conditioning of the ground water, and (g) filler, which can be any material used to fill the space between the other components of the waste package.

2.6. Nuclear Accidents

The majority of individuals who oppose nuclear power plants, and nuclear energy in general, are probably adversely affected by the worst nuclear disasters of the century. These events have raised great public attention and obvious negative publicity. However, it is crucial to mention that science, when used properly and with the required respect, is not disastrous. Irrational decisions, overuse and/or reckless use or other external factors may lead to catastrophic results. Three main events will be described below, where different initial factors led to environmental hazards. However, one of the initial nuclear accidents that occurred was during a fission reaction by Slotin and his colleagues in 1946 leading to Slotin's death and giving us the knowledge regarding the negative results of exposure to radiation. More specifically, on 21 May 1946, at Los Alamos Scientific Laboratory, a nuclear experiment was conducted by Louis Slotin and seven other laboratory members (McLaughlin et al. 2000). Based on McLaughlin et al. (2000), two of the members, Louis Slotin and Alvin Graves, were directly involved in the experiment, while the other six individuals were observing the demonstration. Malenfant (1996), who reported four similar laboratory accidents between 1945 and 1946, reported detailed information regarding the LA-1 accident during Slotin's demonstration, and mentioned that the eight personnel were exposed to excessive levels of radiation. More specifically, Oettingen (2018) underlined that Slotin was exposed to 2100 rem, leading to his death nine days after the occurrence of the accident, while Graves received 360 rem, and recovered several weeks after the experiment. Regarding the six observers, Hempelman et al. (1979) mentioned that they were exposed to lower levels of radiation, and thus there were no major impacts on their health. Malenfant (1996) indicated that this accident, as well as similar accidents to Slotin's, have a great significance because they provide information about the undesirable effects when potentially hazardous experiments were conducted without the appropriate planning, preparation and safety measures. In other words, Slotin's accidents and similar accidents of the past provided us with the knowledge that nuclear energy can be hazardous and disastrous if used without respect for safety measures. This information was not always kept in mind during operations in nuclear power plants. Our review will not focus on the laboratory experiments and failures, not because they are assumed to be insignificant, but due to the fact that the aftermath of these events does not have a catastrophic impact to the environment, society and economy compared to the industrial accidents analyzed in this review. However, these laboratory failures should teach us the lessons of preparedness and respect for nuclear energy in order to avoid or at least minimize the possibility of the occurrence of industrial disasters.

Before continuing further to the brief description of the most known events, we believe that is relevant to describe the International Nuclear and Radiological Event (INES) scale established in 1990 by the International Atomic Energy Agency (IAEA). As described on the IAEA official website, all members of IAEA make use of the 1-to-7 scale when a nuclear event occurs in order to describe its severity, with 1 being the lowest value and referring to a simple anomaly, and 7 being the worst instance of nuclear disaster and officially referred to as a "major accident".

As seen in Figure 1, values 1 to 3 indicate incidents with not so severe impacts on the environment and society in total, while values 4 to 7 indicate nuclear accidents that are catastrophic in most case outcomes. Although the scale was created in 1990, after the occurrence of some of the greatest nuclear accidents, scientists estimated the given INES Scale score for those events as well.

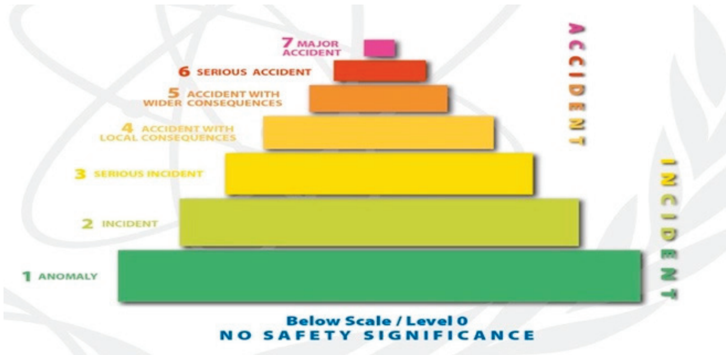


Figure 1. International Nuclear and Radiological Event (INES) Scale. Source: International Atomic Energy Agency official website—https://www-news.iaea.org/InesScale.aspx (accessed on 5 May 2022).

2.6.1. Three Mile Island (1979)

Perrow (1981), just two years after the Three Mile Island event, mentioned that accidents are likely to happen; however, factors that can eliminate their occurrence or impact do exist. The partial meltdown of the Three Mile Island Unit 2 reactor (TMI 2) occurred on 28 March 1979, in Pennsylvania, USA, causing a nuclear accident with high levels of radiation diffusion that experts categorized as a level-5 nuclear accident on the INES scale. As Perrow (1981) mentions, although some industrial accidents can be predicted and their occurrence can be prevented, Three Mile Island was not one of them. The literature tended to describe this event as a "normal accident"; however, Hopkins (2001) criticizes Perrow's "normal accident theory" by explaining that major accidents tend to occur due to management failures, and more specifically due to the lack of the communication of information. More specifically, Hopkins (2001) emphasizes the fact that Perrow's theory was endeavoring to shift the blame away from front line operators.

2.6.2. Chernobyl (1986)

The worst disaster in nuclear history occurred on 26 of April 1986. As Jaworowski (2010) states, the Chernobyl disaster is believed to be the worst possible catastrophe since the advent of nuclear energy. As the records mention, it was the early hours of 26 April when Chernobyl personnel ran a safety test on a flawed reactor design. The fact that this specific test was held during the night shift with insufficiently trained personnel, together with human errors and inaccurate decisions, led to the misuse of the No. 4 reactor, whose power dropped unexpectedly, causing consecutive failures and finally the reactor's shutdown. This moment was the beginning of the deadliest and most catastrophic nuclear disaster in history so far. The explosion following this event destroyed a huge part of the nuclear power plant, and radiation was diffused uncontrollably. The International Atomic Energy Agency (IAEA) categorized this catastrophe as a level-7 major accident on the INES scale.

The Vladimir Lenin Nuclear Power Plant, or Chernobyl Nuclear Power Plant as it is commonly known, is located close to the city of Pripyat, Ukraine, close to the Belarus-Ukraine border (16 km), and 100 Km north of Kyiv, the capital city of Ukraine. After this devastating accident, Pripyat turned into an abandoned city due to high levels of radiation and contamination. It is firmly believed that neighbor cities and countries have faced the negative impact of this disaster mainly due to transboundary pollution. Jaworowski (2010), illustrates that in the days following the accident, the recorded concentration of radiocesium over Poland was 2% to 6% of the maximum level of at the ground level, where the maximum equals 36.1 mBq/m^3. The outcomes of the Chernobyl disaster are uncountable. Not only was it enormous economic disaster due to the collapse of a great part of the reactor, but the whole NPP was abandoned and therefore out of order, which led to a reduction in electricity production. The government had to face the evacuation and relocation of the population from the hazardous areas. Both the health and environmental impacts were tremendous. Jaworowski (2010) examined the cancer risk factor of the nuclear workers, not only in Chernobyl but also in other countries such as Canada, the USA and the United Kingdom. However, it is crucial to mention that the level of cancer in the general population increased in the neighbor areas of the accident due to the uncontrollable diffusion. The most commonly observed cancer in those areas, as an aftermath of the accident, was thyroid cancer and leukemia, and was mainly diagnosed in children (Moysich et al. 2002).

2.6.3. Fukushima (2011)

In contrast to other nuclear disasters, the Fukushima Daiichi Nuclear Power Plant disaster was not an accident caused due to human errors or misconceptions. More specifically, on 11 March 2011, a 9 MW earthquake close to the north-eastern coast of Honshu (Japan) triggered a series of large tsunami waves which devastated all regional coastal areas (Kenta and Managi 2016; Managi and Guan 2017; Behling et al. 2019). As Halkos and Zisiadou (2020b) state, this series of events led to a shutdown of 11 nuclear power plants. Regarding the Fukushima Power Plant, a shutdown of its cooling system due to the flood of the tsunami waves caused the most catastrophic nuclear accident since Chernobyl. The International Atomic Energy Agency (IAEA) initially categorized this catastrophe as a level-5 accident on the INES scale. However, after reassessment, the score was upgraded to level-7 (Halkos and Zisiadou 2020b), a score that to that point had only been given to the Chernobyl disaster (Norio et al. 2011).

2.7. Nuclear Weapons

Nuclear energy can be a blessing or a curse, depending on the way someone uses it. As we have already mentioned, science and scientific achievements are not dangerous for humanity. What makes them dangerous and sometimes catastrophic, however, is the way humans handle them. In other words, the misuse, the overuse or the inappropriate way of adopting and using scientific achievements is what may cause fatalistic consequences. An example of such a case is the use of atomic bombs, also known as nuclear weapons. Einstein's famous equation regarding energy ($E = mc^2$) can turn into the deadliest weapon in the wrong hands.

2.7.1. Nuclear Weapons Used in Military Attacks

Following the latest news regarding the Russo-Ukrainian war, we would like to present the first and only two uses of nuclear weapons as a reminder of. Although nuclear energy has been described above as a source of electricity production, with its benefits and drawbacks, the real case was not always like that. More specifically, during World War II, the world witnessed the first and last time that military authorities decided to use the power of nuclear science as a form of attack. These two attacks became the sharpest memory of World War II, and passed from generation to generation as the most catastrophic military missions. The cities of Hiroshima and Nagasaki became globally famous overnight as the victims of these American attacks.

Hiroshima (1945)

Hiroshima was the first target city of the atomic bombing. As Zolberg (1998) mentions, on 6 August 1945, the American Boeing B-29 Superfortress, also known as the Enola Gay, flown by Paul Tibetts, dropped the first nuclear weapon above Hiroshima. The atomic bomb was called "Little Boy", with a blast yield equal to 15 Kilotons (Kt), and caused more than 90,000 fatalities.

Nagasaki (1945)

Similar to the Hiroshima case, another Japanese city had to face a nuclear attack three days after the initial mission. Specifically, another American Boeing B-29 Superfortress, known as Bockscar, flown by Charles Sweeney, attacked Nagasaki on 9 August 1945. The atomic bomb was called "Fat Man", with a blast yield equal to 21 Kilotons (Kt), and caused more than 39,000 fatalities.

2.7.2. Nuclear Weapons under Testing

As can be seen in Figure 2, "Little Boy" and "Fat Man" were not the only two nuclear weapons created in history. However, the rest of the weapons were not used during warfare but were only tested from laboratories. The representation in Figure 2 was created by Los Alamos National Laboratory, which is the laboratory that designed "Little Boy" and "Fat Man". Since the last atomic bombing, three more confirmed tests were run globally, the "Castle Bravo" in 1951, the "Tsar Bomba" in 1961 and the Licorne in 1970.

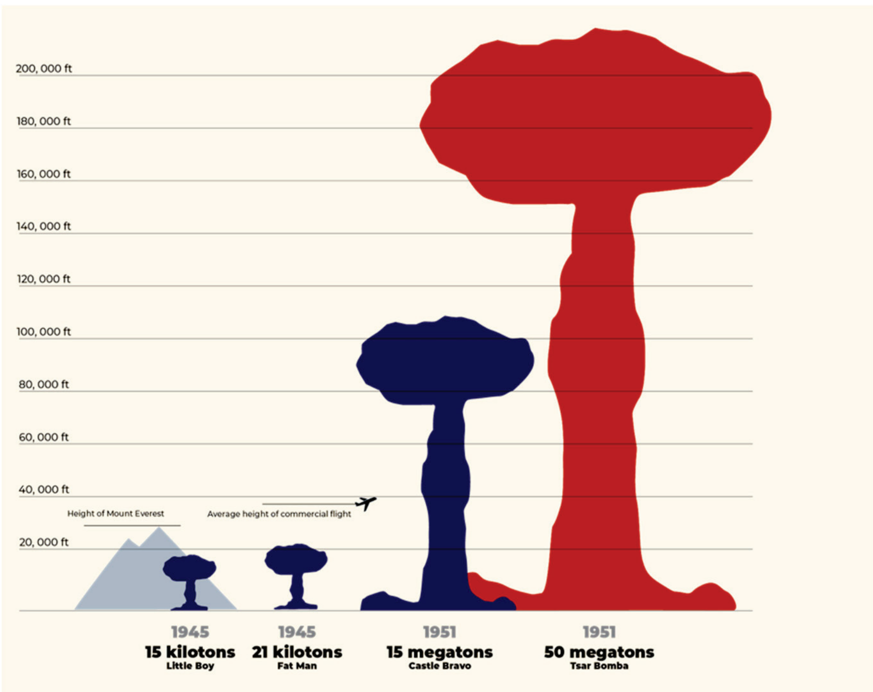


Figure 2. Nuclear Weapons Comparison. Source: Los Alamos National Laboratory (LANL) https://www.lanl.gov/discover/publications/national-security-science/2020-summer/vela-feature.shtml (accessed on 5 May 2022).

In 1951, just 6 years after the atomic bombing during World War II, the US Government decided to create and test another nuclear weapon under the name of "Castle Bravo", which was several times more powerful than "Little Boy" and "Fat Man". As reported, the "Castle

Bravo" blast yield equalled 15 Megatons (Mt). Following the American testing, Russia, then the Soviet Union, launched and tested its own nuclear weapon in 1961, under the name "Tsar Bomba", meaning the King of Bombs, with a blast yield equal to 50 Megatons (Mt), which was the greatest nuclear weapon created to date. Another test that has been ran and not represented in Figure 2 was the French nuclear weapon tested in 1970, under the name of "Licorne", with a blast yield equal to 1 Megaton (Mt), significantly smaller compared to previous tests.

Brook et al. (2014) in their paper make a reference to nuclear weapons. More specifically, they explain that in order for a country or industry to produce nuclear weapons, they should ensure access to the main weapons-grade materials of high purity such as the isotope Pu-239 or the isotope U-235. Regarding the isotope Pu-239, this element is obtained by the irradiation of U-238, while the isotope U-235 is produced by enrichment from the mined natural uranium. Knowing that the availability of those chemical elements is unlimited in nature, we can assume that the production of those weapons is a simple task for countries that already have the know-how, the facilities and the technology required. In other words, countries that have nuclear power plants can easily produce nuclear weapons. The danger that is hidden behind such actions and the increased fear and uncertainty regarding nuclear weapons after the Hiroshima and Nagasaki cases triggered powerful countries to establish an anti-nuclear weapon treaty.

2.7.3. Non-Proliferation Treaty

The Treaty on the Non-Proliferation of Nuclear Weapons (NPT) was initially proposed and signed in 1968, having as its main goal the prevention of nuclear weapons proliferation around the world. Currently, most countries have signed the Non-Proliferation Treaty, which for them is a commitment to reducing the production, spread and use of nuclear weapons (Brook et al. 2014). The responsibility of ensuring that the Treaty will be respected and followed by all members was assigned to the United Nations. More specifically, the International Atomic Energy Agency (IAEA) is responsible for the verification of adherence by all member states to the Non-Proliferation Treaty. As Brook et al. (2014) mention, no production of weapons-grade materials is observed globally, with the only exception being the dual-purpose RBMK-type reactors that have been constructed in the Soviet Union. The Non-Proliferation Treaty consists of 11 Articles (Article I–Article XI) which promote global safety regarding the nuclear weapons. As the Treaty mentiones in Article X, paragraph 2, with the completion of the first 25 years of its effectiveness, a conference should be held in order to evaluate the progress of the Treaty and decide on its continuation, a conference that the United Nations held and where the continuation of the Treaty was agreed upon by the parties involved.

3. The Energy Scheme of the New Millennium

The global community is watching the warfare in Europe between Russian Federation and Ukraine with intense concern since, in addition to the stakes of democracy, it has affected the global economy and the production and supply of energy. Russia is one of the main energy suppliers in terms of natural gas, a source that is used in electricity production as well as a heating fuel in many countries across the world. Moreover, the production line relies on electricity, which nowadays became a mercurial component due to the occurring limitations.

Over the last years, countries across the world, taking into consideration climate change as well as the reduction of the inventory of fossil fuels, have attempted to adopt alternative and specifically renewable forms of energy production. The most common sources of energy have already been mentioned in previous sections. Based on available data, we are aiming to illustrate the current situation since the beginning of the new millennium by using the reported values by BP's *Statistical Review of World Energy* (BP 2022).

Methodology

Before we analyze the results presented in Tables A1 and A2, we need to clarify the methodology used in order to create the tables presented below. Tables present the total values of energy and electricity production globally per year. We calculated the share of production of each type of energy and electricity source by dividing the total value (last column of each table) by the value of each type of energy and electricity source. The summations per year and the percentages based on the total energy mix were calculated by the authors. The number of countries using each specific form of energy source were calculated by the authors simply by subtracting the zero filled countries of the total 79 countries of the sample regarding the energy mix and the 207–209 countries regarding the electricity production.

Initially, we observed the energy mix used since 2000 on an annual basis for the reported countries. Table A1 (Appendix A) presents the summations per energy source per year over the last 22 years (2000–2021). As we can observe, nine different possible energy sources are reported, including biofuels, solar, wind and hydroelectric power, nuclear energy, gas, coal, oil, and finally geothermal, biomass, and other energy sources. As is mentioned in the BP database (2022), the values reported refer to TWh. Following the note of Table A1 (Appendix A), we will initially discuss the results of the total energy mix per year due to the fact that the records in brackets are based on these values. First and foremost, the total energy mix per year is the summation of all forms of energy sources of all 79 included countries. What is illustrated is the fact that, since the beginning of the new millennium and over time, the total amount of energy needed, produced and consumed by the nations globally is steadily increasing. More specifically, there has been a 48.61% increase in the total energy mix over the last 22 years, reaching 158,160.8 TWh of energy in Total. This significant increase indicates the excessive need of energy by society, which can lead us to worst case scenarios regarding the latest energy crisis observed due to the aforementioned warfare. The world demands greater amounts of TWh year by year, amounts that may not be able to be produced due to the lack of natural resources such as gas or fossil fuels.

Another aspect that is crucial to take into consideration is the sources from which all this energy amount is produced. In 2000, the three main sources of energy mix were gas, coal and oil, which covered the 22.131%, 25.364% and 38.529% of total energy mix, respectively. More specifically, in total, 91,551.77 TWh were produced based on these three energy sources, when the total amount of energy produced globally was reaching the 106,425.3 TWh. The majority of the countries examinedwere basing their energy on those sources, whereby 76 countries were using gas, 71 were using coal, and all 79 were using oil. Regarding the share of energy coming from the renewable energy forms including solar, wind and hydroelectric power in 2000, the values reported were significantly low. Only 6.825% of the total energy produced came from solar (0.003%), wind (0.087%) and hydroelectric (6.735%) power. Although 70 countries have already adopted the hydroelectric power as an alternative source of energy, only 7167.738 TWh were produced globally, while solar and wind power were used only by 28 and 39 countries, respectively, with their share of the total energy mix being dramatically low.

As the task here is to review the case of nuclear energy, in 2000, nuclear energy was adopted by 30 countries, providing 7317.65 TWh in total, reaching the share of 6.876% of the total energy mix. The number of countries trusting nuclear energy is low compared to other sources, which have been proven to be pollutants such as coal. A possible reason for the non-use of nuclear energy may be assumed to be the negative theory regarding the nuclear accidents. Indeed, history has shown us that nuclear accidents can be deadly and disastrous for the economy, the society and the environment. However, it is important to weigh the drawbacks of other sources of the energy mix. Oil spills cause great environmental disasters as well; however all 79 countries included in BP's review use oil as one of the main sources of energy. The production of energy using coal has been proven to be harmful both for the environment and human life, and today the world still depends on coal for a significant part of its energy production.

The Energy Mix scheme has drastically changed in 2021. The three main sources of energy mix were gas, coal and oil, which accounted for 24.834%, 27.743% and 30.648% of the total energy mix, respectively. More specifically, in total, 131,629.22 TWh were produced based on these three energy sources, when the total amount of energy produced globally was reaching 158,160.8 TWh. The majority of the countries examined were basing their energy on those sources, and 76 countries were using gas, 75 were using coal, and all 79 were using oil. Regarding the share of energy coming from the renewable energy forms including solar, wind and hydroelectric power in 2021, the values reported are still significantly low, however, there has been a noticeable increase. Only 11.192% of the total energy produced came from solar (1.685%), wind (3.049%) and hydroelectric (6.458%) power. What is really important to mention, however, is the number of countries adopting renewable sources of energy. Compared to 2000, in 2021 the majority of countries have turned to more ecofriendly sources of energy. Currently, solar power is used in 78 out of 79 countries included in the research, indicating a sharp increase compared to the 28 countries in 2000. Similarly, wind power is used in 70 out of 79 countries, where in 2000 only 39 countries were using wind power as part of their energy mix.

The energy produced by the renewable forms has significantly increased. Solar power in 2000 was producing only 3.11549 TWh globally, while in 2021 the energy produced by solar power equals 2664.65 TWh, giving 854.28 times more energy production based on solar power. Similarly, wind power produced 51.35 times more energy in 2021, while hydroelectric power produced 0.42 times more energy. However, regarding the low percentage change of hydroelectric power, we need to take into consideration the fact that solar and wind power were established in many countries as new sources of energy, while hydroelectric power was already prevalent as a source of energy. These changes indicate that communities around the world have started changing, or at least enhancing the energy mix with ecofriendly options, taking into account the climate change phenomenon and its impacts to the environment.

Moving forward, as of 2021 nuclear energy has been adopted by 32 countries, providing 7026.322 TWh in total, reaching the share of 4.443% of the total energy mix. Although there was a slight increase in the number of countries using nuclear power, from 30 to 32 countries, what is observable is the fact that the TWh produced globally in 2021 were 291.33 TWh less than those produced in 2000. More specifically, the global production of nuclear energy in 2021 was 3.98% lower compared to the 2000 values. This decrease may indicate a possible hesitancy with regard to the use of nuclear energy.

Using the same Review by BP (2022) and the data published, it is feasible to display the existing condition regarding the electricity production specifically and not just with regard to the energy mix in general. The difference between those two terms is the fact that electricity constitutes one of the three components of energy, while transport and heating constitute the other two components.

When facing energy crises similar to the one that is currently occurring in the world, it is crucial to know the share of electricity, and not only energy, produced by each source. This knowledge will give us the needed information to create a plan regarding each country's autonomy.

Table A2 (Appendix A) presents the summations per electricity source per year over the last 21 years (2000–2020). As we can observe, eight different possible electricity sources are reported, including solar, wind and hydroelectric power, nuclear energy, gas, coal, oil, and finally renewables including bioenergy and other electricity sources. As is mentioned in the BP database (2022), the values reported refer to TWh. Following the note of Table A2 (Appendix A), we will initially discuss the results of the total electricity per year due to the fact that the records in brackets are based on these values. First and foremost, the total electricity per year is the summation of all forms of energy sources of all included countries. What is illustrated is the fact that, since the beginning of the new millennium and over time, the total amount of electricity needed, produced and consumed by nations globally is steadily increasing. More specifically, there has been a 74.30% increase in total electricity over the last 21 years, reaching 26,106.839 TWh of electricity in total. This significant increase indicates the excessive need for electricity by society, which can worry us once again about the worst case scenarios regarding the latest energy crisis observed due to the present warfare that is taking place.

Figure 3 shows a stacked bar graph which has been constructed in an attempt to visualize the evolution of the global energy mix per year since the beginning of the new millennium. It has been created using the share of energy coming from each specific source to the total energy produced annually.

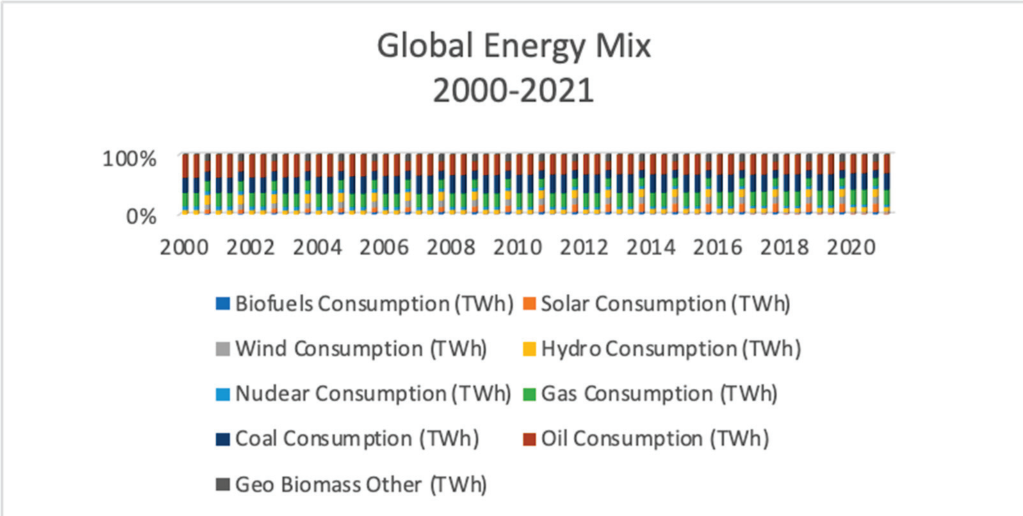


Figure 3. Global Energy Mix (2000–2021). Constructed by the authors.

Another aspect that is crucial to take into consideration is the source from which all of the total electricity amount is produced. In 2000, the main sources of electricity were coal and gas, and the three most common electricity sources were hydroelectric power and oil, which which accounted for 38.153%, 17.969% and 17.527% of total electricity, respectively. More specifically, in total, 11,031.31 TWh were produced based on these three electricity sources, when the total amount of electricity produced globally was reaching 12,978.305 TWh. The majority of the countries studied were basing their energy on those sources, whereas 91 countries were using gas, 70 were using coal, and 146 were using hydroelectric power. Regarding the share of energy coming from the renewable energy forms including solar and wind power in 2000, the values reported were significantly low. Only 0.215% of the total energy produced came from solar (0.007%) and wind (0.208%) power. Another source that has not been discussed yet regarding electricity production is oil.

In contrast to the energy mix analyzed above, oil as a form of electricity source is not a common component. More specifically, in 2000, 175 countries globally were using oil in the electricity production process; however, only 1191.85 TWh were being produced using this source, accounting for a share of 7.957% of total electricity. Moving forward, we once again are aiming to separately discuss the case of nuclear energy. In 2000, the electricity production of nuclear energy was adopted by 31 countries providing 2505.93 TWh in total, reaching the share of 16.730% of total electricity. Similar to the energy mix, the number of countries trusting nuclear energy is low compared to other sources that have been proven to be pollutants, such as coal.

Following the same path of the energy mix scheme, electricity production has drastically changed in 2020. The three main sources of electricity were coal, gas and hydroelectric power, which were covering 35.296%, 23.047% and 16.516% of total electricity, respectively. More specifically, in total, 19,543.22 TWh were produced based on these three electricity sources, while the total amount of electricity produced globally was reaching 26,106.839 TWh. With regard to the share of energy coming from the renewable energy forms including solar and wind power in 2020, the values reported are still significantly low; however, there has been a noticeable increase. Only 9.276% of the total energy produced came from solar (3.197%) and wind (6.079%) power. What is really important to mention, however, is the number of countries adopting renewable sources of energy. Compared to 2000, in 2020 the majority of countries have turned to more ecofriendly sources of energy. Currently, solar power is used in 151 out of 209 countries included in the research, indicating a sharp increase compared to the 15 out of 207 countries in 2000. Similarly, wind power is used in 105 out of 209 countries, where in 2000 only 36 out of 207 countries were using wind power as part of their electricity production.

The electricity produced by the renewable forms has significantly increased. Solar power in 2000 was producing only 1.08 TWh globally, while in 2020 the electricity produced by solar power equalled to 834.63 TWh, providing 771.81 times more electricity production based on solar power. Similarly, wind power produced 49.93 times more energy in 2020. Once again, these changes indicate that communities around the world have started changing, or at least enhancing the energy mix with ecofriendly options, taking into account the climate change phenomenon and its impacts to the environment.

Moving forward, in 2020 energy production resulting from nuclear energy was adopted by 33 countries, providing some 2635.81 TWh in total, accounting for 10.096% of total electricity. Although there was a slight increase in the number of countries using nuclear power, from 31 to 33 countries, what is observable is the fact that the share of electricity produced by nuclear energy globally in 2020 is lower compared to the share from 2000.

Figure 4 is a stacked bar graph which has been constructed in an attempt to visualize the evolution of global electricity per year since the beginning of the new millennium. As is obvious, it has been created using the share of energy coming from each specific source to the total energy produced annually.

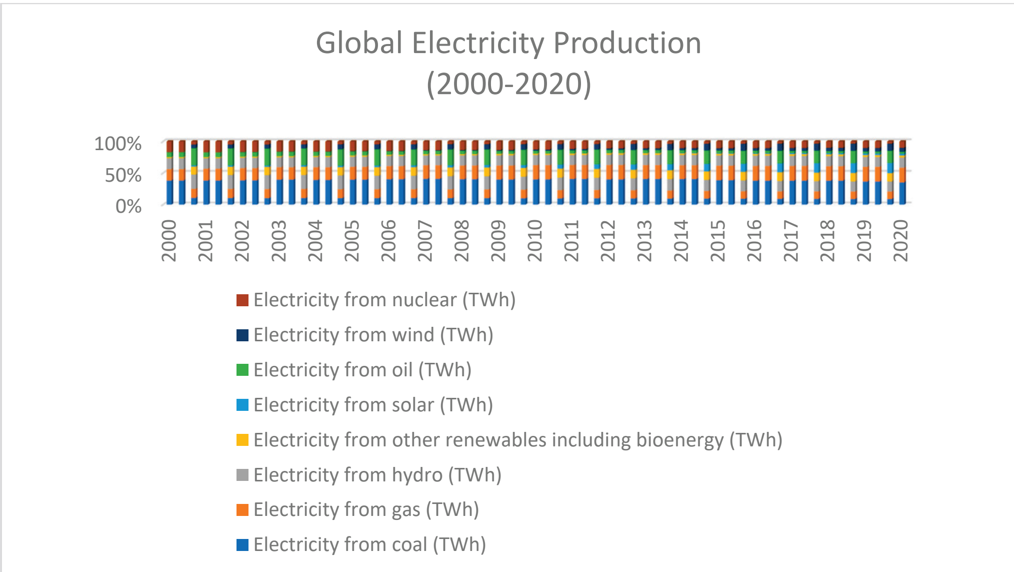


Figure 4. Global Electricity Production (2000–2020). Constructed by the authors.

4. Conclusions

The purpose of this paper was to review the case of nuclear energy and to discuss the issue of coping with energy crisis risks. More specifically, apart from presenting the basics of nuclear energy and nuclear reactors, we attempted to compare this source of electricity with other renewable energy forms such as solar power, wind power and hydroelectric power. It has been seen that, over the last years, there has been an increase in reactor construction globally. Having as an example the countries that gain the majority of their electricity from nuclear power plants, such as France, even more countries have decided to invest in this source of energy.

Supporters of nuclear energy believe that this source of power can be the best solution for sustainable energy production as well as climate change. The fact that during nuclear power production there are no emissions of greenhouse gases and carbon dioxide makes this source of energy a clean and eco-friendly option. The factor that supports its sustainability is uranium, the fuel used during production, which apart from being inexpensive is also unlimited in nature. Renewable energy forms (solar, wind and hydroelectric power) may be great options regarding eco-friendly energy production, but they unfortunately appear to be unsustainable due to their limited availability, making them likely unable to meet daily energy demands.

Nuclear energy, on the other hand, requires expensive investments in equipment and facilities, as well as expensive methods of waste management and disposal. The fact that the waste produced is radioactive and remains radioactive for hundreds of thousands of years, raising great concern about the negative impact that this may have to human beings and/or the environment with regard to potential cases of misuse. Another drawback of nuclear power plants is the fact that they are extremely vulnerable to natural environmental hazards such as earthquakes and tsunamis. The case of the Fukushima Daiichi Nuclear Power Plant Disaster of 2011 serves as a reminder of this.

Regarding the increase of income and the increase of environmental degradation up to a specific level of income and the decrease of environmental pollution afterwards, as proposed by the Environmental Kuznets Curve hypothesis, this has been proven to be accepted as the case in some countries (Halkos 1996, 2003, 2013). Moreover, some countries have proven that there is a causal relationship between economic growth and energy consumption. However, this has not been proven in all cases.

As is presented in the paper, since the beginning of the new millennium the scheme of energy production and electricity production appears to have changed drastically. By using the available data reported by BR, we illustrated that the production of energy and electricity has increased over the last 22 years (2000–2021) due to the excessive demand; however, what is more important to mention is the share of both electricity and energy derived from renewable forms such as solar, wind and hydroelectric power. It has been shown that more and more countries have adopted those sources of energy compared to previous decades.

What is important to examine, however, is whether nuclear energy appears to be trustable as a source of energy production by the nations of the world. As we have seen above, over the last 22 years there has been a slight increase in the number of nations trusting nuclear sources of energy both in terms of energy production and electricity production. Nuclear energy has been adopted by more countries over the last two decades, yet regarding the electricity production only 33 out of the 209 countries included in BP's sample use nuclear energy in order to produce electricity. These values indicate that it is probable that most countries either do not trust this source of energy or there may be an economic reason behind this situation. Knowing that nuclear power plants are an expensive investment, countries may avoid the investment in these facilities. Another possible scenario may be the combination of those two cases, as both the expense of investment and the risk of nuclear power plants may have led to a slight increase of nuclear energy adoption globally. Regarding renewable forms of energy, it has been proven that more and more nations around the world trust and adopt these sources in order to cover a part of their needed energy and electricity production. There has been a sharp increase in the countries creating power plants for renewable sources of energy such as hydroelectric power, solar and wind power. There has been a reported increase in their share of global production. To the best of our knowledge, there is no evidence as to whether there is a better source of energy when comparing the reported forms.

Lastly, when discussing the case of nuclear energy, we should of course mention all the aspects connected to the subject, but we have to always remember the nuclear accidents, which, although they may not happen frequently, the results are disastrous for the economy, the environment and human health. They can either be caused by human error or by acts of nature, but the aftermath is always the same, leaving devastated and abandoned regions with people suffering from chronic illnesses and probable increases in cancer rates, if not immediate death.

Regarding further research, we would like to mention that it is crucial to examine which alternative sources have proven to provide sufficient energy for peak demands and what their impact to the environment is based on their production process. Moreover, it is important to investigate and analyze the taxonomy regulations established by countries and organizations and provide evidence whether nuclear energy and renewable energy sources can be a possible solution to the European or global energy crisis, as well as to use those forms as an instrument to enhance climate protection and provide greener alternatives for a more prosperous future.

Funding: The research received no external funding.

Data Availability Statement: Data available on https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Energy Mix by source (all in TWh).

World	Biofuels	Solar	Wind	Hydro	Nuclear	Gas	Coal	Oil	Geo Biomass Other	Total Energy Mix
2000	114.3912 [0.107%] 8	3.115499 [0.003%] 28	92.11011 [0.087%] 39	7167.738 [6.735%] 70	7317.65 [6.876%] 30	23,552.75 [22.131%] 76	26,994.11 [25.364%] 71	41,004.91 [38.529%] 79	178.5533 [0.168%] 48	106,425.3
2001	110.8154 [0.103%] 8	4.155243 [0.004%] 29	111.9642 [0.104%] 43	6932.966 [6.456%] 70	7475.607 [6.961%] 30	23,839.47 [22.198%] 76	27,420.84 [25.533%] 74	41,312.93 [38.469%] 79	185.3455 [0.173%] 51	107,394.1
2002	140.2483 [0.128%] 13	5.221885 [0.005%] 31	151.7197 [0.138%] 44	6999.372 [6.381%] 70	7545.317 [6.878%] 30	24,529.58 [22.361%] 76	28,508.64 [25.988%] 75	41,617.19 [37.938%] 79	199.8919 [0.182%] 52	109,697.2
2003	159.5752 [0.140%] 14	6.500787 [0.006%] 31	182.5275 [0.161%] 46	6923.877 [6.095%] 70	7345.786 [6.467%] 30	25,198.69 [22.183%] 76	31,040.4 [27.326%] 74	42,525.45 [37.436%] 79	211.0865 [0.186%] 52	113,593.9
2004	191.2764 [0.160%] 17	8.510881 [0.007%] 35	245.5105 [0.206%] 52	7409.625 [6.216%] 70	7630.364 [6.401%] 30	26,148.16 [21.935%] 76	33,196.87 [27.848%] 74	44,149.89 [37.036%] 79	227.9709 [0.191%] 55	119,208.2
2005	230.3407 [0.187%] 19	11.90605 [0.010%] 37	298.5086 [0.243%] 53	7617.337 [6.192%] 70	7600.611 [6.179%] 30	26,850.21 [21.828%] 76	35,695.61 [29.018%] 74	44,458.48 [36.142%] 79	247.37 [0.201%] 55	123,010.4
2006	288.3931 [0.228%] 22	16.28421 [0.013%] 38	378.4353 [0.299%] 57	7878.779 [6.227%] 70	7647.21 [6.044%] 30	27,560.27 [21.782%] 76	37,584.38 [29.704%] 75	44,911.86 [35.495%] 79	262.9176 [0.208%] 57	126,528.5
2007	379.0338 [0.291%] 22	21.95076 [0.017%] 40	483.5379 [0.371%] 59	7958.067 [6.099%] 70	7444.62 [5.706%] 30	28,659.4 [21.966%] 76	39,783.74 [30.492%] 75	45,456.76 [34.840%] 79	286.0014 [0.219%] 58	130,473.1
2008	510.4845 [0.387%] 22	35.42512 [0.027%] 45	620.8337 [0.471%] 60	8406.951 [6.375%] 70	7375.647 [5.593%] 30	29,357.17 [22.261%] 76	40,301.37 [30.560%] 75	44,961.1 [34.094%] 79	305.3945 [0.232%] 60	131,874.4

Table A1. *Cont.*

World	Biofuels	Solar	Wind	Hydro	Nuclear	Gas	Coal	Oil	Geo Biomass Other	Total Energy Mix
2009	579.0933 [0.446%] 23	58.49998 [0.045%] 52	770.5533 [0.594%] 61	8330.405 [6.419%] 70	7226.256 [5.568%] 30	28,762.58 [22.163%] 76	39,744.99 [30.626%] 74	43,977 [33.886%] 79	327.9855 [0.253%] 62	129,777.4
2010	664.4033 [0.488%] 24	93.42596 [0.069%] 59	958.3268 [0.704%] 61	8698.284 [6.393%] 70	7367.101 [5.414%] 29	30,885.52 [22.699%] 76	41,565.2 [30.548%] 75	45,463.5 [33.413%] 79	369.0992 [0.271%] 62	136,064.9
2011	696.9326 [0.500%] 24	179.6672 [0.129%] 66	1210.853 [0.868%] 62	8821.638 [6.327%] 70	7015.404 [5.031%] 30	31,631.98 [22.685%] 76	43,581.74 [31.255%] 75	45,909.24 [32.925%] 79	390.4146 [0.280%] 63	139,437.9
2012	717.1511 [0.508%] 24	276.3121 [0.196%] 71	1448.305 [1.025%] 62	9155.181 [6.480%] 70	6494.912 [4.597%] 30	32,457.17 [22.972%] 76	43,761.25 [30.972%] 76	46,558.83 [32.952%] 79	422.5592 [0.299%] 63	141,291.7
2013	775.1474 [0.539%] 24	375.1109 [0.261%] 76	1722.897 [1.199%] 67	9435.65 [6.565%] 70	6507.32 [4.527%] 30	32,974.96 [22.942%] 76	44,430.93 [30.913%] 76	47,050.79 [32.736%] 79	456.5064 [0.318%] 65	143,729.3
2014	823.7569 [0.567%] 24	528.3576 [0.364%] 76	1898.331 [1.307%] 68	9668.559 [6.657%] 70	6600.271 [4.545%] 29	33,146.83 [22.823%] 76	44,727.26 [30.797%] 76	47,346.81 [32.600%] 79	493.7744 [0.340%] 65	145,233.9
2015	856.9625 [0.586%] 24	679.3229 [0.464%] 76	2216.853 [1.515%] 69	9608.963 [6.569%] 70	6648.553 [4.545%] 30	33,834.85 [23.130%] 76	43,615.22 [29.815%] 76	48,291.22 [33.012%] 79	532.091 [0.364%] 68	146,284
2016	893.1888 [0.602%] 24	865.4356 [0.584%] 77	2549.634 [1.720%] 67	9859.68 [6.650%] 69	6708.681 [4.525%] 30	34,611.59 [23.344%] 76	43,027.98 [29.021%] 76	49,210.73 [33.191%] 79	540.1048 [0.364%] 68	148,267
2017	912.9255 [0.605%] 24	1171.459 [0.776%] 78	3006.434 [1.991%] 67	9928.638 [6.577%] 69	6728.184 [4.457%] 30	35,571.81 [23.562%] 76	43,175.31 [28.599%] 76	49,898.69 [33.052%] 79	576.2781 [0.382%] 69	150,969.7
2018	983.5064 [0.634%] 24	1505.385 [0.971%] 78	3323.316 [2.144%] 68	10,106.86 [6.519%] 69	6850.746 [4.419%] 30	37,376.67 [24.110%] 76	43,720.89 [28.202%] 76	50,538.79 [32.600%] 79	621.7367 [0.401%] 69	155,027.9
2019	1049.003 [0.671%] 24	1829.434 [1.171%] 78	3699.318 [2.367%] 70	10,250.11 [6.558%] 69	7067.73 [4.522%] 30	38,058.26 [24.351%] 76	43,071.78 [27.559%] 76	50,613.4 [32.384%] 79	651.8196 [0.417%] 69	156,290.9
2020	996.0236 [0.664%] 24	2191.175 [1.461%] 78	4145.35 [2.765%] 69	10,492.24 [6.998%] 69	6782.355 [4.524%] 32	37,444.33 [24.975%] 76	41,311.4 [27.554%] 75	45,884.39 [30.604%] 79	681.3334 [0.454%] 69	149,928.6
2021	1065.024 [0.673%] 24	2664.65 [1.685%] 78	4821.799 [3.049%] 70	10,214.18 [6.458%] 69	7026.322 [4.443%] 32	39,277.3 [24.834%] 76	43,878.03 [27.743%] 75	48,473.89 [30.648%] 79	739.5754 [0.468%] 69	158,160.8

Note: The table illustrates the TWh produced by each specific form of energy source, the values in the brackets report the percentage of total energy produced by each specific form of energy source, while the values in the parentheses represent the number of countries each form of energy source by year. Regarding the sample, 79 countries were included in the sample by BP during the time span 2000–2021. The summations per year and the percentages based on the total energy mix were calculated by the authors. The number of countries using each specific form of energy source were calculated by the authors simply by subtracting the zero filled countries of the 79 countries of the sample.

Table A2. Sources of Produced Electricity (all in TWh).

World	Electricity from Coal	Electricity from Gas	Electricity from Hydro	Electricity from Other Renewables Including Bioenergy	Electricity from Solar	Electricity from Oil	Electricity from Wind	Electricityfrom Nuclear (TWh)	Total Electricity
2000	5714.63 38.153% 70	2691.5 17.969% 91	2625.18 17.527% 146	216.975 1.449% 79	1.08 0.007% 15	1191.85 7.957% 175	31.16 0.208% 36	2505.93 16.730% 31	14,978.305
2001	5800.45 38.168% 69	2830.28 18.624% 92	2564.48 16.875% 147	210.898 1.388% 82	1.35 0.009% 17	1179.25 7.760% 174	38.17 0.251% 41	2572.31 16.926% 31	15,197.188
2002	6055.75 38.494% 69	3034.61 19.290% 93	2603.17 16.547% 149	225.229 1.432% 83	1.69 0.011% 18	1158.97 7.367% 177	52.06 0.331% 43	2600.35 16.529% 31	15,731.829
2003	6461.59 39.660% 70	3163.44 19.417% 95	2606.15 15.996% 151	238.931 1.467% 85	2.07 0.013% 18	1180.67 7.247% 177	63.44 0.389% 43	2576.19 15.812% 31	16,292.481
2004	6686.57 39.150% 70	3400.97 19.913% 95	2798.14 16.383% 150	256.838 1.504% 86	2.71 0.016% 19	1167.61 6.836% 176	85.27 0.499% 52	2681.18 15.698% 31	17,079.288
2005	7030.91 39.621% 72	3573.12 20.135% 97	2902.91 16.359% 150	280.742 1.582% 91	3.78 0.021% 20	1164.64 6.563% 177	103.9 0.586% 55	2685.38 15.133% 31	17,745.382
2006	7427.3 40.244% 73	3791.98 20.546% 97	3005.2 16.283% 150	297.064 1.610% 92	5.11 0.028% 22	1076.33 5.832% 176	132.8 0.720% 56	2719.87 14.737% 31	18,455.654
2007	7919.28 40.943% 73	4114.69 21.273% 98	3049.54 15.766% 151	321.125 1.660% 95	6.94 0.036% 27	1094.31 5.658% 178	170.93 0.884% 58	2665.34 13.780% 31	19,342.155
2008	7915.68 40.311% 73	4217.95 21.480% 98	3230.7 16.453% 151	336.825 1.715% 97	11.38 0.058% 30	1049.16 5.343% 176	220.09 1.121% 58	2654.51 13.518% 31	19,636.295
2009	7807.02 39.942% 73	4255.26 21.771% 99	3232.81 16.540% 152	364.058 1.863% 99	19.21 0.098% 36	974.4 4.985% 177	275.9 1.412% 64	2617.32 13.391% 31	19,545.978
2010	8346.96 39.961% 74	4687.57 22.442% 99	3409.23 16.322% 151	408.989 1.958% 103	31.08 0.149% 51	970.96 4.648% 177	346.22 1.658% 68	2686.63 12.862% 30	20,887.639

Table A2. *Cont.*

World	Electricity from Coal	Electricity from Gas	Electricity from Hydro	Electricity from Other Renewables Including Bioenergy	Electricity from Solar	Electricity from Oil	Electricity from Wind	Electricityfrom Nuclear (TWh)	Total Electricity
2011	8807.69 40.702% 78	4773.18 22.058% 102	3476.16 16.064% 152	430.049 1.987% 105	61.93 0.286% 59	1074.46 4.965% 181	439.9 2.033% 69	2576.2 11.905% 31	21,639.569
2012	8827.26 39.961% 78	5011.8 22.688% 102	3641.36 16.484% 152	459.53 2.080% 105	95.43 0.432% 69	1122.18 5.080% 182	528.9 2.394% 75	2403.18 10.879% 31	22,089.64
2013	9284.01 40.838% 79	4912.86 21.611% 102	3768.39 16.576% 152	496.037 2.182% 110	131.45 0.578% 81	1081.27 4.756% 182	640.06 2.815% 81	2419.42 10.643% 31	22,733.497
2014	9453.32 40.576% 81	5047.91 21.667% 102	3860.08 16.568% 152	535.863 2.300% 112	195.9 0.841% 103	1019.78 4.377% 180	716.8 3.077% 91	2468.28 10.594% 30	23,297.933
2015	9134.76 38.765% 83	5379.57 22.829% 102	3869.91 16.423% 152	568.046 2.411% 117	254.23 1.079% 114	1027.27 4.359% 181	829.08 3.518% 96	2501.5 10.616% 31	23,564.366
2016	9186.42 37.965% 83	5625.7 23.250% 103	3999.9 16.531% 151	581.325 2.402% 119	328.48 1.358% 127	982.41 4.060% 181	959.41 3.965% 100	2533.12 10.469% 31	24,196.765
2017	9476.27 38.093% 84	5729.38 23.031% 105	4049.15 16.277% 151	608.876 2.448% 119	443.29 1.782% 140	884.9 3.557% 181	1136.94 4.570% 100	2548.19 10.243% 31	24,876.996
2018	9837.67 38.051% 83	5859.29 22.663% 108	4170.02 16.129% 150	654.59 2.532% 120	567.8 2.196% 145	876.2 3.389% 182	1268.43 4.906% 104	2620.13 10.134% 31	25,854.13
2019	9617.54 36.581% 84	6083.4 23.139% 108	4219.16 16.048% 151	688.912 2.620% 119	694.5 2.642% 151	843.58 3.209% 181	1420.08 5.401% 106	2723.79 10.360% 31	26,290.962
2020	9214.62 35.296% 83	6016.79 23.047% 108	4311.81 16.516% 154	712.899 2.731% 119	834.63 3.197% 151	793.34 3.039% 181	1586.94 6.079% 105	2635.81 10.096% 33	26,106.839

Note: The table illustrates the TWh produced by each specific form of electricity source, the values in the brackets report the percentage of total electricity produced by each specific form of electricity source, while the values in the parentheses represent the number of countries each form of electricity source by year. Regarding the sample, the provided dataset includes 207 countries for the period 2000–2004, 208 countries for the period 2005–2011, and 209 for the period 2012–2020 in the sample by BP during the time span 2000–2020. The summations per year and the percentages based on the total electricity were calculated by the authors. The number of countries using each specific form of energy source were calculated by the authors simply by subtracting the zero filled countries of the total countries of the sample per year.

References

Adamantiades, Achilles, and Ioannis Kessides. 2009. Nuclear power for sustainable development: Current status and future prospects. *Energy Policy* 37: 5149–66. [CrossRef]

Akcay, Behiye. 2009. The case of nuclear energy in Turkey: From Chernobyl to Akkuyu nuclear power plant. *Energy Sources Part B* 4: 347–55. [CrossRef]

Akyuz, Emrah. 2015. Turkey's Nuclear Energy Policy and the Terrorist Threat. *The Journal of Academic Social Science Studies* 40: 523–36.

Akyuz, Emrah. 2017. Advantages and disadvantages of nuclear energy in Turkey: Public perception. *Eurasian Journal of Environmental Research* 1: 1–11.

Alam, Md Kausar, Mosab I. Tabash, Syed Mabruk Billah, Sanjeev Kumar, and Suhaib Anagreh. 2022. The Impacts of the Russia–Ukraine Invasion on Global Markets and Commodities: A Dynamic Connectedness among G7 and BRIC Markets. *Journal of Risk and Financial Management* 15: 352. [CrossRef]

Alkhalidi, Ammar, Tuqa Alrousan, Manal Ishbeytah, Mohammad Ali Abdelkareem, and Olabi Abdul Ghani. 2022. Recommendations for energy storage compartment used in renewable energy project. *International Journal of Thermofluids* 15: 100182. [CrossRef]

Al-Mulali, Usama, Ilhan Ozturk, and Hooi Hooi Lean. 2015. The influence of economic growth, urbanization, trade openness, financial development, and renewable energy on pollution in Europe. *Natural Hazards* 79: 621–44. [CrossRef]

Al-Mulali, Usama, Sakiru Adebola Solarin, and Ozturk Ilhan. 2016. Investigating the presence of the environmental Kuznets curve (EKC) hypothesis in Kenya: An autoregressive distributed lag (ARDL) approach. *Natural Hazards* 80: 1729–47. [CrossRef]

Aneke, Mathew, and Meihong Wang. 2016. Energy storage technologies and real life applications—A state of the art review. *Applied Energy* 179: 350–77. [CrossRef]

Apergis, Nicholas, and James Earl Payne. 2010. A panel study of nuclear energy consumption and economic growth. *Energy Economics* 32: 545–49. [CrossRef]

Aras, Erol. 2013. The necessity of nuclear energy in Turkey: A comparison with hydropower energy. *Energy Sources, Part B: Economics, Planning, and Policy* 8: 107–14. [CrossRef]

Bagher, Amina M., Mazian Abdul Vahid, and Mohsen Mirhabibi. 2014. Geothermal energy. *Journal of Engineering and Technology Research* 6: 146–50.

Barros, Carlos Pestana, and Shunsuke Managi. 2016. French nuclear electricity plants: Productivity and air pollution. *Energy Sources, Part B: Economics, Planning, and Policy* 11: 718–24. [CrossRef]

Barthelmie, Rebecca J., and Sara C. Pryor. 2021. Climate change mitigation potential of wind energy. *Climate* 9: 136.

Basit, Muhammad Abdul, Saad Dilshad, Rabiah Badar, and Syed Muhammad Sami ur Rehman. 2020. Limitations, challenges, and solution approaches in grid-connected renewable energy systems. *International Journal of Energy Research* 44: 4132–62. [CrossRef]

Behling, Noriko, Mark C. Williams, Thomas. G. Behling, and Shunsuke Managi. 2019. Aftermath of Fukushima: Avoiding another major nuclear disaster. *Energy Policy* 126: 411–20. [CrossRef]

BP. 2022. Statistical Review of World Energy. Available online: https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html (accessed on 13 July 2022).

Bridgwater, Tony. 2006. Biomass for energy. *Journal of the Science of Food and Agriculture* 86: 1755–68. [CrossRef]

Brook, Barry W., Agustin Alonso, Daniel A. Meneley, Josef Misak, Tom Blees, and Jan B. van Erp. 2014. Why nuclear energy is sustainable and has to be part of the energy mix. *Sustainable Materials and Technologies* 1: 8–16. [CrossRef]

Cohen, Bernard L. 1983. *Before It Is Too Late*. New York: Springer Publishing.

Cooper, Mark. 2014. Small modular reactors and the future of nuclear power in the United States. *Energy Research & Social Science* 3: 161–77.

Corkhill, Claire, and Neil Hyatt. 2018. *Nuclear Waste Management*. Bristol: IOP Publishing, p. 1.

Coskun, Mehmet Burhanettin, and Banu Tanriover. 2016. An investigation on nuclear energy policy in Turkey and public perception. In *EPJ Web of Conferences*. Paris: EDP Sciences, vol. 128, p. 5001.

da Mata, Jonatas F., Rieder O. Neto, and Amir Z. Mesquita. 2017. Comparison of the Performance, Advantages and Disadvantages of Nuclear Power Generation Compared to Other Clean Sources of Electricity. Available online: https://inis.iaea.org/collection/NCLCollectionStore/_Public/49/009/49009723.pdf?r=1 (accessed on 5 May 2022).

DeCanio, Stephen. J. 2009. The political economy of global carbon emissions reductions. *Ecological Economics* 68: 915–24. [CrossRef]

Deutch, John, Ernest Moniz, Stephen Ansolabehere, Michael Driscoll, Paul Gray, John Holdren, Paul Joskow, Richard Lester, and Neil Todreas. 2003. The Future of Nuclear Power. An MIT Interdisciplinary Study. Available online: http://web.mit.edu/nuclearpower (accessed on 5 May 2022).

Driscoll, Michael J., and Pavel Hejzlar. 2005. Reactor physics challenges in Gen-IV reactor design. *Nuclear Engineering and Technology* 37: 1–10.

EIA (Energy Information Administration). 2006. CANDU Reactors. Available online: http://www.eia.doe.gov/cneaf/nuclear/page/nuc_reactors/china/candu.htmls (accessed on 5 May 2022).

Elliott, David, ed. 2016. *Nuclear Or Not?: Does Nuclear Power Have a Place in a Sustainable Energy Future?* New York: Springer.

Erdogdu, Erkan. 2007. Nuclear power in open energy markets: A case study of Turkey. *Energy Policy* 35: 3061–73. [CrossRef]

Fiore, Karine. 2006. Nuclear energy and sustainability: Understanding ITER. *Energy Policy* 34: 3334–41. [CrossRef]

Flynn, James, James Chalmers, Doug Easterling, Roger Kasperson, Howard Kunreuther, C. K. Mertz, Alvin Mushkatel, David K. Pijawka, Paul Slovic, and Lydia Dotto. 1995. *One Hundred Centuries of Solitude. Redirecting America's High-Level Nuclear Waste Policy*. Boulder: Westview Press.

Forsberg, Charles W. 2009. Sustainability by combining nuclear, fossil, and renewable energy sources. *Progress in Nuclear Energy* 51: 192–200. [CrossRef]

Franco, Sainu, Venkata Ravibabu Mandla, and K. Ram Mohan Rao. 2017. Urbanization, energy consumption and emissions in the Indian context. A review. *Renewable & Sustainable Energy Reviews* 71: 898–907.

Gunay, Defne, and Emre Iseri. 2017. Unexpected Persistence Amidst Enlargement Stasis: Usages of Europe in Turkey's Nuclear Energy Debate. *South European Society and Politics* 22: 101–19. [CrossRef]

Halkos, George, and Argyro Zisiadou. 2019. Examining the natural environmental hazards over the last century. *Economics of Disasters and Climate Change* 3: 119–50. [CrossRef]

Halkos, George, and Argyro Zisiadou. 2020a. An overview of the technological environmental hazards over the last century. *Economics of Disasters and Climate Change* 4: 411–28. [CrossRef]

Halkos, George, and Argyro Zisiadou. 2020b. Is investors' psychology affected due to a potential unexpected environmental disaster? *Journal of Risk and Financial Management* 13: 151. [CrossRef]

Halkos, George. 1996. Incomplete information in the acid rain game. *Empirica* 23: 129–48. [CrossRef]

Halkos, George. 2003. Environmental Kuznets Curve for sulfur: Evidence using GMM estimation and random coefficient panel data models. *Environment and Development Economics* 8: 581–601. [CrossRef]

Halkos, George. 2013. Exploring the economy–environment relationship in the case of sulphur emissions. *Journal of Environmental Planning and Management* 56: 159–77. [CrossRef]

Hamacher, Thomas, and John Sheffield. 2004. *Development of Fusion Power: What Role Could Fusion Power Play in Transitional and Developing Countries*. IPP Garching (with support from EFDA) and Joint Institute for Energy, 2004-04, IIP-Report Max-Planck-Institutfür Plasmaphysik. Available online: https://pure.mpg.de/rest/items/item_2136776/component/file_2136775/content (accessed on 5 May 2022).

Hansen, Kenneth. 2019. Decision-making based on energy costs: Comparing levelized cost of energy and energy system costs. *Energy Strategy Reviews* 24: 68–82. [CrossRef]

Heal, Geoffrey. 2009. Reflections—The economics of renewable energy in the United States. *Review of Environmental Economics and Policy* 4: 139–54. [CrossRef]

Hedenus, Fredrik, Azar Christian, and Daniel J. A. Johansson. 2010. Energy security policies in EU-25—The expected cost of oil supply disruptions. *Energy Policy* 38: 1241–50. [CrossRef]

Hempelman, Louis Henry, Clarence C. Lushbaugh, and George L. Voel. 1979. *What Has Happened to the Survivors of the Early Los Alamos Nuclear Accidents?* LA-UR-79-2802. Los Alamos: Los Alamos Scientific Laboratory.

Hench, Larry L., David E. Clark, and Alan B. Harker. 1986. Nuclear waste solids. *Journal of Materials Science* 21: 1457–78. [CrossRef]

Hopkins, Andrew. 2001. Was three mile island a 'Normal Accident'? *Journal of Contingencies and Crisis Management* 9: 65–72. [CrossRef]

Hossain, Emran, Rej Soumen, Saha Sourav Mohan, Onwe Joshua Chukwuma, Nwulu Nnamdi, Bekun Festus Victor, and Amjad Taha. 2022. Can energy efficiency help in achieving carbon-neutrality pledges? A developing country perspective using dynamic ARDL simulations. *Sustainability* 14: 7537. [CrossRef]

Immonen, Anne, and Maria Kopsakangas-Savolainen. 2022. Capturing Consumers' Awareness and the Intention to Support Carbon Neutrality through Energy Efficient Consumption. *Energies* 15: 4022. [CrossRef]

Irfan, Mohd, and Krishnendu Shaw. 2017. Modeling the effects of energy consumption and urbanization on environmental pollution in South Asian countries: A nonparametric panel approach. *Quality & Quantity* 51: 65–78.

Iwata Hiroki, Okada Keisuke, and Sovannoeun Samreth. 2010. Empirical study on the environmental Kuznets curve for CO_2 in France: The role of nuclear energy. *Energy Policy* 38: 4057–63. [CrossRef]

Jaworowski, Zbigniew. 2010. Observations on the Chernobyl Disaster and LNT. *Dose-Response* 8: 148–71. [CrossRef] [PubMed]

Jebli Mehdi Ben, Youssef Slim Ben, and Ilhan Ozturk. 2016. Testing environmental Kuznets curve hypothesis: The role of renewable and non-renewable energy consumption and trade in OECD countries. *Ecological Indicators* 60: 824–31. [CrossRef]

Jewell, Jessica. 2011. Ready for nuclear energy?: An assessment of capacities and motivations for launching new national nuclear power programs. *Energy Policy* 39: 1041–55. [CrossRef]

Kagel, Alyssa, Diana Bates, and Karl Gawell. 2005. *A Guide to Geothermal Energy and the Environment*; Washington: Geothermal Energy Association. Available online: www.geo-energy.org (accessed on 5 May 2022).

Kebede, Abraham Alem, Theodoros Kalogiannis, Van Mierlo Joeri, and Maitane Berecibar. 2022. A comprehensive review of stationary energy storage devices for large scale renewable energy sources grid integration. *Renewable and Sustainable Energy Reviews* 159: 112213. [CrossRef]

Keeney, Ralph L., and Detlof von Winterfeldt. 1994. Managing nuclear waste from power plants. *Risk Analysis* 14: 107–30. [CrossRef]

Kim, Taek K. 2013. Gen-IV Reactors. In *Nuclear Energy*. New York: Springer, pp. 175–201.

Kok, Besir, and Hüseyin Benli. 2017. Energy diversity and nuclear energy for sustainable development in Turkey. *Renewable Energy* 111: 870–77.

Korkmaz, Ozan, and Bihrat Önöz. 2022. Modelling the Potential Impacts of Nuclear Energy and Renewables in the Turkish Energy System. *Energies* 15: 1392. [CrossRef]

Kumar, Dinesh, Alam Syed Bahauddin, Ridwan Tuhfatur, and Cameron S. Goodwin. 2021. Quantitative risk assessment of a high power density small modular reactor (SMR) core using uncertainty and sensitivity analyses. *Energy* 227: 120400. [CrossRef]

Kurt, Ünal. 2014. The fuzzy TOPSIS and generalized Choquet fuzzy integral algorithm for nuclear power plant site selection—A case study from Turkey. *Journal of Nuclear Science and Technology* 51: 1241–55. [CrossRef]

Liu, Zhitao, and Jihong Fan. 2014. Technology readiness assessment of small modular reactor (SMR) designs. *Progress in Nuclear Energy* 70: 20–28. [CrossRef]

Lorusso, Pierdomenico, Serena Bassini, Del Nevo Alessandro, Di Piazza Ivan, Giannetti Fabio, Tarantino Mariano, and Marco Utili. 2018. GEN-IV LFR development: Status & perspectives. *Progress in Nuclear Energy* 105: 318–31.

Lutze, Werner. 1988. Silicate glasses. In *Radioactive Waste Form for the Future*. Edited by Lutze Werner and Rodney C. Ewing. Amsterdam: Elsevier Science Publishers B.V., pp. 1–159.

Macfarlane, Allison M., and Marvin Miller. 2007. Nuclear energy and uranium resources. *Elements* 3: 185–92. [CrossRef]

Malenfant, Richard E. 1996. *Lessons Learned from Early Criticality Accidents*. No. LA-UR–96-1659. Los Alamos: Los Alamos National Lab.

Managi, Shunsuke, and Dabo Guan. 2017. Multiple disasters management: Lessons from the Fukushima triple events. *Economic Analysis and Policy* 53: 114–22. [CrossRef]

Mastrocinque, Ernesto, F. Javier Ramírez, Honrubia-Escribano Andrés, and Duc T. Pham. 2020. An AHP-based multi-criteria model for sustainable supply chain development in the renewable energy sector. *Expert Systems with Applications* 150: 113321. [CrossRef]

McLaughlin, Thomas P., Shean P. Monahan, Norman L. Pruvost, Vladimir V. Frolov, Boris G. Ryazanov, and Victor I. Sviridov. 2000. *A Review of Criticality Accidents*. 2000 Revision, LA-13638. Los Alamos: Los Alamos National Laboratory.

Menyah, Kojo, and Yemane Wolde-Rufael. 2010. CO_2 emissions, nuclear energy, renewable energy and economic growth in the US. *Energy Policy* 38: 2911–15. [CrossRef]

Mourmouris, John C., and Constantinos Potolias. 2013. A multi-criteria methodology for energy planning and developing renewable energy sources at a regional level: A case study Thassos, Greece. *Energy Policy* 52: 522–30. [CrossRef]

Moysich, Kirsten B., Ravi J. Menezes, and Arthur M. Michalek. 2002. Chernobyl-related ionising radiation exposure and cancer risk: An epidemiological review. *The Lancet Oncology* 3: 269–79. [CrossRef]

Murty, K. Linga, and Indrajit Charit. 2008. Structural materials for Gen-IV nuclear reactors: Challenges and opportunities. *Journal of Nuclear Materials* 383: 189–95. [CrossRef]

NEA. 2005. *Projected Costs of Generating Electricity: Update 2005*. Paris: Publication of the Nuclear Energy Agency (NEA) to the Organisation for Economic Cooperation and Development (OECD) and International Energy Agency (IEA).

NEA. 2008. *Nuclear Energy Outlook*. No. 6348. Paris: OECD.

Norio, Okada, Tao Ye, Yoshio Kajitani, Peijun Shi, and Hirokazu Tatano. 2011. The 2011 eastern Japan great earthquake disaster: Overview and comments. *International Journal of Disaster Risk Science* 2: 34–42. [CrossRef]

Nuttall, William J. 2005. *Nuclear Renaissance: Technologies and Policies for the Future of Nuclear Power*, 1st ed. New York: Taylor and Francis.

OECD. 2001. *Nuclear Power in the OECD, Organisation for Economic Cooperation and Development (OECD) and International Energy Agency (IEA)*. Paris: OECD.

Oettingen, Mikolaj. 2018. Criticality analysis of the Louis Slotin accident. *Nuclear Engineering and Design* 338: 92–101. [CrossRef]

Omri, Anis, Nejah Ben Mabrouk, and Amel Sassi-Tmar. 2015. Modeling the causal linkages between nuclear energy, renewable energy and economic growth in developed and developing countries. *Renewable and Sustainable Energy Reviews* 42: 1012–22. [CrossRef]

Ozcan, Mustafa, Mehmet Yildirim, and Semra Ozturk. 2016. Generation expansion planning scenarios to reduce natural gas dependency of Turkey. *Energy Exploration and Exploitation* 34: 244–61. [CrossRef]

Paramati, Sudharshan Reddy, Avik Sinha, and Eyup Dogan. 2017. The significance of renewable energy use for economic output and environmental protection: Evidence from the Next 11 developing economies. *Environmental Science and Pollution Research* 24: 13546–60. [CrossRef]

Perrow, Charles. 1981. Normal accident at three mile island. *Society* 18: 17–26. [CrossRef]

Prohorovs, Anatolijs. 2022. Russia's war in Ukraine: Consequences for European countries' businesses and economies. *Journal of Risk and Financial Management* 15: 295. [CrossRef]

Rahman, Md Mustafizur, Abayomi Olufemi Oni, Gemechu Eskinder, and Amit Kumar. 2020. Assessment of energy storage technologies: A review. *Energy Conversion and Management* 223: 113295. [CrossRef]

Ramana, Muvva V. 2018. Technical and social problems of nuclear waste. *Wiley Interdisciplinary Reviews: Energy and Environment* 7: e289. [CrossRef]

Reddy, B. Sudhakara, and Gaudenz B. Assenza. 2009. The great climate debate. *Energy Policy* 37: 2997–3008. [CrossRef]

Rowinski, Marcin Karol, Timothy John White, and Jiyun Zhao. 2015. Small and Medium sized Reactors (SMR): A review of technology. *Renewable and Sustainable Energy Reviews* 44: 643–56. [CrossRef]

Rybach, Ladislaus. 2003. Geothermal energy: Sustainability and the environment. *Geothermics* 32: 463–70. [CrossRef]

Sauga, Michael. 2008. The Atomic Ageenters a New Dawn. Spiegel Online International. July 11. Available online: http://www.spiegel.de/international/world/0,1518,565132,00.htmls (accessed on 6 May 2022).

Schiermeier, Quirin, Tollefson Jeff, Scully Tony, Witze Alexandra, and Oliver Morton. 2008. Energy alternatives: Electricity without carbon. *Nature* 454: 816–24. [CrossRef]

Shafiei, Sahar, and Ruhul A. Salim. 2014. Non-renewable and renewable energy consumption and CO_2 emissions in OECD countries: A comparative analysis. *Energy Policy* 66: 547–56. [CrossRef]

Sirin, Selahattin Murat. 2010. An assessment of Turkey's nuclear energy policy in light of South Korea's nuclear experience. *Energy Policy* 38: 6145–52. [CrossRef]
Sokolski, Henry. 2010. The High and Hidden Costs of Nuclear Power. *Policy Review* 162: 53–68.
Song, Bing, and Peter Hall. 2020. Densification of biomass and waste plastic blends as a solid fuel: Hazards, advantages, and perspectives. *Frontiers in Energy Research* 8: 58. [CrossRef]
Stefanovsky, Sergey V., Sergey V. Yudintsev, Reto Gieré, and Gregory R. Lumpkin. 2004. Nuclear waste forms. *Geological Society, London, Special Publications* 236: 37–63. [CrossRef]
Ilhan, Talinli, Emel Topuz, and Mehmet Uygar Akbay. 2010. Comparative analysis for energy production processes (EPPs): Sustainable energy futures for Turkey. *Energy Policy* 38: 4479–88. [CrossRef]
Kenta, Tanaka, and Shunsuke Managi. 2016. Impact of a disaster on land price: Evidence from Fukushima nuclear power plant accident. *The Singapore Economic Review* 61: 1640003. [CrossRef]
Temurçin, Kadir, and Alpaslan Aliağaoğlu. 2003. Nuclear energy and reality of nuclear energy in Turkey in the light of discussions. *Turkish Journal of Geographical Sciences* 1: 25–39.
Tokimatsu, Koji, Jun'ichi Fujino, Konishi Satoshi, Ogawa Yuichi, and Kenji Yamaji. 2003. Role of nuclear fusion in future energy systems and the environment under future uncertainties. *Energy Policy* 31: 775–97. [CrossRef]
Topal-Namli, Hanife, and Suat Sean Namli. 2014. Nuclear Power in Turkey: Pros and Cons. Available online: http://westeastinstitute.com/journals/wp-content/uploads/2015/03/3.Hanife-Topal-Namli-JWEIBE.pdf (accessed on 7 May 2022).
Toth, Ferenc L., and Hans-Holger Rogner. 2006. Oil and nuclear power: Past, present, and future. *Energy Economics* 28: 1–25. [CrossRef]
Vaillancourt, Kathleen, Maryse Labriet, Loulou Richard, and Jean-Philippe Waaub. 2008. The role of nuclear energy in long-term climate scenarios: An analysis with the World-TIMES model. *Energy Policy* 36: 2296–307. [CrossRef]
Vassilev, Stanislav V., Christina G. Vassileva, and Vassil V. Vassilev. 2015. Advantages and disadvantages of composition and properties of biomass in comparison with coal: An overview. *Fuel* 158: 330–50. [CrossRef]
Vujić, Jasmina, Ryan M. Bergmann, Škoda Radek, and Marija Miletić. 2012. Small modular reactors: Simpler, safer, cheaper? *Energy* 45: 288–95. [CrossRef]
Wang, Yuefa, Zhongxi Chao, De Chen, and Hugo A. Jakobsen. 2011. SE-SMR process performance in CFB reactors: Simulation of the CO_2 adsorption/desorption processes with CaO based sorbents. *International Journal of Greenhouse Gas Control* 5: 489–97. [CrossRef]
Weber, William J., Rodney C. Ewing, C. Austen Angell, George W. Arnold, Alastair N. Cormack, Jean Marc Delaye, David L. Griscom, Linn W. Hobbs, Alexandra Navrotsky, David L. Price, and et al. 1997. Radiation effects in glasses used for immobilization of high-level waste and plutonium disposition. *Journal of Materials Research* 12: 1948–78. [CrossRef]
Yim, Man-Sung, and K. Linga Murty. 2000. Materials issues in nuclear-waste management. *Jom* 52: 26–29. [CrossRef]
Zhao, Shujie, Duan Weichao, Zhao Dongfeng, and Qingbin Song. 2022. Identifying the influence factors of residents' low-carbon behavior under the background of "Carbon Neutrality": An empirical study of Qingdao city, China. *Energy Reports* 8: 6876–86. [CrossRef]
Zolberg, Vera L. 1998. Contested remembrance: The Hiroshima exhibit controversy. *Theory and Society* 27: 565–90. [CrossRef]

Journal of
Risk and Financial Management

Communication

Russia's War in Ukraine: Consequences for European Countries' Businesses and Economies

Anatolijs Prohorovs

Faculty of Business and Economics, RISEBA University of Applied Sciences, 3 Meža Street, LV-1048 Riga, Latvia; anatolijs.prohorovs@riseba.lv

Abstract: Companies and countries have needed to adapt their activities to the consequences of the Russian war in Ukraine. The analysis in this article shows that both the Russian war in Ukraine and the subsequent trade restrictions have become a powerful trigger, significantly increasing the level of inflation and exacerbating the existing issues of economies. As a result, the confrontation between the West and Russia has greatly escalated, which will have a long-term, large-scale negative impact on most European companies and economies. There could also be a lasting restructuring of world trade. The article notes that not only the end date of the war in Ukraine may be important for business and economies, but also which of the trade and financial restrictions can be lifted from Russia, and when. The article also makes recommendations that may help company leaders plan, in a timelier and more accurate fashion, the changes necessary to maintain company sustainability.

Keywords: war in Ukraine; Europe; consequences; economy; business; companies; embargo

JEL Classification: G12; G14; G15

Citation: Prohorovs, Anatolijs. 2022. Russia's War in Ukraine: Consequences for European Countries' Businesses and Economies. *Journal of Risk and Financial Management* 15: 295. https://doi.org/10.3390/jrfm15070295

Academic Editor: George Halkos

Received: 8 June 2022
Accepted: 29 June 2022
Published: 2 July 2022

1. Introduction

A little more than two years have passed since the beginning of the COVID-19 pandemic, and it is not clear whether it has finally ended. However, companies and countries face an unexpected new challenge: the need to adapt their activities to the consequences of the war in Ukraine.

To do this, company leaders, entrepreneurs, and policymakers need to know and take into account the possible consequences of the ongoing hostilities on business and the economy.

Yet, as a rule, the studies we found on the possible consequences of the war in Ukraine for the economy constitute "mono-analysis", that is, they do not include the results of various sources of information and are often based on the data from recent months, rather than weeks and days, which can significantly affect the accuracy of the predictions.

As of the beginning of May, only Lim et al. (2022) have stated that the war in Ukraine could have a negative impact on business through limited access to funds, reduced purchasing power, an increasing inflation rate, and a threat to sustainable growth and restrictions on trade as a result of sanctions. It was also noted that the recent COVID-19 crisis, also caused by non-economic factors, has particularly affected small businesses, which have experienced a rapid decline in profitability, combined with difficulties in adjusting costs quickly enough (ECB 2022d).

However, we did not find any publications addressing the main risks that could have a negative impact on the activities of the majority of companies.

Therefore, in this article, based mainly on data from late April to mid-June 2022, we will consider the already occurring and expected consequences of the economic shocks and their possible duration, which will allow company leaders, stakeholders and policymakers to make better decisions.

Our analysis will contribute to a deeper understanding of the economic and business issues that have occurred and may continue to arise as a result of the war in Ukraine, and how long such consequences will continue. The recommendations made may help company leaders, stakeholders, and policymakers to plan and implement, in a timelier and more accurate fashion, the changes necessary to adapt and increase sustainability and competitiveness, in order to respond effectively to the emerging challenges, now and in the future.

The article is structured as follows: in the first section, an analysis of supply, price and supply chain shocks is made; in the second section, the growth prospects and the duration of high inflation are discussed; in the third section, the growth prospects for the economy are addressed; the fourth section analyses the prospects for employment and unemployment; and the fifth section looks at the possible consequences of confrontation between the West and Russia; this is followed by a conclusion and recommendations for companies.

2. Shocks of Price, Supplies, and Supply Chain

As a result of the war in Ukraine, there will be supply and commodity price shocks with wide-ranging long-term consequences. Production, consumption, and trade in commodities will change as countries move towards greater self-sufficiency, creating opportunities for new suppliers. The war is leading to costlier trading patterns and a major diversion in energy trade, and the outlook for commodity markets is highly dependent on the length of the war and the disruption it causes in supply chains (World Bank 2022a). A war in Ukraine will trigger a massive negative supply shock in the global economy (Roubini 2022). The price shocks will have global consequences. In addition, the war in Ukraine will exacerbate the structural problems of post-pandemic Europe (International Monetary Fund 2022a). Georg Zachmann of the Bruegel think tank notes that the sanctions on Russia will have dramatic long-term effects on prices, starting with energy but then spreading to the entire economy. In his opinion, Lithuania, Latvia, and Estonia remain more vulnerable than many other European countries and may be among the first whose economies will feel the blow (Washington Post 2022). Vulnerability can be explained by the share in and structure of EU countries' international trade with Russia. For example, in 2021, Latvian exports to Russia accounted for 7.3% of total exports, while imports accounted for 9.1% (CSB 2022). This is significantly higher than the EU average of 4.1% and 7.5%, respectively (Eurostat 2022b).

A war between Russia and Ukraine will exacerbate the supply chain problems and drive up commodity prices further. The sectors with long global production–supply chains and high reliance on energy and metal supplies will face the strongest hurdles (S&P Global 2022).

The energy-dependent sectors, especially oil refining and power generation, transportation services, metallurgy, and chemicals, will be the first to suffer, as Europe uses Russia as a resource supplier (S&P Global 2022). An embargo on Russian energy resources could lead not only to higher prices, but also to the rationing of energy consumption (Bundesbank 2022a).

According to the world's leading oil and gas analytics agency, Rystad Energy, global LNG demand in 2022 is expected to be 436 million tons and supply is expected to be 410 million tons (GWN 2022). This means that the demand will exceed supply by 6.3%, meaning there is not enough LNG to meet demand. For a specific example, consider how supply shocks can affect prices.

Austria receives about 80% of its gas supplies from Russia and has warned that a shutdown from its largest supplier will lead to a recession in the country; it has begun holding tenders to replenish its storage facilities. In the first tender, Austria paid a 45% premium on European base prices to create a strategic gas reserve before the start of the heating season (Bloomberg 2022b).

Rystad Energy notes that LNG may not be enough to replace Russian gas in freezing weather. They predict that the price of gas in the EU will rise to $3500 per 1000 cubic meters, almost 300% higher than a year ago (GWN 2022). Rystad Energy notes that "replacing a significant part of Russian gas will be extremely difficult, with far-reaching consequences for the population, the economy and the role of gas in the region's energy transition" and that the supply-balanced demand for LNG will only be possible after 2024 (GWN 2022). This means that, as a result of the supply shock, if Russian gas is phased out, LNG prices in Europe will remain high for several years.

Note that the high gas prices will contribute to a significant increase in prices for other energy sources, including affordable ones. For example, this year, firewood prices in Latvia, which is a major exporter of various wood products, have risen from 120 to 200 euros (TVNET 2022), although the heating season is still far away.

Due to the heavy dependence on raw material imports from Russia and Ukraine, Europe is more vulnerable than the other major economies in the world (RaboResearch 2022).

The war is already severely disrupting the narrow food, energy, and financial markets. Persistent disruptions and elongations in the global supply chains and burgeoning transport costs further complicate the situation (United Nations 2022).

To summarize this section, economies and companies will experience price shocks and supply constraints on key commodities, with wide-ranging long-term effects. There will be an increase in the cost of supply chains, and it will also be necessary to reconfigure supply chains to be shorter and more reliable.

3. Inflation

One of the main reasons for the increase in inflation is the increasing demand for natural resources since the economy started to recover in 2021 (Elbahnasawy and Ellis 2022). A supply shock will reduce growth and further increase inflation, at a time when inflation expectations are already unsustainable (Roubini 2022). A surge in oil and food prices will raise an inflation rate that is already too high (Krugman 2022).

According to the forecast of one of the world's largest investment companies, BlackRock, in 2022 the EU countries will spend more than 9% of their GDP on energy. This is the highest share in 40 years and more than double the rate predicted in the US, although, two years ago, the figure for both economies was about 2% (Bloomberg News 2022a). Higher commodity prices exacerbate the already high inflationary pressures around the world (World Bank 2022a). According to the European Central Bank, half of the recent rise in inflation has been driven by higher energy prices (ECB 2022a).

As a result of a sharp increase in commodity prices, the industrial production price growth index in Germany in March 2022 was 30.9%, and in April increased to 33.5% (DeStatis 2022). At the end of March 2022, inflation increased for another month in a row, amounting to 7.8% in the EU, 15.6% in Lithuania, 14.8% in Estonia, 11.5% in Latvia, and 7.6% in Germany (Statista 2022). Consumer prices in Germany in March 2022, compared to February, seasonally adjusted, increased by 2.5%. Such a level of inflation in Germany was last seen in 1981, during the First Gulf War (Bundesbank 2022b).

However, despite already being high and increasing in March 2022, the inflation rate continued to rise in April—to 8.1% (by 0.3 percentage points) in the EU, to 16.6% (by 1 point) in Lithuania, to 19.1% (by 4.3 points) in Estonia, to 13.1% (by 1.6 points) in Latvia, and to 7.8% (by 0.2 points) in Germany (Eurostat 2022a).

Among the Eurozone countries, Estonia, Lithuania, the Czech Republic and Latvia have significantly higher inflation rates than the other EU countries.

However, for policymakers, entrepreneurs, and the upper management of companies, it is important when planning to know not only the current inflation rate, but also what it might be at the end of 2022. Therefore, using the forecast of the Bank of Latvia as an example, we will consider whether the inflation forecasts presented for 2022 are optimistic or realistic.

According to the Bank of Latvia's forecast of 30 March 2022, the inflation rate in Latvia will rise to 9.5% in 2022 and decrease to 3.7% in 2023 (Latvijas Banka 2022a). The commentary to this forecast states that it is based on the fact that energy prices will decline in the second half of the year.

Note that, according to the World Bank (2022b) forecast, the natural gas prices in Europe in 2022 will be twice as high as in 2021, while the coal prices will be 80 percent higher, and will remain at historically high levels until the end of 2024. Taking into account this forecast and other indicators, including the prices for raw materials, the Bank of Latvia forecast for the inflation indicator for 2022–2023 seems too optimistic. It should be noted that on 12 June, the Bank of Latvia raised the inflation forecast to 14.8% for 2022 and to 7% for 2023 (Latvijas Banka 2022b). However, the gas prices for households for the second half of 2022 increased by 60–89% (Latvijas gāze 2022). Although half of the increase in gas prices for households will be compensated by the state (Delfi 2022), the updated inflation forecast of the Bank of Latvia still seems too optimistic, both based on the arguments listed in this section, and based on the fact that inflation in May 2022 in the annual terms in Latvia increased by 3.8 percentage points and amounted to 16.9% (Latvijas Banka 2022c). It is possible that the publication of more optimistic forecasts is justified by the monetary authorities' desire to maintain consumer activity.

The war in Ukraine has exacerbated the dilemma faced by central banks on both sides of the Atlantic, namely, how to tighten financial conditions quickly and intensively in an environment of inflation that can no longer be seen as simply temporary and reversible, and at the same time not reduce the pace of economic activity and not provoke financial turmoil (Canuto 2022). Although the interest rate in America was 0.25 percentage points higher than in the EU, on May 4 the Federal Reserve raised the rate by another 0.5 percentage points to counter inflation (Wall Street Journal 2022). As of 16 June 2022, the rate in America was raised to 1.5% and became 1.5 percentage points higher than in the euro area (Countryeconomy 2022). The transition to a green economy and a possible slowdown in globalization are likely to increase inflation in the medium term (ECB 2022a). It will then have to be decided how high to raise the rate to fight inflation, as this may deepen a possible recession (ECB 2022a). However, in today's environment, with inflation rising, and as central banks are already behind schedule, slower policy tightening could accelerate the weakening of inflation expectations, further exacerbating stagflation (Roubini 2022). High inflation is forcing the region's central banks to prioritize efforts to curb price pressures over growth-supporting policies. The currencies of the countries of Central and Eastern Europe, especially the Hungarian forint, have entered a period of heightened geopolitical instability and are in a precarious position (Oxford Analytica 2022). Although Goldman Sachs notes that the Eurozone economy is shrinking and that inflation is approaching 8%, a rate hike by the European Central Bank in 2022 is unlikely (Bloomberg 2022a), even if the ECB has tentatively authorized a minimum (0.25 percentage point) rate hike in July (ECB 2022b).

Agustin Carstens, CEO of the Bank for International Settlements, says we should not expect inflationary pressures to ease anytime soon, as many of the factors behind high inflation remain in place and new ones are emerging. Moreover, the structural factors that have kept inflation low in recent decades may weaken as globalization retreats, and the inflationary paradigm may change accordingly (Carstens 2022).

It should be noted that increasing inflation may be of a structural nature, and, in this case, its decline will mainly depend not on an increase in the interest rate, but on a significant reduction in energy prices; accordingly, the monetary authorities' ability to reduce inflation may be very limited.

We should also note that the increase in interest rates is leading to an increase in the cost of servicing the debt of both companies and countries.

Corporate debt has been growing worldwide for more than a decade and high levels of corporate debt can make restrictive monetary policy less effective in curbing inflation (Goodhart et al. 2022).

Thus, the ECB notes that the increased level of corporate debt since the COVID-19 pandemic makes firms more vulnerable to increased risks, higher interest rates, or falling profits, while weaker corporate balance sheets pose a risk to banks (ECB 2022d). Some developed countries, as well as a number of low-income countries, are already at high risk or are in a debt crisis (United Nations 2022). For example, Italy's yield on ten-year bonds increased to 3% per annum, although it was less than 0.5% in the summer of 2021 (Tradingeconomics 2022a). Italy's debt/GDP ratio exceeded 150%, and the amount of debt reached 2.75 trillion euros (Tradingeconomics 2022b, 2022c). Since 2008, the amount of Italy's debt has been growing continuously (Tradingeconomics 2022c), but Italy, unlike the US and Japan, cannot "print" money on its own. Therefore, the situation of rising rates and weakening economic growth may pose a threat to countries with high levels of debt, such as Italy, Portugal, Spain, and especially Greece, whose debt is close to 200% of GDP and exceeds the levels of the period when Greece needed financial assistance from the IMF (Tradingeconomics 2022d; International Monetary Fund 2019). Accordingly, under the worst-case scenario, some of the countries with high levels of public debt, high interest rates, and an unwillingness of investors to take on increased risks may require certain financing instruments from the EU; as was the case, for the first time, to finance the reduction in the negative impact of the COVID-19 pandemic on the economies of the EU countries. It should be noted that the design and structure of EU funding, which was first used to reduce the negative impact of the COVID-19 pandemic, is quite complicated and, accordingly, impossible to agree on quickly, since such decisions require the consensus of all countries. Based on this, it should be taken into account that, if decisions on the allocation of financing at the EU level are required to solve the problems of the debt crisis of some countries in the euro area, certain risks will arise for the economies of some countries. The adoption of such decisions will become a new challenge and a test of the strength of both the common European currency (the euro) and the EU as a whole. It should also be taken into account that such decisions may be related to loose monetary policy and, accordingly, not help to reduce inflation.

Summing up this section, we note that the increase in inflation in 2021 and 2022 has mainly been due to an increase in demand and, accordingly, an increase in energy prices. The main factors that can influence the persistence of high inflation, or even its increase, include shocks in the price and supply of energy and other commodities and basic goods. The inflation rate in the EU and a number of other countries may remain high for several years or more and be a significant threat both to the economies of the countries and to companies.

4. Employment and Unemployment Rate

The consequences of the Russian war in Ukraine may complicate and slow down the recovery of the labor market in Europe after the COVID-19 pandemic (International Monetary Fund 2022d). The recovery of the labor market will be negatively affected by a slowdown in the economy and, for some countries, by significant migration flows as well (International Monetary Fund 2022d). Due to economic sanctions, disruption of supply chains and resource flows, as well as the inevitable inflationary consequences, European companies will face a decline in employment (Eurofound 2022). Research carried out by Central Bank of Ireland economist, Reamonn Lydon, notes that job postings in the 21 countries of Europe slowed down after the start of Russia's war in Ukraine (Bloomberg 2022c). Employment was hardest hit in the countries that are most dependent on Russian imports, and energy-intensive sectors, such as manufacturing and transport. The impact could spill over to other sectors if high inflation continues to impact on consumer spending (Bloomberg 2022c).

A slowdown in manufacturing growth will negatively impact hiring activity in Germany, according to a business survey conducted by S&P Global (The Business Times 2022). However, according to Indeed, the number of job postings remains 45% higher than before the COVID-19 pandemic (Bloomberg 2022c).

Since unemployment is one of the most important macroeconomic indicators, it seems appropriate to discuss the possible methods for forecasting employment and the factors that may affect it as a result of the continuation of Russia's war in Ukraine.

Let us consider which of the last two economic crises is more suitable for predicting the level of employment as a result of the war in Ukraine, if we apply the extrapolation method for this purpose.

The COVID-19 pandemic led to the sharpest contraction on record in employment and total hours worked in the second quarter of 2020. There were 5.2 million fewer persons employed in the second quarter of 2020 than at the end of 2019, a fall of 3.2% (ECB 2020). The unemployment rate increased by 1.2 percentage points in half a year and has not risen again (Eurostat 2022c). There has been no increase in the unemployment rate in the EU for almost four months, since the start of the war in Ukraine (Eurostat 2022c). The literature reviewed in this article notes that the war in Ukraine may be protracted and continue to have a negative impact on the economy and business, and, accordingly, on the level of employment, for not months, but years. In addition, the increase in unemployment during the COVID-19 pandemic was significantly affected by the introduction of a number of restrictions, which was especially true for the service and retail sectors, as well as for unvaccinated workers.

Next, let us consider how the unemployment rate changed after the 2008–2009 Global Financial Crisis.

The unemployment rate, which measures the proportion of persons unemployed to the total labor force, rose from 7.2% in March 2008 to 9.5% in May 2009 in the euro area. In the same period, the rate in the EU surged from 6.7% to 8.9% (Eurostat 2009). Unemployment rose in all of the countries of the EU, but the increase was very uneven, ranging from over 10 percentage points in Estonia, Spain, Latvia, and Lithuania to less than one percentage point in the Netherlands and Denmark (Eurostat 2009).

In 2013, the unemployment rate in the euro area reached its highest level in over four years, rising to 12.2%, i.e., by five percentage points (Eurostat 2022c). The growth of unemployment during the years of the Global Financial Crisis was uneven. In the first year, unemployment increased by 2.2 percentage points; after that, for more than a year, the unemployment rate remained at the same level and even decreased slightly, and, over the course of another year and a half, it increased to 11% in the EU and to 12.2% in the euro area (Eurostat 2013). The average annual increase in unemployment since the Global Financial Crisis has been 1.2 percentage points.

Based on a brief analysis of the unemployment rates during the last two crises, in applying the extrapolation method to predict the level of employment in the EU as a result of the war in Ukraine, it is more suitable to analyze the change in the level of employment after the Global Financial Crisis.

There are a number of factors to consider when extrapolating post-Global Financial Crisis unemployment rates to the rise in unemployment as a result of the war in Ukraine.

The main factors that could hold back the increase in the unemployment rate in the EU include the following. Firstly, before the Global Financial Crisis, the unemployment rate in the EU was higher, at 7.1%, and in April 2022 the unemployment rate in the EU reached its lowest level—6.2%. That is, a greater increase in unemployment will be hampered by a record number of vacancies for the EU, which reached 3.1% in the first quarter of 2022 (Eurostat 2022c) due to high demand for labor and the early retirement of older workers (International Monetary Fund 2022d). Secondly, the COVID-19 pandemic has accelerated structural transformations, including through digitalization and automation, and these transformations entail a significant redistribution of workers among sectors and professions (International Monetary Fund 2022d). Thirdly, the experience of remote work gained in recent years allows employees not only to move between sectors and professions, but also to get a job in other countries without changing their place of residence (Prohorovs and Bistrova 2022). In addition, a significant number of migrants requires the creation of additional jobs in the host country in order to provide the migrants with necessary services, which leads to a small increase in employment (and also increases the demand for goods and services in the host country).

Let us consider the main factors that may have an impact on the decline in employment. The first is a higher inflation rate. Thus, in 2008, the highest ever average annual inflation rate, of 3.3%, was recorded for the euro area (Eurostat 2009), while the average annual inflation rate in the euro area was 8.1% in May 2022 (Eurostat 2022a), which is almost 2.5 times higher than in 2008. The second comprises the problems related to economic sanctions, disruption of supply chains, and resource flows (Eurofound 2022). The third is the high level of migrants from Ukraine, especially given the protracted nature of the war in Ukraine. The increase in the hostilities is forcing refugees to stay in their host countries longer than expected and, accordingly, they may seek employment in the labor market of these countries (ILO 2022). Approximately 2.75 million refugees from Ukraine are of working age, of whom 1.2 million worked before the start of the war in Ukraine and, accordingly, could enter the EU labor market. The refugees from Ukraine will have the highest impact on the growth of unemployment in Moldova—unemployment growth may increase from 2.5 to 6.9 percent. In Poland, it may increase from 3 to 5.3 percent, and in Slovakia, Romania, Hungary, and the Czech Republic, unemployment may increase by more than one percentage point, while in other countries, the increase in unemployment will not exceed one percentage point (ILO 2022).

The Polish economist, Sobolewski, notes that in addition to the 185,000 Ukrainian refugees already legally working in Poland, there may be another 100,000 who work in the shadow economy, for example, helping their Polish hosts around the house or in small businesses (BIRN 2022). Based on the data presented by Sobolewski, two conclusions can follow. Firstly, the reception of refugees from Ukraine has an impact on the growth of the shadow economy in the host country. Secondly, in countries with high levels of refugees from Ukraine, the real employment rate will be higher than reflected in the statistics.

The following conclusions can be drawn from this section. It is most likely that the unemployment rate in the EU will grow unevenly both over time and across countries within two to four years. The highest pressure on employment will occur in countries that have had a higher volume of trade relations with Russia and the greatest dependence on Russian resources, that is, through price and supply shocks. In most cases, these are the countries of Central and Eastern Europe and, first and foremost, the Baltic countries, as well as energy-intensive sectors, such as manufacturing and transport.

Unless other countries become directly involved in the war in Ukraine and extremely negative scenarios are realized, for example, a complete cut-off of gas supplies from Russia, it is very unlikely that the unemployment rate in the EU will approach the maximum unemployment rates that were experienced after the Global Financial Crisis in 2013.

5. The Economy Growth Prospects

The JPMorgan Chase CEO, Jamie Dimon, said on 4 May that it is very, very unlikely that the Federal Reserve can tame inflation without triggering a recession (CNN Business 2022). Historical experience shows that once the inflationary genie is out of the bottle, regardless of the policy response, a recession would already be hard to avoid (RaboResearch 2022). The slowdown in business activity is due to the fact that its surge after two years of lockdowns led to a shortage of energy, food, and transport at the end of 2021. This caused disruptions in world trade and accelerated inflation. Consumption is falling because life is becoming more expensive and people are saving, while investment is decreasing as lending rates rise with prices, uncertainty increases, and company profits fall (International Monetary Fund 2022a). The war and the total embargo only exacerbate these problems, since Russia is the largest exporter of oil and gas, and, together with Ukraine, it is also the leading supplier of food (BBC News 2022). According to experts, the Introduction of fiscal support measures in Germany compensates for only a small part of the losses of households and companies (Bundesbank 2022a).

Both America and the Eurozone will face recession in 2022 and 2023 (McKinsey 2022). The recession forecasts have already begun to come true. Thus, in the first quarter of 2022, America's GDP decreased by 1.4 percent year-on-year, while personal income, adjusted for taxes and inflation, decreased by 2.0 percent (Bureau of Economic Analysis U.S. 2022). The Eurozone economy grew by 0.2 percent QOQ in the first quarter of 2022, below the forecasts and the lowest indicator since the EU emerged from recession last year (Trading Economics 2022). There is also a risk that some EU countries may enter a technical recession in 2022 (International Monetary Fund 2022a). Thus, at the end of April, the Bundesbank warned that, in the event of an embargo on Russian gas, Germany's real GDP would fall by almost 2% this year (The Guardian 2022). The researchers of the Institute for Monetary and Financial Stability gave a more radical assessment of a possible fall in Germany's GDP in the event of a suspension of energy supplies from Russia—3–6%. At the same time, they note that the impact on GDP of a complete cessation of Russian energy supplies to Germany may be comparable to the impact of the Global Financial Crisis or the COVID-19 crisis, but neither the overall consequences nor the political recipes that follow from them can be compared with the consequences of the cessation of energy imports from Russia (Berger et al. 2022).

Under unfavorable scenarios, the decline in the Eurozone economy in 2022 could range from 1.2 to 1.4%, and the recession could continue into 2023 (ECB 2022c).

The decline in economic growth in Western and Eastern Europe will also be affected by the fact that many companies of these countries have stopped or restricted international trade and other operations in Russia (Canuto 2022). According to Eurostat (2022b), in 2021, Russia was the EU's fifth largest merchandise export partner (4.1%), and third largest merchandise import partner (7.5%). Accordingly, the countries whose companies had a higher level of international trade with Russia may suffer more if the companies fail to reorient themselves to other markets. In addition, it must be taken into account that sanctions also exist for Belarus, and trade with Ukrainian companies is limited by the circumstances related to the war. If companies fail to reorient themselves to other markets, the above restrictions on trade will lead to a decrease in net turnover, income, and, respectively, the profits of numerous companies. This will negatively affect economic growth indicators and will have a negative impact on countries' budget revenues, which will also limit their ability to support households and businesses, which is especially important in the context of both an economic downturn and high inflation.

There is a risk that persistently high commodity prices, which are likely to continue until the end of 2024, will lead to stagflation—sluggish economic activity combined with strong pressure on the cost of living (World Bank 2022c). According to BlackRock researchers, there is a risk of stagflation in Europe (Bloomberg News 2022a). The European Central Bank is facing such a crisis and potential stagflation for the first time. Given the nature of the supply shock, tightening policy will not reduce inflation. Conversely, additional stimulus will not be able to prevent an economic downturn if it is caused by a lack of supply (RaboResearch 2022).

As a result of the Russian war in Ukraine, the impact on countries' economies will be heterogeneous and asymmetric and will also differ greatly depending on the economies' gas intensity (International Monetary Fund 2022b). The blow to Europe will be one and a half times more tangible than to the United States, and the energy war will cause the greatest damage to the European economy (BBC News 2022). The economic and financial impact of the war and the associated stagflation shock will be greatest in Russia and Ukraine, followed by the European Union, due to its heavy dependence on Russian gas (Roubini 2022). Europe and the fragile emerging markets will be hit hardest (Oxford Analytica 2022). The head of the Center for Macroeconomics and Development, former Executive Director of the World Bank (Canuto 2022), has the same opinion. The economies of European countries are the most negatively affected, of which the countries of Central and Eastern Europe will be the most negatively affected (Coface 2022).

Summing up this section, we can state that there is a significant likelihood that in 2022–2023, a recession or even stagflation could begin in a number of EU countries.

6. Possible Consequences of Confrontation between the West and Russia

Due to the Russian war in Ukraine and the subsequent economic sanctions against Russia, a process of transformation of the architecture of the world economy has begun, according to Papava (2022). In his opinion, the hallmark of this process is the relative fragmentation of the world economy into countries that adhere to the economic sanctions against Russia and countries that are allies of Russia or maintain a certain neutrality in this confrontational economic scheme.

The Russian–Ukrainian conflict will reshape international trade and redefine global value chains, leading to decentralized value chains and improved trade efficiency (Estrada and Koutronas 2022).

The S&P Chief Economist, Paul Gruenwald, notes that there could be a strong realignment of global trade and that the global economy is becoming more regionalized—shorter supply chains with reliable partners (CNBC 2022). The director of research at global consulting firm Wood Mackenzie, Peter Martin, notes that the global economy may be undergoing more constant change, thereby increasing the importance of reliable trading partners (CNBC 2022). As a result of the war in Ukraine, a geopolitical division will occur, fragmenting the world economy (Oxford Analytica 2022). The push for deglobalization, started during the pandemic, will intensify and one can expect an increase in the influence of geopolitics in international payments and in access to certain types of commodities (Canuto 2022). If energy trade changes, the supply chains reconfigure, payment networks fragment, and countries rethink their reserve currency holdings; this could fundamentally change the global economic and geopolitical order in the long term (International Monetary Fund 2022c). Growing geopolitical tensions further increase the risks of economic fragmentation, especially in trade and technology. War between Russia and Ukraine could be a prelude to a new geopolitical era, in which businesses may need to be able navigate an altered map of globalization (KPMG 2022). The consequences of the Ukrainian crisis and the wider geopolitical depression it portends could be anything, but there is no going back to the original state of things (Roubini 2022). However, Paul Krugman (2022) finds that, although the consequences of the Russian war in Ukraine for the global economy will be bad, they will not be catastrophic. In his opinion, the shock of the confrontation between Russia and the West is unlikely to be as serious as the oil shock that shook the world economy in the 1970s.

Ukraine's location has drawn more attention to it compared with other conflicts, making the prospect of a protracted war more likely, not least because the EU and NATO have a strong interest in maintaining Ukraine's sovereignty (Serhan 2022). Thus, the British Foreign Secretary, Liz Truss, has said that the war in Ukraine could last up to 10 years (NDTV 2022). US and European officials increasingly believe that the war will not end in the short term (CNN 2022). European countries need to prepare for the fact that the war will last for a long time (Gressel 2022). The U.S. National Security Adviser, Jake Sullivan, has also said that the fighting in Ukraine is likely to be protracted (CNN 2022). According to NATO Secretary, General Jens Stoltenberg, the war could last for years, and the military alliance should not stop supporting Ukraine (Bloomberg News 2022b).

The periodical literature reports that Russia's war in Ukraine could end in a relatively short time, due to the very significant restrictions on Russian exports, but Pestova et al. (2022) find that the Russian economy is difficult to undermine through oil and gas embargoes.

For business, however, not only will the end date of the war in Ukraine be important, but also what trade and financial restrictions can be lifted from Russia and Belarus and within what time frame. Sanctions are usually imposed much faster than they are lifted (Blinken 2015; Grauvogel and Attia 2019). The rapid lifting of economic sanctions and a return to the pre-war situation seems completely illusory, even if a peace agreement is signed between Russia and Ukraine (Coface 2022). The possible duration of the sanctions imposed against Russia may be influenced by a number of factors, some of which are difficult to predict. For example, at what rate will there be a sustainable increase in the capacity of alternative energy sources and its storage, what will be the cost of these processes, and what will be the price and availability of some raw materials. In addition, it should be taken into account that both within the European Union and in other countries that have imposed sanctions, there may be different economic and geopolitical goals and priorities, which may complicate and therefore lengthen the period for lifting sanctions. We should keep in mind that adverse economic shocks may lead to a wider dissemination of ideological politics (Ash et al. 2021); therefore, the parties to the conflict may consider the imposed sanctions for geopolitical purposes, as well as for the purposes of economic and/or technological dominance, which may increase the timing of the lifting of some of the sanctions. As Roubini (2022) points out, one must bear in mind that sanctions against Russia, no matter how large or necessary for the future containment of Russia, will inevitably hurt not only Russia, but also the US, the West, and the emerging markets.

From the analysis carried out in this section, we can conclude that, firstly, the war in Ukraine has transformed into a confrontation between the West and Russia and will have wide-ranging long-term consequences for the countries' economies and for business. Secondly, one can expect an increase in the influence of geopolitics in the banking and financial sphere and in international trade.

Thirdly, there could be a lasting realignment of world trade: the global economy is becoming more regionalized—shorter supply chains with reliable partners.

Fourthly, the world economy may be subject to more constant change, which in the long run could change the global economic and geopolitical order.

7. Conclusions and Recommendations

Russia's war in Ukraine has further exacerbated the economic problems arising from the American–Chinese trade confrontation, the impact on the economy of the COVID-19 pandemic, and 14 years of loose monetary policy. The trade restrictions imposed against Russia and Belarus and the countersanctions became a powerful trigger which has significantly raised the inflation rate and negatively impacted economic growth, creating the threat of stagflation. Among Western countries, European countries will be the hardest hit. Among the European countries, the countries of Central and Eastern Europe, especially those that had closer trade relations with Russia, will suffer more severely.

The war in Ukraine has greatly increased the level of uncertainty in the economy and business. Thus, almost half of British firms identified the war in Ukraine as the most important source of uncertainty for their business. A survey of UK firms shows that energy use, demand, trade, and ownership are the most important determinants of business uncertainty (Anayi et al. 2022).

As a result of the war in Ukraine, the confrontation between the West and Russia has greatly escalated, which will have a long-term negative impact on most European companies and economies. The negative impact of the consequences that have already occurred cannot be fully measured yet, and some of the consequences are not yet known. This means that companies and countries will have to make and implement decisions in the face of a very high degree of uncertainty, adjusting action plans in a timely manner if the changing circumstances require it.

In our opinion, as a result of the shocks taking place in the economy, the majority of companies have faced or will face at least two big problems at once, that will put pressure on profits. Firstly, there is a decrease in sales volumes and marginality, which in itself will lead to a decrease in income. Secondly, for the reasons stated in our article, there will be an increase in costs. A significant decline in revenues, along with a significant increase in expenses, will lead to a large decrease in profits and/or increased losses for many companies.

The negative pressure on companies' profitability will also influence the projected increase in central bank rates and, accordingly, an increase in the price of credit resources, as well as the fact that, due to a decrease in companies' performance, their ability to attract financing may be limited. Based on the above factors, many companies will not be able to maintain the ratio of income and expenses necessary to continue business without restructuring and/or changing their business models.

Therefore, company management needs to assess how much income their companies will be able to receive in a few months, six months, or maybe a year, and bring expenses to a level that corresponds to current and forecasted circumstances and updated company development plans. In the case of restructuring planning, the company's management should take into account that there is a risk of a further decline in demand in many sectors, and the risk of rising prices for raw materials and logistics also remains. If costs are reduced without taking this factor into account, companies will have a fairly high risk of re-restructuring. The companies that resume operations with a profit will find it easier to negotiate with banks to continue lending beyond certain covenants, for example, tying the lending limit to EBITDA.

When planning restructuring, it should be taken into account that, due to the loss of the Russian and Belarusian markets, and, in some cases, the logistical and other complications of working with companies from Ukraine, exporting companies will have to look for opportunities to enter new markets. To maintain sales volume, this will need to be completed by almost all of the export companies of all of the countries that have exported their products to the aforementioned countries. Therefore, it can be expected that competition will increase both in the foreign and domestic markets, and because of this, marginality may decrease significantly. In addition, if the company has plans to enter new markets, restructuring must take into account that such entry will incur additional costs.

According to a survey of 2300 IT directors by the American consulting company, Gartner, destruction is the new normal (Gartner 2022). Specialists of the Boston Consulting Group (BCG) believe that resilience is at the core of a company's ability to absorb stress, recover critical functions, and thrive in changing circumstances (Reeves et al. 2020).

After a very long break, and with a two-year interval, two crises occurred in the world economy, caused by non-economic factors. Based on this, the companies that want to prosper in the future need to organize their activities so that they are ready for the fact that destruction is the new normal, and that resilience is the ability to absorb stress, as stated in the aforementioned BCG and Gartner theses.

Therefore, in order to maintain and, better yet, strengthen their competitiveness, entrepreneurs, companies, and countries need to adjust their strategies, business models, and management models in such a way as to be ready for an effective response to crises, including the shocks caused by non-economic matters.

8. Suggestions for Future Research

In future research, it would be advisable to develop a quantitative econometric model that could be used to quantify the consequences of Russia's war in Ukraine. The restoration of the Ukrainian economy and sources of its financing might also be a relevant topic for future research.

Funding: The APC was funded by Krišjānis Valdemārs Foundation.

Acknowledgments: We express our deep gratitude to the three anonymous reviewers for their valuable advice and comments. We are also grateful for the advice and comments we received from Vulfs Kozlinskis (RISEBA University); Tatjana Vasiljeva (RISEBA University), who, to our great regret, passed away at the end of April; Julija Bistrova from Riga Technical University. We are also grateful Levs Fainglozs from RISEBA University, who helped in selecting materials and provided other technical support. We also express our gratitude to the Krišjānis Valdemārs Foundation for financial support of this study.

Conflicts of Interest: The author declares no conflict of interest.

References

Anayi, Lena, Nicholas Bloom, Philip Bunn, Paul Mizen, Gregory Thwaites, and Ivan Yotzov. 2022. The Impact of the War in Ukraine on Economic Uncertainty. Available online: https://voxeu.org/article/impact-war-ukraine-economic-uncertainty#:~{}:text=The%20war%20in%20Ukraine%20has%20raised%20uncertainty%20about%20future%20sales,2019%20average%20to%20almost%20140%25 (accessed on 24 May 2022).

Ash, Elliott, Sharun Mukand, and Dani Rodrik. 2021. *Economic Interests, Worldviews, and Identities: Theory and Evidence on Ideational Politics*. Working Paper 29474. Cambridge, MA: National Bureau of Economic Research.

BBC News. 2022. The Western Economy Is Slowing Down Because of Russia's War with Ukraine. And This Is just the Beginning. Available online: https://www.bbc.com/russian/news-61275658 (accessed on 28 May 2022).

Berger, Eva, Sylwia Bialek, Niklas Garnadt, Veronika Grimm, Lars Other, Leonard Salzmann, Monika Schnitzer, Achim Truger, and Volker Wieland. 2022. *A Potential Sudden Stop of Energy Imports from Russia: Effects on Energy Security and Economic Output in Germany and the EU*. IMFS Working Paper Series, No. 166; Frankfurt a. M: Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

BIRN. 2022. Labour Pains in Central Europe. Balkan Investigative Reporting Network. June 7. Available online: https://balkaninsight.com/2022/06/07/labour-pains-in-central-europe/ (accessed on 25 June 2022).

Blinken, Antony. 2015. Iran Nuclear Negotiations: Status of Talks and the Role of Congress. In *Written Statement, Senate Committee on Foreign Relations*; January 21. Available online: https://www.foreign.senate.gov/imo/media/doc/Blinken_Testimony1.pdf (accessed on 24 May 2022).

Bloomberg. 2022a. Goldman Sees Euro-Area Economy Shrinking, Inflation Close to 8%. Available online: https://www.bloomberg.com/news/articles/2022-03-10/goldman-sees-euro-area-economy-shrinking-inflation-close-to-8?utm_medium=cpc_search&utm_campaign=NB_ENG_DSAXX_DSAXXXXXXXXXXX_EVG_XXXX_XXX_Y0469_EN_EN_X_BLOM_GO_SE_XXX_XXXXXXXXXX&gclid=Cj0KCQjwma6TBhDIARIsAOKuANx9zgsKj5rIo_YgYsQeR6lHveoVjhSRSasU3FnTPy3MesW1N3byez8aApEBEALw_wcB&gclsrc=aw.ds (accessed on 24 May 2022).

Bloomberg. 2022b. Central Europe's Huge Gas Depot Risks Being Empty Next Winter. Available online: https://www.bnnbloomberg.ca/central-europe-s-huge-gas-depot-risks-being-empty-next-winter-1.1771327 (accessed on 29 May 2022).

Bloomberg. 2022c. Ukraine War Slows European Job Growth, Indeed Data Say. *By Lizzy Burden*. May 4. Available online: https://www.bloomberg.com/news/articles/2022-05-03/war-in-ukraine-stalls-growth-in-european-jobs-indeed-data-show (accessed on 25 June 2022).

Bloomberg News. 2022a. Energy Shock Is Huge for EU, Manageable in U.S., Blackrock Says. Available online: https://financialpost.com/pmn/business-pmn/energy-shock-is-huge-for-eu-manageable-in-u-s-blackrock-says (accessed on 24 May 2022).

Bloomberg News. 2022b. Ukraine Latest: Zelenskiy Awaits Fateful EU Vote. 11:38 AM GMT+3Updated on June 19, 5:35 PM GMT+3. *Harvest Alarm*. June 19. Available online: https://www.bloomberg.com/news/articles/2022-06-19/ukraine-latest-nato-chief-and-uk-s-johnson-warn-of-long-war?utm_source=google&utm_medium=bd&cmpId=google (accessed on 25 June 2022).

Bundesbank. 2022a. War against Ukraine: Energy Embargo Could Significantly Weaken German Economy. *Bundesbank*. April 22. Available online: https://www.bundesbank.de/en/tasks/topics/war-against-ukraine-energy-embargo-could-significantly-weaken-german-economy-889696 (accessed on 24 May 2022).

Bundesbank. 2022b. Monthly Report: Economic Activity May Have More or Less Stagnated in the First Quarter. *Bundesbank*. April 22. Available online: https://www.bundesbank.de/en/tasks/topics/monthly-report-economic-activity-may-have-more-or-less-stagnated-in-the-first-quarter-889702 (accessed on 24 May 2022).

Bureau of Economic Analysis U.S. 2022. Gross Domestic Product, First Quarter 2022. April 28. Available online: https://www.bea.gov/news/blog/2022-04-28/gross-domestic-product-first-quarter-2022 (accessed on 24 May 2022).

Canuto, Otaviano. 2022. War in Ukraine and Risks of Stagflation. Available online: https://www.policycenter.ma/sites/default/files/2022-03/PB_18-22_Canuto_0.pdf (accessed on 24 May 2022).

Carstens, Agustín. 2022. The Return of Inflation. Available online: https://www.bis.org/speeches/sp220405.htm (accessed on 24 May 2022).

CNBC. 2022. A Russia-Germany 'Trade Rupture' Could Trigger a Financial Shock, Says S&P Chief Economist. Available online: https://www.cnbc.com/2022/04/13/trade-rupture-between-russia-germany-could-cause-financial-shock-sp-global.html (accessed on 24 May 2022).

CNN. 2022. No Short-Term End in Sight in Russia's Invasion of Ukraine, Western Officials Say. By Jennifer Hansler and Kylie Atwood, Updated 1005 GMT (1805 HKT). April 15. Available online: https://edition.cnn.com/2022/04/15/politics/ukraine-russia-protracted-war-us-european-officials/index.html (accessed on 24 May 2022).

CNN Business. 2022. Jamie Dimon: The Cold War Is Back and a Recession Is a Real Possibility. Updated 1239 GMT (2039 HKT). May 4. Available online: https://edition.cnn.com/2022/05/04/economy/jamie-dimon-fed-recession/index.html (accessed on 24 May 2022).

Coface. 2022. War in Ukraine: Many (Big) Losers, Few (Real) Winners. *Coface Economic Publications*. May 3. Available online: https://www.coface.com/News-Publications/News/War-in-Ukraine-Many-big-losers-few-real-winners/ (accessed on 24 May 2022).

Countryeconomy. 2022. Central Bank Key Rates. Available online: https://countryeconomy.com/key-rates (accessed on 25 June 2022).

CSB. 2022. Exports and Imports by Countries (CN at 2-Digit Level) 2005–2022. Available online: https://data.stat.gov.lv/pxweb/en/OSP_PUB/START__TIR__AT__ATD/ATD020/ (accessed on 24 May 2022).

Delfi. 2022. Valdība lems par 350 miljonu eiro atbalsta pakotni iedzīvotājiem. Available online: https://www.delfi.lv/news/national/politics/valdiba-lems-par-350-miljonu-eiro-atbalsta-pakotni-iedzivotajiem.d?id=54472250 (accessed on 25 June 2022).

DeStatis. 2022. Pressrelease #172 from 20 April 2022. Available online: https://www.destatis.de/EN/Press/2022/04/PE22_172_61241.html (accessed on 24 May 2022).

ECB. 2020. The Impact of the COVID-19 Pandemic on the Euro Area Labour Market. Available online: https://www.ecb.europa.eu/pub/economic-bulletin/articles/2021/html/ecb.ebart202008_02~{}bc749d90e7.en.html (accessed on 25 June 2022).

ECB. 2022a. Inflation in the Near-Term and the Medium-Term. Available online: https://www.ecb.europa.eu/press/key/date/2022/html/ecb.sp220217_1~{}592ac6ec12.en.html (accessed on 24 May 2022).

ECB. 2022b. Interview with Isabel Schnabel, Member of the Executive Board of the ECB, Conducted by Jan Mallien and Frank Wiebe on 29 April 2022. Available online: https://www.ecb.europa.eu/press/inter/date/2022/html/ecb.in220503~{}10f5563a10.en.html (accessed on 24 May 2022).

ECB. 2022c. ECB Staff Macroeconomic Projections for the Euro Area. Available online: https://www.ecb.europa.eu/pub/projections/html/ecb.projections202203_ecbstaff~{}44f998dfd7.en.html#toc2 (accessed on 24 May 2022).

ECB. 2022d. Assessing Corporate Vulnerabilities in the Euro Area. Available online: https://www.ecb.europa.eu/pub/economic-bulletin/articles/2022/html/ecb.ebart202202_02~{}7a61e442be.en.html (accessed on 28 May 2022).

Elbahnasawy, Nasr G., and Michael A. Ellis. 2022. Inflation and the Structure of Economic and Political Systems. *Structural Change and Economic Dynamics* 60: 59–74. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0954349X2100151X (accessed on 24 May 2022). [CrossRef]

Estrada, Mario Arturo Ruiz, and Evangelos Koutronas. 2022. The Trade Suffocation and Investment Desgrowth in Europe from the Russian Military Aggression on Ukraine. Project: An Alternative Multidimensional Economic Modelling for Economics. Available online: https://www.researchgate.net/profile/Mario-Ruiz-Estrada/publication/361119172_The_Trade_Suffocation_and_Investment_Desgrowth_in_Europe_from_the_Russian_Military_Aggression_on_Ukraine/links/629e8a436886635d5cc578fc/The-Trade-Suffocation-and-Investment-Desgrowth-in-Europe-from-the-Russian-Military-Aggression-on-Ukraine.pdf (accessed on 25 June 2022).

Eurofound. 2022. First Impacts of the Ukrainian Crisis on Employment in the EU. Available online: https://www.eurofound.europa.eu/publications/article/2022/first-impacts-of-the-ukrainian-crisis-on-employment-in-the-eu (accessed on 25 June 2022).

Eurostat. 2009. Trends in Consumer Price Inflation. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Trends_in_consumer_prices&oldid=15540#:~{}:text=In%202008%2C%20the%20highest%20ever,recorded%20for%20the%20euro%20area (accessed on 25 June 2022).

Eurostat. 2013. September 2013. Euro Area Unemployment Rate at 12.2%, EU28 at 11.0%. Available online: https://ec.europa.eu/eurostat/documents/2995521/5164586/3-31102013-BP-EN.PDF/ac1c16fc-35ba-4d32-a0a0-91e9409eda8f (accessed on 25 June 2022).

Eurostat. 2022a. Annual Inflation Stable at 7.4% in the Euro Area. Up to 8.1% in the EU. Available online: https://ec.europa.eu/eurostat/documents/2995521/14636041/2-18052022-AP-EN.pdf/9ac63755-4fec-2a70-c149-32edbda92849#:~{}:text=European%20Union%20annual%20inflation%20was,office%20of%20the%20European%20Union (accessed on 24 May 2022).

Eurostat. 2022b. Russia-EU—International Trade in Goods Statistics. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Russia-EU_%E2%80%93_international_trade_in_goods_statistics#:~{}:text=Highlights&text=In%202021%2C%20Russia%20was%20the,of%20goods%20(7.5%20%25) (accessed on 24 May 2022).

Eurostat. 2022c. Euro Area Unemployment at 6.8%, EU at 6.2%. Available online: https://ec.europa.eu/eurostat/documents/2995521/14613608/3-03052022-AP-EN.pdf/36631a07-778c-efb0-01f2-8a052bde985e?t=1651561306689 (accessed on 25 June 2022).

Gartner. 2022. 2022 CIO Agenda: Create an Action Plan to Master Business Composability. Available online: https://www.gartner.com/en/information-technology/insights/cio-agenda/cio-agenda-ebook (accessed on 24 May 2022).

Goodhart, Charles A., Udara Peiris, Dimitrios P. Tsomocos, and Xuan Wang. 2022. Corporate Legacy Debt, Inflation, and The Efficacy of Monetary Policy. CEPR Discussion Papers 16799, C.E.P.R. Discussion Papers. Available online: http://eprints.lse.ac.uk/112955/1/Goodhart_corporate_legacy_debt_inflation_published.pdf (accessed on 28 May 2022).

Grauvogel, Julia, and Hana Attia. 2019. Easier in Than Out: The Protracted Process of Ending Sanctions, Number 5 | 2019 | ISSN: 1862-3581. Available online: https://www.giga-hamburg.de/en/publications/giga-focus/easier-in-than-out-the-protracted-process-of-ending-sanctions (accessed on 24 May 2022).

Gressel, Gustav. 2022. Ukraine: Time for the West to Prepare for the Long War. European Council on Foreign Relation. April 19. Available online: https://ecfr.eu/article/ukraine-time-for-the-west-to-prepare-for-the-long-war/ (accessed on 24 May 2022).

GWN. 2022. Gas Prices in Europe Are Expected to Triple. Available online: https://goodwordnews.com/gas-prices-in-europe-are-expected-to-triple/ (accessed on 28 May 2022).

ILO. 2022. The Impact of the Ukraine Crisis on the World of Work: Initial Assessments. May 11. Available online: https://www.ilo.org/europe/publications/WCMS_844295/lang--en/index.htm (accessed on 25 June 2022).

International Monetary Fund. 2019. The IMF and the Greek Crisis: Myths and Realities. September 30. Available online: https://www.imf.org/en/News/Articles/2019/10/01/sp093019-The-IMF-and-the-Greek-Crisis-Myths-and-Realities (accessed on 25 June 2022).

International Monetary Fund. 2022a. Regional Economic Outlook Europe ALFRED KAMMER Director. European Department IMF. April 22. Available online: https://www.imf.org/en/News/Articles/2022/04/22/tr04222022-european-department-regional-economic-outlook-press-briefing (accessed on 24 May 2022).

International Monetary Fund. 2022b. Regional Economic Outlook, April 2022, Europe: War Sets Back the European Recovery. International Monetary Fund. Available online: https://www.elibrary.imf.org/view/book/9798400208904/9798400208904.xml (accessed on 9 May 2022).

International Monetary Fund. 2022c. How War in Ukraine Is Reverberating Across World's Regions. March 15. Available online: https://blogs.imf.org/2022/03/15/how-war-in-ukraine-is-reverberating-across-worlds-regions/ (accessed on 24 May 2022).

International Monetary Fund. 2022d. Europe's Job Retention Schemes Contained Unemployment, But Challenges Remain. By IMF European Department. March 30. Available online: https://www.imf.org/en/News/Articles/2022/03/30/cf-europe-job-retention-schemes-contained-unemployment (accessed on 25 June 2022).

KPMG. 2022. Global Economic Outlook. The Global Outlook: Without Respite. KPMG April 2022. Available online: https://home.kpmg/xx/en/home/insights/2022/04/global-economic-outlook.html (accessed on 24 May 2022).

Krugman, Paul. 2022. How the Putin shock might affect the world economy. *The New Jork Times*. March 8. Available online: https://www.nytimes.com/2022/03/08/opinion/putin-russia-ukraine-economy-sanctions.html (accessed on 25 June 2022).

Latvijas Banka. 2022a. Latvijas Bankas Prognozes. Aktualizēts: 30.03.2022. Available online: https://www.bank.lv/darbibas-jomas/monetaras-politikas-istenosana/prognozes (accessed on 24 May 2022).

Latvijas Banka. 2022b. Forecasts of Latvijas Banka. Updated: 10.06.2022. Available online: https://www.bank.lv/en/tasks/task-monetary-policy/forecasts (accessed on 25 June 2022).

Latvijas Banka. 2022c. Key Macroeconomic Indicators. Updated: 10.06.2022. Available online: https://www.bank.lv/en/statistics/stat-data/key-macroeconomic-indicators (accessed on 25 June 2022).

Latvijas gāze. 2022. The New Natural Gas Tariffs for Households Enter into Force on 1 July. Available online: https://lg.lv/en/news/the-new-natural-gas-tariffs-for-households-enter-into-force-on-1-july (accessed on 25 June 2022).

Lim, Weng Marc, Markson Wee Chien Chin, Yaw Seng Ee, Chorng Yuan Fung, Carolina Sandra Giang, Kiat Sing Heng, Melinda Lian Fah Kong, Agnes Siang Siew Lim, Bibiana Chiu Yiong Lim, Rodney Thiam Hock Lim, and et al. 2022. What is at stake in a war? A prospective evaluation of the Ukraine and Russia conflict for business and society. *Global Business and Organizational Excellence*, 1–14. [CrossRef]

McKinsey. 2022. War in Ukraine: Lives and Livelihoods, Lost and Disrupted. Available online: https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/war-in-ukraine-lives-and-livelihoods-lost-and-disrupted (accessed on 28 May 2022).

NDTV. 2022. Russia-Ukraine War Could Last 10 Years: UK Foreign Secretary. April 28. Available online: https://www.ndtv.com/world-news/russia-ukraine-war-could-last-10-years-uk-foreign-secretary-2927511 (accessed on 24 May 2022).

Oxford Analytica. 2022. Geopolitical Divisions Will Lower Potential GDP Growth, Oxford Analytica—Emerald Expert Briefings. Available online: https://www.emerald.com/insight/content/doi/10.1108/OXAN-DB268718/full/html (accessed on 24 May 2022).

Papava, Vladimer. 2022. On The Main Challenges of the Architecture Transformation of the World Economy–OpEd. *Eurasia Review*. March 31. Available online: https://www.eurasiareview.com/31032022-on-the-main-challenges-of-the-architecture-transformation-of-the-world-economy-oped/ (accessed on 25 June 2022).

Pestova, Anna, Mikhail Mamonov, and Steven Ongena. 2022. *The Price of War: Macroeconomic Effects of the 2022 Sanctions on Russia*. London: VoxEU, CEPR Policy Portal, Available online: https://www.zora.uzh.ch/id/eprint/218249/1/Macroeconomic_effects_of_the_2022_sanctions_on_Russia.pdf (accessed on 25 June 2022).

Prohorovs, Anatolijs, and Julija Bistrova. 2022. Labour Share Convergence in the European Union. *Journal Economies*. (the article is under reviewing).

RaboResearch. 2022. War in Europe: Is Recession Now Inevitable? March 3. Available online: https://economics.rabobank.com/globalassets/documents/2022/20220303_war_in_europe_recession_inevitable.pdf (accessed on 24 May 2022).

Reeves, Martin, Saumeet Nanda, Kevin Whitaker, and Edzard Wesselink. 2020. Becoming an All-Weather Company. BCG Henderson Institute. Available online: https://bcghendersoninstitute.com/becoming-an-all-weather-company-35d819aa99fe (accessed on 24 May 2022).

Roubini, Nouriel. 2022. Russia's War and the Global Economy. Available online: https://www.project-syndicate.org/onpoint/russias-war-and-the-global-economy-by-nouriel-roubini-2022-02 (accessed on 25 June 2022).

S&P Global. 2022. European Supply Shock: Sectors and Countries at Most Risk (17 March 2022 Vaiva Seckute). Available online: https://ihsmarkit.com/research-analysis/european-supply-shock-sectors-and-countries-at-most-risk.html (accessed on 24 May 2022).

Serhan, Yasmeen. 2022. The War in Ukraine Is Just Beginning, By Yasmeen Serhan. Available online: https://www.theatlantic.com/international/archive/2022/03/how-long-will-ukraine-russia-war-last/627036/ (accessed on 24 May 2022).

Statista. 2022. Harmonized Index of Consumer Prices (HICP) Inflation Rate of the European Union in March 2022, by Country. Available online: https://www.statista.com/statistics/225698/monthly-inflation-rate-in-eu-countries/ (accessed on 24 May 2022).

The Business Times. 2022. German Unemployment Drops Least in Year Amid War in Ukraine. Available online: https://www.businesstimes.com.sg/government-economy/german-unemployment-drops-least-in-year-amid-war-in-ukraine (accessed on 25 June 2022).

The Guardian. 2022. Ban on Russian Gas Would Plunge Germany into Recession, Warns Bundesbank. Available online: https://www.theguardian.com/business/2022/apr/22/russian-gas-ban-germany-recession-bundesbank (accessed on 24 May 2022).

Trading Economics. 2022. Euro Area GDP Growth Rate. Available online: https://tradingeconomics.com/euro-area/gdp-growth (accessed on 24 May 2022).

Tradingeconomics. 2022a. Italy Government Bond 10Y. Available online: https://tradingeconomics.com/italy/government-bond-yield (accessed on 25 June 2022).
Tradingeconomics. 2022b. Italy Government Debt to GDP. Available online: https://tradingeconomics.com/italy/government-debt-to-gdp (accessed on 25 June 2022).
Tradingeconomics. 2022c. Italy General Government Debt. Available online: https://tradingeconomics.com/italy/government-debt (accessed on 25 June 2022).
Tradingeconomics. 2022d. Greece Government Debt to GDP. Available online: https://tradingeconomics.com/greece/government-debt-to-gdp (accessed on 25 June 2022).
TVNET. 2022. Raidījums: Malkas cena turpina pieaugt; pircējiem nākas gaidīt nedēļām ilgi. Available online: https://www.tvnet.lv/7514832/raidijums-malkas-cena-turpina-pieaugt-pircejiem-nakas-gaidit-nedelam-ilgi (accessed on 25 June 2022).
United Nations. 2022. Global Impact of War in Ukraine on Food, Energy and Finance Systems. BRIEF NO.1. April 13. Available online: https://news.un.org/pages/wp-content/uploads/2022/04/UN-GCRG-Brief-1.pdf (accessed on 24 May 2022).
Wall Street Journal. 2022. The Wall Street Journal. May 4. Available online: https://www.wsj.com/livecoverage/federal-reserve-meeting-inflation-rate-may-2022 (accessed on 24 May 2022).
Washington Post. 2022. European Sanctions on Russia Will Cost Europe, Too, Early Signs Show. Available online: https://www.washingtonpost.com/world/2022/03/02/europe-russia-sanctions-backlash/ (accessed on 24 May 2022).
World Bank. 2022a. Ukraine War to Cause Biggest Price Shock in 50 Years–World Bank. Available online: https://www.bbc.com/news/business-61235528 (accessed on 24 May 2022).
World Bank. 2022b. Food and Energy Price Shocks from Ukraine War Could Last for Years, 2022, WB. Available online: https://www.worldbank.org/en/news/press-release/2022/04/26/food-and-energy-price-shocks-from-ukraine-war (accessed on 24 May 2022).
World Bank. 2022c. Ukraine war 'will mean high food and energy prices for three years'. *The Guardian*. April 26. Available online: https://www.theguardian.com/business/2022/apr/26/ukraine-war-food-energy-prices-world-bank (accessed on 24 May 2022).

Journal of
Risk and Financial Management

MDPI

Article

Tracing the Optimal Level of Political and Social Change under Risks and Uncertainties: Some Lessons from Ancient Sparta and Athens

George E. Halkos *, Emmanouil M. L. Economou * and Nicholas C. Kyriazis

Department of Economics, University of Thessaly, 38333 Volos, Greece
* Correspondence: halkos@uth.gr (G.E.H.); emmoikon@uth.gr (E.M.L.E.)

Abstract: The present paper is inspired by the notions of "financial risk" and "financial uncertainties" and transfers their basic reasoning to social science analysis; that is, it develops a theoretical analysis in order to explain social and political change. We know that the degree of social and political change depends on the set of established institutions in a society. Societies can face two extremes: volatility, e.g., rapid changes that lead to instability, which increases the risk of a system or regime collapsing, or rigidity, which does not permit necessary adaptation and change and thus may again increase the risk of the regime collapsing. Thus, an optimal (or ideal) point of change is between the two extremes, permitting change that is neither too sudden and fast nor too slow and inflexible. To illustrate this, we analyze two cases from ancient Greece: Sparta, as a society and state with too many institutional checks and balances that led to rigidity and collapse, and Athens, which in the 5th century BCE had an institutional setting with very limited checks and balances, which again led to near collapse until the late 5th century BCE, when new institutions that were related to some efficient checks and balances were introduced that enabled the state to survive in a world of changing circumstances and balances of power.

Keywords: institutional volatility; institutional rigidity; political and social change; regime risks; Sparta; Athens

JEL Classification: D72; H50; N13; P26

Citation: Halkos, George E., Emmanouil M. L. Economou, and Nicholas C. Kyriazis. 2022. Tracing the Optimal Level of Political and Social Change under Risks and Uncertainties: Some Lessons from Ancient Sparta and Athens. *Journal of Risk and Financial Management* 15: 416. https://doi.org/10.3390/jrfm15090416

Academic Editor: Thanasis Stengos

Received: 26 August 2022
Accepted: 15 September 2022
Published: 19 September 2022

1. Introduction

The issue of the risks that are associated with the survivability (or not) of a regime when social and political change occurs has occupied a central place in the writings of philosophers, historians, political scientists and economists, since at least the time of the ancient Greek thinkers such as Plato, Aristotle and Thucydides. Such questions were revived by 17th–18th-century authors such as John Locke, David Hume, Adam Smith, Montesquieu, Voltaire and Jean Jacques Rousseau, and then by Karl Marx, his followers and his opponents, as well as institutional economists. Among the latter, North (1981, 1990, p. 2) argues that inadequate institutions are uprooted while effective ones survive and so there is a gradual evolution of more adequate forms of economic, political and social organization. This actually means that, when a sociopolitical regime fails to provide effective policy measures, for example, on economic policy issues, the risk of its final collapse increases significantly, as Toynbee (1951), Acemoglu and Robinson (2013), Economou and Kyriazis (2019) and others have observed (from different perspectives).

In general, an environment that promotes the evolution of the institutions of a society, that is, institutional change, through an innovative spirit is crucial for economic growth (North 1990; Grossman and Helpman 1991). Societies with economic institutions that encourage innovative ideas can achieve prosperity in the long run (Acemoglu and Robinson 2013). This is because according to Kuznets (1966), innovations that are related to structural change are directly related to economic growth. This also reduces the risk of a regime collapsing due to internal or external factors. Regarding the internal factors, this, for example, is related to failures in establishing efficient institutions that are necessary for long-term economic growth such as property right protection, the rule of law in general and sound money (Rodrik 2000, 2007; Acemoglu and Robinson 2013; Economou and Kyriazis 2019). Regarding external factors, this is related to failures in properly handling an external threat such as a military invasion that actually tests the endurance and the ability of reaction or survivability of a system or regime, as Toynbee (1951) argued in his seminal work, *The Study of History*.

For example, the radical transformation of the Athenian economy of the Classical period (508–322 BCE) from an agrarian economy of primitive structures to an extroverted maritime economy where market forces (by the standards of the time) and international trade values were present was the result of the need to face an imminent external threat, the Persian invasion. This was achieved by the Athenians, under their democratic leader, Themistocles, by introducing a vast shipbuilding program of warships so as to repel the enemy. The shipbuilding program was related to many risks and financial uncertainties, but in the end, it was crowned with success, saving the Athenian city-state itself and transforming the structure of the Athenian economy from an agrarian into a maritime and commercially extroverted one (Kyriazis and Zouboulakis 2004; Halkos and Kyriazis 2010; Economou and Kyriazis 2019; Bitros et al. 2020; Economou et al. 2022).

In this paper, being inspired by the notions of “financial risk” and “financial uncertainties”, we transfer their basic reasoning to social science analysis, that is, give them a sociopolitical diastasis, so as to explain how and why the need for institutional change may lead to the survival and the evolution of a sociopolitical regime or, alternatively, to its collapse. To the best of our knowledge, this is an innovative approach. As will be analyzed in Section 2, there is a vast bibliography that connects financial risk to sociopolitical risk. For example, in their important contributions, Alesina et al. (1992) and Alesina and Perotti (1996) have analyzed the nexus between economic growth (due to, among other factors, efficient financial institutions) and a political regime that safeguards political stability. In this paper we expand the above findings by transferring the basic reasoning of the well-established notions of financial uncertainties, risk and volatility to social science analysis.

In particular, in our paper, we focus on the ancient Athenian and the Spartan sociopolitical systems during the Classical times (508–323 BCE), as case studies, and then draw some inferences for modern societies. We argue that there are both benefits and drawbacks for any society characterized by flexible political and economic institutions such as the ancient Athenian one. The benefits are related to the ability of the sociopolitical regime to evolve and progress through time. The drawbacks are that flexibility may slip to extreme volatility[1], leading to detrimental policy outcomes and the risk that finally the system and the regime may collapse as a whole. On the other hand, a rigid political system, like that of Sparta, may ostensibly look strong enough during a specific, even long, period of time, but the risk of collapse may prove high when the socioeconomic circumstances that demand significant institutional changes cannot be met by such a rigid regime.

Thus, the ideal scenario is an “optimal” level of change, that is, the efficient level of change which permits a society’s adaptation to new ideas and institutional proposals without a breakdown of the regime, which ensures its durability in the long run. In other words, we argue that the best option that significantly diminishes the risk of a regime collapsing is a system which is neither too volatile and flexible regarding sociopolitical change nor too rigid, tending to appear ossified.

Our paper is organized as follows: Section 2 explains the nexus between financial and political risks for a state. In principle, there is a strong correlation between the two. This section links the findings of the international bibliography regarding the nexus between financial and political risks to the discussion that follows in the next sections. Furthermore, Section 3 defines the concepts of “rigidity”, “volatility” and the “optimal level of change” regarding the structure of a sociopolitical regime by relating them to the logic of "risk" and "uncertainties" as concepts that are transferred from the field of financial economics into sociopolitical analysis.

Based on the logic of Sections 2 and 3, Sections 4–7 analyze as case studies the political institutions of Sparta and Athens. They explain how these institutions functioned in practice and, furthermore, why Sparta failed to introduce institutional changes and finally collapsed, while Athens managed to effectively transform its 5th century BCE institutions so as to successfully handle the challenges of the 4th century. It is further argued that the Athenian political system came to an end, not because of internal pressure and collapse, but because of defeat by an "external" factor, the superior Macedonian navy in 322 BCE, in the Lamian War.

Our argument is that during the period 508/7 to 416 BCE, Athens provides an example of a sociopolitical change that illustrates the extreme of too high volatility, while during the period 415–322 BCE it illustrates the ideal level of change, defined as "optimal". Sparta, on the other hand, illustrates the other extreme, that of rigidity during the period 508–372, followed by its inability to reform in the period 371–192, which led to its final collapse. Section 8 provides a synopsis of the previous sections. It also discusses the limitations and the opportunities for further research under the logic that was described throughout this paper.

2. The Nexus between Financial and Sociopolitical Risks: A Review

There is a vast bibliography that connects financial risk (FR) to political risk (PR).[2] For example, Kirikkaleli (2020), by using the wavelet coherence technique with quarterly data from 1984 to 2018, showed how PR affected the economic and financial risk in Venezuela. He found that PR had a strong power for explaining economic risk from 1995 to 2005, while between 1984 and 2010, economic risk and PR were positively correlated at different frequency levels. In general, he found that in the long run, changes in PR significantly led to changes in FR in Venezuela.

The strong nexus between PR and the financial sector is highlighted in the definition provided by Bekaert et al. (2014) where the relation of PR with the value of investments is highlighted. In general, academics argue that PR is influential on financial markets (Stiglitz 1993) as well as concerning firms (Faccio et al. 2006; Tran et al. 2021). Barro (1991) has found that in periods of political turbulence, entrepreneurs with a risk-averse mentality are likely to shift their investments to other countries that offer more relative safety and less risk. Thus, when PR increases in general, investors try to diversify their portfolios between different countries so as to minimize their FRs (Kinsella and Rubins 2005). In the same vein, Tabassam et al. (2016) have found that political uncertainties in a country change the nature of investment and the patterns of spending, which has negative effects on economic growth. Thus, in principle, political institutions that induce policy uncertainties are harmful to investments in the long run (North 1990; Weingast 1993; Georgiou et al. 2015).

PR also reduces the likelihood of entry in a particular market of a multinational company for making a foreign direct investment, as various authors have found so far (Henisz and Delios 2001; Delios and Henisz 2003; Uhlenbruck et al. 2006). PR is also related to central banks and the stock market. Major central banks such as the Federal Reserve and the European Central Bank support that PR factors are found to be influential on volatility embedded in financial intermediaries and have exerted major shocks on commercial banks. Furthermore, authors such as Charfeddine and Al Refai (2019) have found that stock markets are adversely affected by rising political tension in both domestic markets and in the international arena. In particular, Wang (2019), utilizing synthetic control methods (SCMs), estimated the causal impact of a political shock, the Taiwan Strait Crisis in 1995–1996, on Taiwan's stock market. His major findings are consistent with existing literature which reveals the causality between political uncertainty and stock prices in emerging markets.

Another part of the bibliography relates FR to the quality of governance. Such risks derive from uncertainties such as the extent to which a country is governed by the rule of law, the effectiveness of a country's court system, the issue of enforcement (or not) of contracts between firms and their local trading partners and violation of intellectual property rights (Uhlenbruck et al. 2006).

Another sub-category that connects FR to PR is related to the so-called *geopolitical risk* of a country in the international system. According to the important contribution of Caldara and Iacoviello (2022), geopolitical risk is associated with wars, terrorist acts and tensions between states that affect the normal and peaceful course of international relations. Relevant research on the topic includes the work of Kyriazis and Economou (2021) who explored whether the Turkish geopolitical uncertainty exhibits significant linkages with fluctuations in the Turkish lira values against other currencies during the Erdoğan administration period (March 2003 up to the present). Their econometric outcomes revealed that geopolitical risk uncertainty led to the devaluation of the Turkish lira against the US dollar, the Swiss franc and the Swedish krona but in a non-significant manner, whereas mixed results emerged concerning the impacts of macroeconomic or financial variables. In a similar study, Economou and Kyriazis (2021) have explored the nexus between Russia and Turkey regarding their geopolitical uncertainty during the Putin administration era in Russia. They implemented the innovative Caldara and Iacoviello indices and the vector autoregressive (VAR) methodology. The nexus between geopolitical risk in Turkey and Russia was found to be unstable. They found that uncertainty in Turkey constitutes both a negative and a positive determinant of geopolitical stability in Russia, depending on the time horizon of the impact. Russia could take advantage of Turkish positive effects in the medium run. This could be alarming for investors but could also prove beneficial as they should not invest in Russian assets when the country's geopolitical risk is elevated due to Turkey's geopolitical instability. Table 1 summarizes the above analysis.

Table 1. The nexus between financial and sociopolitical risks.

Financial Risk (FR) and Political Risk (PR)	Bibliographical References
PR negatively affects financial markets	(Stiglitz 1993)
A strong nexus between PR, economic risks and FR in general	(Alesina et al. 1992; Alesina and Perotti 1996; Georgiou et al. 2015; Kirikkaleli 2020)
A strong nexus between PR and economic governance	(North 1990; Weingast 1993; Uhlenbruck et al. 2006)
A strong nexus between PR and the financial sector in general	(Bekaert et al. 2014)
PR negatively affects firms' profitability and survivability	(Faccio et al. 2006; Tran et al. 2021)
PR negatively affects investments	(Barro 1991; Kinsella and Rubins 2005; Tabassam et al. 2016)
PR negatively affects the stock market	(Charfeddine and Al Refai 2019; Wang 2019)
PR discourages multinational companies from entering a market	(Henisz and Delios 2001; Delios and Henisz 2003; Uhlenbruck et al. 2006)
Geopolitical risks, which are directly related to PR, affect the financial sector and markets	(Caldara and Iacoviello 2022; Economou and Kyriazis 2021; Kyriazis and Economou 2021)

As a general comment, the above findings certify that there is a strong and negative correlation between FR and PR on a global scale. As a final comment, the nexus between political instability and financial crisis applies also vice versa. This is verified by Geithner (2014), among other authors, who found that financial crises have led to changes in governments and smaller or greater political instability, as in the cases of Mexico, Thailand, Indonesia and South Korea.

3. Defining Further Political and Social Change in Democratic Regimes

The question of the appropriate level of political and social change for a society, not too sudden and fast, not too slow and inflexible, is a central issue of all polities, old and new, democratic and authoritarian.

It can be argued in general that the institutional set-up of democracies (either representative and, perhaps, with "direct" elements) is more adaptable to institutional change, which may be related to the adoption of new constitutional arrangements, because the alteration of parties in government introduces new strategies, some of which are adopted and may prove to be successful. This is a kind of continuous trial and error procedure. It cannot be excluded that new strategies may be introduced in other forms of government (political regimes) such as oligarchies or monarchies. However, what historical examples seem to show is that authoritarian regimes (oligarchies, centralized monarchies and empires, or 20th-century communist states) are less reluctant to introduce new strategies, because this may trigger a chain of events that cannot be controlled, thus increasing the risk for the collapse of such regimes. This built-in rigidity proved in most cases to be short-sighted, leading to the regimes' eventual breakdown. Such cases, of course, in the modern era include both the Soviet and the Eastern European socialist regimes during and after the Cold War (Kornai 1990; Gorlizki and Khlevniuk 2020). On the other hand, dictatorships or kingships may appear to be stable, but in many cases have proved to be unable to adapt, increasing the risk to their long-term survivability and, finally, leading to their overthrow by revolutions (for example the French and the Russian revolutions) or internal collapse (as in the case of the communist states after 1989). Thus, the appropriate level of sociopolitical change of institutions is difficult to specify in actual terms: we therefore introduce it as a theoretical benchmark. We focus on the rate of change of the institutions of governance.

Political decision-makers of a state society face a set of choices in each period of time. According to the institutional setting of each state, the decision makers are different. They may be a king with his supporting elite; a more diffuse oligarchic body; an all-powerful citizen body in ancient direct democracies; or what Achen (2022) and Hodgson (2022), among others, define as "mixed" cases of regimes. We postulate that the decision makers try to maximize their individual welfare, which may be a mixture of economic values and "intangible" ones such as prestige, religion, a particular way of life and freedom.

Thus, according to who the decision makers are in each case, a different welfare function is maximized, but in all cases, the decision makers are postulated as being rational, trying to maximize their individual welfare as they perceive it. Thus, the function to be maximized at each period is:

$$\max\,(S_1, S_2, \ldots, S_n) \tag{1}$$

where $S_1 \ldots S_n$ are the various strategies on the decision makers choice set, linked to different welfare levels. During each period of time, new strategies may be introduced and old ones may drop out, as no longer practicable. External and internal circumstances modify the set of strategies available, a point that will be made clear in the following discussion of the case studies.

During each period of time, the decision makers adopt one of the available strategies, and this shapes the decision tree probabilities as illustrated in Figure 1 below. Figure 1 shows that if strategy S_2 is chosen at time period 1, to the exclusion of strategies S_2 and S_3, then at time period 2 the strategies S_4, S_5, S_6 and S_7 are given as options, and if S_5 is chosen, then at time period 3 strategies S_8 to S_{10} are available, etc.

Thus, in theory, there may be two extremes: In the first case, it is possible that many new strategies can be introduced. New ones can be adopted with high frequency and older ones may be discarded. This could lead to a political system that is very adaptable, but also too fickle and variable, with a high degree of uncertainty and low predictability which increases the risk of its collapse.

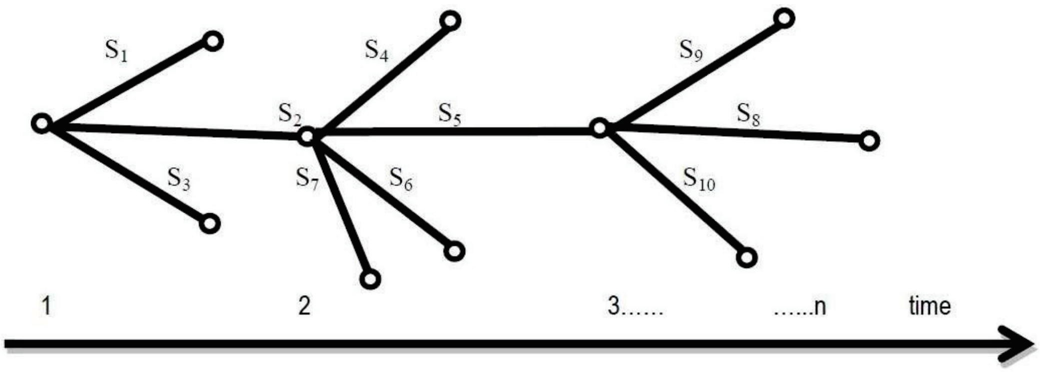


Figure 1. Decision tree probabilities and the adoption of new strategies.

In the second case, the system seems to be characterized by great stability and predictability, such as the strategies S_1 and S_3 with very few new strategies being introduced, and even fewer being adopted, no breaks, low adaptability to new ideas, being unable to successfully face new challenges, a system that tends to be "ossified". Such a system resembles a deterministic model with perfect predictability in future steps. Such a regime may become too unstable if circumstances require significant institutional changes that the system is not prepared to perform. In such a scenario, the risk of collapse of the regime increases significantly.

The above discretion resembles the structure of the Athenian and Spartan regimes. During the 5th century BCE, Athenian democracy's institutional arrangements and post-constitutional policies were decided directly by the Assembly of citizens, and the distinction between rules for policy making and policy measures was sometimes blurred. As will be argued later in Section 5, in the late 5th century (from 415 and later), some new institutional mechanisms were introduced that could be seen, to some degree, as "checks and balances" mechanisms to the Assembly's decision. For example, Themistocles' Naval Law of 483/2 BCE inaugurated a "turning to the sea" policy, which, as already mentioned in Section 1, in practical terms, had to do with the introduction of a vast ship-building program of 200 warships so as to face the imminent Persian threat. According to North's (1990, 1991) spectacles, this "turning to the sea" strategy could be seen as the driving force for institutional change through creating opportunities for the players in the society to change the rules of the game. On the other hand, it may be argued that Sparta had a 'constitution', known as the *Great Rhetra*, which included rigid checks and balances (Cartledge 1987; Rahe 2016).

Thus, Figure 2 below actually describes three cases of sociopolitical regimes. Vertical axes exhibits the 'regime change' and horizontal axis the 'time period' that is, how much time is need in the long run for a regime to change. Regime 1 (red line) denotes great stability, almost rigidity over time. Such a sociopolitical regime, where few new strategies are introduced and even fewer are adopted, results in a system characterized by stability and predictability, policy continuity and low adaptability to new ideas. This system tends to be "ossified" and may not be able to succeed against new challenges which may cause its collapse. Such a system may give the impression that the risk of collapse in the long run is extremely low.

Regime 1 describes Sparta whose institutional structures proved to be rigid. Although this had benefits for a considerable period of time, 750–372 BCE where Sparta was the mightiest city-state in metropolitan Greece, as will be further explained in Section 3, the system of governance ultimately failed to adapt after 371 when the historical necessity for institutional sociopolitical changes arose. The Spartan regime faced a combination of shocks which eventually led to its collapse. Cases of rigidity are found in many other historical cases, such as the Kingdom of Croesus in Lydia in the 6th century BCE, Spain in the 16th century, France during the 16th-18th centuries where the *ancien regime* broke down with the French Revolution, the Indian Mughal Empire which broke down at the beginning of the 18th century and the Soviet regime in 1989 (Economou and Kyriazis 2019; Gorlizki and Khlevniuk 2020; Ober 2022).

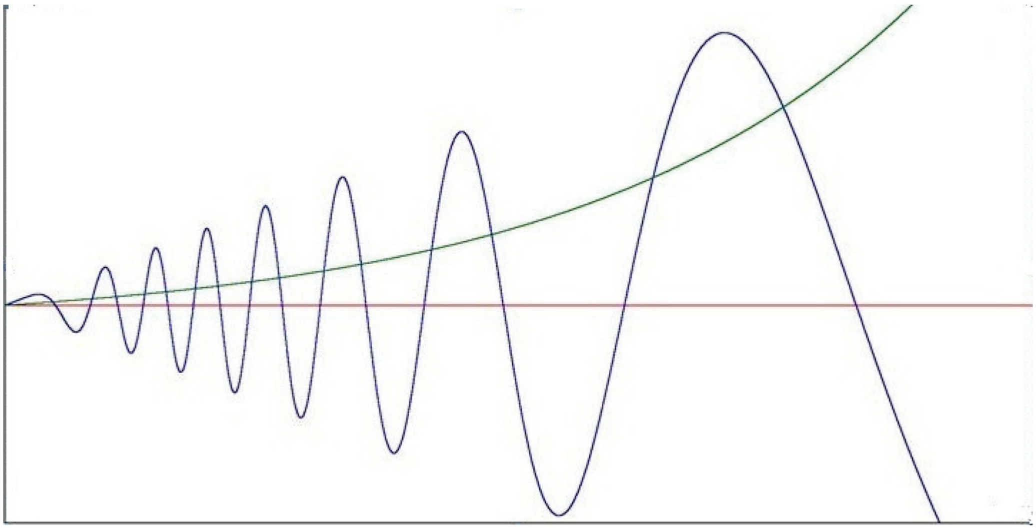


Figure 2. The three different cases of a regime change. Figure based on Economou et al. (2016).

Regime 2, as depicted by the ongoing series of (blue color) curves in Figure 2, denotes extreme volatility, where its intensity increases through time. The Regime 1 case is characterized by great certainty and predictability. By contrast, Regime 2 is characterized by great uncertainty and no predictability, where a large number of new strategies are introduced and adopted with high frequency, while older arrangements and institutions are discarded. This could lead to a political system that is adaptable but is subject to the risk of becoming too fickle, with a high degree of uncertainty and low predictability. This regime is characterized by extreme volatility resulting in great uncertainty, and the risk of collapse in the long run is extremely high. As it will be further argued in Sections 6 and 7, 5th century BCE Athens comes close to Regime 2, where many new ideas and proposals were offered and implemented which caused the regime to make wrong decisions (e.g., the Sicily campaign) leading to the verge of collapse (after the loss of the Peloponnesian War).

At this point, we have to clarify that we do not argue that a sociopolitical regime introducing a multitude of proposals always leads to instability and collapse. This further means that if a political regime is hit by a multitude of shocks (in our case, proposals as presented in Figure 1), it may make good sense to respond with a range of policy measures, each one addressing the source of a different shock. If these policy measures prove effective, the political–social regime avoids collapse. Thus, what is for sure is that the introduction of a large number of proposals by a regime may increase its volatility and the risk of collapse, but this does not (necessarily) mean that the regime will finally become unstable, or even collapse.

The third type, Regime 3 in Figure 2 (green exponential curve), is the intermediate case, the efficient one, which displays gradual (institutional) change through time; new ideas are introduced and accommodated, leading to an evolutionary regime change. Thus, the Regime 3 case represents the "optimal" or ideal rate of change that prevents collapse, which is a high probability in the previous two cases due both to internal and external shocks. Faced with external and internal shocks, the systems of the first and second cases are more likely to collapse than the system of the third case which is more adaptable. Figure 2 illustrates the three cases. We consider the Regime 3 type as the ideal case scenario. It satisfies the following criteria: (i) it makes the regime durable and prevents collapse; (ii) it is adaptable to external and internal shocks (accommodating them and thus again preventing collapse); and thus, as a result, (iii) it prevents failure. An "optimal" or ideal level of change is thus a benchmark that guarantees the longevity of a regime and safeguards it against the risk of collapse when facing internal or external challenges. A system of checks and balances (see Buchanan and Tullock 1962), in modern terminology, would thus be considered as successful if it comes close to the ideal level of social and political change. As will be further explained in Sections 6 and 7, 4th century Athens resembles the third case, when Athens introduced a number of institutional reforms that restricted, to some degree, the unfettered power of the *demos* (the people) which then resulted in less volatile policy changes.

4. Sparta: The Structure of Political and Economic Institutions

This section explains the institutional rigidness of the Spartan sociopolitical regime by taking into account various aspects.

4.1. Political Institutions

Sparta may claim to have had the first ever constitution, the so-called *Great Rhetra*, introduced allegedly in approximately 750 BCE by the lawgiver Lycurgus. The institutional setting of Sparta incorporated a complex system of checks and balances: two kings provided by two royal families, the Agiadai and the Eurypodidai; the Council of the *Five Ephors* who were high-ranking officials, elected annually and sharing power with the Spartan kings; the *gerousia* of 30 (the Senate) which included the two kings; and the *Apella*, or popular Assembly of all male Spartan citizens, the so-called *homoioi*, meaning "equals", that is, full citizens (Cartledge 1987; Gagarin 2013, pp. 231–32).

By the mid-10th century BCE, Sparta on the Eurotas River, in the southwest part of the Peloponnese, had been settled by the Dorian branch of Greeks, which enslaved the earlier population, turning them into *helots*. They then expanded to neighboring Messenia, which they conquered after war, also turning the Messenians into helots. The word helot is rendered by modern historians as slave, but some see it as a kind of serf (Cartledge 2001). The population of *helots* outnumbered the Spartans by an unknown multiple. The Spartans lived in fear of a revolt by the helots (which happened, for example, in the 460s BCE), and in the face of this threat, transformed themselves into "the first hoplite republic in Greece through a unique package of reforms" (Cartledge 1987, p. 15). *Hoplites* were free citizen soldiers of ancient Greek city-states. To defend their city-states, they financed their military equipment (primarily spears, breastplates, greaves, helmets and shields) themselves and participated in the *phalanx* formations, which were the main military organization under which *hoplites* fought on the battlefields (Kagan and Viggiano 2013).

The essence of the social reforms that were adopted after the emergence of the phalanx type of warfare and organization was as follows: firstly, Spartan citizenship depended on the possession of an allotment (Greek: *klaros*) from the production of which a Spartan "equal" (citizen) was required to pay a fixed minimum contribution on a monthly basis to a common mess (common meal, Greek: *syssition*). Thus, citizenship depended on economic qualifications, something which was true of most of the city-states of the period, where full citizenship depended on *hoplite* status. What was different was that Spartans provided not only presumably their hoplite equipment, but also the fixed "common mess contribution". Secondly, all Spartans served as *hoplites* in the army. Thirdly, all Spartans participated in the popular Assembly known as *Apella*, one of the main institutional bodies, and they had the right to vote on peace or war, by acclamation. Fourth, socially, in order to produce the desired uniformity, solidarity and unquestioning obedience and discipline, a comprehensive system of public education, the *agoge*, was rigidly enforced by the central authorities. Sparta was the only Greek city-state with such a kind of public education system.

The Spartan social classification included the additional class of *perioikoi* (those living around the main city-state, Sparta) who lived in about 80 settlements scattered all over Lacedaemon and Messene. The *perioikoi* were free and had local autonomy but had no public political rights whatsoever. They probably made some financial contribution, possibly in the form of taxes, in kind, to the Spartan state, and they participated in the army. By the 4th century, they served as *hoplites*, brigaded in the same units as the Spartan *hoplites*, and probably had to provide their arms and armor by their own means (Cartledge 1987, pp. 15–16; 2001; Cartledge and Spawforth 2002).

The Spartan political system (constitution), was considered already by Aristotle (1944, pp. 20–35) and modern authors such Rahe (2016, chp. 2) as a mixed system, characterized by Rhodes (2007, pp. 58–60) as a peculiar kind of oligarchy. One of the many particularities of the Spartan constitution was that during the Classical period, it retained a system of two kings when monarchies no longer existed except on the fringes of the Greek world, as in Macedonia and Epirus. Sparta inaugurated an early type of constitutional monarchy since each king acted as a check on the other but more importantly they were both checked by the *ephors*. The kings were the military commanders during war. Once back in Sparta, the king could, in effect, be called to account by trial before his peers, in which case, the 30-member *gerousia* sat as a mixed ad hoc supreme court together with the five *ephors*. Through their exclusive power of jurisdiction, the *gerontes* (old men, being over 60 years of age) in these and other capital cases wielded a strong influence on the conduct of foreign and domestic policy. Even strong kings such as Cleomenes I and Pausanias (in 403 BCE) faced trials. The *gerousia* was the most senior political body, and its members served for life. The *gerontes* were elected by the Assembly of the Spartan citizens.

Probably only those from aristocratic families could be elected as *gerontes*. Apart from functioning as a Supreme Court, empowered to put even kings on trial, the *gerousia* had two other crucial functions. Together with the Council of the *ephors*, it had general supervision of Spartan laws, which were unwritten. This entailed two separate functions: (i) interpretation of the laws, and since they were unwritten, the *gerousia* had a wide latitude in sharing power as far as the legal system is concerned, which was somewhat akin to common law practice; (ii) according to the *Great Rhetra*, the *gerousia* was empowered to both introduce and refuse to introduce proposals at the Assembly, thus controlling its agenda and policy making in general. In modern terminology, the *gerousia* enjoyed the right of both prospective and retrospective veto (Cartledge 1987, pp. 121–25; Rahe 2016).

The five *ephors* were elected for a year and probably could not be re-elected. They were possibly elected, in the same manner as the *gerousia*, or they were somehow nominated or co-opted by the *gerousia*. The *ephors* had executive, judicial and religious duties. They controlled contacts between Spartans and foreigners, received foreign embassies, had the power to send away foreigners residing in Sparta (a procedure known as *xenelasia*), granted or withheld permission to Spartans to reside abroad and engaged in negotiations with other states, having more or less the functions of a modern foreign office.

In times of war, two *ephors* accompanied a king on campaigns, supervising him, and they were empowered to give orders to overseas commanders. They also determined the age-groups of the army to go on campaigns and were responsible for logistics, e.g., provisions and technical staff, and presumably, siege engines and their servants. At home, they supervised the *agoge* (the military training of boys until they reached manhood) and policy functions, including the supervision of the *helots*' to ensure that no revolts took place. The *crypteia* (hidden force), the Spartan "Secret Service", was also administered by them. Together with the *gerousia*, they had the power to bring kings to trial. Their judicial powers included preliminary hearings of capital cases in order to decide whether a case should come to court. They also judged all civil suits individually. In non-capital criminal cases, they might impose and exact fines ad hoc, and the proceeds were supposed to go to the public treasury, of which they were the comptrollers. They supervised the few taxes raised on Spartans and foreign tribute. One of the *ephors* presided at the Assembly's meetings. Thus, in today's terminology, the *ephors* combined elements, of the executive branch (foreign affairs, finance, defense) and the judiciary.

The Spartan Assembly had very limited powers: it made the final decisions concerning war and peace and legislation, but without the power of initiative, since the agendas and the proposals were introduced by the *gerousia* placed before it by the presiding *ephor*. No ordinary citizen had the right to speak (Cartledge 1987, pp. 128–29).

In addition to the above institutional setting, Spartan society was characterized by respect for the laws and the officials, military discipline, and the system of common messes to which all citizens were obliged to participate (and contribute) up to the age of 60. The system of *agoge* was deemed to create a uniformity of values and societal norms, which placed the state above the individual. Thus, it is probably no exaggeration to call the Spartan state and society both collectivist and militarist.

4.2. Economic Institutions

Sparta's peculiar institutional system extended to the economy, which was also unique among Greek city-states.:Being professional soldiers, Spartans did not have civilian occupations as in other city-states, such as farmers, handicraft workers, employment in services (e.g., trade, banking, shipping like in Athens) or entrepreneurs (bankers, factory owners, ship-owners, etc.). Sparta was, to a higher degree than the average city-state, an agricultural state, but agriculture was carried out by *helots* and *perioikoi* (Cartledge and Spawforth 2002). We do not have sufficient extant comparative evidence to see if this absenteeism led to lower efficiency and productivity in Sparta, compared to other states, where farmers, being owners and laborers at the same time, had incentives to maximize output and efficiency.

The services sector was very small, mainly concerning trade, probably again in the hands of *perioikoi* and possibly some foreigners. There were no Spartan banking services and almost no commercial shipping activity. Foreign trade was also relatively small since the Spartan way of life frowned at luxuries or works of art and had autarky as its ideal. Manufacturing and handicrafts were again in the hands of *helots* and *perioikoi*, and possibly some foreigners, but catered mainly to basic needs, such as weapons, utensils and furniture, rather than luxuries and goods for export. An even greater peculiarity was the aversion to a monetized economy. Sparta originally had only unwieldy iron coins instead of silver or gold. However, this changed during the Hellenistic period, and the Spartan economy was monetized (Hodkinson 2009).

Once more, we do not have sufficient extant evidence, but the inference we can draw from ancient sources and modern interpretations seems to indicate that, similarly to the political institutions, the Spartan economy too was stationary. In contrast to Athens and other city-states, it recorded no growth and was ill-adapted to face serious challenges.

5. Evaluation of Sparta's Political and Social System

In this section, we argue that Sparta, with its systemic checks and balances, failed to undertake an efficient management of geopolitics and internal affairs; failed to follow the new trends in military organization, though a militarist society itself; and finally failed to make the necessary socioeconomic changes so as to avoid its final collapse.

5.1. Rigidity in Managing Geopolitics and Internal Political Affairs

The Spartan politico-economic system was probably in place already by the second half of the 7th century BCE and endured to the end of the 3rd century BCE, when it broke down. During this period, Sparta faced the challenges of the second Persian invasion 480/479, the third Messenian War (the uprising of *helots* in 464), the Peloponnesian War during 431–404 BCE and the so-called Corinthian (395–387 BCE) and Theban wars (378–375 BCE), among other challenges.

After the long-drawn Peloponnesian War (431–404 BCE), Sparta was at the apex of its power and took over the "Athenian empire".[3] It established Spartan supervisors (*harmostes*) in many city-states. This opened Sparta to foreign influences. In general, *harmostes* mismanaged their tasks, so much so that less than a decade from the end of the Peloponnesian War, old allies such as Corinth and Thebes took arms against Sparta, together with a revitalized Athens. The wars continued with changing alliances up to 362 BCE, when the Spartan army suffered a crushing defeat at Mantineia (after two previous defeats at Tegeira and Leuctra in 371 BCE), where the Spartan king and commander Kleombrotos died (Cartledge 2001). After Leuctra, the Theban commander, Epaminondas, freed the

Messenian *helots* from Spartan rule. Sparta thus lost half its territory, which it never regained, and became a second-rate power. During the following years, Sparta chose a policy of isolation instead of participation in the coalition of southern Greeks against the Macedonians. Next, Sparta chose not to participate in the Macedonian-led Greek coalition against Persia, established by Philip II and Alexander the Great, but in 330 BCE chose to challenge alone the Macedonian regent in Greece (Alexander fighting against Persia in Asia).

During the second half of the 3rd century BCE, King Agis IV tried to introduce reforms (including cancelation of debts), but the *ephors* had him murdered. His successor, King Cleomenes III, introduced sweeping reforms that abolished the old system. He abolished the ephors and the *gerousia*, but he governed with almost absolute power, much like a Roman emperor, at the same time introducing social reforms by granting Spartan citizenship to many non-citizens such as the *perioikoi*, as well as land reform. However, he again confronted the Macedonians alone and suffered a crushing defeat by King Antigonos Doson, again at Sellasia in 222. As a result, he was forced into exile. Continuing its isolationist policy, Sparta chose not to participate in the reborn Achaean League. It became a member of the Achaean League only by force in 191 BCE (Economou 2020). Finally, the system completely collapsed when Sparta was converted into a kind of dictatorship under Machanidas in 211 and under Nabis in 207 and subsequently fell under Roman rule.

We believe that this mismanagement of foreign policy and geopolitics which affected Sparta's geopolitical decline during the 4th and 3rd centuries BCE must be attributed to the rigidity of the Spartan political system which did not accommodate change after external and internal shocks. Spartan policymakers proved unable to handle Sparta's grand strategy after 403 BCE. The reason is that the Spartan political regime was characterized by a peculiar system of checks and balances, which discouraged the implementation of new ideas, protected the ruling elite and prevented reformers from challenging the status quo. Only small, occasional shifts of power could take place when an innovative king reigned (such as Cleomenes I).

5.2. Failure to Adapt to New Trends in Military Organization

At the same time, Sparta showed an inability to adopt military innovations. It did not introduce military reforms, for example equipping itself with a strong cavalry, which proved to be necessary during the Peloponnesian War. It neglected the new, lightly equipped skirmishers (*peltastai*), at whose hands it suffered a grave defeat at Lechaion during the Corinthian war, in 390 BCE, where the Athenian general Iphicrates, who introduced innovative military tactics, managed to crush a whole Spartan unit. Of course, Sparta developed its navy during the Peloponnesian War, but in the tactical field, in which the Spartans were supreme to the end of the 5th century, it failed to adapt during the 4th century, thus suffering some crucial defeats at the hands of the outnumbered Thebans and Boeotians, when General Epaminondas introduced his so-called "oblique" approach military tactic.

During the 4th century, Greek armies combined heavy and light infantrymen with cavalry, while the Spartan army remained mainly a heavy infantry one. However, even in this, we believe that it suffered critical defeats because the percentage of professional Spartan soldiers in its ranks diminished sharply, as the number of Spartan *homoioi* fell dramatically from about 8000 in the 480s BCE to about 3000 at the beginning of the 4th century. So, in order to maintain the numbers of the *phalanx*, *perioikoi* and *neodamodeis* increasingly served in the ranks of the phalanx, alongside Spartans (Cartledge 2001).[4] However, *perioikoi* and *neodamodeis* were not professional soldiers, and thus we presume that they were no better than the average hoplite of other Greek city-states.

In the end, the Spartans through the decision taken by king Cleomenes III adopted the superior Macedonian type of phalanx, but this came too late to change the fate of Sparta. The Spartan army was crushed in Sellasia by the Macedonian and the Achaean League armies, and then Sparta, until its subjugation by Rome in 146 BCE, declined almost completely and played only a minor role in Greek affairs (Economou 2020).

5.3. Failure to Establish a Series of Necessary Socioeconomic Changes

In addition, the Spartan institutional setting failed to accommodate internal pressures. A common theme among contemporary writers was the diminishing number of the *homoioi* citizens, which did not happen, at least not to the same degree, in other city-states. This shortage of manpower (*oliganthropia*) intrigued not only ancient writers such as Herodotus and Aristotle but also modern writers (Crawford and Whitehead 1983, pp. 503–4; Hodkinson 2009). The Peloponnesian War forcefully brought Sparta into contact with the mainstream of Greek socioeconomic developments, thus creating tensions beneath the constitutional facade, which emphasized the equality of the Spartan citizens and their ideal of concord (same-mindedness, *homonoia*) among themselves.

In fact, there always had been rich and poor Spartans, but the Peloponnesian War increased this disparity: the wealthy upper class possessed land in addition to the *klaros* (allotment), the minimal land plot attributed to every Spartan citizen, which enabled it to contribute to the common mess and provide itself with *hoplite* equipment. Rich Spartans, in order to preserve their social status, apparently aimed to limit their male offspring to a single son and would seek to marry off their daughters to the sons of other wealthy families. Poor Spartans also tended to restrict their offspring, because if they had many sons, then their *klaros* would have to be divided among them, thereby becoming too small to permit any of them to retain the necessary minimal mess contribution. These social forces of the institutional setting brought about a gradual decrease in population. Another source of demographic decrease was manpower losses, which were severe also for Sparta, especially during the last period of the Peloponnesian War in the naval battles which they lost.

During the Peloponnesian War, Sparta had to acquire fleets in order to combat Athens in the Aegean. However, fleets were very costly, both to build and to maintain, and the rudimentary non-monetized (at that time) Spartan economy had to turn to contributions, in money (coins), to their allies and even more to the Persian Empire. So, in a decade, Sparta became de facto a monetized economy, a tendency which continued after the war with the Spartan governors and garrisons in many city-states, paid in coins. However, the distribution of this monetary inflow was uneven: admirals, ship captains, governors and officers serving abroad received the lion's share, increasing the gulf between rich and poor.

Thus, wealthy Spartans had now a surplus of liquid funds to invest. However, since the backward Spartan economy did not offer a diversity of investment opportunities, the only investment outlet they had within the Spartan state was the acquisition of additional land. Poor Spartans had fallen into debt by raising a "mortgage" on the security of the crops raised on their land, or on the land itself, in return for coined money which the rich Spartans had now in surplus. This led to land concentration and downward social mobility from Spartan status to *ypomeiones* (those who did not have sufficient land for the mess contributions and their hoplite equipment) which then explains the diminishing number of Spartan citizens during and after the Peloponnesian War (Cartledge 1987, pp. 169–70). This is why we stated above that the monetization of the Spartan economy "corrupted" the Spartan citizens, which could be seen as a "curse in disguise".

5.4. A Comprehensive Assessment of the Socioeconomic and Political System of Sparta

We conclude that the Spartan regime, with its checks and balances, diffusion of power, respect for authority and tradition, lack of incentives and alternatives for economic activity, the stifling of the initiative of outstanding individuals and the mismanagement of geopolitics and grand strategy, was the outcome of an endogenous systemic rigidity. As a result, the lack of adaptability both to internal change and external shocks increased the risk of the collapse of the regime and finally led to its breakdown at the end of the 3rd century BCE.

Under A. Toynbee (1951).logic it could be argued that Spartan history shows a rigid regime that fails to adapt when facing pressures from new, unexpected, external and internal conditions. Sparta won the Peloponnesian War but proved ill-adapted to lead the Greek world. An interesting counterfactual would have been for Sparta to have willingly decided to participate as a member state of the Achaean League during the second half of the 2nd century BCE or later during the first half of the 1st century BCE. However, Sparta chose to continue its isolationist policy, and it finally joined the Achaean League only for a short period of time and only after it was forced to do so, during the 192–146 BCE period (Economou 2020).

In terms of our model, this would have meant the adoption of a new strategy after 404 BCE. Additional strategies to address internal social problems would have been to prohibit rich Spartans from buying land belonging to the poor, thus avoiding downward social mobility, and to grant full citizenship to *perioikoi* and *helots* before the Cleomenes III reforms in the 240s BCE, by which time it was too late. Thus, as our analysis indicates, having failed to introduce new strategies at crucial periods, Sparta declined, and its political system broke down by the end of the 3rd century, when Machanidas and Nabis established dictatorships. This is not to deny the earlier success of the Spartan institutions. During the period 650–372 BCE, Sparta was the most powerful city-state in ancient Greece. What can be argued is that when circumstances changed in certain important ways, the same institutional rigidity that had been Sparta's strength became, instead, a liability.

6. Athens: An Adaptable Society with Collective Decision-Making Procedures

6.1. The Political Structure

Since the political history of Classical Athens is relatively well-known (see, among others, Ober 2008, 2017; Cartledge 2018), we will not analyze it in detail. The Athenian democracy emerged in 510–507 BCE through the reforms of Cleisthenes and was strengthened gradually through the introduction of the Athenian Naval Law (*Decree of Themistocles*) which initiated Athens' "turn to the sea" shipbuilding program in order to repel the Persian invasion and opened the way to full political rights for all citizens. This "turn to the sea" strategy along with the reforms of Ephialtes and Pericles gradually reshaped Athens' socioeconomic and political institutions from an aristocratic dispensation during the 6th century BCE to a direct democracy and a free market type of economy during the 5th and 4th centuries (Economou and Kyriazis 2019; Bitros et al. 2020).

Ober (2008) explains the emergence of democracy in Athens as a shift of power between, on the one side, the aristocracy and, on the other side, the middle (*hoplites*) and poor (*thetes*) classes, leading to a bottom-up democratic and perhaps spontaneous revolution in 510–507 BCE. The aristocracy found the expansion of democratic rights to be in its interest. In contrast to Athens, Sparta had a militarily focused, aristocratic government that had little incentives to grant voting rights to the agricultural workers (Fleck and Hanssen 2006)

By the mid-5th century BCE, Athens was the most complete direct democracy the world had known. Major decisions about all aspects of external and domestic policy were made by the Assembly of the demos through regular voting. Judicial powers were vested in the popular courts to which jurors were appointed by lot, as was also the case with the Council of Five Hundred which set the agenda for the Assembly's meetings. The Athenian political system was based on equality of political rights. Citizens over 20 had the right to vote in the people's Assembly. Full political rights were obtained at the age of 30 when a citizen was allowed to present himself at the annual election of magistrates (Bitros et al. 2020).[5] All citizens had the same opportunity to occupy state posts through selection by lot.

The only posts not filled by sortition (selection by lot) were those deemed to demand a proven ability and experience by the occupant, such as those of the 10 generals (leaders of the army, requiring military ability) and judges who judged special commercial cases known as *dikai emporikai* (Cohen 1992). Equality was encapsulated in the principles of isonomia (equality in front of the law), *isegoria* (equality of all to speak, e.g., to introduce proposals for policy, laws and decrees to be voted by the Assembly) and *isocratia* (equality of political rights, to be appointed to public posts including the courts) (Bitros et al. 2020; Cartledge 2022; Ober 2022).

Contrary to the Spartan system, the 5th century Athenian political setting until 415 BCE had almost no provisions for checks and balances, since the people (*demos*) were supreme in political, military, judiciary and public finance policy making. As it is further argued in Section 6, this entailed the risk of too high volatility in decision making.

6.2. The Economy

The Athenian economy was flourishing and growing during the 5th and 4th centuries BCE. It was highly diversified and export-oriented. It had established a regime of property rights protection and implementation of contracts, and for the first time in history, the contribution of the manufacturing (e.g., metallurgy; carpentry; shipyards and dockyards for warships, the *triremes*, and merchant ships; pottery; artifacts) and services sectors (in areas such as trade, shipping, insurance services for commerce, maritime loans and banking services, even primitive versions of joint-stock companies) was higher in GDP terms than the overall annual agricultural production, giving thus to Athens the characteristics of the "first modern economy".[6]

During the 5th century, Athens benefited from the revenues of its empire. Remarkably, during the next century, after Athens lost its empire, it made an economic and financial recovery by introducing various economic policy measures promoting trade. Under these policy measures, Piraeus became the entrepôt of the Mediterranean world and Athens one of the main manufacturing centers, mainly for high-quality and value-added goods, such as jewelry, works of art, utensils, pottery, honey and figs (Cohen 1992; Halkos and Kyriazis 2010; Halkos et al. 2021). After the reforms of Eubulus and Lycurgus, who introduced extensive public work programs as opposed to a continuing war strategy, public revenues increased from a low of 130 *talents*[7] in the 350s to 400 *talents* in the 340s and 1200 *talents* in the 330s; the latter figure is comparable to the imperial revenue of the 440s (Economou and Kyriazis 2016; Bitros et al. 2020).

7. Reducing Systemic Risks When Institutional and Political Change Occurs: Athens during the 4th Century

Contrary to the Spartan system of diffusion of power, the Athenian system was characterized by a concentration of power in the hands of the citizens, through the Assembly (mainly) and the courts (Ober 2008, 2017). Possibly, there was some kind of a relative blocking (or vetoing) power as far as the courts are concerned during the 5th century BCE (the extreme volatility period according to our analysis). However, what is for sure is that the veto power of the courts increased significantly during the 4th century BCE. This became a necessity due to some very serious and costly strategic mistakes that the Athenians made, becoming also victims of demagogy, such as, first, the Sicilian campaign which lead to the Athenian disaster at Syracuse in 415–413; second, the condemnation of the victorious generals after the Battle of Arginousae in 406; and third, the volatility in decision making concerning foreign policy which is related to a high propensity of the Athenians to shift their alliances very often. According to Schwartzberg (2004, pp. 316–17), this made them look quite unreliable, in comparison to the Spartans, and this led to the loss of many important allies since the Athenians were considered untrustworthy.

Recognizing the weaknesses of the democratic constitution of the previous period, with its too high volatility in decision making and abrupt policy swings, the Athenians introduced institutional changes which made decision making more stable.

The most important institutional reforms were the following three: The first, which took the form of a checks and balances mechanism, was the *graphe paranomon* procedure, introduced in 415 BCE, through which any citizen could sue (within a year) the proposer (known as *ho boulomenos*) of a decree with a charge that his proposal was unconstitutional, that is, it contradicted with the existing laws of the state, or that his proposal did not follow the correct procedures when it was submitted to the Athenian Assembly (Schwartzberg 2004, pp. 319–20). Under the same reasoning, any citizen could sue the proposer of a law through the *graphe nomon me epitedeion theinai* procedure, with a charge that the proposal was unconstitutional, or that it did not follow the correct procedures when it was submitted to the Athenian Assembly (Canevaro 2018). Fines were very severe for those who proposed such laws or decrees. For a *graphe paranomon* that was rejected by the *nomothetai* (see below), the proposer could be fined with 1000 drachmae, while for a rejected *graphe nomon me epitedeion theinai*, the proposer could even face a death penalty.[8]

Linked to these two kinds of *graphai*, at the end of the 5th century, possibly in 404 BCE, two new special boards were introduced, the *nomothetai* (meaning those who are establishing new laws) and the *thesmomothetai*. The *nomothetai* board was drawn from the annual group of judges serving in the *Heliaia* courts. It could be a large gathering of judges. For example, Demosthenes (*Against Timocrates*, 24.27) mentions a case of 1001 judges.

The *nomothetai* could be summoned (i) if the Assembly called them for revisions to the existing laws, (ii) if an Athenian citizen proposed a change in the laws or (iii) if they were asked by the *thesmothetai* (see below). Thus, the main duty of the *nomothetai* was to certify that a decree or law proposed by a citizen did not contradict any other existing laws (Harris 2013, pp. 244–45; Canevaro 2018). If the proposed laws did contradict any other existing laws, they were rejected. It was actually a two-tier board where the first board (elected by the Council of the Five Hundred) had also the right to propose new laws while the second (elected from the Athenian municipalities—demes) voted on whether to ratify each law and add it to the existing laws (Harris 2013). Thus, this body, in addition to the role of inspecting the "constitutionality" of the proposed laws against the current legislation, also had the power to propose new laws.

The *thesmothetai* were an institution with similar or supplementary duties to those of *nomothetai*. According to the orator Aeschines (*Against Ctesiphon*, 3.38–3.40), the *thesmothetai* were a board of six *archons*[9], whose duties were to review the existing laws annually to determine whether there were inconsistencies, duplications or invalid laws in force and, if so, to inform the *nomothetai* to rectify them. Schwartzberg (2004, p. 320), among others, accepts this interpretation. These boards were connected with the restoration of democracy in 404 BCE.

The above description shows that the *nomothetai* actually had a veto right on law proposals to be decided by the Assembly. Even if a proposal was accepted and became state law, it could be annulled at a later time by procedures run (i) by the Assembly, (ii) the citizens (*graphe paranomon*, *graphe nomon me epitedeion theinai*) and (iii) the *nomothetai*. Harris (2013, p. 244) argues that in the 5th century, the constitution of the Athenians placed that legislative power in the hands of the Assembly, but in the 4th century, that legislative power was placed in the hands of the Assembly and the *nomothetai*. The introduction of these checks and balances institutions does not mean that the Athenian citizens (the *demos*) were deprived of their ability to make decisions. *Demos* remained the superior political institution concerning decision making. Harris (2013, p. 244 ftn. 78) clarifies this by arguing that:

> "Even though the *nomothetai* gave final approval to the laws in the fourth century, it was the Assembly that initiated all laws and submitted them to the *nomothetai*. Their authority derived from the citizens of Athens, which is why they were called the laws of the Athenians (not the laws of the *nomothetai*)."

What we are arguing is that democracy became more moderate, mature and less volatile in the 4th century in comparison to the previous century, which proved beneficial for the survival of the political regime, and Athens avoided the risk of internal collapse at the end.

Returning to the narrative of the main political/military events that took place in the 4th century, after the unsuccessful attempts of the Athenians to establish a viable Second Athenian League which culminated with the so-called Social War during 357–355, Athenian public revenues were falling to 130 *talents* per year, a very low sum in comparison, say, to the Periclean era. However, the innovative spirit of the Athenians then again led them to a new key strategic choice that proved beneficial in the long run. They introduced a pacifist grand strategy by avoiding new war engagements, and they replaced large war expenses with large fiscal expansionary programs of a Keynesian nature, and this led them to impressive economic growth; in 322, after just 33 years, their public revenues rose to 1200 *talents* from 130 in 355 (Economou and Kyriazis 2016; Bitros et al. 2020).

The Athenian democracy was abolished in 322 BCE after Athens was defeated by the Macedonians at sea at the battle of Amorgos. This means that the system did not "fail" or collapse internally like Sparta's, as was argued in Section 4, but that it failed to effectively repel a stronger enemy in military terms, the Kingdom of Macedon. The Athenians also made an unsuccessful effort to expel the Macedonian garrison in the 3rd century during the Chraemonidean War (267–261 BCE) and later managed to do so, paying off the garrison, as an inducement for them to leave Athens. Showing its strong roots, Athenian democracy more or less functioned (but with interruptions) until about the middle of the 2nd century and the Roman conquest (but we lack the information to trace institutional changes), even though Athens was no longer a major power in Greece.[10]

The above historical facts prove that although Athens failed in military terms, it still continued to be a successful example of socioeconomic organization during the Hellenistic period. Thus, in the end, it did not "collapse" as it happened with the Spartan case. Thus, Athenian society during the period 508–404 BCE comes close to the case of Regime 2 in our model, the high-volatility case, where the risk of collapse is extremely high, but after 403 and more profoundly during the period 355–322, it seems to be closer to the case of Regime 3, that is, a system with an optimal level of change that successfully accommodates external shocks and internal changes so that the risk of internal collapse diminishes significantly. By contrast, Sparta, although it managed to remain the greatest power in geopolitical terms in ancient Greece from 650 to 372 BCE, due to the reasons that were described in Sections 4 and 6, failed to adapt.

8. Concluding Remarks

The objective of this paper was to transfer the well-established notions of "financial risk" and "financial uncertainties" to social science analysis, that is, to give them a sociopolitical diastasis. The theoretical implications of the paper are that it defines what kind of institutions are needed so that an efficient level of social and political change takes place so that, on the one hand, the risk of a regime's collapse is avoided and, on the other hand, social prosperity and progress continue in the long run. We have presented an analysis of sociopolitical rigidity versus volatility of a regime that permits a sociopolitical regime to avoid the risk of collapse and adapt both to external shocks and internal change, making it thus durable. The practical implications of the paper are that we have presented Sparta and Athens as two examples, the first for rigidity, the second for high volatility during almost the 5th century (until 415 BCE) and then, during the 4th century, as approaching the ideal or "optimal" level of change.

According to our argumentation, Athens during the 5th century was a very volatile regime, but it avoided collapse during the 4th century due to institutional changes which made the socioeconomic and political structure more moderate regarding decision making. By contrast, Sparta remained in a state of rigidity due to the structure of its institutions, and when some visionary leaders such as Cleomenes III tried to introduce changes, it was too late for Sparta to become successful; this resembles what happened with the Soviet regime when Michael Gorbachev tried to introduce a radical package of institutional reforms of socioeconomic and political nature, *perestroika* and *glasnost*, but he failed since, obviously, the regime could not be saved due to its chronic pathogens and structural weaknesses.

To the best of our knowledge, this is the first attempt to interpret sociopolitical change by simultaneously comparing Sparta and Athens from such an analysis perspective. The limitations regarding this particular case study, Sparta vs. Athens, are related to the absence of cliometric–statistical data for testing our model and hypotheses. We however believe (and hope) that this analysis could be further used by social scientists to interpret not only historical case studies, as happens with this paper, but also recent cases, such as the current Chinese regime or the European Union. For example, based on the logic of the above analysis, an interesting issue for future research would be to try to define the optimal level of economic and sociopolitical reforms that the current Chinese regime needs to introduce so as to further expand its economic might in the future combined with a more democratically oriented package of political institutions. Equally interesting would be to explain how any institutional changes regarding sociopolitical matters may occur in China in the long run based on factors such as the authoritarian characteristics of the Chinese Communist regime.

Another interesting case study for future research is the European Union's further political integration (as a process of sociopolitical change), which is strongly related to the achievement (or not) of high financial performance throughout the union, since this will increase prosperity among the member states in the long run.

We hope that this article sets the appropriate stimuli for further research on related topics in the near future.

Author Contributions: All authors have contributed to all stages of this research. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank the Editor and the referees for their very fruitful comments and suggestions that led to the improvement of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1. A notion that derives from financial economics.
2. Due to space limitations, we cannot retrieve all of this bibliography here, but only a representative part of it.
3. The term refers to the so-called Delian League or Athenian Alliance of the 478–404 period. It was an association/military alliance of more than 300 Greek city-states, under the leadership of Athens, whose initial purpose was to continue fighting the Persian Empire. Gradually, Athens acquired a hegemonic role within it. For a financial aspect regarding the functioning of the Alliance, see Figueira and Jensen (2021).
4. *Neodamodeis* (new citizens) were given citizenship whenever the state deemed it necessary, mainly, we guess, when the city needed to replenish the ranks of the citizens after severe manpower losses, such as after the earthquake of about 461 BCE, the losses from the Peloponnesian War, and the crushing defeats of the Spartan army by the Thebans and Boeotians in 371 and 362 BCE.
5. Only Athenian male citizens enjoyed full political rights. Slaves, *metics* (citizens that came to Athens from other city-states for work) and women were excluded from political participation.
6. See, among others, Cohen (1992), Ober (2008), Halkos and Kyriazis (2010), Lyttkens (2013), Bergh and Lyttkens (2014), Bitros et al. (2020), Economou et al. (2021a, 2021b), Halkos et al. (2021, 2022) and the references cited therein.
7. One *talent* was equal to 6000 drachmae. As a measure of comparison, the daily wage of an unskilled worker during the 4th century BCE was 1.5 drachmae.
8. Due to space limitations, we cannot retrieve all this evidence here, but a detailed analysis of such legal and judicial procedures is provided by Schwartzberg (2004), Harris (2013), Lyttkens et al. (2018) and Canevaro (2018), among others.
9. They were members of the *nine archons*, an ex-aristocratic body.
10. A very interesting historical analysis regarding Athens during the Hellenistic and Roman periods until the time of Mark Antony is provided by Habicht (1999).

References

Acemoglu, Daron, and James Robinson. 2013. *Why Nations Fail*. New York: Crown Business.

Achen, Christopher H. 2022. The Crisis of Democracy: A Self-Inflicted Wound. In *Democracy in Times of Crises Challenges, Problems and Policy Proposals*. Edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis and Athanasios Platias. Cham: Switzerland, pp. 21–40.

Alesina, Alberto, and Roberto Perotti. 1996. Income distribution, political instability, and investment. *European Economic Review* 40: 1203–28. [CrossRef]

Alesina, Alberto, Sule Ozler, Nouriel Roubini, and Philip Swagel. 1992. *Political Instability and Economic Growth*. NBER Working Paper, No. 4173. Cambridge: National Bureau of Economic Research.

Aristotle. 1944. *Politics, vol. 21, Engl. Trnsl. H. Rackham*. Cambridge: Harvard University Press. London: William Heinemann Ltd.

Barro, Robert J. 1991. A Cross-country Study of Growth, Saving, and Government. In *National Saving and Economic Performance*. Edited by B. Douglas Bernheim and John B. Shoven. Chicago: University of Chicago Press, pp. 271–304.

Bekaert, Geert, Campbell R. Harvey, Christian T. Lundblad, and Stephan Siegel. 2014. Political risk spreads. *Journal of International Business Studies* 45: 471–93. [CrossRef]

Bergh, Andreas, and Carl Hampus Lyttkens. 2014. Measuring institutional quality in Ancient Athens. *Journal of Institutional Economics* 10: 279–310. [CrossRef]

Bitros, Geprge C., Emmanouil M. L. Economou, and Nicholas C. Kyriazis. 2020. *Democracy and Money: Lessons for Today from Athens in Classical Times*. London and New York: Routledge.

Buchanan, James, and Gordon Tullock. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Caldara, Dario, and Mateo Iacoviello. 2022. Measuring geopolitical risk. *American Economic Review* 112: 1194–225. [CrossRef]

Canevaro, Mirko. 2018. What was the law of Leptines' really about? Reflections on Athenian public economy and legislation in the fourth century BCE. *Constitutional Political Economy* 29: 440–64. [CrossRef]

Cartledge, Paul. 1987. *Agesilaos and the Crisis of Sparta*. Baltimore: The John Hopkins University Press.

Cartledge, Paul. 2001. *Sparta and Laconia. A Regional History, 1300–362 BC*. London and New York: Routledge.

Cartledge, Paul. 2018. *Democracy: A Life*, 2nd ed. Oxford: Oxford University Press.

Cartledge, Paul. 2022. Greece's Finest Hour? The Democratic Implications of the Battle of Salamis. In *Democracy and Salamis. 2500 Years After the Battle that Saved Greece And the Western World*. Edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis and Athanasios Platias. Cham: Springer, pp. 57–74.

Cartledge, Paul, and Anthony Spawforth. 2002. *Sparta and Lakonia. A Regional History 1300–362 BC*. London and New York: Routledge.

Charfeddine, Lanouar, and Hisham Al Refai. 2019. Political tensions, stock market dependence and volatility spillover: Evidence from the recent intra-GCC crises. *North American Journal of Economics and Finance* 50: 101032. [CrossRef]

Cohen, Edward E. 1992. *Athenian Economy and Society: A Banking perspective*. Princeton: Princeton University Press.

Crawford, Michael H., and David Whitehead. 1983. *Archaic and Classical Greece. A Selection of Ancient Sources*. Cambridge: Cambridge University Press.

Delios, Andrew, and Witold J. Henisz. 2003. Political hazards, experience, and sequential entry strategies: The international expansion of Japanese firms, 1980–1998. *Strategic Management Journal* 24: 1153–64. [CrossRef]

Economou, Emmanouil M. L. 2020. *The Achaean Federation in Ancient Greece: History, Political and Economic Organisation, Warfare and Strategy*. Cham: Springer International Publishing.

Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2016. Choosing peace against war strategy. A history from the ancient Athenian democracy. *Peace Economics, Peace Science and Public Policy* 22: 191–212. [CrossRef]

Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2019. *Democracy and Economy: An Inseparable Relationship Since Ancient Times to Today*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Economou, Emmanouil M. L., and Nikolaos A. Kyriazis. 2021. Spillovers between Russia's and Turkey's geopolitical risk During the 2000–2021 Putin administration. *Peace Economics, Peace Sciense and Public Policy* 28: 29–50. [CrossRef]

Economou, Emmanouil M. L., Nicholas C. Kyriazis, and Loukas Zachilas. 2016. Interpreting Sociopolitical Change by Using Chaos Theory: A Lesson from Sparta and Athens. MPRA Paper No. 76117. Available online: https://mpra.ub.uni-muenchen.de/76117/ (accessed on 10 September 2022).

Economou, Emmanouil-Marios L., Nicholas C. Kyriazis, and Nikolaos A. Kyriazis. 2021a. Money decentralization under direct democracy procedures. The case of Classical Athens. *Journal of Risk and Financial Management* 14: 30. [CrossRef]

Economou, Emmanouil M. L., Nikolaos A. Kyriazis, and Nicholas C. Kyriazis. 2021b. Managing financial risk while performing international commercial transactions. Intertemporal lessons from Athens in Classical times. *Journal of Risk and Financial Management* 14: 509. [CrossRef]

Economou, Emmanouil M. L., Nicholas C. Kyriazis, and Athanasios Platias, eds. 2022. *Democracy and Salamis. 2500 Years After the Battle that Saved Greece And the Western World*. Cham: Springer.

Faccio, Mara, Ronald W. Masulis, and John J. McConnell. 2006. Political connections and corporate bailouts. *The Journal of Finance* 61: 2597–635. [CrossRef]

Figueira, Thomas J., and Sean R. Jensen. 2021. *Hegemonic Finances. Funding Athenian Domination in the 5th Centuries BC*. London: Bloomsbury.

Fleck, Robert K., and F. Andrew Hanssen. 2006. The origins of democracy: A model with application to Ancient Greece. *Journal of Law and Economics* XLIX: 115–46. [CrossRef]

Gagarin, Michael. 2013. Laws and Legislation in Ancient Greece. In *A Companion to Ancient Greek Government*. Edited by Hans Beck. Malden: Willey-Blackwell, pp. 221–34.

Geithner, Timothy. F. 2014. *Stress Test: Reflections on Financial Crises*. New York: Crown Publisher.

Georgiou, Miltiades, Nicholas C. Kyriazis, and Emmanouil M. L. Economou. 2015. Democracy, Political Stability and Economic performance. A panel data Analysis. *Journal of Risk & Control* 2: 1–18.

Gorlizki, Yoram, and Oleg Khlevniuk. 2020. *Substate Dictatorship: Networks, Loyalty, and Institutional Change in the Soviet Union*. New Haven and London: Yale University Press.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.

Habicht, Christian. 1999. *Athens from Alexander to Anthony*. Cambridge, MA and London: Harvard University Press.

Halkos, George E., and Nicholas C. Kyriazis. 2010. The Athenian economy in the age of Demosthenes: Path dependence and change. *European Journal of Law and Economics* 29: 255–77. [CrossRef]

Halkos, George E., Nicholas C. Kyriazis, and Emmanouil M. L. Economou. 2021. Plato as a game theorist towards an international trade policy. *Journal of Risk and Financial Management* 14: 115. [CrossRef]

Halkos, George E., Emmanouil M. L. Economou, and Nicholas C. Kyriazis. 2022. Environmental economics in Classical Athens. *Journal of Institutional Economics*, 1–17. [CrossRef]

Harris, Edward D. 2013. *The Rule of Law in Action in Democratic Athens*. Oxford: Oxford University Press.

Henisz, Witold J., and Andrew Delios. 2001. Uncertainty, imitation, and plant location: Japanese multinational corporations, 1990–1996. *Administrative Science Quarterly* 46: 443–75. [CrossRef]

Hodgson, Geoffrey M. 2022. On the Limits of Democracy. In *Democracy in Times of Crises Challenges, Problems and Policy Proposals*. Edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis and Athanasios Platias. Cham: Springer Nature, pp. 71–92.

Hodkinson, Stephen. 2009. *Property and Wealth in Classical Sparta*. Swansea: Classical Press of Wales.

Kagan, Donald, and Gregory F. Viggiano. 2013. *Men of Bronze: Hoplite Warfare in Ancient Greece*. Princeton: Princeton University Press.

Kinsella, Stephan, and Noah D. Rubins. 2005. *International Investment, Political Risk and Dispute Resolution*. Oxford: Oxford University Press.

Kirikkaleli, Dervis. 2020. Does political risk matter for economic and financial risks in Venezuela? *Economic Structures* 9: 1–10. [CrossRef]

Kornai, Janos. 1990. *The Road to a Free Economy*. New York: W.W. Norton & Company.

Kuznets, Simon. 1966. *Modern Economic Growth: Rate, Structure, and Spread*. New Haven and London: Yale University Press.

Kyriazis, Nikolaos A., and Emmanouil M. L. Economou. 2021. The Impacts of geopolitical uncertainty on turkish lira during the Erdoğan administration. *Defence and Peace Economics* 33: 731–50. [CrossRef]

Kyriazis, Nicholas C., and Michel Zouboulakis. 2004. Democracy, sea power and institutional change: An economic analysis of the Athenian Naval Law. *European Journal of Law and Economics* 17: 117–32. [CrossRef]

Lyttkens, Carl Hampus. 2013. *Economic Analysis of Institutional Change in Ancient Greece. Politics, Taxation and Rational Behaviour*. London and New York: Routledge.

Lyttkens, Carl Hampus, George Tridimas, and Anna Lindgren. 2018. Making direct democracy work: A rational-actor perspective on the grapheparanomon in ancient Athens. *Constitutional Political Economy* 29: 389–412. [CrossRef]

North, Douglass C. 1981. *Structure and Change in Economic History*. New York: W.W. Norton and Company.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

North, Douglass C. 1991. Institutions. *The Journal of Economic Perspectives* 5: 97–112. [CrossRef]

Ober, Josiah. 2008. *Democracy and Knowledge. Innovation and Learning in Classical Athens*. Princeton: Princeton University Press.

Ober, Josiah. 2017. *Demopolis: Democracy before Liberalism in Theory and Practice*. Cambridge: Cambridge University Press.

Ober, Josiah. 2022. Herodotus on Rationality and Cooperation Before Salamis. In *Democracy and Salamis. 2500 Years After the Battle that Saved Greece And the Western World*. Edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis and Athanasios Platias. Cham: Springer Verlag, pp. 33–56.

Rahe, Paul A. 2016. *The Spartan Regime. Its Character, Origins, and Grand Strategy*. New Haven and London: Yale University Press.

Rhodes, Paul J. 2007. *The Greek City States: A Source Book*. Cambridge: Cambridge University Press.

Rodrik, Dani. 2000. Institutions for high-quality growth: What they are and how to acquire them. *Studies in Comparative International Development* 35: 3–31. [CrossRef]

Rodrik, Dani. 2007. *One Economics. Many Recipes. Globalization, Institutions and Economic Growth*. Princeton and Oxford: Princeton University Press.

Schwartzberg, Melissa. 2004. Athenian democracy and legal change. *American Political Science Review* 98: 311–25. [CrossRef]

Stiglitz, Joseph E. 1993. The role of the state in financial markets. *The World Bank Economic Review* 7: 19–52. [CrossRef]

Tabassam, Aftab H., Shujahat H. Hashmi, and Faiz U. Rehman. 2016. Nexus between political instability and economic growth in Pakistan. *Procedia-Social and Behavioral Sciences* 230: 325–34. [CrossRef]

Toynbee, Arnold J. 1951. *A Study of History. Vol. 4, Breakdowns of Civilizations*. Oxford: Oxford University Press.

Tran, Dung Viet, M. Kabir Hassan, Ahmed W. Alam, Luca Pezzo, and Mariani Abdul-Majid. 2021. Economic policy uncertainty, agency problem, and funding structure: Evidence from US banking industry. *Research in International Business and Finance* 58: 101470. [CrossRef]

Uhlenbruck, Klaus, Peter Rodriguez, Jonathan Doh, and Lorraine Eden. 2006. The impact of corruption on entry strategy: Evidence from telecommunications projects in emerging economies. *Organization Science* 17: 402–14. [CrossRef]

Wang, Huiqiang. 2019. The causality link between political risk and stock prices: A counterfactual study in an emerging market. *Journal of Financial Economic Policy* 11: 338–67. [CrossRef]

Weingast, Barry. 1993. Constitutions as governance structures: The political foundations of secure markets. *Journal of Institutional and Theoretical Economics* 149: 286–311.

Journal of
Risk and Financial Management

Article

Copula Modelling to Analyse Financial Data

Paul R. Dewick * and Shuangzhe Liu

Faculty of Science and Technology, University of Canberra, Canberra 2617, Australia; shuangzhe.liu@canberra.edu.au
* Correspondence: paul.dewick@canberra.edu.au

Abstract: Copula modelling is a popular tool in analysing the dependencies between variables. Copula modelling allows the investigation of tail dependencies, which is of particular interest in risk and survival applications. Copula modelling is also of specific interest to economic and financial modelling as it can help in the prediction of financial contagion and periods of "boom" or "bust". Bivariate copula modelling has a rich variety of copulas that may be chosen to represent the modelled dataset dependencies and possible extreme events that may lie within the dataset tails. Financial copula modelling tends to diverge as this richness of copula types within the literature may not be well realised with the two different types of modelling, one being non-time-series and the other being time-series, being undertaken differently. This paper investigates standard copula modelling and financial copula modelling and shows why the modelling strategies in using time-series and non-time-series copula modelling is undertaken using different methods. This difference, apart from the issues surrounding the time-series component, is mostly due to standard copula modelling having the ability to use empirical CDFs for the probability integral transformation. Financial time-series copula modelling uses pseudo-CDFs due to the standardized time-series residuals being centred around zero. The standardized residuals inhibit the estimation of the possible distributions required for constructing the copula model in the usual manner.

Keywords: financial; non-stationary; time-series; copula; dependence; risk; univariate; bivariate

Citation: Dewick, Paul R., and Shuangzhe Liu. 2022. Copula Modelling to Analyse Financial Data. *Journal of Risk and Financial Management* 15: 104. https://doi.org/10.3390/jrfm15030104

Academic Editor: George Halkos

Received: 27 January 2022
Accepted: 22 February 2022
Published: 25 February 2022

1. Introduction

In investigating copula modelling, it is often stated within the literature that the Gaussian (Normal) copula is to blame for the financial crash of 2007 and 2008. This general statement is a bit of an oddity, as copula modelling can be a virtually automated process. It can be a semi-automated, or even a totally manual process, based on goodness of fit estimates in copula selection. Which method is used is up to the modeller by following a copula modelling process. Blame is squarely put in using the Gaussian copula, see Salmon (2009), by financial traders in selling vast quantities of new securities and expanding financial markets to unimaginable levels. Zimmer (2012) highlights that the Gaussian copula's inability to account for tail dependence limits its use in estimating relationships in housing price movements that can lead to model miss-specification, which is what happened in the financial crash of 2007 and 2008.

However, the problem is not with using the Gaussian copula, just with the modellers for not using a properly specified copula model, see Salmon (2009). A properly specified model may have resulted in another copula shape apart from the Gaussian copula. It is interesting to think how a totally wrong copula model can be selected that does not represent the datasets dependencies when a copula model can be generated quickly and easily with just a few lines of code, with many examples to follow within the literature.

Parametric copula modelling can be undertaken following a clear process, from identifying and determining the model's distributions, transforming the distributions to marginal distributions (usually uniform), through to copula selection. This process is widely used and recommended, see Hofert et al. (2018) and Vuolo (2017), with other examples scattered throughout the copula modelling literature.

Upon investigating copula modelling using financial data that contain a time-series component, even though non-time-series copula and time-series copula modelling are based on the same underlying theorem, Sklar's Theorem, Section 3, different modelling strategies must be employed. These differences are that the time-series and the volatility of the time-series data must be modelled separately before copula modelling methods can be used, see Patton (2012) and Zhang and Singh (2019). It is this time-series and volatile nature of financial data that makes copula modelling more challenging compared to the standard copula modelling process.

Financial copula modelling, should it contain a time-series component, needs to have the time-series and possible volatility parameters estimated. Then, the copula model can only be undertaken using a pseudo-CDF. Standard non-time-series copula modelling can be undertaken using either an empirical CDF or pseudo-CDF for the estimation of the marginal distributions, see Hofert et al. (2018). Using an empirical CDF for the probability integral transformation allows for determining outliers, influential data points and insight on how well the distributions fit the modelled dataset. Using an empirical CDF allows for the copula model to be estimated using actual observations, not the pseudo-observations generated from a pseudo-CDF.

Standard, non-time-series copula models consist of a rich variety of copula types, see Stamatatou et al. (2018). Popular copula model types include, but are not limited to, Gaussian, Student-t, Clayton, Frank, Gumbel, Joe and an assortment of BB copula models, see Schepsmeier et al. (2015). The variety of these copula models allows for the modelling of no tail dependence, some tail dependence or even extreme tail dependence in either survival or risk modelling Hofert et al. (2018). This also allows for a range of available copula models that can be chosen to suit the dependence structure of the modelled data, allowing for the correct model specification.

The aim of this paper is to highlight the difference between time-series copula modelling and non-time-series copula modelling and where the modelling methods diverge and why. In addition, we use the rich diversity of possible copula models Schepsmeier et al. (2015) using time-series data, such as financial data—as the literature is limited mostly focused on the common Gaussian, Gumbel, Clayton, Frank and Joe type copula models.

We proceed as follows: Section 2 introduces the background in financial copula modelling; Section 3 is copula theory; Section 4, an overview of time-series modelling; Section 5, a worked example of a financial copula model; Section 6, a brief discussion about financial copula modelling; and Section 7, makes our conclusions.

2. Background in Financial Copula Modelling

The financial crisis and subsequent worldwide economic and financial developments have made clear the importance of the analysis of financial data, see Guharay et al. (2013). Research in financial economics has been centred on modelling and simulation, both to gain insight into mechanisms of events (e.g., such as bubbles or crashes) and to characterise the system's stability Guharay et al. (2013).

Within economics and finance, there is a particular difficulty in copula modelling Fermanian (2017) due to the time-dependence; significant advances have been observed in terms of copula modelling of univariate and multivariate time-series. Although advances have been made, financial copula modelling can be awkward as there are many different modelling strategies and using a defined process can be difficult.

The dependence of financial markets during a period of extreme fluctuations has received considerable attention within the literature, see Mensah and Adam (2020). The copula co-movements capture how shocks in a specific market may transcend to other currency markets. Therefore, measuring the co-movements and tail dependence structures and determining the volatility spill-over and evolution over time is essential in risk management, diversification and pricing, see Mensah and Adam (2020).

When modelling financial data that contain time-series, the serial correlations within the time-series play a fundamental role in the statistical process. The observed data may not be independent since the present conditions rely on the past. Appropriate models for serial correlation dependence are necessary Emura et al. (2017). Financial institutions and investors are concerned about volatility and global currency markets, which can dominate the stock and bond markets, see Mensah and Adam (2020). Copula models that contain time-series have been employed extensively to study the co-movement and tail dependence structure of financial data.

Overall, copula modelling has shown to be a worthy tool in analysing highly dependent phenomena Gródek-Szostak et al. (2019), such as financial market performance, market confidence, market speculation, drought, floods, fire, rainfall or even crop oversupply, to name a few. Overall, studies have found Bhatti and Do (2019) that copula models are good at characterising joint dependence among variables, especially when extreme values are clearly evident. Nevertheless, copula modelling is sometimes used in a "black box" fashion Haugh (2016).

Copulas allow the compounding of joint distributions when only marginal distributions are known with certainty. An important advantage is that they allow marginal distributions from different families, such as the Elliptical and Archimedean distributions Kayalar et al. (2017). Copulas also provide a flexible methodology for understanding associations between related phenomena and their joint probabilities. The copula models ask a fundamentally different question than typical techniques modelling conditional values. Rather than how does variable X influence variable Y, copulas ask, "How do two variables move together in unison and how strong is that con-current movement at various points in the distribution" Vuolo (2017).

Copula modelling can be undertaken to explore extreme events, with the main advantage of extreme value copulas Gródek-Szostak et al. (2019) being the possibility of analysing above-average losses or profits in the fields of finance and insurance, but also in the case of examining future agricultural products. Although there is an array of divergent definitions that can cause ambiguity when it comes to the definition of extreme events, see Broska et al. (2020), in the context of finance and insurance, extreme events relate to unexpected, abnormal or extreme outcomes. Common extreme value and survival copulas are the Student-t, Clayton, Galambos and BB8 (Joe-Frank) copulas.

Financial copula modelling is about identifying the time-series within the data, and there are standard families of multivariate models for financial time series. These models are the GARCH-type and/or stochastic volatility type models Fermanian (2017), and there is a dominant use of GARCH models Mensah and Adam (2020). Econometric literature contains a preponderance of evidence that the conditional volatility of economic time-series changes through time, see Patton (2012).

3. Copula Theory

The basic definition of a bivariate copula Beare (2010) is a bivariate probability distribution function on $[0,1]^2$ for which the two univariate marginal distribution functions are uniform on $[0,1]$. Suppose that X and Y are real-valued random variables with joint distribution function $F_{(X,Y)}$ and marginal distribution functions F_X and F_Y. We say that X and Y admit the copula C if $C(F_X(x), F_Y(y)) = F_(X,Y)(x,y)$ for all $x, y \in R$.

Further to the above, suppose that X and Y are real-valued random variables with joint distribution function $F_{(X,Y)}$ and marginal distribution functions F_X and F_Y. We say we can construct a copula $C(F_X(x), F_Y(y)) = F(X,Y)(x,y)$ for all $x, y \in R$ if and only if the Fréchet–Hoffding bounds hold true.

Sklar's theorem ensures that for any random variable X and Y, there exists a copula C such that X and Y admit C Beare (2010). Moreover, C is uniquely defined on the product of the ranges of the marginal distribution functions of X and Y. Hence, C is unique if X and Y are continuous random variables. If X or Y is not continuous, C may nevertheless be uniquely defined by bilinear interpolation between uniquely defined values. The practical implication of Sklar's theorem is that modelling of the marginal distributions can be conveniently separated from the dependence modelling in terms of the copula Brechmann and Schepsmeier (2013).

Based on the Sklar's Theorem, the copula is a mathematical function, see Li et al. (2018), of univariate marginal distribution functions for constructing the joint distribution. For a bivariate case, if the marginal distribution function of the continuous random vectors (X, Y) are F_x and F_y, the joint probability distribution function can be expressed as:

$$P(X \leq x, Y \leq y) = C[F_x(x), F_y(y); \theta] = C(u_1, u_2) \tag{1}$$

where C is a copula, θ is the copula parameter that summarizes the dependence structure and $u_1 = F_x(x)$, $u_2 = F_y(y)$ are marginal probabilities, x and y are the realizations of X and Y. Copulas are functions Haugh (2016) that enable the separation of the marginal distributions from the dependency structure of a given multivariate distribution.

Sklar's Theorem (1959)

Theorem 1 (Sklar's Theorem)**.** *For any d-dimensional dfH with univariate margins $F_1, \ldots, F_d$, there exists a d-dimensional copula C such that:*

$$H(x) = C(F_1(x_1), ..., F_d(x_d)), x \in R \tag{2}$$

The complete Sklar's theorem and proof of the theorem can be found throughout the literature. A proportion is provided there to highlight some of the important parts of the theorem. According to Sklar's theorem (1959), for any joint distribution $F(y_1, y_2)$ with marginal cumulative distributions $F_1(y_1)$ and $F_2(y_2)$ there exists a copula function:

$$F(y_1, y_2) = C_\theta(F_1(y_1), F(y)) \tag{3}$$

An important aspect of Sklar's theorem is that the joint distribution of any two outcomes y_1 and y_2 can be expressed as a copula function that is determined by the individual marginal CDFs $F_1(y_1)$ and $F_2(y_2)$ and an association parameter θ that binds them together Vuolo (2017). Given that CDFs are bound between 0 and 1, by definition, the function takes a value on the square and product value on the unit value. For any two continuous marginals, the CDF is uniform from 0 to 1, or $F_1, F_2 \sim U(0, 1)$.

A transformation on the dataset to make the standard uniform margins is a probability transformation. Figure 1, as an example, shows the distributions that may be suited for a probability integral transformation to obtain the empirical CDFs. Figure 2 shows the uniform distribution generated from the chosen distributions by a probability integral transformation.

Although the probability integral transformation usually transforms the distributions to uniform marginal distributions, other distributions can be used, as long they are the same type of distributions Vuolo (2017)—either two uniform or two normal distributions, for example.

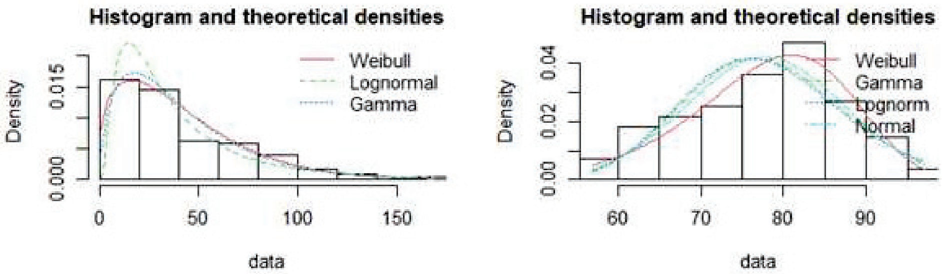


Figure 1. Example Distribution Estimation of a Datasets Distribution.

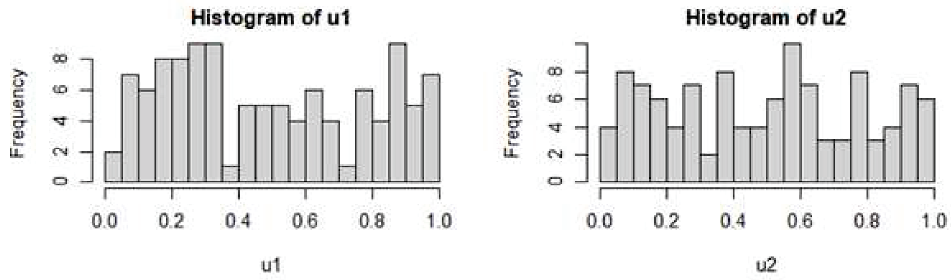


Figure 2. Example uniform marginal distributions estimated from an empirical CDF.

This is a common method in copula modelling in the absence of time-series. Financial copula modelling diverges from standard bivariate copula modelling in obtaining the uniform marginal distributions due to the time-series component, and we proceed as follows.

4. Time Series Modelling

Time-series can be defined Li et al. (2020) as a sequence of observations on one or more variables over time. Time is an important dimension because past events can influence future events. The two main features of time-series Li et al. (2020) are the data frequency and autocorrelation (the correlation between the observations of the same variable over successive time intervals). These features must be carefully modelled to allow for correctly specified residuals to be used within the copula model.

Time series modelling with no volatility can be undertaken using ARMA and ARIMA methods. The ARIMA equation is where the stationary time series is linear, in which the predictors consist of lags of the dependent variable. A nonseasonal ARIMA model is classified as am ARIMA(p,d,q) model, where p is the number of autoregressive terms, d is the number of nonseasonal differences needed for stationarity, and q is the number of lagged forecast errors in the prediction equation.

The difference between an ARMA and the ARIMA is the number of needed differences in a series of observations to achieve stationarity. ARMA(p,q) is equivalent to ARIMA$(p,0,q)$ and if $d > 0$, an ARMA(p,q) can be used after the differencing of the original time-series has achieved stationarity.

Time-series may be identified by using sample autocorrelation (ACF) and partial autocorrelation (PACF) functions for time-series, $X_t : t = 1, ..., n$. If the sample ACF falls into the 95% confidence bound quickly, then the time series X_t may be considered stationary. Otherwise, the time series is non-stationary and differencing is required to convert it to a stationary time series, see Zhang and Singh (2019).

Time-series modelling can also include ARCH and GARCH methods when there is the presents of volatility within the data. ARCH models are an AR model with conditional heteroscedasticity. If the volatility does not necessarily happen at particular times, the variance itself can be modelled with an AR(p) model. A GARCH (generalized ARCH) model is a better fit for modeling time-series data when the data exhibits heteroscedasticity and also volatility clustering.

The ARMA-GARCH time series models, model the time-series along with the volatility contained within the data. The ARMA-GARCH models are used to remove the serial correlations and conditional heteroscedasticity, see Albulescu et al. (2020).

5. Financial Copula Modelling

Financial copula modelling includes a time-dependent sequence, see Zhang and Singh (2019). The copula modelling process must account for each variables' univariate time-series structure. The model's residuals must be estimated from the fitted univariate time-series model and then applied to the copula to model using standardized residuals.

Financial copula modelling uses pseudo-CDFs due to the standardized residuals being centred around zero, with the fitting of gamma or lognormal distributions, as an example, to undertake the probability integral transformation not being possible. The pseudo-observations from the standardized residuals by way of a pseudo-CDF is used.

Figure 3 shows the Australian Crude Oil WTI and Natural Gas weekly Futures prices between the dates of 4 December 2012 and 28 November 2021. As an example, these time-series datasets will be used to undertake a copula model, in highlighting the copula modelling process for financial data that contains time-series.

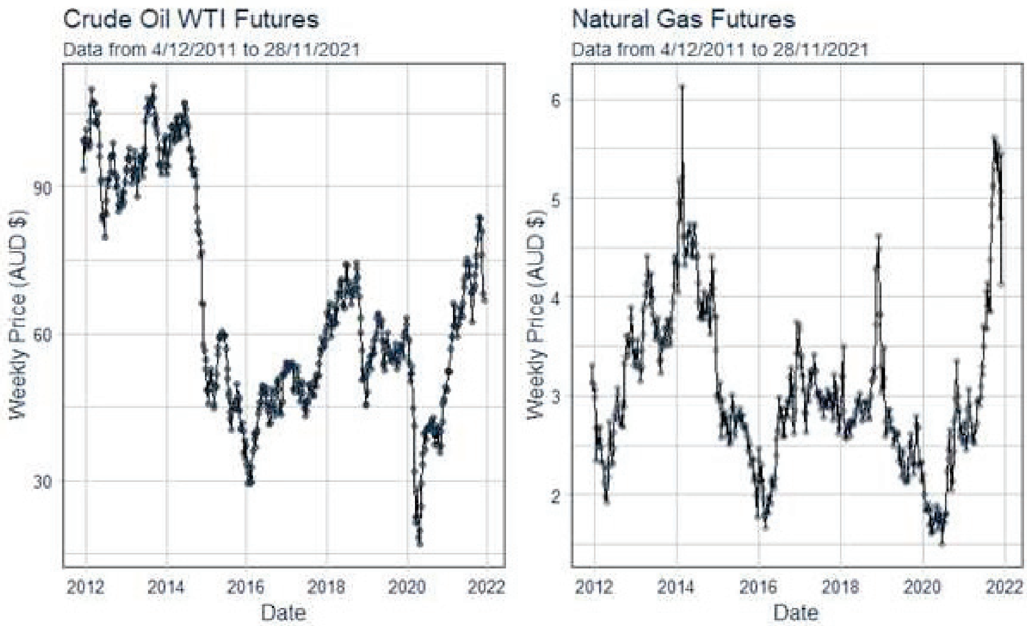


Figure 3. Crude oil and natural gas time series plots.

The first steps in undertaking a financial copula modelling are determining the ARMA model, then determining the GARCH component of the time-series, or if the GARCH component is required. Once these have been obtained, the standardized residuals can be modelled into a copula. This is the most critical part of the copula modelling process, as determining the correct time-series model will allow for correct tail dependence to be modelled by way of the standardized residuals.

Literature for copula time-series modelling using ARIMA-type models can be sparse, whereas, the literature for copula time-series modelling using ARMA-GARCH models is far more abundant Mensah and Adam (2020), as the ARMA or ARMA-GARCH process models the volatility within the dataset, making them suitable for financial data. Testing whether an ARMA or ARMA-GARCH model is appropriate should be undertaken. For short time periods, an ARIMA may be appropriate, and a long time period may require the ARMA-GARCH model.

Estimating the ARIMA and ARMA parameters for Crude Oil and Natural Gas was undertaken using the *auto.arima* package within R software, https://CRAN.R-project.org/ (accessed on 10 April 2020), see Hyndman et al. (2020). This gave the results for Crude Oil an ARIMA model of order (2,1,1) and Natural Gas as an ARIMA model (1,1,1). After differencing, the ARMA estimated model for Crude Oil was of order (2,1) and Natural Gas was (1,1). These values and the differenced data were used for the GARCH modelling.

Fitting a GARCH model to the data was used by the *ugarchfit* package within the R software, https://CRAN.R-project.org/ (accessed on 10 April 2020), see Ghalanos (2020). GARCH modelling for both Crude Oil and Natural Gas gave a GARCH model result of (1,1). Using the GARCH, *ugarchfit* package, allowed for the fitting of the ARMA parameters for the Crude Oil and the Natural Gas datasets and the obtained residuals were then standardized.

Both datasets contained heteroscedasticity, therefore a GARCH process was used to model the volatility. The residuals conditional variance equation for Crude Oil was an ARMA(2,1)-GARCH(1,1), and Natural Gas was an ARMA(1,1)-GARCH(1,1). The ARMA-GARCH modelled the non-constant variance of the two datasets, which will produce residuals $N \approx (0,1)$. These models are used within the rest of the paper.

Figure 4, shows the histograms of the standardized residuals from the time-series modelling results, which are overlaid with the normal distribution. The results show that the x-axis values contain negative values. Estimation of log-normal and gamma distributions that may better fit the residuals cannot be undertaken to estimate the copulas marginal distributions. This highlights the need for copula time-series models to be based on pseudo-CDFs.

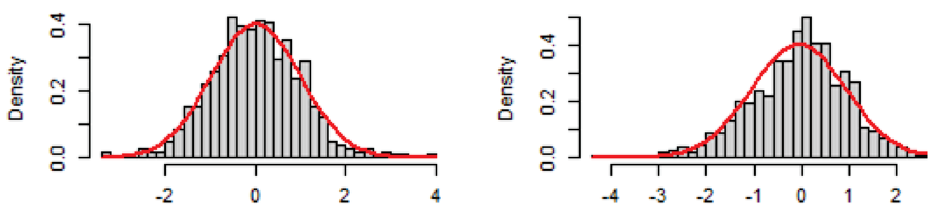


Figure 4. Standardized residuals for the crude oil and natural gas datasets.

A graphical method to determine if any tail dependence is present can be represented in a Q-Q plot of the two standardized residuals for crude oil and natural gas. The Q-Q plots give an indication if the tails are not Gaussian, indicating possible tail dependence. The Q-Q plots have been plotted and shown in Figure 5. Overall, the standardized residuals for both datasets follow the Gaussian distribution, with the exemption that only a few points at the tails fall outside a Gaussian distributions.

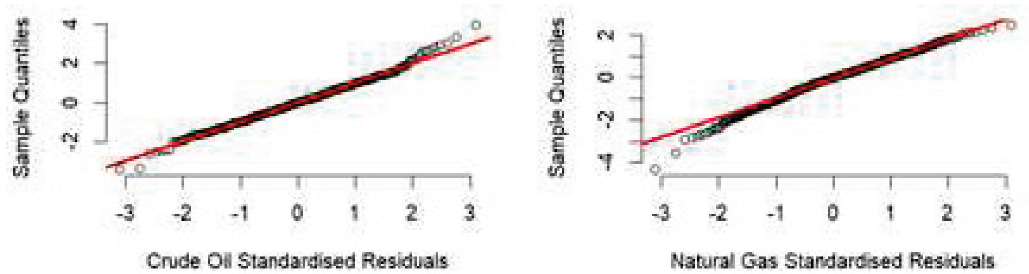


Figure 5. Q-Q plots for the standardized residuals.

The standardized residuals were then applied to the *BiCopSelect* function, see Schepsmeier et al. (2015), with R software package, https://CRAN.R-project.org/ (accessed on 10 April 2020), to generate (estimate) the copula type and parameters. Figure 6, shows the results, which produced a Gaussian (Normal) copula with a τ value of 0.15. The τ value of only 0.15, which is a measure of dependence, a Pearson's linear correlation coefficient, shows that there is very little dependence between the crude oil and natural gas prices.

Figure 6, shows the estimated Gaussian copula which had a τ value of 0.15, with Figure 7 being included to show the Gaussian copula dependence with a τ value of 0.15 and a τ value of 0.7, which is considered a strong association value. A worthwhile endeavour in any copula modelling situation is to plot the copulas dependence. The estimated copula with the τ value = 0.15, shows significantly less dependence with no tail dependence.

Figures 8 and 9 show the comparison between the estimated Gaussian distribution of τ = 0.15, to a strong dependence value of τ = 0.7. In Figure 8, the generated copula distribution, shows a distribution that almost represents an independent copula, being that u1 and u2, are almost independent with a τ = 0.15. Figure 9, shows a distribution with another set of u1 and u2 values, having a strong dependency value of $\tau = 0.7$.

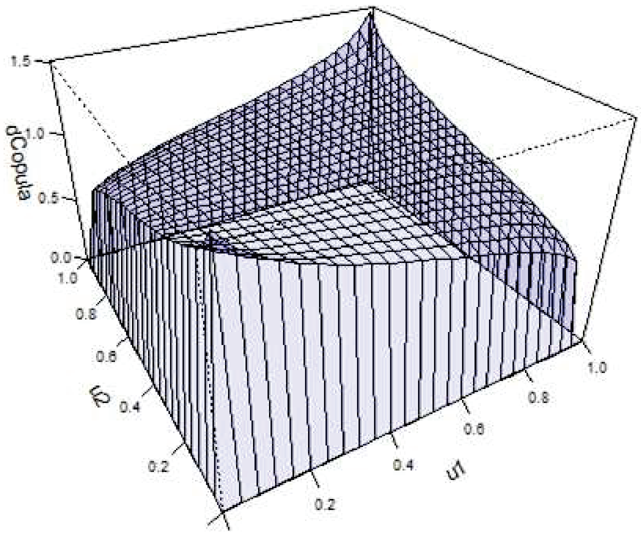


Figure 6. The estimated Gaussian copula.

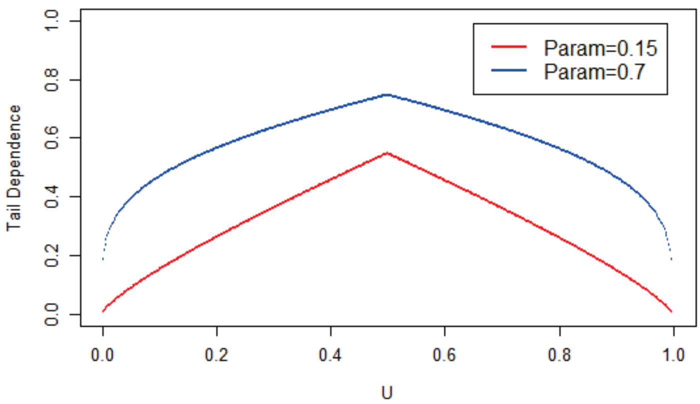


Figure 7. Gaussian copula dependence.

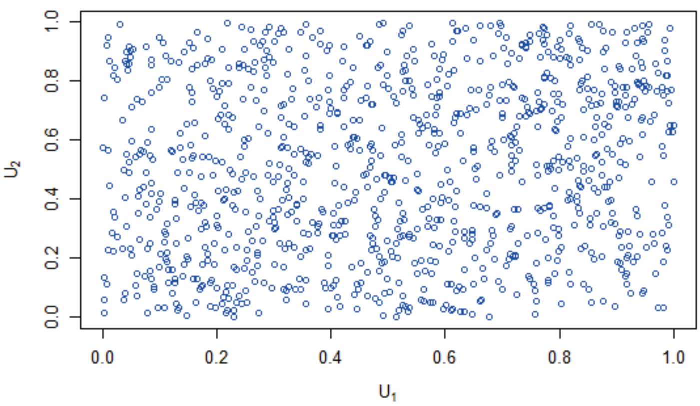


Figure 8. The estimated Gaussian distribution, Tau = 0.15.

Therefore, we can conclude there was no dependence between the crude oil and natural gas prices, as the estimated Gaussian copula had a $\tau = 0.15$. Figure 8, confirms this, as the estimated copula model represents an independent copula distribution, see Hofert et al. (2018).

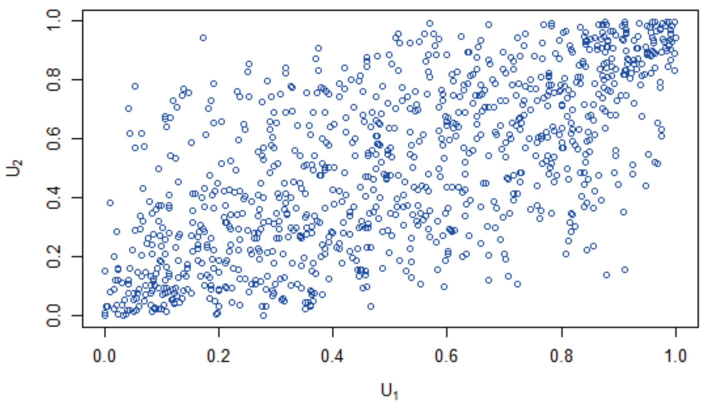


Figure 9. Gaussian distribution, Tau = 0.7.

6. Discussion

In returning to the original statement that the Gaussian copula can be blamed for the 2007-2008 financial crisis, Section 1, can be appreciated now. An easy approach was undertaken by the financial traders in using a Gaussian copula model Zimmer (2012), which inhibited any ability for tail dependence to be modelled and identified. This allowed for missed opportunities in detecting the "up and coming" financial bust of 2007 and 2008.

With standard copula modelling, choices can be made to use CDFs or pseudo-CDFs to model the copula. Using CDFs has the advantage of the copula model being modelled using distributions with the actual values from a dataset. Undertaking the inverse probability integral transformation using a CDF highlights potential issues within the dataset that may allow for further exploration. Smaller datasets will benefit from using a CDF, as the CDF and pseudo-CDF converge with large datasets.

Financial copula modelling requires the added process of extracting the standardised residuals from a time-series model that will allow for a financial copula model to be modelled. Unifying the time-series component and copula modelling process can be an awkward process, as there are many different modelling methods and coding options available and little guidance within the literature in having these working together coherently.

This paper has highlighted the importance of the time-series parameters to be correctly specified that will allow for the standardized residuals to be used within the copula model that will allow for a correctly-specified copula model to be estimated.

Regardless, no dependence was found between the crude oil and natural gas prices. The generated copula model is valid, assuming the correct time-series parameters were estimated. This process should be used in the copula modelling using financial data when time-series is involved, which allows for the copula modelling process in determining the copula shape, allowing for the various copula types to be included within copula selection that will allow for correct model specification Schepsmeier et al. (2015).

Not to be overlooked, a good source of information and coding when undertaking financial copula time-series modelling can come from the ecological field, as an example. Many ecological studies use time-series copula modelling, which can be taken advantage of within the financial fields.

7. Conclusions

Both standard copula modelling and financial copula modelling follow the same process, with financial copula modelling requiring the correct specification of the time-series to be identified. Once the time-series is identified, copula modelling can proceed using a pseudo-CDF to model the uniform marginal distribution to produce the copula model. Only by correct specification of the time-series components can a correctly specified copula model be produced.

Using MA, ARMA-type models will only model constant variance. The ARMA-GARCH types model the non-constant variance. Through the process of determining the time-series components within the data, will a correctly specified copula model be produced.

Using the *BCopSelect* function, see Schepsmeier et al. (2015), within the R software, https://CRAN.R-project.org/ (accessed on 10 April 2020), allowed for the opportunity for many copula model forms to be considered for the copula model. This quick or exploratory method allows for a valid copula model to be estimated.

Copula models that are popular, such as the Gumbel copula, and produce similar likelihood, AIC or BIC values to a BB copula, as an example, could also be considered as a final model. This may allow for a better audience understanding of the copula model, as the literature is focused on these popular copula types.

Author Contributions: P.R.D.: Data curation, formal analysis, methodology, writing original draft, writing review and editing; S.L.: Investigation, formal analysis, methodology, writing, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analysed in this study. This data can be found here: https://au.investing.com/commodities/crude-oil (accessed on 22 December 2021), and at: https://au.investing.com/commodities/natural-gas (accessed on 22 December 2021).

Acknowledgments: We would like to thank the Editors and Reviewers for their constructive comments which led to an improved presentation of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACF	Auto Correlation Function
AIC	Akaike Information Criterion
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroscedasticity
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
BIC	Bayesian Information Criterion
CDF	Cumulative Density Function
GARCH	Generalized AutoRegressive Conditional Heteroscedasticity
MA	Moving Average
PACF	Partial Autocorrelation Function

References

Albulescu, Claudiu Tiberiu, Aviral Kumar Tiwari, and Qiang Ji. 2020. Copula-based local dependence among energy, agriculture and metal commodities markets. *Energy* 202: 117762. [CrossRef]

Beare, Brendan K. 2010. Copulas and temporal dependence. *Econometrica* 78: 395–410

Bhatti, Ishaq M., and Hung Quang Do. 2019. Recent development in copula and its applications to the energy, forestry and environmental sciences. *International Journal of Hydrogen Energy* 44: 19453–73. [CrossRef]

Brechmann, Eike Christian, and Ulf Schepsmeier. 2013. Cdvine: Modeling dependence with c-and d-vine copulas in r. *Journal of Statistical Software* 52: 1–27. [CrossRef]

Broska, Lisa Hanna, Witold-Roger Poganietz, and Stefan Vögele. 2020. Extreme events defined—A conceptual discussion applying a complex systems approach. *Futures* 115: 102490. [CrossRef]

Emura, Takeshi, Ting-Hsuan Long, and Li-Hsien Sun. 2017. R routines for performing estimation and statistical process control under copula-based time series models. *Communications in Statistics-Simulation and Computation* 46: 3067–87 [CrossRef]

Fermanian, Jean-David. 2017. Recent developments in copula models. *Econometrics* 5: 34. [CrossRef]

Ghalanos, Alexios. 2020. Rugarch: Univariate GARCH Models. R Package Version 1.4-4. Available online: https://cran.r-project.org/web/packages/rugarch/index.html (accessed on 10 January 2022).

Gródek-Szostak, Zofia, Gabriela Malik, Danuta Kajrunajtys, Anna Szeląg-Sikora, Jakub Sikora, Maciej Kuboń, Marcin Niemiec, and Joanna Kapusta-Duch. 2019. Modeling the dependency between extreme prices of selected agricultural products on the derivatives market using the linkage function. *Sustainability* 11: 4144. [CrossRef]

Guharay, Samar K., Gaurav S. Thakur, Fred J. Goodman, Scott L. Rosen, and Daniel Houser. 2013. Analysis of non-stationary dynamics in the financial system. *Economics Letters* 121: 454–57. [CrossRef]

Haugh, Martin. 2016. An introduction to copulas In *IEOR E4602: Quantitative Risk Management*. Lecture Notes. New York: Columbia University.

Hofert, Marius, Ivan Kojadinovic, Martin Mächler, and Jun Yan. 2018. *Elements of Copula Modeling with R*. Berlin/Heidelberg: Springer International Publishing AG.

Hyndman, Rob J., George Athanasopoulos, Christoph Bergmeir, Gabriel Caceres, Leanne Chhay, Mitchell O'Hara-Wild, Fotios Petropoulos, Slava Razbash, and Earo Wang. 2020 Package 'Forecast'. Available online: https://cran.r-project.org/web/packages/forecast/forecast.pdf (accessed on 10 January 2022).

Kayalar, Derya Ezgi, Coşkun C. Küçüközmen, and Sevtap A. Selcuk-Kestel. 2017. The impact of crude oil prices on financial market indicators: Copula approach. *Energy Economics* 61: 162–73. [CrossRef]

Li, Fan, Jiren Zhou, and Chen Liu. 2018. Statistical modelling of extreme storms using copulas: A comparison study. *Coastal Engineering* 142: 52–61. [CrossRef]

Liu, Timina, Shuangzhe Liu, and Lei Shi. 2020. *Time Series Analysis Using SAS Enterprise Guide*. Springer Nature: Singapore.

Mensah, Prince Osei, and Anokye M. Adam. 2020. Copula-Based Assessment of Co-Movement and Tail Dependence Structure Among Major Trading Foreign Currencies in Ghana. *Risks* 8: 55. [CrossRef]

Patton, Andrew J. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* 110: 4–18. [CrossRef]

Salmon, Felix. 2009. Recipe for disaster: The formula that killed Wall Street. *Wired Magazine* 17: 3.

Schepsmeier, Ulf, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, Thomas Nagler, Tobias Erhardt, Carlos Almeida, Aleksey Min, Claudia Czado, Mathias Hofmann, and et al. 2015. Package 'VineCopula'. *R Package Version*. Available online: https://cran.microsoft.com/snapshot/2018-07-26/web/packages/VineCopula/VineCopula.pdf (accessed on 10 April 2020).

Stamatatou, Nikoletta, Lampros Vasiliades, and Athanasios Loukas. 2018. Bivariate Flood Frequency Analysis Using Copulas. *Proceedings* 2: 635. [CrossRef]

Vuolo, Mike. 2017. Copula models for sociology: Measures of dependence and probabilities for joint distributions. *Sociological Methods and Research* 46: 604–48. [CrossRef]

Zhang, Lan, and Vijay P. Singh. 2019. *Copulas and Their Applications in Water Resources Engineering*. Cambridge: Cambridge University Press.

Zimmer, David M. 2012. The role of copulas in the housing crisis. *Review of Economics and Statistics* 94: 607–20. [CrossRef]

Journal of
Risk and Financial Management

Article

Fight Alone or Together? The Influence of Risk Perception on Helping Behavior

Liping Yin [1] and Yen-Chun Jim Wu [2,3,*]

[1] School of Cultural Industry and Tourism Management, Henan University, Kaifeng 475001, China; yinliping@henu.edu.cn
[2] College of Humanities and Arts, National Taipei University of Education, Taipei 106, Taiwan
[3] Department of Hospitality Management, Ming Chuan University, Taipei 111, Taiwan
* Correspondence: wuyenchun@gmail.com

Abstract: Will there be a greater sense of solidarity and friendship during public crises? This study aims to determine whether risk perception influences employees' willingness to assist in times of public crisis, taking COVID-19 as a specific research scenario and based on the theory of "tend and befriend". This study hypothesized that risk perception will influence employees' helping behavior via the in-group identity, with the degree of impact dependent on the COVID-19 pandemic's severity. A questionnaire survey of 925 practitioners from various industries in the pandemic area revealed that: risk perception has a positive influence on employees' helping behavior; in-group identity plays a certain mediating role in the process of risk perception that influences employees' helping behavior; and the severity of a local pandemic negatively moderates the relationship between risk perception and helping behavior, but positively moderates the relationship between risk perception and in-group identity. Specifically, employees in high-risk areas are more likely to "align" (higher degree of recognition by the in-group) but demonstrate less helping behavior, compared with those in areas with moderate and low risk from the COVID-19. By contrast, employees in low-risk areas display more helping behavior but have less in-group identity, compared with those in areas with moderate and high risk from the COVID-19. This study expands the research on the relationship between risk perception and helping behavior, enriches the research results on risk management theory, and provides a practical reference for risk governance.

Keywords: COVID-19; helping behavior; in-group identity; risk perception; severity of a local pandemic

Citation: Yin, Liping, and Yen-Chun Jim Wu. 2022. Fight Alone or Together? The Influence of Risk Perception on Helping Behavior. *Journal of Risk and Financial Management* 15: 78. https://doi.org/10.3390/jrfm15020078

Academic Editor: Stephen Satchell

Received: 22 December 2021
Accepted: 11 February 2022
Published: 13 February 2022

1. Introduction

Risk perception describes the attitude and intuitive judgment of people toward uncertain events under specific scenarios and is highly variable and dependent on scenarios. It comprises not only people's intuitive judgment in the face of risks but also subjective awareness based on rational and complex analysis. The subjective nature of risk perception explains why people have different responses to the same sources of risk. Risk perception concerning COVID-19 is a relatively new research topic. Severity of a local pandemic and different coping measures by health departments may lead to public differences in perceptions of COVID-19 risk (Wei 2020). However, as found by Zhang and Zhou (2021), compared with other potential health threats, people have higher risk perceptions about COVID-19. Excessive risk perception leads people to adopt excessive protective measures (Savadori and Lauriola 2021), such as irrational behavior, including panic buying of foods and medical supplies and discrimination against people in areas with more severe pandemic conditions (Zhang and Zhou 2021). Further confirmed, is a significant correlation between perceptions of COVID-19 risk and negative emotions, such as loneliness and anxiety (Han et al. 2021); it might also cause symptoms of depression among employees (Yan et al. 2021). Obviously, most of the existing research on psychology and behavior in the aftermath of the COVID-19 pandemic focuses on negative emotions and self-centered behavior.

However, recent studies have state in that in the event of disasters, such as earthquakes, floods, and illness, individuals with an instinct for survival positively seek external support and comfort and demonstrate prosocial behaviors to enhance their social linkage with groups (Miao et al. 2021). (Taylor et al. 2000; Taylor 2006) call this response a "tend and befriend." As they state, stressed people long for social support to relieve their anxiety and help them obtain comfort and protection, etc. Therefore, people engage in more friendly and prosocial behaviors to maintain good social relations. COVID-19 features strong infectivity and a high morbidity, and 90% of Chinese residents have been influenced by the COVID-19 pandemic. Among them, two-thirds have expressed concern about infection, whereas one-third believe COVID-19 infection has high mortality. The degree of such mental panic reaches a high level of stress (Xue and Wang 2021). According to the "tend and befriend" response mode, employees tend to perform in a prosocial manner after perceiving the risks of COVID-19 pandemic. In fact, during the COVID-19 pandemic outbreak, the population demonstrated a high level of prosocial behavior.

Based on these ideas, we take the COVID-19 pandemic as the starting point, and introduces the theory of "tend and befriend" to establish a theoretical mechanism explaining how employees' risk perceptions influence helping behaviors. This phenomenon is examined not only due to its prevalence during the COVID-19 pandemic outbreak, but also because it contributes to the resolution of the following critical theoretical issues: (1) Do individuals exhibit prosocial tendencies in the face of a significant public crisis, such as demonstrating helping behaviors in organizations, and does the "tend and befriend" theory hold true in public crises? (2) What effect does risk perception have on an individual's willingness to assist others, and what is the underlying mechanism? (3) Is risk (i.e., the severity of the COVID-19 pandemic) a factor in the relationship between risk perception and prosocial behavior? Are there distinctions between high and low risk situations in terms of prosocial motivation and behavior? This study establishes a theoretical model for risk perception that influenced employees' helping behavior, with the in-group identified playing a mediating role and the severity of local pandemic acting as a moderating variable. The research hypothesis model in this paper is summarized in Figure 1. The research results can more comprehensively and accurately reflect real awareness and behaviors involved in employees' risk perception from the theoretical perspective, thereby providing beneficial references for the improvement of risk management and risk governance.

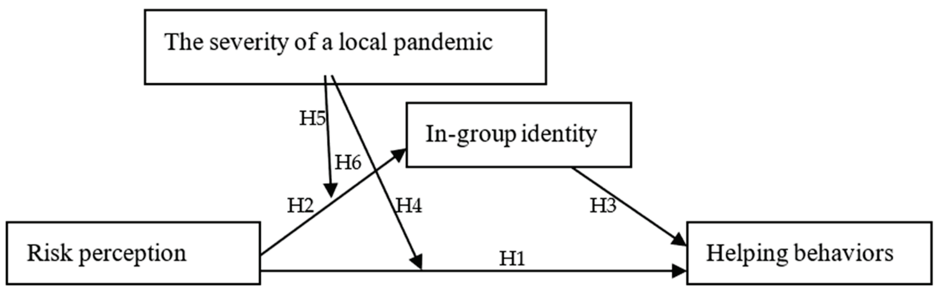


Figure 1. Research model.

2. Literature Review and Hypotheses Development

2.1. Risk Perception and Helping Behavior

The COVID-19 pandemic features rapid transmission, high infection rates, and high morbidity. The theory of "tend and befriend" assumes that people want social support to relieve their anxiety and comfort and protection through "alliances" under stress (Mai et al. 2021), thus people engage in more friendly behaviors in order to maintain good social relations (Taylor et al. 2000; Taylor 2006). This theory was first confirmed in the laboratory: Von Dawans et al. (2012) used a number of tasks to investigate individuals' prosocial behaviors under stress and found that stress groups showed more trust in one another, more willingness to share, and placed a higher value on being trusted. Maier et al. (2015) also verified that subjects in stress groups were more generous than those in the control group. Levy et al. (2019) compared other reactions by men and women under stress and found that women and men were both more likely to demonstrate caring and make friends under stress. Not only adults but also children donate more money under stress than the non-stress group (Alen et al. 2021).

The theory of "tend and befriend" has also been confirmed in the public crisis scenario: through research on earthquake victims, Rao et al. (2011) found that they are more likely to demonstrate prosocial behavior, and the more serious the disaster was, the more money they gave to responders in the game. Wang and Sun (2021) also confirmed that people who have suffered in major disasters, such as wars, terrorist attacks, and natural disasters, display more helping behaviors, such as the donation of goods, money, and volunteering. These research results demonstrate that when people perceive risks and threats, they generally show more prosocial behaviors, and the "tend and befriend" reaction after stress may be common. Recently, Schneider et al. (2021) investigated the relationship between the perception of COVID-19 risk and prosocial behavior in the British public. They found that in the initial stage of the COVID-19 pandemic, people performed more prosocial behaviors to work together in fighting the spread of disease. Wang and Sun (2021) also compared the differences in volunteer activities between Chinese and American people during COVID-19 and found that respondents in China generally participated in more volunteer activities than those in the United States, while respondents in China participated in slightly fewer volunteer activities than those in the United States in areas where the pandemic was more severe.

Obviously, the relationship between perceived COVID-19 risk and prosocial behavior has been preliminarily confirmed. Employee helping behavior is typical prosocial behavior, which is embodied in helping colleagues solve problems and sharing work actively, sharing their professional knowledge and experience with colleagues, helping new colleagues adapt to the environment, and actively providing extra help to customers (Farh et al. 2007). The helping behavior represents a kind of cooperation-oriented, attribution-oriented behavior outside the role, which helps to establish efficient interpersonal relationships and strengthen social connections in a complex organizational environment (Tu et al. 2020). Based on these research results, it can be predicted that employees are likely to implement more helping behaviors after they perceive risks, so as to establish good social relations and withstand risks. Therefore, we posit the following hypothesis:

Hypothesis 1 (H1). *There is a positive relationship between risk perception and helping behavior.*

2.2. Mediating Effect of In-Group Identity

When individuals are threatened by natural disasters or pandemic diseases, such as earthquakes, typhoons, or floods, individuals often feel helpless and powerless. At this time, in order to survive, human beings must choose to cooperate with others to withstand external threats (Fowler and Christakis 2010). (Taylor et al. 2000; Taylor 2006) questioned the appropriateness of the "fight or flight" response model and broke through it from the perspective of social evolution and neuroendocrine, proposing, instead, the "tend and befriend" response model of individuals under stress. They think that positive social interaction after stress is conducive to mutual trust, dependability, and sharing behaviors, which play an important role in controlling adverse reactions after stress. Therefore, "tend and befriend" may be a protective reaction after stress, which can provide internal social support for individuals, enable individuals to better self-regulate under stress, weaken the sense of threat after stress, and cope with risk in a more positive way (Von Dawans et al. 2012). Rao et al. (2011) investigated the cooperative behavior of residents in disaster areas and nondisaster areas after the year of the Wenchuan earthquake and found that residents in disaster areas had a higher tendency toward cooperative behaviors than those in nondisaster areas; the more serious the disaster was, the stronger was their cooperation tendency. Based on this, Rao et al. (2011) put forward the view that disadvantages make people more cooperative. When individuals encounter a crisis, their own strength is often limited. At this time, seeking the cooperation of others is the best option. It can be said that external crisis enhances people's sense of belonging and recognition of the group to a certain extent, as they hope that "we" can jointly withstand risks, instead of facing of the crisis alone.

Prior researchers have also conducted examined the public crisis still ongoing and found that in the face of sudden major threats, members of a group have greater cohesion, which enhances internal effectiveness and group efficacy, thus maintaining the continuity of ethnic groups (Xie et al. 2017). In this case, reducing interpersonal interaction leads to more serious psychological problems. Zhu et al. (2020) found through investigation that COVID-19 caused negative emotions, such as public anxiety, panic, and helplessness. Interpersonal alienation played a positive role in regulating these emotions and posttraumatic stress disorder (PTSD). A high level of interpersonal alienation leads to more serious PTSD, so they suggest that families, enterprises, and communities actively build social support systems and provide social support resources to individuals in various ways to reduce their interpersonal alienation and expand people's psychological energy. Among some survivors of natural disasters, we find that their need for belonging increased significantly compared with those in normal situations (Miao et al. 2021). Based on this analysis, we infer that when employees perceive risks, they have a stronger in-group identity, and they are more eager to obtain the warmth and support of the group, so as to reduce their perception of real threats; thus we propose Hypothesis 2.

Hypothesis 2 (H2). *There is a positive relationship between risk perceptions and in-group identity.*

In-group identity refers to an individual's recognition of the identity of a certain social group to which he belongs and the values and emotions associated with the group identity (Cikara and Bavel 2014). According to social identity theory, employees distinguish their own group from other groups and pay more attention to the norms and values of their own group, and think more about problems related to the interests of the group and act in ways that are beneficial to the group, which result in more behavior that is consistent with the group and conform to its norms (Meng et al. 2021). When the members of a group have a higher recognition of the group, they maintain a higher-quality social exchange relationship and are more willing to share resources, and there are more prosocial behaviors in social interaction, such as cooperation and reciprocity (Hitlin et al. 2021). When individuals strongly identify with the group to which they belong, the phenomenon of in-group favoritism develops, which shows a more positive evaluation of the inner group and helps the members of the inner group in a targeted way (Leyens et al. 2016). For example, Vezzali et al. (2015) found that the higher an individual's identity with an in-group, the more intergroup helping behaviors were performed. In a further investigation of the types of willingness to help, it was found that group identity could positively predict self-directed willingness to help (Zhou et al. 2018). Moreover, a benign interaction arises between employees' identify with the group and altruistic behaviors. The stronger an individual's identification with the group is, the closer the connection between that person and the group. In the face of threatening situations, people with high identity are more inclined to adopt collective strategies, regarding group interests as the goal of emotional and behavioral responses to promote more altruistic behaviors (Miao et al. 2021).

Xie et al. (2017) thinks that the implementation of intragroup helping behavior can promote positive interaction between their bodies and minds, reduce the sense of body load, and help to improve their adaptability. Externally, altruism is a signal to an individual, which reflects that person's individual qualities and shows other members that this person is trustworthy and reliable, thus driving others in the group to choose opportunities to cooperate with one another and gain status within the group. In-group identity is also an important way to satisfy people's sense of belonging, which can make individuals realize that they belong to a specific group and the emotional and value significance brought by a group identity (Miao et al. 2021). Human beings belong to a social species that lives in groups. Therefore, they offer assistance to others when they need help, perhaps through a direct or indirect reciprocity mechanism, so that personal interests can be enhanced (Rao et al. 2011). Therefore, mutual assistance and cooperation among groups affect one's entire social life and are key factors in human survival, development, and reproduction. In particular, when individuals perceive huge external risks, they withstand them through alliance and cooperation and perform more helping behaviors in the group, so as to strengthen cooperation among group members and withstand risks together. Based on this inference, we propose the following hypothesis:

Hypothesis 3 (H3). *In-group identity mediates the positive relationship between risk perceptions and helping behaviors.*

2.3. Moderating Effect of the Severity of a Local Pandemic

Rao et al. (2011) used a dictator game and volunteer questions to evaluate the prosocial behaviors of residents in earthquake-stricken areas and non-earthquake areas. The survey results demonstrate that residents in non-disaster areas allocated fewer funds to strangers and took part in volunteer activities less than those in disaster-stricken areas, while those in areas stricken by a less serious disaster did so less than those in areas affected by a moderate disaster. The most funds and volunteer activities came from people living in areas with either moderate or severe disasters. The results demonstrate that the degree of prosocial behaviors by residents increased with the degree of seriousness of a disaster. The research on the relationship between stress and sympathy in daily life demonstrates that the greater the severity of adversity experienced by an individual was, the more sympathy he received, which leads to more prosocial behaviors (Lim and DeSteno 2016). Xu (2018) believes that people who have experienced disasters change their feelings, such as becoming more compassionate, seeking meaning, social status, and factors, such as reciprocal social learning and a change in the environment, which lead to more prosocial behaviors.

The theory that altruism comes out of suffering (Staub and Vollhardt 2008) and the disadvantage hypothesis (Rao et al. 2011) both demonstrate that suffering makes individuals more motivated to help others. Research on common experience and common fate demonstrate that people tend to establish social connections with others who have the same experience (Nakayachi and Ozaki 2014). Having the same experiences or common risks put individuals and those who experience them in the same predicament, and they expect each other to want to establish friendly relations with others like themselves, so as to help each other and be together (Leyens et al. 2016). The COVID-19 continues to spread all over the world, and mankind has encountered an unprecedented public health crisis. Under those circumstances, the more serious the pandemic is, the more likely local employees are to bring themselves into the same group as colleagues, customers, and other stakeholders around them, and are willing to strengthen interaction and cooperation with them in order to seek mutual comfort and support. The more serious the pandemic is, the stronger people's identity with their specific group identification will be, and the more helping behaviors will be generated. Based on this, we propose the following hypothesis:

Hypothesis 4 (H4). *The severity of a local pandemic moderates the relationship between risk perceptions and helping behaviors such that the relationship is stronger when the severity of a local pandemic is high, rather than low.*

Margittai et al. (2015) found in the research that people's friendly behaviors under stress are conditional, not for everyone, and only for those who are closely related. Debono et al. (2020) hold the same view: humans choose objects discriminatively to establish social relations with them and tend to choose objects that are more likely to establish connections. Turner (1981) put forward the self-categorization theory on the basis of social identity theory, which holds that people automatically divide themselves and others into internal groups and external groups according to some clues, even some insignificant clues, thus forming membership in different groups. Self-classification is an important process of social identity, and social identity is an important mechanism for explaining the connection between group psychology and individual psychology (Crane and Ruebottom 2011). Group identity is activated by a certain stimulus, and individuals will realize the value and emotional meaning of being group members. Specific identity cognition forms the identity of the group to which they belong (Vesa 2020).

Without external group comparison, common destiny and similarity among members can also enhance in-group identity (Miao et al. 2021). The experience of "sharing weal and woe" is an important adhesive in interpersonal relations. Miao et al. (2021) found that sharing negative events, such as pain and disaster, can stimulate people's need for belonging, thus enhancing cooperation among those who share it. In the face of a common threat, individuals turn greater attention to group identity (Kovoor-Misra 2009). This shift in attention is an instinctive reaction by human beings in the long-term struggle with nature. In the face of danger, people naturally band together and become more united. COVID-19 features rapid transmission and great harm after infection, which makes people avoid it. When a COVID-19 infection occurs in a certain place, disease prevention agencies usually quickly isolate the area and enforce measures, such as isolation and screening. As a result, many companies are seriously affected, and employees face not only the problem of whether their working resources can be sustained but also the risk of infection. This common experience is a very important and dominant social clue, according to which employees are likely to classify themselves and those who have experienced it together, showing greater in-group identity. Based on this analysis, we propose Hypothesis 5.

Hypothesis 5 (H5). *The severity of a local pandemic moderates the relationship between risk perception and in-group identity such that the relationship is stronger when the severity is high, rather than low.*

Furthermore, the severity of a local pandemic moderates not only the relationship between risk perception and in-group identity but also the indirect effect of risk perception on helping behavior through in-group identity. According to this reasoning, we posit the following hypothesis:

Hypothesis 6 (H6). *The severity of a local pandemic moderates the mediating relationship between risk perception and helping behaviors, such that in-group identity has a stronger positive effect when the severity is high, rather than low.*

3. Procedure

3.1. Sample and Procedure

In July 2021, the COVID-19 pandemic with the Delta variant were found in China. Because this strain is highly contagious, concealed, and harmful, a new round of COVID-19 pandemic broke out in many places in China. As of 10 August 2021, China has 224 moderate- and high-risk areas, mainly concentrated in the Hunan, Henan, and Sichuan provinces. The entire country immediately went into a state of preparedness, strictly locked down the pandemic areas, and prohibited unnecessary commercial activities to slow the rapid spread of the virus. Under these circumstances, we talked with the heads of 20 enterprises or departments engaged in retail, catering, tourism, consulting, education and training, health care, and other industries in Henan and Sichuan, explained the purpose of the survey, and obtained their support. Afterward, we conducted a survey of the employees at these enterprises through an online questionnaire platform from 1 to 20 August 2021. This research was ethically approved by the Academic Committee of Henan University and by the appropriate department at the enterprise. All participants were promised that the survey results would be used only for academic research, and their answers would be kept strictly confidential. After completing the questionnaire, participants received cash compensation of 5 yuan. Before the questionnaire was distributed, we explained the purpose and requirements of this survey to the respondents in detail, and then the person in charge of the enterprise sent the link for the questionnaire to the employees. After the employees completed the questionnaire, they submitted it anonymously through the online questionnaire platform.

In this way, we collected a total of 952 questionnaires. Excluding 27 invalid questionnaires, we received 925 valid questionnaires, an effective rate of 97%. Among the 95 respondents, 55.3% were men (n = 512) and 44.7% were women (n = 413). The average age of the respondents was about 36 years; 48.43% of the respondents have a college degree or more (n = 448), and 2.4% have a postgraduate degree or more (n = 22); 92% of the employees worked at the company for more than one year (n = 851); 83.8% (n = 775) of the respondents are ordinary employees, and 16.2% (n = 150) hold management positions. Among the respondents, 10.3% (n = 95) were unmarried, and 84.6% (n = 783) were married or had children; 9.4% of the respondents lived or worked in areas at high risk of COVID-19 pandemic (n = 87). Among these respondents, 5% (n = 46) reported that they had been quarantined at home, and 95% (n = 838) said that their work and life were affected by the COVID-19 pandemic. On the whole, the distribution of the samples is suitable for subsequent data analysis.

3.2. Study Variables

The questionnaire is written in Chinese. According to the standard "translation-back translation" procedure, these scales were translated from the original English into Chinese and then translated back into English. After repeated revisions, 15 employees were used as pretest subjects, and the Chinese version of the questionnaire was preliminarily tested to evaluate the reliability, validity, and usability of the scale. According to the results of the pretest, the language used in the questionnaire was modified to make it more consistent with the perspective of the interviewee. All items were measured from 1 (very inconsistent) to 5 (very consistent). Risk perceptions are subjected to the psychological measurement paradigm, based on the risk perception scale designed by Williams and Voon (1999), and they were slightly modified to be suitable for the specific conditions in the COVID-19 pandemic. We used eight items, such as "COVID-19 pandemic makes me worried" and "I think I am likely to be infected with COVID-19 pandemic." In-group identity was subjected to the scale compiled by Zheng et al. (2021), which consists of three items, including "I am very similar to everyone else's living conditions under the pandemic" and "I have similar experiences with everyone under the pandemic." The helping behaviors were subjected to the scale compiled by Yue et al. (2017), involving five items, such as "Help other employees when it is clear their workload is too high" and "Lends a helping hand to coworkers when needed." According to the COVID-19 pandemic prevention and control requirements issued by the National Health Commission of the People's Republic of China, the severity of the local pandemic is divided into low-risk (no verified cases or no newly verified cases for 14 consecutive days), moderate-risk (newly verified cases within 14 days, no more than 50 cumulative verified cases, or more than 50 cumulative verified cases, with no pandemic cluster within 14 days), and high-risk (more than 50 cumulative cases, with an pandemic cluster within 14 days), which became the standard for measuring the severity of the COVID-19 pandemic in China.

4. Results

4.1. Reliability and Validity Tests and Correlation Analysis

In this study, Cronbach's α coefficient was used to verify the reliability of each scale. In general, the Cronbach's α coefficient greater than 0.7 is considered acceptable, and a coefficient greater than 0.8 is ideal. The test results demonstrate that Cronbach's α coefficients of risk perception, in-group identity, and helping behavior are 0.76, 0.86, and 0.94, respectively, which indicates that the internal consistency of each variable is high and meets the measurement requirements. Then, the composite reliability (CR) and average variance extracted (AVE) were used to test the validity of the scale. In general, the CR value of a latent variable should exceed 0.7, but is acceptable over 0.6. The normal AVE should be greater than or equal to 0.5 and is acceptable at 0.36–0.5. The test results demonstrate that the CR values of the three variables are all above 0.8, indicating that the aggregate validity of each variable is high. The AVE values of the three variables are all above 0.5, which indicates that each scale has high differential validity.

Because the research conducted measurement at the same time point and adopted the method of employee self-evaluation, it is necessary to further test for discrimination between these variables. Mplus7.0 was used for confirmatory factor analysis. The fitting effect of the three-factor measurement model (risk perception, in-group identity, and helping behavior) is better (χ^2 = 223.997, df = 41, χ^2/df = 5.46, CFI = 0.956, TLI = 0.941, $RMSEA$ = 0.086). Furthermore, this study uses two-factor and one-factor measurement models to test the discriminant validity. The test results show that the fitting effect is significantly worse with one and two factors than with the three-factor model. The results of this analysis demonstrate that the three variables in this study have significant differences, which also demonstrates the absence of any serious problem with common method deviation in this study.

Then, Pearson correlation analysis is used to test the correlation among the variables. Table 1 shows the mean, standard deviation, and correlation coefficient of each variable. Risk perceptions are positively correlated with helping behaviors (r = 0.20, p < 0.01), and the risk perceptions are positively correlated with in-group identity (r = 0.42, p < 0.01). There is a significant positive correlation between in-group identity and helping behavior (r = 0.31, p < 0.01).

Table 1. Means, standard deviation, and correlations among variables (n = 925).

Variables	Mean	SD	1	2	3	4	5	6	7	8	9
Gender	1.39	0.49									
Age	2.84	0.99	−0.45 **								
Education	2.58	1.03	0.25 **	−0.45 **							
Tenure of employment	4.30	1.29	−0.21 **	0.42 **	−0.20 **						
Marital status	2.79	0.65	−0.11 **	0.41 **	−0.31 **	0.48 **					
Experience with the pandemic	3.95	0.24	0.02	0.05	0.02	−0.01	−0.01				
Severity of the local pandemic	2.88	0.40	−0.04	0.11 **	−0.01	0.05	0.12 **	0.03			
Risk perceptions	4.13	0.89	−0.02	−0.05	−0.01	0.03	0.02	−0.03	−0.03		
In-group identity	3.86	0.94	−0.08 *	0.09 **	−0.12 **	0.10 **	0.11 **	−0.01	−0.02	0.42 **	
Helping behaviors	4.69	0.58	0.01	0.09 **	−0.15 **	0.09 **	0.17 **	0.01	0.06	0.20 **	0.31 **

Note: ** p < 0.01, * p < 0.05.

4.2. Hypothesis Testing

H1 test. As shown in Table 2, after controlling for gender, age, education level, marital status, type of company, and experience with the pandemic, the standardized path coefficient of risk perceptions to helping behaviors is 0.230 (p < 0.001); the test results demonstrate that risk perceptions have a significant positive impact on helping behaviors, thus H1 is confirmed. The results of the influence of control variables on helping behaviors are shown in Table 2, with gender (β = 0.092; p < 0.05) having a significant influence on helping behaviors, which proves that women demonstrate more helping behaviors than men. The education level (β = −0.136; p < 0.01) has a significant influence on helping behaviors, indicating that a higher education level is associated with less helping behavior. The type of company (β = −0.101; $p < 0.01$) has a significant influence on helping behaviors, which proves that employees at state-owned enterprises demonstrate more helping behaviors than employees at private enterprises; experience with the pandemic (β = 0.112; $p < 0.01$) has a positive effect on helping behaviors, which demonstrates that employees affected by COVID-19 engaged in more helping behaviors than those who are unaffected. Other variables, such as age (β = 0.031; $p > 0.05$), work tenure at the organization (β = 0.008; $p > 0.05$), and marital status (β = 0.021; $p > 0.05$) have no significant effect on helping behaviors.

Table 2. Direct effect and mediating role.

Variables	Estimate	S.E.	Est./S.E.	*p*-Value
Gender	0.092 *	0.037	2.531	0.011
Age	0.031	0.043	0.708	0.479
Education	−0.136 **	0.040	−3.418	0.001
Tenure of employment	−0.008	0.039	−0.217	0.828
Type of company	−0.101 **	0.033	−3.091	0.002
Marital status	0.112 **	0.039	2.886	0.004
Experience with the pandemic	0.021	0.077	0.651	0.515
RP → HB	0.230 **	0.036	6.467	0.001
RP → IGI	0.558 ***	0.068	8.249	0.000
Mediating effect of in-group identity	0.097 ***	0.021	4.636	0.000

Note: RP is risk perceptions, IGI is in-group identity, and HB is helping behaviors. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

H2 and H3 test. The latent variable modeling method was used to estimate the mediating effect of in-group identity between risk perception and helping behavior with Mplus7.0. The model tests the significance of the coefficient product directly with the bootstrap method, setting bootstrap resampling to 5000 times to test H2 and H3. The test results are shown in Table 2. Risk perceptions have a significantly positive impact on in-group identity (β = 0.558, $p < 0.001$), confirming H2. The coefficient for the indirect impact of risk perception on helping behavior through in-group identity was 0.097 ($p < 0.001$), and the 95% bias-corrected confidence interval ranged from LLCI = 0.084 to ULCI = 0.178, excluding 0, which proves that the mediating effect of in-group identity between risk perception and helping behavior was established, confirming H3.

H4 test. The severity in the local pandemic is a variable divided into high, moderate, and low-risk. In this study, the structural equation model was constructed with Mplus7.0 to compare whether significant differences exist between employees' risk perception and helping behavior in areas with high, moderate, and low-risk, and if significant differences emerged, the moderating effect was confirmed. The test results are shown in Table 3. The regression slope of employees' risk perception to helping behavior in high-risk areas is 0.015 ($p > 0.05$), which fails to satisfy the test for significance. The regression slope of employees' risk perception to helping behavior in areas with moderate risk is 0.064 ($p > 0.05$), which also does not satisfy the significance standard. The regression slope of employees' risk perception to helping behavior in low-risk areas is 0.135 ($p < 0.001$), which satisfies the significance standard. A further comparison demonstrates that the difference in the slopes among the three groups is −0.072 ($p < 0.001$), satisfying the significance standard, and the 95% bias-corrected confidence interval ranged from LLCI = −0.114 to ULCI = −0.038 (5000 bootstrap resamples); it does not contain 0, which confirms that the severity in the local pandemic has a negative moderating effect on the relationship between risk perception and helping behavior—that is, the higher the risk level, the less helping behavior employees perform after perceived risk; the empirical results are the opposite of our research hypothesis, thus H4 is rejected.

Table 3. Results of moderated mediation analysis for the severity of the local pandemic.

Model		Estimate	S.E.	Est./S.E.	p-Value	95% CI	
						Lower	Upper
H4	High-risk areas	0.015	0.082	0.185	0.853	−0.120	0.217
	Areas with moderate risk	0.064	0.036	1.751	0.080	−0.003	0.144
	Low-risk areas	0.135	0.032	4.164	0.000	0.080	0.204
	Differences between groups	−0.072	0.019	−3.679	0.000	−0.114	−0.038
H5	High-risk areas	0.790	0.151	5.245	0.000	0.430	1.048
	Areas with moderate risk	0.413	0.047	8.845	0.003	0.318	0.501
	Low-risk areas	0.428	0.040	10.645	0.001	0.325	0.506
	Differences between groups	−0.805	0.153	−5.255	0.002	−1.070	−0.441
H6	High-risk areas	0.056	0.067	0.830	0.406	−0.028	0.238
	Areas with moderate risk	0.070	0.013	5.211	0.000	0.047	0.100
	Low-risk areas	0.245	0.177	1.381	0.167	0.090	0.737
	Differences between groups	0.119	0.189	0.627	0.531	−0.118	0.623

H5 test. The structural equation model was constructed using Mplus7.0 to compare whether the employees' risk perception in areas with high, moderate, and low risk was significant for in-group identity. The test results are shown in Table 3. The regression slope of the risk perception of employees in high-risk areas for in-group identity is 0.790 ($p < 0.001$), which satisfies the significance standard. The regression slope of employees' risk perception for in-group identity in areas with moderate risk was 0.413 ($p < 0.01$), which satisfies the significance standard; and the regression slope of employees' risk perception for in-group identity in low-risk areas was 0.428 ($p < 0.01$), which also satisfies the significance standard. A further comparison shows that the difference in slope among the three groups is −0.805 ($p < 0.01$), which satisfies the significance standard, and the 95% bias-corrected confidence interval ranges from LLCI = −1.07 to ULCI = −0.44 (5000 bootstrap resamples); it excludes 0, which indicates significant differences among the three groups. H5 is confirmed—that is, the more severe the local pandemic is, the higher the employee's risk perception of in-group identity will be.

H6 test. Because the severity of a local pandemic is a variable divided into three groups, multigroup analysis is used to compare whether significant differences exist in the mediating effect of in-group identity in the three conditions, and the regression model is constructed using Mplus7.0. The test results are shown in Table 3: In high-risk areas, the mediating effectiveness is not significant ($\beta = 0.056$, $p > 0.05$); in areas with moderate risk, the mediating effectiveness ($\beta = 0.070$, $p < 0.001$) satisfies the significance standard; in low-risk areas, the mediating effectiveness is not significant ($\beta = 0.245$, $p > 0.05$). The intergroup difference coefficient of three-group mediating effectiveness is 0.119 ($p > 0.05$), which fails to satisfy the significance standard, and the 95% bias-corrected confidence interval ranges from LLCI = −0.118 to ULCI = 0.623 (5000 bootstrap resamples) and includes 0; therefore, H6 is rejected.

5. Discussion

First, the results discovered that risk perception caused by the COVID-19 pandemic would motivate employees to engage in helping behaviors within the organization, thereby confirming the "tend and befriend" theory's applicability to public crises. Employees actively help colleagues, customers, and other stakeholders around them because of their perception of risk due to an external public crisis. Cannon first proposed the "fight or flight" response pattern in individual behavior due to stress, stating that individuals under stress react by either fighting or fleeing (Taylor et al. 2000). Therefore, existing research results mainly discuss risk perception and negative emotion related to the COVID-19 pandemic, such as fear, loneliness, anxiety, and depression (Han et al. 2021; Yan et al. 2021), and is significantly related to self-interested protective behavior (Zhang and Zhou 2021). However, the theory of "tend and befriend" assumes that people want social support to relieve their anxiety, comfort, and protection through "alliances" under stress, so people engage in more friendly behaviors in order to maintain good social relations (Taylor et al. 2000; Taylor 2006). During the outbreak of COVID-19 pandemic in China, many stories emerged about a large number of ordinary heroes and touching interactions, and behaviors such as helping and cooperating became commonplace. In view of these facts, the research introduces the "tend and befriend" theory, which explains the mechanism of risk perception in helping behavior, helps to enrich the explanatory framework of risk perception in altruistic behavior, and expands the research on risk perception.

Second, this paper verified that risk perception influences helping behavior via the mediating mechanism of in-group identity. This indicates that when a public crisis increases individuals' risk perception, they will first turn their attention to the organization in the hope of obtaining the organization's support and protection to jointly resist risks and crises. As a result, their trust in the organization and sense of belonging are enhanced. They become more willing to assist their colleagues and other stakeholders when they have a stronger sense of identity and belonging to the organization. While the "tend and befriend" theory has been widely accepted, the underlying mechanism underlying the relationship between stress and prosocial behavior is primarily based on physiological indicators. For example, Taylor thinks that humans demonstrate tending and befriending responses to stress responses underpinned by the hormone oxytocin, by opioids, and by dopaminergic pathways. Based on cognitive psychology, this paper elucidates the internal mechanisms underlying risk perception and prosocial behavior. This is in line with views of Yang et al. (2021): the prosocial behaviors displayed by stressed individuals are more likely to be protective mechanisms. Additionally, individuals can acquire additional instrumental resources by engaging in more prosocial behaviors that mitigate threats and pressures. The research results enrich the theoretical research on "tend and befriend" to a certain extent.

Third, this paper discovered that the severity of a local pandemic modifies the relationship between risk perception and helping behavior in a negative manner. Specifically, employees in high-risk areas engage in less helping behavior than those in moderate and low-risk areas, while employees in low-risk areas engage in the most helping behavior. This contradicts the research hypothesis, but it makes sense. This might be because employees in areas with high and moderate risk are at constant risk of infection because of the severity of the pandemic. This anxiety and concern might cause them to pay more attention to their own needs and less attention to that of others. At the same time, because of the measures intended to prevent and control the spread of infection, such as reducing unnecessary contact and isolation at home, the opportunities to help others are reduced to a certain extent. However, employees in low-risk areas live in a relatively safe environment, so they are more likely to display altruistic behavior in groups after they perceive risks and assist colleagues and customers, so as to build good social connections and collectively withstand risk. The findings corroborate those of Wang and Sun (2021) regarding the voluntary activities of Chinese citizens during COVID-19: the voluntary activities in high-risk areas decreased. That is, the relationship between individual risk perception and prosocial behaviors might be influenced by the intensity of risk stimuli. The boundary further demonstrates that employees' helping behaviors are not produced out of thin air, and the relationship between risk perception and helping behaviors cannot be separated from specific situational factors.

Forth, this paper verified that the severity of the local pandemic positively moderates the relationship between risk perception and in-group identity. In other words, employees in low-risk areas have a lower in-group identity than those in medium- and high-risk areas, while employees in high-risk areas have the highest in-group identity. Through a combined analysis of the research results on H4 and H5, we find an interesting phenomenon: when risk is low, individuals fall into the "tend and befriend" response mode, they are not only willing to "align" (i.e., improve their in-group identity), but also to engage in more helpful behaviors. However, as the risk level increases, individuals do not fully engage in the "tend and befriend" response mode. They are fervent in their desire to "align" (i.e., in-group identity is significantly improved), but significantly reduce helping behaviors. According to the "tend and befriend" theory, individuals experiencing stress have two primary behavioral responses: caring and alliances. This study adds to the body of knowledge regarding the "tend and befriend" theory and further states that individuals exhibit distinct behavioral responses to varying degrees of perceived risk. That is, "alliance" and "caring" do not have to occur concurrently, and both can occur in low- and moderate-risk scenarios. People are more likely to "forge an alliance" and less likely to "care" in high-risk situations. As such, the prosocial behaviors of individuals under stress (e.g., helpful behaviors) might also be a protective mode, and individuals can mitigate the threats and pressures they face by participating in more prosocial behaviors (Steinbeis et al. 2015). The research in this paper makes up for the lack of research on the role of the degree of risk between risk perception and behavioral response in previous empirical studies and can further deepen the discussion on the mechanism in the relationship between risk perception and altruistic behavior.

6. Conclusions

This paper focuses on the COVID-19 pandemic, a major risk event. We obtained data on 925 employees of 20 enterprises in August 2021, through an investigation of enterprises in the Henan and Sichuan provinces, which experienced severe pandemic conditions. Using structural equation modeling, we tested our research hypotheses. The results confirm, first, that risk perceptions have a positive impact on helping behaviors. Second, risk perception has a positive impact on in-group identity, and in-group identity plays a mediating role in the relationship between risk perception and helping behavior. Third, the severity of the local pandemic negatively moderates the relationship between risk perception and helping behavior, but positively moderates the relationship between risk perception and in-group identity.

In real life, people face various crises and risks, such as earthquakes, typhoons, floods, and infectious disease. Studying the influencing mechanism in employees' risk perceptions on helping behaviors in public crises and explaining employees' psychological and behavioral performance after risk perception under different risk levels can give managers experience in risk management: in the face of huge external risks, employees often tend to "huddle together for warmth" and engage in altruistic behaviors. At this time, managers should strengthen the management of employees' organizational identity and prosocial behavior, form a good atmosphere for teamwork in the company, jointly withstand risks, and help enterprises to survive difficulties. The research in this paper also provides an explanation for various prosocial and altruistic behaviors, such as charitable donations and volunteerism, which offers theoretical guidance to relevant government agencies on how to effectively organize social forces and jointly confront risks in a time of crisis, and help to further improve the national risk management system.

Our study has some limitations. First, the sample comes mainly from China, and no similar work has been conducted in other countries around the world with a COVID-19 outbreak. Countries and regions have different cultural backgrounds, which vary from that of China, whose culture is collectivist and subordinates individual interests to collective interests. By contrast, the culture in Western countries emphasizes individualism and advocates independence and freedom. Such differences in cultural background could affect the results, so the generalizability of this study needs to be tested. Second, the data used come from employees' self-reporting, which could limit their effectiveness, as self-reporting is easily influenced by variations in self-expression (self-deception; impression management). Although our results pass tests of reliability and validity, the problem of the common method deviation is inevitable. In the future, the empirical sampling method, experimental method, multipoint investigation method, and so on can be used to improve the accuracy and external validity of the results. Finally, this paper discusses the influence of employees' risk perception on helping behaviors, as well as the corresponding mediating mechanism and boundary conditions, which have a certain theoretical value. However, the survey respondents are mainly employees of a company, and their helping behaviors mainly involve stakeholders, such as colleagues and customers. It is not clear whether they would display the same helping behavior in dealing with strangers; therefore, that offers an avenue for future research. Thus, the study could be expanded to encompass the general public to explore whether they also would help strangers, so as to further test the generalizability of the "tend and befriend" response model across risk scenarios.

Author Contributions: Conceptualization, L.Y. and Y.-C.J.W.; methodology, L.Y.; formal analysis, L.Y.; resources, Y.-C.J.W.; writing—original draft preparation, L.Y. and Y.-C.J.W.; writing—review and editing, L.Y. and Y.-C.J.W.; supervision, Y.-C.J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

Alen, Nicholas V., LillyBelle K. Deer, Mona Karimi, Elis Feyzieva, Paul D. Hastings, and Camelia E. Hostinar. 2021. Children's altruism following acute stress: The role of autonomic nervous system activity and social support. *Developmental Science* 24: e13099. [CrossRef] [PubMed]

Cikara, Mina, and Jay J. Van Bavel. 2014. The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science* 9: 245–74. [CrossRef] [PubMed]

Crane, Andrew, and Trish Ruebottom. 2011. Stakeholder theory and social identity: Rethinking stakeholder identification. *Journal of Business Ethics* 102: 77–87. [CrossRef]

Debono, Amber, Natarshia Corley, and Mark Muraven. 2020. Why am i left out? interpretations of exclusion affect antisocial and prosocial behaviors. *The American Journal of Psychology* 133: 63–79. [CrossRef]

Farh, Jiing-Lih, Rick D. Hackett, and Jian Liang. 2007. Individual-level cultural values as moderators of perceived organizational support-employee outcome relationships in China: Comparing the effects of power distance and traditionality. *Academy of Management Journal* 50: 715–29. [CrossRef]

Fowler, James H., and Nicholas A. Christakis. 2010. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences* 107: 5334–38. [CrossRef]

Han, Qing, Bang Zheng, Maximilian Agostini, Jocelyn J. Bélanger, Ben Gützkow, Jannis Kreienkamp, Anne Margit Reitsema, Jolien A. van Breen, N. Pontus Leander, and PsyCorona Collaboration. 2021. Associations of risk perception of COVID-19 with emotion and mental health during the pandemic. *Journal of Affective Disorders* 284: 247–55. [CrossRef]

Hitlin, Steven, Hye Won Kwon, and Rengin Firat. 2021. In- and out-groups across cultures: Identities and perceived group values. *Social Science Research* 97: 102569. [CrossRef]

Kovoor-Misra, Sarah. 2009. Understanding perceived organizational identity during crisis and change. *Journal of Organizational Change Management* 22: 494–510. [CrossRef]

Levy, Kenneth N., Jessica K. Hlay, Benjamin N. Johnson, and Courtney P. Witmer. 2019. An attachment theoretical perspective on tend-and-befriend stress reactions. *Evolutionary Psychological Science* 5: 426–39. [CrossRef]

Leyens, Jacques-Philippe, Paola M. Paladino, Ramon Rodriguez-Torres, Jeroen Vaes, Stephanie Demoulin, Armando Rodriguez-Perez, and Ruth Gaunt. 2016. The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality & Social Psychology Review* 4: 186–97.

Lim, Daniel, and David DeSteno. 2016. Suffering and compassion: The links among adverse life experiences, empathy, compassion, and prosocial behavior. *Emotion* 16: 175–82. [CrossRef] [PubMed]

Maier, Silvia U., Aidan B. Makwana, and Todd A. Hare. 2015. Acute stress impairs self-control in goal-directed choice by altering multiple functional connections within the brain's decision circuits. *Neuron* 87: 621–31. [CrossRef] [PubMed]

Mai, Yingping, Yenchun Jim Wu, and Yanni Huang. 2021. What type of social support is important for student resilience during COVID-19? a latent profile analysis. *Frontiers in Psychology* 12: 2463. [CrossRef] [PubMed]

Margittai, Zsofia, Tina Strombach, M. van Wingerden, M. Joëls, Lars Schwabe, and Tobias Kalenscher. 2015. A friend in need: Time-dependent effects of stress on social discounting in men. *Hormones and Behavior* 73: 75–82. [CrossRef] [PubMed]

Meng, Lu, Fengjun Liu, Kun Duan, Xi Li, and Tongmao Li. 2021. Impact of intensity of shared human events on citizen behavior in the novel coronavirus pneumonia: Based on the perspective of corporate social responsibility perception. *Nankai Business Review* 5: 1–28.

Miao, Xiaoya, Xin Sun, Yi Kuang, and Zuojun Wang. 2021. Co-experiencing the same negative emotional events promotes cooperation. *Acta Psychologica Sinica* 53: 81–94. [CrossRef]

Nakayachi, Kazuya, and Taku Ozaki. 2014. A method to improve trust in disaster risk managers: Voluntary action to share a common fate. *International Journal of Disaster Risk Reduction* 10: 59–66. [CrossRef]

Rao, Li lin, Ru Han, Xiaopeng Ren, Xinwen Bai, Rui Zhang, Huan Liu, Zuojun Wang, Kan Zhang, and Shu Li. 2011. Disadvantage and prosocial behavior: The effects of the Wenchuan earthquake. *Evolution and Human Behavior* 32: 63–69. [CrossRef]

Savadori, Lucia, and Marco Lauriola. 2021. Risk perception and protective behaviors during the rise of the Covid-19 outbreak in Italy. *Frontiers in Psychology* 11: 1–20. [CrossRef]

Schneider, Claudia R, Sarah Dryhurst, John Kerr, Alexandra L. J. Freeman, and Sander Van Der Linden. 2021. COVID-19 risk perception: A longitudinal analysis of its predictors and associations with health protective behaviours in the United Kingdom. *Journal of Risk Research* 24: 294–313. [CrossRef]

Staub, Ervin, and Johanna Vollhardt. 2008. Altruism born of suffering: The roots of caring and helping after victimization and other trauma. *American Journal of Orthopsychiatry* 78: 267–80. [CrossRef] [PubMed]

Steinbeis, Nikolaus, Veronika Engert, Roman Linz, and Tania Singer. 2015. The effects of stress and affiliation on social decision-making: Investigating the tend-and-befriend pattern. *Psychoneuroendocrinology* 62: 138–48. [CrossRef] [PubMed]

Taylor, Shelley E. 2006. Tend and befriend: Biobehavioral bases of affiliation under stress. *Current Directions in Psychological Science* 15: 273–77. [CrossRef]

Taylor, Shelley E., Laura Cousino Klein, Brian P. Lewis, Tara L. Gruenewald, Regan AR Gurung, and John A. Updegraff. 2000. Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review* 107: 411–29. [CrossRef]

Turner, John C. 1981. Toward a cognitive redefinition of the social group. *Cahiers De Psychologie Cognitive* 1: 93–118.

Tu, Xingyong, Leijie Liu, Yiping Zhang, Yaya Peng, and Chenglu Lin. 2020. How prosocial motivation and helping behavior influence contextual performance: The mechanism and context situation of cognitive trust. *Nankai Business Review* 23: 203–13.

Vesa, Peltokorpi. 2020. Host country national employees' prosocial behavior toward expatriates in foreign subsidiaries: A common ingroup identity model perspective. *International Business Review* 29: 1184–99.

Vezzali, Loris, Alessia Cadamuro, Annalisa Versari, Dino Giovannini, and Elena Trifiletti. 2015. Feeling like a group after a natural disaster: Common ingroup identity and relations with outgroup victims among majority and minority young children. *British Journal of Social Psychology* 54: 519–38. [CrossRef]

Von Dawans, Bernadette, Urs Fischbacher, Clemens Kirschbaum, Ernst Fehr, and Markus Heinrichs. 2012. The social dimension of stress reactivity: Acute stress increases prosocial behavior in humans. *Psychological Science* 23: 651–60. [CrossRef]

Wang, Yiying, and Yingying Sun. 2021. Discussion on voluntary behavior during the prevention and control of COVID-19 from the perspective of international comparison. *Statistics and Management* 36: 36–42.

Wei, Jiuchang. 2020. The present and future research on public risk perception evolution and protective action decision amid public health emergencies. *Bulletin of National Natural Science Foundation of China* 34: 776–85.

Williams, Steve, and Yvonne Wong Wee Voon. 1999. The effects of mood on managerial risk perceptions: Exploring affect and the dimensions of risk. *Journal of Social Psychology* 139: 268–87. [CrossRef] [PubMed]

Xie, Xiaofei, Yilu Wang, Siyi Gu, and Wei Li. 2017. Is altruism just other-benefiting? A dual pathway model from an evolutionary perspective. *Advances in Psychological Science* 25: 1441–55. [CrossRef]

Xue, Yali, and Yalin Wang. 2021. The emerging emotions and life changes in the risk situation. *Journal of HIT (Social science editon)* 23: 50–58.

Xu, Dan. 2018. The relationship between traumatic experience and prosocial behavior, and its mechanism. *Chinese Social Psychological Review* 2: 198–214, 249–250.

Yang, Seong Won, Scott M Soltis, Jason R. Ross, and Giuseppe Joe Labianca. 2021. Dormant tie reactivation as an affiliative coping response to stressors during the COVID-19 crisis. *Journal of Applied Psychology* 106: 489–500. [CrossRef]

Yan, Jiaqi, Sunghoon Kim, Stephen X. Zhang, Maw-Der Foo, Aldo Alvarez-Risco, Shyla Del-Aguila-Arcentales, and Jaime A. Yáñez. 2021. Hospitality workers' covid-19 risk perception and depression: A contingent model based on transactional theory of stress model. *International Journal of Hospitality Management* 95: 102935. [CrossRef]

Yue, Yumeng, Karyn L. Wang, and Markus Groth. 2017. Feeling bad and doing good: The effect of customer mistreatment on service employee's daily display of helping behaviors. *Personnel Psychology* 70: 769–808. [CrossRef]

Zhang, Yaping, and Ronggang Zhou. 2021. Promoting social distancing and preventing panic buying during the pandemic of COVID-19: The contributions of people's psychological and behavioural factors. *Journal of Public Health* 4: 1–15.

Zheng, Chun dong, Ning Liu, Nan Feng, and Minqiang Li. 2021. Generating mechanism of individual prosocial behavior in the context major public threats: A case of the novel coronavirus pandemic. *Journal of Management Sciences in China* 24: 63–79.

Zhou, Tianshuang, Qing Hu, and Lijuan Cui. 2018. Common ingroup identity and intergroup helping: The mediating effect of intergroup threat. *Psychological Research* 11: 333–42.

Zhu, Yue, Yimo Shen, Xia Zhou, and Dong Yang. 2020. A conditional process model of public negative emotions and mental health under the COVID 19: The moderating role of interpersonal alienation. *Journal of Southwest University (Natural Science Edition)* 42: 1–10.

Journal of
Risk and Financial Management

MDPI

Article

Managing Financial Risks while Performing International Commercial Transactions. Intertemporal Lessons from Athens in Classical Times

Emmanouil M. L. Economou , Nikolaos A. Kyriazis and Nicholas C. Kyriazis *

Department of Economics, University of Thessaly, 38333 Volos, Greece; emmoikon@uth.gr (E.M.L.E.); knikolaos@uth.gr (N.A.K.)
* Correspondence: nkyr@uth.gr

Abstract: In this paper, we mainly focus on two institutional aspects that are related to financial risk, that is, profiteering and the use of non-fraudulent coins when performing financial transactions. We argue that these two prerequisites were important for the success of the commercially oriented economy of the Athenian state in comparison with its allies in the East Mediterranean during the classical period. In particular, we briefly explain the structure of the Athenian economy, and then we focus on the *agoranomoi* and the *dokimastai*, the two main financial institutions related to (i) measures against profiteering and (ii) ensuring the purity of the currency when performing commercial transactions. Then, following a game theoretical approach, we provide a fictional example as to how the two institutions functioned in practice. Our findings confirm that these institutions were crucial in reducing financial risk when performing international commercial transactions, since they provided symmetrical information on the quality and purity of the currencies circulating in the Athenian economy. In the case of the Athenian state, we further convey that measures against profiteering and the use of unadulterated currency comprise intertemporal axioms, in the sense that their importance is not merely a phenomenon of modern times, but rather, on the contrary, one that dates back to much earlier times.

Keywords: financial risk; financial institutions; international commerce; profiteering; non-fraudulent currency; classical Athens

JEL Classification: D81; G28; N23; N43

Citation: Economou, Emmanouil M. L., Nikolaos A. Kyriazis, and Nicholas C. Kyriazis. 2021. Managing Financial Risks while Performing International Commercial Transactions. Intertemporal Lessons from Athens in Classical Times. *Journal of Risk and Financial Management* 14: 509. https://doi.org/10.3390/jrfm14110509

Academic Editor: Thanasis Stengos

Received: 11 August 2021
Accepted: 12 October 2021
Published: 22 October 2021

1. Introduction

In this paper, inspired by the Athenian model of economic organization in classical times (508–323 BCE), we provide further evidence to the international bibliography regarding the main institutional mechanisms introduced by the ancient Athenians for the purpose of managing financial transactions and reducing financial risk, thus establishing their state as the dominant economy of the times in the East Mediterranean, as verified by a multitude of scholars nowadays (here, due to space limitations, we will name only Figueira (1998); Amemiya (2007); Gabrielsen (2014); Bresson (2016a, 2016b); Harris and Lewis (2016); Woolmer (2016); O'Halloran (2018); Bitros et al. (2020); and Economou et al. (2021), but there are many others).

We mainly focus on two institutional mechanisms related to financial risk, that is, the avoidance of market failures and the use of non-fraudulent coins when carrying out local or international financial transactions. In particular, we argue that a key reason for the success of the Athenian economy and the great commercial expansion dynamics it exhibited was, among other things, a combination of market supervision against profiteering and the use of a genuine currency (i.e., non-fraudulent coinage) that was to be circulated throughout the Athenian economy, as well as throughout the economies with which the Athenians were trading in the East Mediterranean.

The modern economic theory and evidence reveals the importance of such institutions to the success (or not) of economic systems in the long run. Regulation, allowing the supply and demand mechanism to work smoothly, and permitting no discrepancies to appear in such profiteering or the use of adulterated currency, provides the foundation of a healthy economic environment. On the other hand, market failures, including profiteering and the use of fraudulent currency, are factors of financial risk. This situation leads to inferior Pareto outcomes that are not efficient from a societal point of view.

North (1990) and Hodgson (2015), among others, argue, from an institutionalist point of view, that markets, in order to flourish, need robust institutions that protect the legal aspect of the transactions between different parties, guaranteeing that the parties are being treated fairly (North 1990; Hodgson 2015). Moreover, Mishkin and Eakings (2015, p. 7), among others, add that financial institutions are what make financial markets work, because without them, financial markets would not be able to move funds from people who save to people who seek productive investment opportunities. According to Arrow (1969), it is for this reason that market failures are often the reason that self-regulatory organizations, governments, or supra-national institutions intervene in a particular market.

Profiteering, that is, economic situations in which individuals pursue self-interest only, is a factor of market failures. Counterfeiting, for its part, can negatively affect a currency's functions as a store of value and as a medium of exchange. It can affect the demand for currency through a loss of confidence in the use of currency, as described by the famous Gresham Law. This can lead to financial risk, since it can erode confidence in the market. Market confidence is weakened if the public perceives there to be a greater risk that they could unknowingly receive a counterfeit as payment (Viles et al. 2015). Counterfeiting, in general, causes substantial financial losses and opportunity costs to businesses worldwide (Martinez and Jaeger 2016). In some (extreme) cases, counterfeiting can even severely destabilize financial systems (Nosal and Wallace 2007).

There are a plethora of cases worldwide that can serve as paradigmatic examples of counterfeiting actions; for example, from the late 1980s until the early 2000s, so-called *Superdollars* were being circulated worldwide, printed by unknown entities. These fake banknotes were almost indiscernible from genuine USD 100 bills. The US government accused, among others, North Korea of producing and passing them off in various countries.[1] Another paradigm is related to a recent (June 2020) attempt to introduce several hundred counterfeit banknotes into the markets of Spain and Romania, which was finally effectively prevented by their prosecuting authorities.[2]

The methodology in our study follows a theoretical approach supported by a game theoretical analysis. Our methodology is basically theoretical in nature because we are dealing with an ancient economy in which cliometric/statistical data are absent. An empirical analysis on the basis of which to test our hypotheses using tangible statistical data is (unfortunately) impossible to carry out. Thus, in order to explain how the institutions we describe below functioned in practice, we have to rely mainly on the findings of historians, further elaborated through our own synthesis of this evidence, as well as by using game theoretical analysis as a further methodology for the interpretation of this evidence.

Our paper is organized as follows: In Section 2, we provide a brief historical analysis regarding the institutional status of the Athenian economy during the classical times. We perform this analysis since the institutions we describe in Sections 3 and 4 could not have flourished unless a sophisticated (for the era) package of financial institutions was present. In Section 3, we focus on the description of two key institutions that are related to the reduction in financial risks and are greatly involved in the success of the Athenian market, the *agoranomoi* and the *dokimastai*. These institutions were related to the prevention of profiteering and of any other kind of fraud in the market, in combination with performing financial transactions through non-fraudulent coinage.

In Section 4, through a fictional case, which is based on historical evidence, we link these two important institutions with a fictional financial transaction between two merchants, and we analyze how they functioned in practice, as a means of preventing delinquent behavior regarding market transactions. In particular, we develop a two-player game where mixed equilibria are derived, and indicate that higher payoff for the society as a whole is achieved when institutional mechanisms motivate players to act honestly. Section 5 concludes by arguing that the success of commercial cooperation between two partiers can be achieved by auditing institutional mechanisms to minimize the risks of performing financial transactions; we found that this does exist and is present in practice.

Section 5 also provides some proposals that could be useful for economies nowadays that come as an inspiration, which derives from the functioning of the Athenian auditing institutions.

2. The Athenian Economy in Brief

The discussion regarding the status of the ancient Greek economy dates back to the late 19th century when two prominent German historians, Karl Bücher (1847–1930) and Eduard Meyer (1855–1930), started a debate, which, today, is known as the *Bücher–Meyer controversy*. Obviously, due to the limitations of space, it is not our intention to reproduce all this evidence in a literature review, but readers could consult Tridimas (2019), among others. This discussion was revived after WWII, mainly due to the influential works of a series of Marxist-oriented historians, such as Moses Finley (1973, 1983), Austin and Vidal-Naquet (1980), and de Ste Croix (1981). These historians were advocates of the *primitiveness* of the ancient Greek and Roman economies.

However, during the last 30 years, a new and convincing trend towards the *formalist* approaches (re)appeared in the international academic literature, regarding the structure, the institutions, and the organization of ancient economies, focusing on those of ancient Greece and Rome. *Formalists* argued that societies such as those of the ancient Greeks and Romans functioned with their own economic autonomy, in a way that was based on a primitive version of what we characterize as a free market economy nowadays. They further argue that modern concepts, such as rational choice and utility maximization, were present in these economies, similarly to nowadays (see, for example, O'Halloran 2018). Formalist approaches include methodological concepts, such as *new institutional economics*, which, according to Lyttkens (2013), is a very efficient methodology to interpret ancient history. This includes not only the economic history of ancient Greece (by focusing mainly on the history of Athens), but also the history of Rome (see, among others, Temin (2012) and Verboven (2021)). This new trend, which is supported by eminent historians, economists, and political scientists, has almost prevailed in the literature in the last 30 years, and has significantly refuted the older Finleyan approaches.

Due to the limited space in the analysis that follows in this section, we can only retrieve a part of such evidence and argumentation. However, this discussion, regarding the *primitivist* vs. modern or *formalist* characteristics of the ancient Greek economy, is important to the analysis that follows, since, as argued in the Introduction, the kind of institutions we describe in Section 3 could not have flourished unless a sophisticated (for the era) package of financial institutions was present.

To be more precise, Amemiya (2007); Lyttkens (2013); Bresson (2016a, 2016b); and Bitros et al. (2020), among other authors, analyzed the structure of the institutions and markets in the Athenian economy (mainly), while Bresson (2016a, 2016b); O'Halloran (2018); and Bitros et al. (2020) shed more light on the critical issue that, in actuality, the Athenian economy functioned as a primitive version of a market type of an economy where institutions such as public magistrates, who protected the market against profiteering, known as the *agoranomoi*, did truly exist. Bitros and Karayannis (2008) further analyzed the related issue of entrepreneurship in Athens, under free market economy principles. Economou and Kyriazis (2017, 2019a, 2019b) argued that an efficient regime of property rights protection and protection of commercial contracts, through established law, juries, and courts, did function very well in classical Athens and later on. Of course, there is an intertemporal axiom that property rights protection is a very important prerequisite, enabling commercial transactions to become credible (Hodgson 2015). Ober (2008) adds that this attitude in Athens effectively reduced transactional costs under a Coasian logic. In addition, E. Cohen (1992), in his seminal book the *Athenian Economy and Society: A Banking Perspective,* exhaustively analyzed the sophisticated way in which banks functioned in classical Athens. The Athenians had established, among others, the following banking service institutions (also see Bitros et al. 2020, p. 110). They performed the following tasks:

- Exchanged coins and foreign currencies;
- Accepted deposits and carried out payments on behalf of their customers;
- Extended loans to various business operations, including bottomry loans in shipping, and even financing of consumer credit;
- Provided sureties, negotiated claims, and offered guarantees and personal advice to important customers;
- Accepted documents and valuables for safekeeping;
- Facilitated export–import activities by settling payments among importers from, and ex-porters to, merchants abroad.

At its peak, the banking industry numbered not less than 30 bankers, such as the famous case of Passion (Cohen 1992, p. 31; Bitros et al. 2020). Cohen (1992), and Homer and Sylla (2005), among others, provide various known cases of given loans. The ordinary interest rate for someone who borrowed money from a bank varied between 12 and 18%. There were also cases of even lower interest rates. Loans contracted at an interest rate of 10% were considered to be very favorable for the borrower. Financially riskier maritime loans were also offered. Acton (2014, pp. 252–53) writes that contracts specified the route and timing, and interest was charged according to the length of the ship's voyage and the level of danger inhibited, which often depended on the time of year. Those merchants, when they needed to take loans to finance such activities, paid higher interest rates, from 15 to 35%, depending on the degree of risk of a voyage (cargo with or without a return trip, etc.). In exceptional cases, interest rates increased to as high as 100% (Schefold 2011; Bitros et al. 2020). A typical two- to three-month journey, such as from Athens to the Bosporus and back, was accompanied by an interest rate of 20–30%, while a shorter, one-way voyage, such as from Sestos (an ancient city in Thrace) to Athens, might be worth 12.5% (Acton 2014, pp. 252–53).

The above evidence reveals an intertemporal axiom that was also applied in the cases of Athens and its allies. The level of interest rate was directly and strongly related to the level of financial risk of these loans; for example, a risky voyage, which entailed uncertainties and dangers, such as the ship being attacked by pirates or facing a storm on the high seas, with the danger of losing the cargo, entailed much higher interest rates.

Amemiya (2007) and Acton (2014) provided evidence regarding insurance services that were linked to the loans provided by the banks, for performing efficient international commercial transactions. In other words, to reduce the financial risks of maritime trade, the Athenians introduced insurance, similarly to many other wealthy maritime societies of later times, such as Venice during the Middle Ages, and the United Provinces (Dutch Republic) and England in early modern Europe (1549–1789 AD).

Acton (2014) and Gabrielsen (2014), among others, provide evidence regarding the existence of a primitive version of Athenian joint stock companies. These companies were important for the promotion of very extensive international commercial transactions between the city-state of Athens and other states, and, of course, among Athens and its more than 316 allies in the Delian League during the classical period (Figueira 1998). Bitros et al. (2020) and Economou et al. (2021) analyze how decisions on public finance and public spending were taken by the two most important Athenian policymaking institutions, the Athenian Assembly of citizens (known as *Ecclesia*) and the *Council of the Five Hundred* (known as *Boule*). Economou and Kyriazis (2019a) argue that the process of the transformation of the Athenian economy from an agrarian into a maritime economy, during the classical period, also led to "industrialization" in a variety of areas, such as the development of primitive versions of small industries, with some examples being metalwork, the shipping industry, silver mining, ceramics, carpentry, paints, fabric, etc.

All these transformational procedures rendered Athens the first economy in history in which "services" and "industrial" (handicrafts) sectors contributed more to (what we nowadays call) GDP than agriculture (Kyriazis and Michael 2004; Halkos and Nicholas 2010; Economou and Kyriazis 2019b).

Arguably, the discussion regarding the achievements of the Athenian economy is also related to the issue of economic growth. Even though there are no cliometric data available, authors such as Bergh and Lyttkens (2011); Ober (2015); Harris and Lewis (2016); Carugati et al. (2019); Economou and Kyriazis (2019a); and Bitros et al. (2020), using different methodologies, have convincingly argued that economic growth was achieved in Athens, at least during the periods of the famous political figures of Pericles in the 5th century BCE, and Eubulus and Lycurgus during the 4th.

The above references consist only a small part of the international bibliography in the last 30 years that revisits the older, though influential, views of M. Finley and his followers, regarding the *primitiveness* of the ancient Greek (by mainly focusing on the Athenian) economy; for example, Mackil (2013, p. 238) characteristically writes that 'Finley's model has been gradually discredited by specific arguments, and that historians now have begun to return to questions of political economy and the intervention of states in economic activities'. Moreover, Harris and Lewis (2016, p. 8), among others, add the following:

> 'In the past fifteen years, however, some ancient historians have shown a willingness to pay more attention to the role of markets in the economy of the ancient Greek polis'.

3. The *agoranomoi*, the *dokimastai* and the Issue of Using Non-Fraudulent Currency

In this section, we argue that a key reason for the success of the Athenian economy and the great commercial expansion dynamics it exhibited was, among others, a combination of two qualitative elements—market supervision against profiteering, and using genuine, non-fraudulent currency to be circulated in the Athenian economy and in parallel with the economies with which the Athenians were trading.

As was already mentioned, profiteering is related to market failures, which are related to financial risks, and is often the reason that governments intervene in a particular market, so as to remedy this critical discrepancy. This is exactly what happened with the case of the institution of the *agoranomoi* in ancient Greece and in classical Athens in particular. The *agoranomoi* (in singular, *agoranomos*) was a group of public magistrates responsible for supervising the Athenian market, the so-called *Agora*, ensuring, among other duties, that the prices of the goods in the market were not excessive and, in general, for securing the smooth functioning of the market. Opportunistic price increases, through market manipulation or exploitation of sudden changes in supply and demand, were considered unfair and abusive by consumers. In cases of treachery or attempts to sell fraudulent products, the *agoranomoi,* who caught those sellers on the spot, imposed heavy fines on them through their own initiative and judgement, after, in many cases, also hearing complaints from consumers who believed that they had been cheated in some way; for example, that they had been sold adulterated or defective goods (Lanni 2018, p. 63).

Thus, heavy fines were imposed on those merchants who were trying to benefit themselves through profiteering. The *agoranomoi* had to ensure that the quality of the products that were circulating in the market met their specifications (Amemiya 2007; Ober 2008, p. 253; Fröhlich 2013, p. 258; Bresson 2016a; O'Halloran 2018, p. 273; Rahyab 2019; Bitros et al. 2020). The *agoranomoi* further solved disputes among merchants or between buyers and sellers in the market. Moreover, this intervention, according to Harris and Lewis (2016, p. 30), ensured that transactions were free of fear of violence or intimidation. The intervention of the *agoranomoi* also lowered what we know as *transactional costs* nowadays (Ober 2008, pp. 106–8). It also lowered *information asymmetries* during the transactions (Bresson 2016a).

The office to which the *agoranomoi* belonged, known as the Athenian *Agoranomia* (literally meaning 'market police'), was located in the so-called *Stoa of Zeus*, near to the

Poikille Stoa and the *Basileios Stoa* on the east side of the famous Athenian Agora, and, more specifically, on the hill of the *Agoraios Kolonos*. According to the Athenian laws, in order to perform commercial activity, each seller of goods in the market was obliged to use a special scale that he had to procure or hire from the *Agoranomia*, with which he would weigh the goods in front of his/her customers, so that cases of fraud were avoided. Special scales bore the official stamp, with a seal of the state on them (Elliott 2018, p. 9; Bitros et al. 2020). In passage 24.112 of his work *Against Timocrates,* Demosthenes informs us that when an *agoranomos* had no experience, or he was ignorant regarding an issue concerning his post, or was found guilty of theft, he was punished, but with not a very severe penalty (Demosthenes 1939).

It is obvious that, thanks to the supervisory role of the *agoranomoi*, transactions between two or more parties could take place without the fear of financial loss for any party, due to a possible attempt by one member of the transaction to defraud the other. Furthermore, the supervisory role of the *agoranomoi* reduced the financial risks not only between two parties, but also in favor of the financial system of the Athenian city-state as a whole, in which a multitude of transactions were taking place. This means that through their intervention in the market, information asymmetries regarding real prices in the market, as well as the purity of the currency, were significantly diminished, to the benefit of the consumers as a whole.

As a final comment, the importance of the *agoranomoi* as an institution can be also deduced by the fact that other states introduced them too, and their existence is also attested for the Hellenistic period (323–146 BCE). Mackil (2013, pp. 268–69), who wrote an important book regarding ancient Greek federalism, with the title *The Greek Koinon,* has argued that the *agoranomoi* were widely attested throughout the Greek world, from the fourth century onwards. Economou (2020) conveys this by providing evidence that is focused exclusively on the Achaean League, a federal state in the Peloponnese (South Greece). Manning (2012) and Mackil (2013) provide further evidence for the Ptolemaic Kingdom of Egypt. Rahyab (2019) further conveys the existence of the *agoranomoi* throughout the areas of Greek influence (such as Anatolia and Egypt), even if this institution did not function in exactly the same way in each of these areas.

Regarding the second important prerequisite that our paper deals with, using pure currency in the market, it is important to bear in mind that no fraudulent coins were allowed to be circulated in the Athenian economy. In particular, a special law, known as Nicophon's Law, was introduced in 375/4 BCE, by the name of its initiator in the Athenian Assembly of citizens, which, among other things, required that all foreign-made imitations of the Athenian drachma that were found with a lower silver content, or that had a bronze or lead core, had to be confiscated immediately from the market, then cut, and their metal was then dedicated to a special temple, which also functioned as a state treasury, the so-called *Treasury-Temple of the Mother of Gods*. However, good imitations of Athenian drachmae, made by mints from other states, could be circulated in the city of Athens. Furthermore, other "national" currencies issued from other city-states, such as Aegina or Corinth, could also be circulated in the Athenian economy, providing that they were not fraudulent, that is, they contained the proper quantity of silver or gold (Figueira 1998; Bitros et al. 2020; Economou et al. 2021).

To implement Nicophon's Law, two public slaves were introduced, known as *dokimastai* ('testers'–in singular, the *dokimastes*), who had a bench in the Agora of Athens and at the harbor of Piraeus, and examined whether foreign currencies were fraudulent. Their tasks were related to great responsibilities and had direct effects on the lives of citizens. An earlier version of the institution of the *dokimastai* was dated back to 398/7 (Stroud 1974, pp. 166–67, 176–77).

In passage 24.212 of his work *Against Timocrates,* Demosthenes further argues that whoever tried to promote fraudulent coins in the market was put to death by law, which was the most severe penalty in the Athenian state. This actually implies that the Athenians had understood the importance of not using fraudulent currency because this would have caused serious malfunctions, such as the disorganization of the Athenian market, and the erosion of the character of the twin city of Athens and Piraeus as the greatest commercial center in the East Mediterranean during the classical period. Additionally, since this privilege of the Athenians was directly related to their overall welfare and progress, anyone who tried to undermine this achievement deserved, according to their mentality, the worst and most severe punishment of all, death.

The rationale behind the protective law of Nicophon lies, according to Ober (2008, p. 233), in the very essence of the famous Gresham's Law. It appears that the perception of the importance of protecting the credibility of the currency was so ingrained in the Athenians that even a comedian, the famous Aristophanes, strongly criticized the phenomenon of fraudulent and debased coins that was observed in his era during the Peloponnesian war, even if this was only a temporal phenomenon that only appeared due to the extraordinary circumstances that Athens faced during the war. In verses 718–724 of his work *Frogs*, Aristophanes wrote the following:

> 'Many times, it seems to us the city has done the same thing with the best and the brightest of its citizens as with the old coinage and the new gold currency. For these, not counterfeit at all, but the finest it seems of all coins, and the only ones of the proper stamp, of resounding metal amongst Greeks and foreigners everywhere, we never use but the inferior bronze ones instead, minted just yesterday or the day before with the basest stamp . . . ..' (Aristophanes 1994).

There is no doubt that, due to the economic financial institutions, as mentioned in Section 2, as well as the *agoranomoi* and the *dokimastai*, a very productive and extensive commercial network was established, with Piraeus functioning, according to Cohen (1992, p. 141) and others, as an international entrepot of her times, as was Alexandria during the Hellenistic period, Amsterdam during the 17th century AD, and, among others, Hong Kong, Shanghai and Rotterdam nowadays. The Delian League functioned during 478–404 BCE. Except the geostrategic military aspect of the alliance, Athens, the leader of this alliance, actually also managed to integrate a large part of the East Mediterranean region into an ad hoc unified area of economic cooperation, where a network of parallel currencies was established, where the famous Athenian *drachma* was the dominant currency among other competitive coins, such as the Aeginetan *stater* (Figueira 1998; Bresson 2016b; Harris and Lewis 2016; Woolmer 2016; Bitros et al. 2020; Economou et al. 2021). This means that the Athenian drachma functioned somewhat similarly to the 'dollar of the classical times'.

Bitros et al. (2020); Bitros (2021); and Economou et al. (2021) argue that, essentially, the international success of the Athenian currency should be attributed, among other things, to the following three main reasons: (i) the perfect quality of its construction; (ii) keeping the *seigniorage* charge at a relatively low cost of 5% or 8%, according to the case; (iii) achieving economies of scale in the fabrication of these coins. O'Halloran (2018, p. 130) adds that the widespread usage of owls by Athens and its allies created an enormous conglomeration of interdependent markets that led to regional economic integration, with Athens being the paramount commercial metropolis among them. In his work *Panegyricus*, in verses 4.42, Isocrates proudly proclaimed that there was no good produced anywhere in the world that could not be found in the port of Athens (Isocrates 1980). Similar views were expressed by other ancient authors too, such as Thucydides (1993) in passage 2.38 of his *Histories*, Plato (1969) in passages 2.370e to 2.371a in the *Republic,* and Polybius (1972) in passage 4.38.9 of his *Histories*. The existence of such a unified economic area is described by further modern authors, such as Jarde [1926] (Jarde [1926] 1996, p. 255), who wrote the following:

> 'When a single city exercised hegemony over a whole group it was careful to obtain, by persuasion or force, the adoption of its own means. Thus the Attic system became that of the whole maritime confederacy. Money, in the same way, acquired an international value . . . . Each city had its mint which struck coins bearing the emblem of the city. But, as in the case of weights and measures, there was a movement towards unification . . . . Above all the importance assumed by the trade of certain cities and the good alloy of their currency caused certain

coins to be accepted in every market. The money of Aegina, Corinth, Phocaea, Cyzicus and Lampsacos had international value . . . ..'

Now, after having analyzed the crucial role of the *agoranomoi* and the *dokimastai* in making the Athenian economy the dominant economy in the East Mediterranean, in the following section, we provide a fictional example of how these two institutions functioned in practice.

4. A Fictional Scenario of International Commerce in Classical Greek Antiquity

At this point, in order to exhibit the importance of the *dokimastai,* regarding the promotion of international commerce, we relate this institution to that of the *agoranomoi* through a fictional paradigm between two merchants, Artemios of Piraeus, and Nikolaos, a merchant from Syracuse.

Nikolaos is an architect in Syracuse and has been assigned to construct a new luxurious house for a rich merchant in Syracuse, Hermogenes, whose intention is to build the new house and offer it as dowry for his beloved daughter Irene's forthcoming marriage. Hermogenes wishes to please his daughter and her prospective husband, and, for this purpose, he is determined to build the house with the best quality materials of the time. He also wishes to decorate the house beautifully. Thus, he asks for Nikolaos, an experienced architect, to handle this. Nikolaos responds that, among other things, high-quality marble is required and the most famous is the Athenian marble from Mt. Pentelikon, from which the Parthenon of the Acropolis was also built in the past. However, it is also expensive. Hermogenes agrees and asks for Nikolaos to buy the required quantity of marble.

Nikolaos then travels to Piraeus and finds Artemios of Piraeus, one of the most competent businessmen in Athens, famous for his capabilities of satisfying his clients. Nikolaos meets Artemios, and they agree on a transaction for a specific quantity of marble, and the price is to be paid in Syracuse *tetradrachms*. However, for an unknown reason, neither Artemios nor Nikolaos trust each other. Nikolaos has fears that the quality of the marble is not the proper quality. If it was not, he would return to Syracuse and he could not please his employer. On the other hand, Artemios does not trust the purity of the currency that Nikolaos brings with him on behalf of Hermogenes.

At this point, the financial risk between the two parties is quite high. Thus, they first apply to an *agoranomos* and then to a *dokimastes*. Being asked by Artemios and Nikolaos, one of the *agoranomoi*, with the aid of a connoisseur hired by the state, confirms the high quality of the marble. Being satisfied, Nikolaos then agrees to finish the transaction by paying Artemios. However, the latter denies payment unless the purity of the currency is confirmed by a *dokimastes*. This is performed by a *dokimastes.* The currency is found to be of the proper silver content, and then Artemios is also happy and he agrees to finish the transaction. The intervention of the two state institutions was the means by which the risk regarding the transaction of the two parties was almost neutralized.

Nikolaos and Artemios now become friends and trust each other. In fact, Artemios decides to offer a meal to Nikolaos at a famous *kapileion* (restaurant) in Athens, located in the Agora, just below the Acropolis, known as Athenais. There, they have a good time tasting nice seafood and amazing wines from the international markets of the Aegean, and delicious and juicy wines from Samos and Chios.[3] Nikolaos, being excited by the character and the integrity of Artemios, trusts him and assures him that he wishes to establish a long-term cooperation with him and his firm[4]. Furthermore, he will introduce him to all the architects and contractors of Syracuse, a privilege for him, since Syracuse is a very populous and prosperous polis, a mega-polis of its times, similarly to Athens. Thus, perhaps, new orders for building materials for the firm of Artemios may come from abroad in the next period. Artemios, on his side, is also very satisfied with Nikolaos' initiative to introduce him to the other architects in Syracuse.

5. Commerce and Financial Risk. Preferring Fairness than Cheating. A Game Theoretical Approach

The fictional example that was described in Section 4 is also related to the discussion of another important market failure, which is *asymmetry of information*, as defined by Akerlof (1970) in his seminal paper, *The Market for 'Lemons'*, as well as by other authors (Tirole 2017, pp. 117, 120–21). Before applying to both the *agoranomoi* and the *dokimastai*, neither Artemios nor Nikolaos had any information regarding the quality and purity of the currency, and the character of the market participants. This, of course, posed a high financial risk to both traders.

Thus, what we argue here is that the Athenian institutional setup was successful in the following two particular directions: through its auditing institutions, it gave symmetrical information on the quality and purity of the different currencies, and information on the character of the market participants, if they were honest or dishonest. Artemios does not know if the foreign currency is good and Nikolaos does not know if the construction materials he was about to buy were of the proper quality. This problem is solved as soon as both parties apply to both the *agoranomoi* and the *dokimastai*. After the evaluation of the latter, the information of the former becomes symmetrical and the financial risk they both face significantly reduces (or is even nullified in the ideal scenario).

This argumentation is further elaborated with the following game theoretical approach. In Table 1, we present four scenarios of 'cheat' and 'fair game' between Artemios and Nikolaos, followed by their payoffs in each case.

Table 1. Payoffs of Nikolaos under alternative strategies.

		Artemios	
		Strategy D (y)	**Strategy E (1-y)**
Nikolaos	**Strategy A (x_1)**	**35**	**30**
	Strategy B (x_2)	**30**	**40**
	Strategy C (x_3)	**40**	**25**

Nikolaos, who holds the money, could either use half of the money of good quality and half of the money of bad quality (strategy A), all the money of bad quality (strategy B), or all the money of good quality (strategy C). The numbers in Table 1 represent the payoff for Nikolaos. We suppose that the maximum utility that can be achieved by this game equals 100, as is used in these types of strategic games. Let us denote, with x_1, the probability that Nikolaos follows strategy A. Moreover, x_2 would be the probability that Nikolaos follows strategy B, while x_3 the probability that Nikolaos follows strategy C. It should be noted that $x_1 + x_2 + x_3 = 1$.

On the other hand, Artemios, who holds the marble, could either give Nikolaos marble of good quality (strategy D) or of bad quality (strategy E). Let us denote, with $0 < y < 1$, the probability that Artemios follows strategy D, whereas 1-y is the probability that he follows strategy E.

The expected payoffs for Artemios, given the strategies of Nikolaos, are as follows:

$$V(Art, A) = 35y + 30(1 - y) = 30 + 5y \quad (1)$$

$$V(Art, B) = 30y + 40(1 - y) = 40 - 10y \quad (2)$$

$$V(Art, C) = 40y + 25(1 - y) = 25 + 15y \quad (3)$$

If one wished to make a graphical representation of strategies A, B, and C of Nikolaos, it would be made by setting y = 0 and y = 1, and the derived lines could be used in order to discern the *minimax point* (the minimum of the maxima), which would be given by the crossing point of the lines that represent strategies B and C. Therefore, strategy A would no longer be needed for finding the solution of the game, as Nikolaos would never use this. So, the following applies:

$$V(\text{Art}, B) = V(\text{Art}, C) = 40 - 10y = 25 + 15y. \text{ Therefore } y = 0.6. \quad (4)$$

Thereby, the mixed strategy of Artemios is (0.6, 0, 0.4, 0).

By replacing y = 0.6 with either V(Art, B) or V(Art, C), it can be derived that Nikolaos will acquire a payoff that is equal to 34% of the total payoff from the transaction. It can easily be found that Artemios will obtain the remainder of the total payoff, which is equal to 66% of the total payoff from the transaction.

The optimal mixed strategy for Nikolaos (as a maximin point) is as follows:

$$V(\text{Nik}, D) = V(\text{Nik}, E) = 30x + 40(1 - x) = 40x + 25(1 - x) \text{ so } x = 0.6 \quad (5)$$

Consequently, the mixed strategy of Nikolaos is (0, 0, 0.6, 0.4). By replacing x = 0.6 with V(Nik, D) or V(Nik, E), it is, once again, calculated that V = 34.

Thereby, this mixed strategy game has no unique Nash equilibrium, but leads to the optimal solution where Nikolaos takes 34% of the maximum utility of the transaction, while Artemios takes 66%. These payoffs represent the weighted average of the potential payoffs, derived from the possible strategies undertaken by both players when the optimal responses to the other player's strategy are employed. Thereby, given that Nikolaos and Artemios have set their reaction factors, as indicated by their best interests given by the possible outcomes, these factors are replaced by their utility functions, and their weighted optimal payoffs (in the form of utility) are calculated. As the total utility of the game is supposed to equal 100, we exhibit that one player takes 1/3 of the utility and the second player takes two thirds of the utility. This solution is reached after the dominated strategies have been deleted, and maximization of the utility has taken place by graphical and numerical methods.

Based on the above, it can be argued that the inclusion of the scenario about mitigating risk due to the institutions of *agoranomoi* and *dokimastai* is beneficial for the society as a whole, as it serves to create higher overall wealth that can be achieved by mixed strategies. This is valid, as the honesty perspective for both players is reinforced due to the existence of auditing mechanisms. The existence of such mechanisms gives higher possibilities of both players being honest. Thereby, the optimal positive outcome for the economy as a whole is significantly more feasible. Such a situation obviously benefits society as a whole, since such institutions are supportive of what Tirole (2017) characterizes as the 'common good'.

What is also important is that it appears that through institutions such as the *agoranomoi* and the *dokimastai*, the Athenian economic system, during its historical heyday, had managed to create the conditions for policymaking to be carried out by citizens who, as a general trend, were 'educated' by the functioning of the institutions themselves to behave through a 'generalized morality', according to Tabellini's (Tabellini 2008) perspective, that is, the ethical attitudes and beliefs of people about how to interact with each other, beyond their family and kinship or social group. On this issue, James (2015) presents an empirical examination of the relationship between the public morality of people and economic growth, and he finds a positive relationship between the two. It, of course, also applies that if such values of 'generalized morality' are absent, this is detrimental for a society, since phenomena such as corruption, graft, and bribery may appear, which are further linked to economic inefficiencies. Additionally, as North (1990, p. 33) argues, from an institutionalist point of view, achieving a state of ethical behavior in a society also reduces transaction costs, thus laying the foundations for economic growth.

Two further aspects of the success of these two Athenian key market auditing institutions are directly related to the following two facts: The first has to do with the immediate punishment of a merchant or citizen who behaved illegally; the *agoranomoi* were public magistrates, who were authorized by the state authorities to impose fines on illegal financial/commercial transactions on the spot. The *dokimastai*, also being public magistrates, were authorized to immediately confiscate any fraudulent form of currency that they found through their inspections. We believe that these immediate procedures of punishment were an important factor in discouraging illegal commercial activities in the market, since the punishment was severe. Secondly, as mentioned in Section 3, according to the Athenian law, any forger who tried to distribute fraudulent coins in the market was punished by death, the most severe punishment of all. This was obviously a very discouraging factor to those who planned to make illegal profit through injecting fraudulent currency into the market.

In fact, according to our view, the above two elements, immediate punishment and increasing the degree of punishment, also come as practical and intertemporal proposals from Athens, in the classical times, to current policymakers worldwide, as a means of deterring those who try to undermine global markets through profiteering or selling fraudulent products, or illegally trying to inject large (or even small) amounts of fraudulent currency into global markets to make illegal profit.

Furthermore, there is no homogeneous system of law regarding punishing such illegal activities at the global level nowadays. The penalties that are imposed to legal frauds are quite different; for example, we indicatively mention that in the USA, the penalties for creating and injecting counterfeit money into the market include 20 years of imprisonment, with a fine with or in lieu of imprisonment, whereas in the UK, the penalty is 10 years, with a fine with or in lieu of imprisonment.[5] We believe that to achieve a more homogeneous system at the international level, further cooperation is needed among governments in these areas.

6. Conclusions

In this study, we have undertaken the strenuous task of investigating the interaction between two important financial institutions that proved to be crucial for the success of the internationally oriented Athenian commercial strategy, the *agoranomoi* and the *dokimastai*.

Based on the findings of prominent historians, referred to at appropriate points in Sections 2 and 3, and through a fictional scenario that we developed in Section 4, we simulated a typical commercial transaction in Athens during the classical period, which, in our case, included two trading parties, the fictional characters of Nikolaos and Artemios, and the presence of two state auditing mechanisms, the *agoranomoi* and the *dokimastai*. Finally, in Section 5, we also employed a mixed strategy game that examined the case of honesty, semi-honesty, and dishonesty during this kind of transaction.

Our study leads to some specific conclusions that we believe are of scientific value. We argue that the Athenian institutional setup, through its auditing institutions, proved to be successful in the following directions: Firstly, we know that the existence of robust auditing institutions, so as to ensure transparency while performing financial transactions, is widely accepted nowadays as a basic prerequisite for the establishment of a healthy international commercial environment. The modern economic theory and evidence reveals that the existence and the efficient functioning, in practice, of auditing institutions against profiteering, and for ensuring that only pure currency is circulated in the market, is an important prerequisite that can guarantee the success of an economic system in the long run. In this paper, we argue that such an axiom is not only a concept of modern times, but, on the contrary, has an intertemporal nature, since it has origins in Greek antiquity. Thus, such an axiom has an intertemporal historical orientation.

Secondly, a key element of the success of these two Athenian institutions is that they provided *symmetrical* information on the quality and purity of the different currencies, and on the character of the market participants, if they were honest or dishonest. This was important so that transactions between two or more parties could not cause financial losses for any party, due to a possible attempt by one member of the transaction to defraud the other. The supervisory role of the *agoranomoi* reduced the financial risks not only between two parties, but also regarding the Athenian financial system as a whole, in which a multitude of transactions were taking place.

Thirdly, through the game theoretical approach we performed, we argued that the institutions of the *agoranomoi* and the *dokimastai* enhanced the economy to reach a socially optimum level that better served the interests of the society as a whole. This happens because these institutions motivated both players to act honestly. Thereby, the mixed equilibria that we have calculated, based on the best-response parameters that were estimated according to the game structure, represent high utility for each individual, in comparison with the maximum utility that one player could achieve. Moreover, a higher potential payoff can be achieved for the overall economy in relation to a hypothetical game where the *agoranomoi* and *dokimastai* interventions would not be available options, and this would lead to a smaller space for mixed strategies.

The current practical utility from the Athenian auditing institutions, as described in this paper, that could come as an inspiration for policymakers nowadays lies in the three following directions: (i) immediate punishments to those who attempt to undermine global markets through profiteering, selling fraudulent products, or illegally trying to inject large (or even small) amounts of fraudulent currency into global markets; (ii) imposing much heavier punishments for those illegal activities, as described by (i); (iii) further international cooperation among governments, to achieve a more homogeneous system of surveillance against the illegal activities, as mentioned in (i).

There were, of course, some limitations regarding this particular research. The most important was the absence of cliometric data, but this applies not only for classical Athens, but also for every other ancient economy in general. In this case, we could only rely on the findings of history by using them cautiously, through interpretative tools from disciplines such as financial history and historical political economy, as well as game theory, as our paper does.

As a final comment, we hope that the methodology we follow in this paper, which combines financial history, issues of risk while performing financial transactions, and game theory as an interpretative tool for financial transactions, will further stimulate the interest of the academic community on related issues.

Author Contributions: All authors have contributed to all stages of this research. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors wish to thank the Editor and the referees for their very fruitful comments and suggestions that led to the improvement of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1. https://www.uscurrency.gov/media/news/federal-reserve-announces-day-issue-redesigned-100-note (accessed: 25 July 2021).
2. https://www.europol.europa.eu/newsroom/news/counterfeit-currencies-worth-millions-of-euros-prevented-entering-eu-economy-in-romania-and-spain (accessed: 25 July 2021).
3. For the confirmed existence of restaurants in Classical Athens see Davidson (1997).
4. Efficient cooperation mechanisms based on trust are necessary for the efficient function of modern firms and business associations (Fink and Kessler 2010). Among others, trust itself is a coordination mechanism that lowers transaction costs between firms (Adler 2001).
5. https://www.law.cornell.edu/uscode/text/18/471 and https://www.legislation.gov.uk/ukpga/1981/45/section/22 (accessed: 22 July 2021).

References

Acton, Peter. 2014. *Poiesis: Manufacturing in Classical Athens*. Oxford: Oxford University Press.

Adler, Paul S. 2001. Market, Hierarchy and Trust: The Knowledge Economy and the Future of Capitalism. *Organisation Science* 12: 215–34. [CrossRef]

Akerlof, George. 1970. The market for lemons: Quality, uncertainty and market mechanism. *Quarterly Journal of Economics* 84: 488–500. [CrossRef]

Amemiya, Takeshi. 2007. *Economy and Economics in Ancient Greece*. London and New York: Routledge.

Aristophanes. 1994. *Frogs*. English Translated by John M. Dillon. London and New York: Routledge.

Arrow, Kenneth. J. 1969. The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-market Allocations. In *Analysis and Evaluation of Public Expenditures: The PPP System*. Washington, DC: Joint Economic Committee of Congress.

Austin, Michael. M., and Pierre Emmanuel Vidal-Naquet. 1980. *Economic and Social History of Ancient Greece*. London: Batsford Academic and Educational Ltd.

Bergh, A., and Carl Hampus Lyttkens. 2011. *Between Burkina Faso and Brazil? Measuring Institutional Quality in Ancient Athens*. Lund, Sweden: Department of Economics, Lund University, Available online: http://swopec.hhs.se/lunewp/abs/lunewp2011_010.htm (accessed on 17 July 2021).

Bitros, George C. 2021. *From the Athenian Silver to the Bitcoin Standard: Private Money in a State-Enforced Free Banking Model*. Working Paper no. 14-2021. Athens: Department of Economics, Athens University of Economics and Business.

Bitros, George C., and Anastassios Karayannis. 2008. Values and Institutions as Determinants of Entrepreneurship in Ancient Athens. *Journal of Institutional Economics* 4: 205–30. [CrossRef]

Bitros, George C., Emmanouil M. L. Economou, and Nicholas C. Kyriazis. 2020. *Democracy and Money: Lessons for Today from Athens in Classical Times*. London and New York: Routledge.

Bresson, Alain. 2016a. *The Making of the Ancient Greek Economy: Institutions, Markets, and Growth in the City-States*. Princeton: Princeton University Press.

Bresson, A. 2016b. Aristotle and Foreign Trade. In *The Ancient Greek Economy: Markets, Households and City States*. Edited by Edward M. Harris, David M. Lewis and Mark Woolmer. New York: Cambridge University Press, pp. 41–65.

Carugati, Federica, Josiah Ober, and Barry R. Weingast. 2019. Is Development Uniquely Modern? Ancient Athens on the Doorstep. *Public Choice* 181: 29–47. [CrossRef]

Cohen, Edward E. 1992. *Athenian Economy and Society: A Banking Perspective*. Princeton: Princeton University Press.

Davidson, James. 1997. *Courtesans and Fishcakes*. The Consuming Passions of Classical Athens. London: Fontana Press.

de Ste Croix, Geoffrey Ernest Maurice. 1981. *The Class Struggle in the Ancient Greek World from the Archaic Age to the Arab Conquests*. Ithaca: Cornell University Press.

Demosthenes. 1939. *Against Timocrates*. English Translated by A. T. Murray. Cambridge: Harvard University Press, London: William Heinemann Ltd. Perseus Digital Library.

Economou, Emmanouil M. L. 2020. *The Achaean Federation in Ancient Greece: History, Political and Economic Organisation, Warfare and Strategy*. Cham: Springer.

Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2017. The Emergence and the Evolution of Property Rights in Ancient Greece. *Journal of Institutional Economics* 13: 53–77. [CrossRef]

Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2019a. *Democracy and Economy: An Inseparable Relationship Since Ancient Times to Today*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2019b. The evolution of property rights in Hellenistic Greece and the Ptolemaic Kingdom of Egypt. *Journal of Institutional Economics* 15: 827–43. [CrossRef]

Economou, Emmanouil M. L., Nicholas C. Kyriazis, and Nikolaos A. Kyriazis. 2021. Money Decentralization Under Direct Democracy Procedures. The Case of Classical Athens. *Journal of Risk and Financial Management* 14: 30. [CrossRef]

Elliott, Colin P. 2018. The Role of Money in the Economies of Ancient Greece and Rome. In *Handbook of the History of Money and Currency*. Edited by Stephano Battilossi, Youssef Cassis and Yago Kazuhiko. Singapore: Springer Nature Ltd., pp. 67–86.

Figueira, Thomas J. 1998. *The Power of Money: Coinage and Politics in the Athenian Empire*. Philadelphia: University of Pennsylvania Press.

Fink, K. Matthias, and Alexander Kessler. 2010. Cooperation, Trust and Performance–Empirical Results from Three Countries. *British Journal of Management* 21: 469–83. [CrossRef]

Finley, Moses I. 1973. *The Ancient Economy*. Berkeley: University of California Press.

Finley, Moses I. 1983. *Economy and Society in Ancient Greece*. New York: Penguin Non-Classics.

Fröhlich, Pierre. 2013. Governmental Checks and Balances. In *A Companion to Ancient Greek Government*. Edited by Hans Beck. Malden: Willey-Blackwell, pp. 252–66.

Gabrielsen, Vincent. 2014. Mankind's Most Secure and Durable Institution: State, Credit, Trade and Capital Accumulation in the Classical–Early Hellenistic Aegean. In *Infrastructure and Distribution in Ancient Economies*. Vienna: Austrian Academy of Sciences, pp. 25–44.

Halkos, George, and Kyriazis Nicholas. 2010. The Athenian Economy in the Age of Demosthenes. *European Journal of Law and Economics* 29: 255–77. [CrossRef]

Harris, Edward M., and David M. Lewis. 2016. Introduction: Markets in Classical and Hellenistic Greece. In *The Ancient Greek Economy: Markets, Households and City States*. Edited by Edward M. Harris, David M. Lewis and Mark Woolmer. New York: Cambridge University Press, pp. 1–37.

Hodgson, Geoffrey. 2015. *Conceptualizing Capitalism: Institutions, Evolution, Future*. Chicago and London: University of Chicago Press.

Homer, Sydney, and Richard Sylla. 2005. *A History of Interest Rates*. Hoboken: Wiley.

Isocrates. 1980. *Panegyricus*. English Translated by George Norlin. Cambridge: Harvard University Press, London: William Heinemann Ltd.

James, Harvey. S. 2015. Generalized Morality, Institutions and Economic Growth, and the Intermediating Role of Generalized Trust. *Kyklos* 68: 165–96. [CrossRef]

Jarde, A. 1996. *The Formation of the Greek People*. London and New York: Routledge. First published 1926.

Kyriazis, Nicholas, and Zouboulakis Michael. 2004. Democracy, Sea Power and Institutional Change: An Economic Analysis of the Athenian Naval Law. *European Journal of Law and Economics* 17: 117–32. [CrossRef]

Lanni, Adrianne. 2018. *Law and Order in Ancient Athens*, 2nd ed. Cambridge: Cambridge University Press.

Lyttkens, Carl Hampus H. 2013. *Economic Analysis of Institutional Change in Ancient Greece. Politics, Taxation and Rational Behaviour*. London and New York: Routledge.

Mackil, Emily. 2013. *Creating a Common Polity: Religion, Economy, and Politics in the Making of the Greek Koinon*. Berkeley and Los Angeles: University of California Press.

Manning, Joseph G. 2012. *The Last Pharaohs: Egypt Under the Ptolemies, 305–30 BC*. Princeton: Princeton University Press.

Martinez, L. F., and D. S. Jaeger. 2016. Ethical Decision Making in Counterfeit Purchase Situations: The Influence of Moral Awareness and Moral Emotions on Moral Judgement and Purchase Intentions. *Journal of Consumer Marketing* 33: 213–23. [CrossRef]

Mishkin, Frederick S., and Stanley G. Eakings. 2015. *Financial Markets and Institution*, 8th ed. Boston: Pearson.

North, Douglas Cecil. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Nosal, Ed, and Neil Wallace. 2007. A Model of (the Threat of) Counterfeiting. *Journal of Monetary Economics* 54: 994–1001. [CrossRef]

O'Halloran, Barry. 2018. *The Political Economy of Classical Athens: A Naval Perspective*. Leiden: Brill.

Ober, Josiah. 2008. *Democracy and Knowledge. Innovation and Learning in Classical Athens*. Princeton and Oxford: Princeton University Press.

Ober, Josiah. 2015. *The Rise and the Fall of Classical Greece*. Princeton and Oxford: Princeton University Press.

Plato. 1969. *Republic. Plato in Twelve Volumes*. Translated by Paul Shorey. Cambridge: Harvard University Press, London: William Heinemann Ltd., vol. 5–6, Available online: https://www.perseus.tufts.edu/hopper/text?doc=Plat.+Rep.+2.370&fromdoc=Perseus%3Atext%3A1999.01.0168 (accessed on 20 July 2021).

Polybius. 1972. *Histories*. Berkeley and Los Angeles: University of California Press.

Rahyab, Susan. 2019. The Rise and Development of the Office of Agoranomos in Greco-Roman Egypt. *New England Classical Journal* 46: 37–61.

Schefold, Bertram. 2011. The Applicability of Modern Economics to Forms of Capitalism in Antiquity. Some Theoretical Considerations and Textual Evidence. *Journal of Economic Asymmetries* 8: 131–63. [CrossRef]

Stroud, Ronald S. 1974. An Athenian Law on Silver Coinage. *Hesperia* XLIII: 157–88. [CrossRef]

Tabellini, G. 2008. The Scope of Cooperation: Values and Incentives. *Quarterly Journal of Economics* 123: 905–50. [CrossRef]

Temin, Peter. 2012. *The Roman Market Economy*. Princeton: Princeton University Press.

Thucydides. 1993. *History of the Peloponnesian War*. London: Penguin Classics.

Tirole, Jean. 2017. *Economics For the Common Good*. Princeton: Princeton University Press.

Tridimas, George. 2019. The Debate About the Nature of the Ancient Greek Economy: 19th Century German Scholarship and Modern Economic Theory. In *German-Greek Yearbook of Political Economy*. Edited by Manfred J. Holler and George Tridimas. München: Verlag Holler, vol. 2, pp. 4–25.

Verboven, Koenraad. 2021. *Complexity Economics. Building a New Approach to Ancient Economic History*. Cham: Palgrave MacMillan.

Viles, Nathan, Rush Alexandra, and Rohling Thomas. 2015. The Social Costs of Currency Counterfeiting. Research Discussion Paper. RDP 2015-05. Reserve Bank of Australia. Available online: https://www.rba.gov.au/publications/rdp/2015/pdf/rdp2015-05.pdf (accessed on 15 July 2021).

Woolmer, Mark. 2016. Forging Links Between Regions: Trade Policy in Classical Athens. In *The Ancient Greek Economy: Markets, Households and City States*. Edited by Edward M. Harris, David M. Lewis and Mark Woolmer. New York: Cambridge University Press, pp. 66–89.

Journal of
Risk and Financial Management

Article

Utility Indifference Option Pricing Model with a Non-Constant Risk-Aversion under Transaction Costs and Its Numerical Approximation

Pedro Pólvora and Daniel Ševčovič *

Department of Applied Mathematics and Statistics, Comenius University, 842 48 Bratislava, Slovakia; pedro.polvora@gmail.com
* Correspondence: sevcovic@fmph.uniba.sk

Abstract: Our goal is to analyze the system of Hamilton-Jacobi-Bellman equations arising in derivative securities pricing models. The European style of an option price is constructed as a difference of the certainty equivalents to the value functions solving the system of HJB equations. We introduce the transformation method for solving the penalized nonlinear partial differential equation. The transformed equation involves possibly non-constant the risk aversion function containing the negative ratio between the second and first derivatives of the utility function. Using comparison principles we derive useful bounds on the option price. We also propose a finite difference numerical discretization scheme with some computational examples.

Keywords: option pricing; utility indifference pricing; transaction costs; Hamilton-Jacobi-Bellman equation; penalty methods; finite difference approximation

Citation: Pólvora, Pedro, and Daniel Ševčovič. 2021. Utility Indifference Option Pricing Model with a Non-Constant Risk-Aversion under Transaction Costs and Its Numerical Approximation. *Journal of Risk and Financial Management* 14: 399. https://doi.org/10.3390/jrfm14090399

Academic Editor: George Halkos

Received: 1 July 2021
Accepted: 23 August 2021
Published: 25 August 2021

1. Introduction

In the last century the world witnessed a tremendous change and evolution in almost every industry and the financial one is no exception. One of the aspects that evolved greatly in finance was the financial derivatives, which saw their usage grow exponentially.

A financial derivative is a contract between two parties where they agree to future financial exchanges and whose value depends on one more underlying assets. There are multiple types of these contracts and they are used extensively in many industries, both for hedging and speculation. Depending on the type of derivative and on the position (buyer vs. seller) they can be used to either limit or increase the financial exposure to a particular financial asset. Examples of uses of financial derivatives include: financial institutions transforming a pool of equally risky mortgages into multiple contracts with different specific risk profiles, international enterprises reducing their foreign exchange risk, investors increasing their exposure to the increase in price of a stock by buying financial derivatives.

Financial options, a particular type of financial derivatives, are contracts where the buyer has the option but not the obligation to transact an asset at predefined conditions such as price or time. A key aspect of financial options is that their value, or price, is dependent on the underlying assets and finding it is fundamental for trading and managing the option and requires some type of mathematical modeling. Since the future payoff of the option is uncertain at the time of the trade it is required to price it with probabilistic and statistical considerations, and different approaches have been developed, examples include: Binomial Trees, Monte Carlo simulations, Black-Scholes/PDEs. This was a key piece in the birth of what is now called financial mathematics.

The usage of these type of contracts is very diverse, and they are important not only for the financial industry but for virtually every industry. They make it possible to manage risk in a way that gives great financial flexibility to enterprises, consequently promoting economic and business growth and development. For these reasons, it is not surprising that the number and type of financial options issued and traded has grown immensely over the last decades. This growth required more and more elaborate models to accommodate the new complexity of the contracts. Also, the greater size and impact of the usage of these contracts made evident the first models widely used were not good enough. Recent very unfortunate economic events such as the sub-prime crisis in the United States in 2008 were partially caused by the misuse of financial derivatives, and this demonstrated the importance of properly modeling and understand these contracts. The research proposed in this document aims to study new mathematical models that take into account some often neglected features of financial markets.

The most well known model for pricing financial options is called the Black-Scholes (BS) model and, although still largely used, it has multiple shortfalls like the fact that it does not account for feedback effects or transaction costs. The BS model does not consider that a trade can have an impact on the price, however, it has been empirically verified that a very large trader (such as an investment bank) can affect the assets' prices upon performing a large trade. Also, virtually every market has transaction costs or a different price for buying versus selling an asset (bid-ask spreads), which is not considered by the BS model as it assumes continuous cost-free trading to perfectly hedge the portfolio. Due to these shortfalls, in many situations, the Black-Scholes model is not sufficient for a robust application. Consequently a lot of research has been made on new models that extend Black-Scholes considering at least one of the previous characteristics. These extended models often result in non-linear pricing equations, which we introduce below.

The Leland (1985) model was one of the first and most popular extensions of the Black-Scholes to accommodate transaction costs. This model assumes only discrete trading at a pre-specified time intervals as opposed to the Black-Scholes model where trading is made continuously. Following Leland's approach, a model involving variable transaction costs has been introduced and analyzed in Ševčovič and Žitňanská (2016). For an overview of nonlinear option pricing models of the Leland type under transaction costs we refer to Ševčovič (2017).

A model that considers proportional transaction costs was introduced by Barles and Soner (1998). The authors apply a utility maximization approach, where they consider an economic agent with a constant risk-aversion. Using asymptotic analysis where the transaction costs were taken to zero and the risk-aversion to infinity they found that, in the limit, price of an European style option is given by a PDE of the Black-Scholes type where the volatility nonlinearly depends on the Gamma of the option price.

The concept of a utility function and even expected utility has been known and used for several decades in economics in general and its usage to price financial derivatives has gained a lot of momentum in recent years. The idea was first formulated by Hodges and Neuberger (1989), however, their work not fully formalized and mathematically proved which was then done by other authors such as Davis et al. (1993) that proposed a numerical scheme for solving the equation. Barles and Soner worked on that same model and provided an analytical study by take asymptotic limits on the transaction costs and risk-aversion coefficient (c.f. (Barles and Soner 1998)). This has become a very well-known model due to many practical reasons. Indifference pricing theory was presented in the book Carmona and Çinlar (2009) which covers, in a very deep and comprehensive matter, the subject of pricing via utility maximization.

In Davis et al. (1993) the authors investigated the problem of pricing European options in a Black-Scholes model with proportional costs on stock transactions and they defined the option writing price as the difference between the utilities achievable by going into the market to hedge the option and by going into the market on one's own account. Without transaction costs, this definition is shown to yield the usual Black-Scholes price. To compute the option price under transaction costs, one has to solve two stochastic control problems, corresponding to the two utilities compared above. The value functions of these problems are shown to be the unique viscosity solutions of one fully nonlinear quasi-variational inequality, with two different boundary and terminal conditions. They constructed a stable and convergent discretization scheme based on the binomial approximation of the stock price process. A generalization of this model was done by Cantarutti et al. (2017), where besides having proportional transaction costs, the underlying stock price dynamics was considered to have the form of a general exponential Lévy process. Numerical results are obtained by Markov chain approximation methods when the returns follow a Brownian motion and a variance gamma process.

In Monoyios (2004), an efficient algorithm is developed to price European options in the presence of proportional transaction costs, using the optimal portfolio framework of Davis et al. in Dempster and Pliska (1997). In this approach, the fair option price is determined by requiring that an infinitesimal diversion of funds into the purchase or sale of options has a neutral effect on achievable utility. This results in a general option pricing formula, in which option prices are computed from the solution of the investor's basic portfolio selection problem, without the need to solve a more complex optimisation problem involving the insertion of the option payoff into the terminal value function. The option prices are computed numerically using a Markov chain approximation to the continuous time singular stochastic optimal control problem, for the case of exponential utility. Comparisons with approximately replicating strategies are made. The method results in a uniquely specified option price for every initial holding of stock, and the price lies within bounds which are tight even as transaction costs become large. A general definition of an option hedging strategy for a utility maximising investor is developed that involved calculating the perturbation to the optimal portfolio strategy when an option trade is executed.

In Kallsen and Muhle-Karbe (2015), asymptotic formulas for utility indifference prices and hedging strategies in the presence of small transaction costs were obtained. Perrakis and Lefoll (2000) derived optimal perfect hedging portfolios in the presence of transaction costs. In the paper Dong and Lu (2021) the price of a European option with proportional transaction costs has been determined using a utility indifference approach where the resulting Hamilton-Jacobi-Bellman equation for the portfolio without option is two-dimensional instead of three-dimensional as in standard utility indifference approaches (c.f. (Davis et al. 1993)).

Furthermore, Li and Wang (2009) study the application of the penalty method to solve the resulting variational inequality. Song Wang and Wen Li have published numerical results of an implementation of the penalty method to price for both American and European style options (c.f. (Li and Wang 2014)). They considered exponential utility which is by far one of the most studied utilities but still slightly restrictive.

Our goal is to analyze the system of two Hamilton-Jacobi-Bellman (HJB) equations.The option price is constructed as a difference of the certainty equivalents to the value functions solving the system of HJB equations. We introduce a transformation method for solving the penalized nonlinear partial differential equation. The transformed equation involves possibly non-constant and non-zero risk aversion function containing the negative ratio between the second and first derivatives of the utility function. Using the parabolic comparison principles we derive useful bounds on the option price. We also propose a finite difference numerical discretization scheme with some computational examples.

The paper is organized as follows. The next section is focused on generalization of the utility indifference option pricing model. We consider a general class of concave utility functions. A system of two Hamilton-Jacobi-Bellman equations is derived. The option price is then obtained in terms of a difference of their solutions. In Section 3 we present a transformation method for solving the penalized nonlinear partial differential equation. The penalized equation involves the risk aversion function. Section 4 is devoted to construction of a numerical scheme which is based on time implicit backward Euler method in combination with an upwind finite difference method for spatial discretizations. The Hamilton-Jacobi-Bellman equations are solved by means of the penalty method utilizing the policy iteration method. It contains numerical examples of option prices for various concave utility functions.

2. Utility Indifference Option Pricing Model

In economics, a utility function is a function measuring the economic agent's preferences on different goods. In a financial context the utility function is usually applied to monetary quantities, and it can be used to measure the agent's risk aversion when in a context of uncertainty.

The usual requirements for an utility function are that it is continuous and non-decreasing function. Additional properties such as concavity or convexity can be shown to be directly related with the investor's risk aversion (see below).

2.1. Risk Aversion and the Concept of a Certainty Equivalent

If an investor's wealth at a future time T is affected by a source of uncertainty then it can be modeled by a random variable, say W_T which we assume to have finite expectation $\mathbb{E}[W_T] =: w$. Then, we know from Jensen's inequality that if $U : \mathbb{R} \to \mathbb{R}$ is a concave function then

$$\mathbb{E}[U(W_T)] - U(w) \leq 0.$$

This difference can be seen as how much an investor prefers to hold an uncertain amount (which can turn out to be greater or lower that its average) or its *average*. The greater the concavity of U the greater that difference, and that leads one to define the Arrow-Pratt measure of absolute risk-aversion (also referred to as the coefficient of risk-aversion),

$$R(\xi) \equiv -U''(\xi)/U'(\xi).$$

For a concave increasing utility function U we have $R \geq 0$. Now, for pricing financial options one still needs one more concept, the concept of certainty equivalent which we denoted by v. It is defined as follows:

$$v : U(v) = \mathbb{E}[U(W_T)], \quad v = U^{-1}(\mathbb{E}[U(W_T)]).$$

Throughout the paper we shall consider various types of utility functions with different risk-aversion profiles. Their profiles $U(x)$ and the risk aversion coefficients $R(\xi) = -U''(\xi)/U'(\xi)$ are shown in Table 1.

Table 1. Utility functions, their inverse functions and risk aversion functions.

Type	Utility Function	Parameter	Inverse Utility Function	Risk Aversion
Linear	$U(\xi) = \xi$	—	$U^{-1}(y) = y$	$R(\xi) = 0$
Exp.	$U(\xi) = 1 - e^{-\gamma\xi}$	$\gamma > 0$	$U^{-1}(y) = -\ln(1-y)/\gamma$	$R(\xi) = \gamma$
Power	$U(\xi) = \xi^a$	$a < 1$	$U^{-1}(y) = y^{1/a}$	$R(\xi) = (1-a)\xi^{-1}$
Log.	$U(\xi) = \ln(b\xi + 1)$	$b > 0$	$U^{-1}(y) = (e^y - 1)/b$	$R(\xi) = b/(b\xi + 1)$

The power utility function $U(\xi) = \xi^a$ and the logarithmic utility function $U(\xi) = \ln(b\xi + 1)$ belong to the wide class of the so-called decreasing absolute risk aversion (DARA) utility function characterized by a decreasing risk aversion coefficient $R(\xi)$ considered as a function of the wealth ξ. In the context of dynamic stochastic portfolio optimization the importance of DARA utility functions has been investigated in papers by Post et al. (2015); Kilianová and Ševčovič (2018).

2.2. Utility Indifference Option Pricing Model under Transaction Costs

To price a derivative in a market with proportional transaction costs using this framework we proceed as follows.

Firstly, we consider that we have an investor that can invest in either a risky asset or a risk-free asset following the dynamics of a geometric Brownian motion and exponential deterministic process, respectively. That is

$$dS = \mu S \mathrm{d}t + \sigma S \mathrm{d}w_t, \qquad dB = rBdt, \tag{1}$$

where $\{w_t, t \geq 0\}$, stands for the standard Wiener stochastic process.

We model the transactions costs by introducing two different prices for the risky asset, depending on weather the investor is buying (ask price) or selling (bid price) the underlying asset,

$$S_{ask} = (1+\theta)S,\ S_{bid} = (1-\theta)S,$$

where $\theta = (S_{ask} - S_{bid})/(2S)$ represents the bid-ask spread factor.

We assume that the investor can buy or sell shares of the risky asset (shares account α_t) by increasing or decreasing his holdings in the risk-free asset (money account β_t). We model the cumulative purchase of risky assets by a process L_t and the sale by the process M_t. The goal of an investor is to maximize his terminal utility and will chose his trades L_t, M_t accordingly. The resulting dynamics for the investors portfolio are,

$$d\alpha = dL - dM, \qquad d\beta = r\beta dt - (1+\theta)SdL + (1-\theta)SdM. \tag{2}$$

Now, we can define the liquid wealth W_t of the investor as follows:

$$W_t = \beta_t + S_t(\alpha_t - \theta|\alpha_t|).$$

Then the option pricing problem can be formulated as a stochastic control problem. Firstly, we introduce a portfolio in which the investor is optimizing their expected utility by trading either stock or bonds,

$$v^0(\alpha, \beta, s, t) = \sup_{L,M} \mathbb{E}[U(W_T)|\alpha, \beta, s]. \tag{3}$$

Secondly, we consider that the investor has at his disposal another portfolio, which is equally comprised of risk-free and risky assets but also a short buyer position with $\delta = -1$ or long seller position with $\delta = +1$ on a derivative with the terminal payoff C_T. The value v^δ of this second portfolio is given by,

$$v^\delta(\alpha, \beta, s, t) = \sup_{L,M} \mathbb{E}[U(W_T + \delta C_T(s))|\alpha, \beta, s]. \tag{4}$$

Let us denote the certainty equivalents of portfolios by $z^\delta = z^\delta(\alpha, \beta, s, t)$ and $z^0 = z^0(\alpha, \beta, s, t)$, respectively. The functions z^δ, z^0 satisfy the system of equations:

$$U(w - z^0) = v^0, \qquad U(w - z^\delta) = v^\delta, \quad w = \beta + s(\alpha - \theta|\alpha|). \tag{5}$$

For further details of utility indifference option pricing we refer to the book Carmona and Çinlar (2009), the price V of an option with a payoff diagram C_T is given as the discounted difference of certainty equivalents, i.e.,

$$V = e^{-r(T-t)}(z^\delta - z^0), \quad \text{where } z^\delta - z^0 = U^{-1}(v^0) - U^{-1}(v^\delta). \tag{6}$$

In order to determine solutions z^δ and z^0 we have to solve a pair of stochastic optimal control problems for v^δ, and v^0, by means of the dynamic programming principle. Mathematical representation leads to a system of two Hamilton-Jacobi-Bellman (HJB) equations that we introduce and analyze in the next subsection.

2.3. Hamilton-Jacobi-Bellman Equations for Value Functions

Following Davis et al. (1993) the functions v^δ and v^0 satisfy the system of variational inequalities of the form:

$$\min(\mathcal{V}_A(v), \mathcal{V}_B(v), \mathcal{V}_C(v)) = 0, \tag{7}$$

where the linear differential operators $\mathcal{V}_A, \mathcal{V}_B, \mathcal{V}_C$ are defined as follows:

$$\mathcal{V}_A \equiv \partial_t + \frac{\sigma^2}{2}s^2\partial_s^2 + \mu s\partial_s + r\beta\partial_\beta, \;\; \mathcal{V}_B \equiv -\partial_\alpha + s(1+\theta)\partial_\beta, \;\; \mathcal{V}_C \equiv \partial_\alpha - s(1-\theta)\partial_\beta,$$

and terminal conditions,

$$v^0(\alpha,\beta,s,T) = U(w(\alpha,\beta,s)), \qquad v^\delta(\alpha,\beta,s,T) = U(w(\alpha,\beta,s) + \delta C_T(s)). \tag{8}$$

Here $C_T(S)$ denotes the prescribed pay-off diagram, i.e., $C_T(S) = (S-K)^+$ in the case of a plain vanilla call option, or $C_T(S) = (K-S)^+$ in the case of a put option. Here $K > 0$ denotes the strike price, $(\beta)^+ = \max(\beta, 0)$, and $(\beta)^- = \min(\beta, 0)$ denote the positive and negative parts of a real number β. Recall that the no-transaction region of values (α, β, s, t) is characterized by the equation $\mathcal{V}_A(v) = 0$. The Buy and Sell regions correspond to equations $\mathcal{V}_B(v) = 0$ and $\mathcal{V}_C(v) = 0$, respectively (c.f. (Davis et al. 1993)).

The minimal Equation (7) is equivalent to the following linear complementarity problem for the functions $v = v^0$, and $v = v^\delta$:

$$\mathcal{V}_A(v) \geq 0, \; \mathcal{V}_B(v) \geq 0, \; \mathcal{V}_C(v) \geq 0, \qquad \mathcal{V}_A(v)\cdot\mathcal{V}_B(v)\cdot\mathcal{V}_C(v) = 0. \tag{9}$$

2.4. Penalty Method for Solving HJB Equations

With penalty methods, the initial variational inequality is replaced by one single equation which has a term parameterized by a small parameter. One should prove that the solution of this new equation will converge to the initial one. Besides convergence to the initial problem, the perturbed equation's solution will always respect the constrains posed by the initial problem. Next, introduce the specific implementation of the penalty method for our pricing model following Li and Wang (2009, 2014). The penalty method has been successfully adopted for solving various nonlinear option pricing model by Lesmana and Wang (2015), or Chernogorova et al. (2018), and others. The optimal time dependent penalty function has been proposed recently by Clevenhaus et al. (2020).

Next we introduce the penalty method in a more detail. Let us define the following penalized perturbed equation for the function $v = v_{\lambda_B,\lambda_C}(\alpha,\beta,s,t)$,

$$\mathcal{V}_A(v) + \lambda_B[\mathcal{V}_B(v)]^- + \lambda_C[\mathcal{V}_C(v)]^- = 0. \tag{10}$$

Here $\lambda_B, \lambda_C \gg 0$ are sufficiently large penalty parameters. In what follows, we will drop the subscripts for the sake of simplicity $v := v_{\lambda_B,\lambda_C}$. In the limit $\lambda_B, \lambda_C \to \infty$ we formally deduce that the limiting solution v solves the linear complementarity problem. Indeed, $\mathcal{V}_A(v) = -\lambda_B[\mathcal{V}_B(v)]^- - \lambda_C[\mathcal{V}_C(v)]^- \geq 0$, and $\mathcal{V}_B(v), \mathcal{V}_C(v) \geq 0$ in the limit $\lambda_B, \lambda_C \to \infty$. Taking $\lambda_B, \lambda_C \to \infty$ such that $\lambda_B/\lambda_C \to \infty$ we obtain $\mathcal{V}_A(v)\mathcal{V}_B(v) = 0$. Similarly, $\mathcal{V}_A(v)\mathcal{V}_C(v) = 0$.

3. Transformation of the HJB Equation Involving Risk Aversion Function

For a general utility function U we search the solution $v = v^\delta, \delta = 0, \pm 1$, in the following form:

$$v(\alpha,\beta,s,t) = U(e^{\mu(T-t)}(w(\alpha,\beta,s) + \mathscr{A}(t) + \mathscr{V}(\alpha,\beta,s,t))),$$

where

$$\mathscr{A}(t) = \frac{\beta}{\mu}(r-\mu)(1 - e^{-\mu(T-t)})$$

is a time dependent shift function such that $\mathscr{A}(T) = 0$. Notice that $\partial_s^2 w = 0, \partial_\beta w = 1$, and $s\partial_s w = w - \beta$. Hence

$$\begin{aligned} \mathcal{V}_A(e^{\mu(T-t)}(w + \mathscr{A})) &= e^{\mu(T-t)}(-\mu(w + \mathscr{A}) + \mathscr{A}' + \mu(w - \beta) + r\beta) \\ &= e^{\mu(T-t)}(\mathscr{A}' - \mu\mathscr{A} + \beta(r - \mu)) = 0, \end{aligned}$$

because of the definition of the auxiliary function $\mathscr{A}(t)$. For any function $z = z(\alpha, \beta, s, t)$ we have

$$\begin{aligned} \mathcal{V}_A(U(z)) &= U'(z)(\partial_t z + \mu s\partial_s z + r\beta\partial_\beta z + \frac{\sigma^2}{2}s^2\partial_s(U'(z)\partial_s z)) \\ &= U'(z)\Big(\mathcal{V}_A(z) - \frac{\sigma^2}{2}R(z)(s\partial_s z))^2\Big), \end{aligned}$$

where $R(z) = -U''(z)/U(z) \geq 0$ is the risk aversion function. Taking $z = e^{\mu(T-t)}(w + \mathscr{A} + \mathscr{V})$ we obtain

$$\mathcal{V}_A(v) = \mathscr{F}(\alpha, \beta, s, t)\Big(\mathcal{V}_A(\mathscr{V}) - \mu\mathscr{V} - \frac{\sigma^2}{2}R(z)e^{\mu(T-t)}(s\partial_s w + s\partial_s\mathscr{V})^2\Big),$$

where $\mathscr{F}(\alpha, \beta, s, t) = U'(z)e^{\mu(T-t)} > 0$ is a positive factor. For $\mathcal{V}_B(w)$, and $\mathcal{V}_C(w)$ we have

$$\mathcal{V}_B(w) = \begin{cases} 2s\theta, & \text{if } \alpha > 0, \\ 0, & \text{if } \alpha \leq 0, \end{cases} \qquad \mathcal{V}_C(w) = \begin{cases} 0, & \text{if } \alpha > 0, \\ 2s\theta, & \text{if } \alpha \leq 0. \end{cases}$$

Furthermore, as $\mathcal{V}_B(\mathscr{A}) = \mathcal{V}_C(\mathscr{A}) = 0$, we have

$$\mathcal{V}_B(v) = \mathscr{F}(\alpha, \beta, s, t)(\mathcal{V}_B(w) + \mathcal{V}_B(\mathscr{V})), \quad \mathcal{V}_C(v) = \mathscr{F}(\alpha, \beta, s, t)(\mathcal{V}_C(w) + \mathcal{V}_C(\mathscr{V})).$$

A solution $\mathscr{V} = \mathscr{V}^\delta$ is subject to the terminal condition at $t = T$:

$$\mathscr{V}(\alpha, \beta, s, T) = \begin{cases} 0, & \text{if } \delta = 0, \\ C_T(s), & \text{if } \delta = 1, \text{ (long seller position)}, \\ -C_T(s), & \text{if } \delta = -1, \text{ (short buyer position)}. \end{cases} \tag{11}$$

Summarizing, we deduce that the penalized problem (10) can be reformulated in terms of the function $\mathscr{V}$ as follows:

$$\begin{aligned} \mathcal{V}_A(\mathscr{V}) - \mu\mathscr{V} &- \frac{\sigma^2}{2}R(z)e^{\mu(T-t)}(s\partial_s w + s\partial_s\mathscr{V})^2 \\ &+ \lambda_B[\mathcal{V}_B(w + \mathscr{V})]^- + \lambda_C[\mathcal{V}_C(w + \mathscr{V})]^- = 0. \end{aligned} \tag{12}$$

Recall that the option price V is obtained as the difference between certainity equivalents z^0 and z^δ. It means that

$$V = e^{-r(T-t)}(z^\delta - z^0) = e^{-r(T-t)}(U^{-1}(v^0) - U^{-1}(v^\delta)) = e^{(\mu-r)(T-t)}(\mathscr{V}^0 - \mathscr{V}^\delta).$$

In the next proposition we compare a solution to the system of transformed HJB equations with the explicit solution to the linear Black-Scholes equation:

$$\partial_t\mathscr{V} + \frac{\sigma^2}{2}s^2\partial_s^2\mathscr{V} + \mu s\partial_s\mathscr{V} - \mu\mathscr{V} = 0, \qquad \mathscr{V}(s, T) = \delta C_T(s), \quad \delta = 0, \pm 1.$$

In the call option case where $C_T(s) = (s - K)^+$ the price $\mathscr{V}(s, t) = \mathscr{V}_{BS}(s, t)$ is given by an explicit formula:

$$\mathscr{V}_{BS}(s, t) = \delta\Big(s\Phi(d_1) - Ke^{-\mu(T-t)}\Phi(d_2)\Big),$$

where $d_{1,2} = (\ln(s/K) + (\mu \pm \sigma^2/2)(T-t))/(\sigma\sqrt{T-t})$ and $\Phi(d) = (2\pi)^{-1/2} \int_{-\infty}^{d} e^{-\xi^2/2} d\xi$. A similar formula is available for pricing of put options.

Proposition 1. *Assume U is an exponential ($\gamma > 0$) or linear ($\gamma = 0$) utility function, i.e., its risk aversion function $R(\xi) = -U''(\xi)/U'(\xi) \equiv \gamma$ where $\gamma \geq 0$ is a non-negative constant. Then the solution $\mathscr{V}$ of the penalized problem (12) satisfying the terminal condition (11) is independent of the factor β, i.e., $\mathscr{V} = \mathscr{V}(\alpha, s, t)$. Consequently, the option price $V = V(\alpha, s, t)$ is independent of β. Moreover,*

$$\mathscr{V}(\alpha, s, t) \leq \mathscr{V}_{BS}(s, t), \quad \text{for all } \alpha, s > 0, t \in [0, T].$$

Proof. Notice that $R(\xi) \equiv \gamma$, and $s\partial_s w, \mathcal{V}_B(w), \mathcal{V}_C(w)$, as well as the terminal condition (11) are independent functions of the factor β. The penalized Equation (12) can be rewritten as follows:

$$\begin{aligned} &\partial_t \mathscr{V} + \frac{\sigma^2}{2} s^2 \partial_s^2 \mathscr{V} + \mu s \partial_s \mathscr{V} + \mu \beta \partial_\beta \mathscr{V} - \mu \mathscr{V} \\ &= \frac{\sigma^2}{2} \gamma e^{\mu(T-t)} (s\partial_s w + s\partial_s \mathscr{V})^2 - \lambda_B [\mathcal{V}_B(w) + \mathcal{V}_B(\mathscr{V})]^- - \lambda_C [\mathcal{V}_C(w) + \mathcal{V}_C(\mathscr{V})]^-. \end{aligned} \tag{13}$$

The right-hand side of (13) is nonnegative and it does not explicitly depend on β, so does the solution $\mathscr{V} = \mathscr{V}(\alpha, s, t)$. Furthermore,

$$\partial_t \mathscr{V} + \frac{\sigma^2}{2} s^2 \partial_s^2 \mathscr{V} + \mu s \partial_s \mathscr{V} - \mu \mathscr{V} \geq 0.$$

As the Black-Scholes solution $\mathscr{V}_{BS}$ satisfies

$$\partial_t \mathscr{V}_{BS} + \frac{\sigma^2}{2} s^2 \partial_s^2 \mathscr{V}_{BS} + \mu s \partial_s \mathscr{V}_{BS} - \mu \mathscr{V}_{BS} = 0,$$

then, taking into account $\mathscr{V}(\alpha, s, T) = \mathscr{V}_{BS}(s, T)$, applying the maximum principle for parabolic equations on unbounded domains due to ((Meyer and Needham 2014), Theorem 3.4), we obtain the inequality $\mathscr{V}(\alpha, s, t) \leq \mathscr{V}_{BS}(s, t)$ for a given parameter α and all $s > 0, t \in [0, T]$, as claimed. □

The following proposition is a direct consequence of Proposition 1.

Proposition 2. *Assume U is the linear utility function, i.e., its risk aversion function $R(\xi) \equiv 0$. Suppose that there are no transaction costs, i.e., $\theta = 0$. Then the solution $\mathscr{V}$ of the penalized problem (12) satisfying the terminal condition (11) is independent of the factors α, β, i.e., $\mathscr{V} = \mathscr{V}(s, t) = \mathscr{V}_{BS}(s, t)$, i.e., $\mathscr{V}$ is the Black-Scholes price of a European style option.*

Proof. Since $\theta = 0$ we have $\mathcal{V}_B(v) = -\mathcal{V}_C(v)$ for any function v. Hence v is a solution to (9) if and only if $\mathcal{V}_B(v) = \mathcal{V}_C(v) = 0$ and $\mathcal{V}_A(v) \geq 0$. Therefore, a solution $\mathscr{V}$ to the penalized problem (12) satisfies the linear Black-Scholes equation. Hence, $\mathscr{V} = \mathscr{V}(s, t) = \mathscr{V}_{BS}(s, t)$, as claimed. □

4. Construction of a Numerical Discretization Upwind Finite Difference Scheme

In this section we propose a numerical discretisation scheme and several computational examples. The scheme is based on the finite difference method proposed in Li and Wang (2009). The resulting scheme is of upwind type in the space discretization and the backward Euler implicit scheme in time.

4.1. Finite Difference Approximation of a Solution to the Penalized Problem

We first introduce Ω^b the truncated domain corresponding to the solvency region where $\beta + S(\alpha - \theta|\alpha|) > 0$ as follows:

$$\Omega^b = \{(\alpha, \beta, S) \in (L_\alpha^-, L_\alpha^+) \times (L_\beta^-, L_\beta^+) \times (0, S^+) : \beta + S(\alpha - \theta|\alpha|) > 0\}.$$

We consider a simple 3D uniform mesh grid:

$$(\alpha_i, \beta_j, S_k) \in \Omega^b, \quad i = 0, \cdots, N_\alpha, \quad j = 0, \cdots, N_\beta, \quad k = 0, \cdots, N_S,$$

$$\alpha_i = L_\alpha^- + ih_\alpha, \quad \beta_j = L_\beta^- + jh_\beta, \quad S_k = kh_S,$$

$$h_\alpha = (L_\alpha^+ - L_\alpha^-)/N_\alpha, \quad h_\beta = (L_\beta^+ - L_\beta^-)/N_\beta, \quad h_S = S^+/N_S,$$

where N_α, N_β, and N_S are the numbers of discretization steps in the α, β, and S variables. Notice that the spatial discretization can be easily adopted to a non-uniform grid, e.g., by considering uniform discretization for the logarithmic variable $x_k = \ln(S_k/K)$. We consider a uniform time discretization with time steps $n\Delta t$ for $n = N, \cdots, 1, 0$, where $\Delta t = T/N$, and N is the number of time discretization steps. The solution $v = v(\alpha, \beta, S, t)$ will be discretized by the value V_{ijk}^n at (α_i, β_j, S_k) and time $t = n\Delta t$.

We define the following finite difference discretization operators:

$$\begin{aligned} D_t V_{ijk}^n &= \frac{V_{ijk}^{n+1} - V_{ijk}^n}{\Delta t}, \quad D_\alpha^\pm V_{ijk}^n = \pm\frac{V_{(i\pm1)jk}^n - V_{ijk}^n}{h_\alpha}, \quad D_\beta^\pm V_{ijk}^n = \pm\frac{V_{i(j\pm1)k}^n - V_{ijk}^n}{h_\beta}. \\ D_S V_{ijk}^n &= \frac{V_{ij(k+1)}^n - V_{ijk}^n}{h_S}, \; D_{SS} V_{ijk}^n = \frac{V_{ij(k+1)}^n - 2V_{ijk}^n + V_{ij(k-1)}^n}{h_S^2}, \end{aligned} \tag{14}$$

$$\begin{aligned} \mathcal{L}_\mathcal{A} V_{ijk}^n &= -\left(D_t + r(\beta_j)^+ D_\beta^+ + r(\beta_j)^- D_\beta^- + \mu S_k D_S + \frac{\sigma^2}{2} S_k^2 D_{SS}\right) V_{ijk}^n, \\ \mathcal{L}_\mathcal{B} V_{ijk}^n &= \left(-D_\alpha^+ + (1+\theta) S_k D_\beta^-\right) V_{ijk}^n, \quad \mathcal{L}_\mathcal{C} V_{ijk}^n = \left(D_\alpha^- - (1-\theta) S_k D_\beta^+\right) V_{ijk}^n. \end{aligned} \tag{15}$$

Clearly, for any $\lambda > 0$, and $\mathcal{L} \in \mathbb{R}$, we have

$$\lambda[\mathcal{L}]^- = \min_{m\in[0,\lambda]} m\mathcal{L} = \begin{cases} 0, & \text{if } \mathcal{L} > 0, \\ \lambda\mathcal{L}, & \text{if } \mathcal{L} \le 0. \end{cases}$$

The numerical discretization scheme is then given by:

$$\mathcal{L}_\mathcal{A} V_{ijk}^n + \min_{\bar{m}\in[0,\lambda_B]} \bar{m}\mathcal{L}_\mathcal{B} V_{ijk}^n + \min_{\bar{n}\in[0,\lambda_C]} \bar{n}\mathcal{L}_\mathcal{C} V_{ijk}^n = 0, \quad \forall i, j, k. \tag{16}$$

Terminal conditions.
For the last terminal time level $n = N$ we have, for the call (put) option case with the pay-off diagram $C_T(S) = (S-K)^+$ $(C_T(S) = (K-S)^+)$,

$$V_{ijk}^N = U(\beta_j + S_k(\alpha_i - \theta|\alpha_i|) + \delta C_T(S_k)), \text{ for } (\alpha_i, \beta_j, S_k) \in \Omega^b.$$

Boundary conditions. We apply the Dirichlet boundary conditions, i.e.,

$$V_{ijk}^n = U(\beta_j + S_k(\alpha_i - \theta|\alpha_i|) + \delta C_T(S_k)), \text{ for } (\alpha_i, \beta_j, S_k) \in \partial\Omega^b, \; n = N-1, \cdots, 1, 0.$$

Here we set $\delta = 0$ in the case of numerical approximation of the value function v^0, and $\delta = \pm 1$ in the case of approximation of the value function $v^\delta, \delta = \pm 1$.

Next, we present the full numerical discretization algorithm involving the policy iteration method for solving the penalized PDE (16). Notice that $(\beta_j)^+ + (\beta_j)^- = \beta_j$. It yields the following system of linear equations for the unknown vector V^n for $n = N-1, \cdots, 1, 0$,

$$
\begin{aligned}
&\left[1 + \Delta t\left(\frac{r}{h_\beta}\beta_j + \frac{\mu}{h_S}S_k + \frac{\sigma^2}{h_S^2}S_k^2\right)\right] V_{ijk}^{n,p+1} \\
&-\left[\frac{\mu\Delta t}{h_S}S_k + \frac{\sigma^2\Delta t}{2h_S^2}S_k^2\right] V_{ij(k+1)}^{n,p+1} - \frac{\sigma^2\Delta t}{2h_S^2}S_k^2 V_{ij(k-1)}^{n,p+1} \\
&+\Delta t\left[\tilde{m}\left(\frac{1}{h_\alpha} + \frac{(1+\theta)}{h_\beta}S_k\right) + \tilde{n}\left(\frac{1}{h_\alpha} + \frac{(1-\theta)}{h_\beta}S_k\right)\right] V_{ijk}^{n,p} \\
&-\Delta t\left[\frac{r\beta_j^+}{h_\beta} + \tilde{n}\frac{(1-\theta)}{h_\beta}S_k\right] V_{i(j+1)k}^{n,p} - \Delta t\left[\frac{r\beta_j^-}{h_\beta} + \tilde{m}\frac{(1+\theta)}{h_\beta}S_k\right] V_{i(j-1)k}^{n,p} \\
&-\tilde{m}\frac{\Delta t}{h_\alpha} V_{(i+1)jk}^{n,p} - \tilde{n}\frac{\Delta t}{h_\alpha} V_{(i-1)jk}^{n,p} = V_{ijk}^{n+1},
\end{aligned}
\tag{17}
$$

where $p = 0, \cdots, p_{max}$ is the policy iteration parameter, $\tilde{m} = \tilde{m}_{ijk}^{n,p}, \tilde{n} = \tilde{n}_{ijk}^{n,p}$ are arguments of the minimum in (16). The above system of linear equations for the unknown stacked vector $V = (V_{ijk}^{n,p+1})$ can be rewritten as a system of linear equations of the form $\mathcal{A}V = b$ where $\mathcal{A}$ is a sparse matrix with at most 3 nonzero elements in each row. The right-hand side vector b consists of the known vector V^{n+1} complemented by the boundary conditions. It is important to note that the coefficients of the matrix $\mathcal{A}$ depend on the coefficients $\tilde{m}, \tilde{n}$ which has to be computed within each policy iteration step. The full algorithm for the computation of the value function is as in the Algorithm 1.

Algorithm 1: The algorithm for computing the value function V for $\delta = 0, \pm 1$.

```
Initialization of model parameters and numerical parameters p_max, tol_max > 0;
Compute the terminal conditions for n = N.
Set V^N_{i,j,k} = U(w_{i,j,k} + δC_T(S_k)) for each i, j, k;
Set n = N − 1;
while n > 0 do
    Initiate policy iteration p = 0 with V^{n,0}_{ijk} = V^{n+1}_{ijk};
    Compute the right-hand side vector b from V^{n+1}_{ijk} and boundary conditions ;
    while p < p_max and tol ≥ tol_max do
        Compute the penalty terms. Set m̃^{n,p}_{ijk} = 0, ñ^{n,p}_{ijk} = 0
        if V_B(V^{n,p}_{ijk}) < 0 then
            m̃^{n,p}_{ijk} = λ_B
        end
        if V_C(V^{n,p}_{ijk}) < 0 then
            ñ^{n,p}_{ijk} = λ_C
        end
        Compute the elements of the matrix 𝒜;
        Solve the linear system of equations 𝒜V = b;
        Compute the difference tol = max |V^{n,p}_{ijk} − V_{ijk}|;
        V^{n,p+1}_{ijk} ← V_{ijk};
        p ← p + 1
    end
    n ← n − 1;
end
```

Remark 1. *Our numerical scheme is based on solving the system of linear Equation (16) for the unknown stacked vector (V^n_{ijk}). Its matrix representation contains at most 3 nonzero elements in each row. It has a block matrix structure with $N_\alpha \times N_\beta$ tridiagonal $N_S \times N_S$ matrices on the block diagonal. For each $N_S \times N_S$ tridiagonal we can employ the fast Thomas algorithm with time complexity $O(N_S)$. The overall complexity of computation of the system is therefore $O(N_\alpha \times N_\beta \times N_S \times p_{max})$ where p_{max} is the maximal number of policy iterations. Recall that there are other fast and robust numerical methods for solving problems of the form (16). Among them there are alternating direction explicit (ADI) methods for linear, nonlinear and multi-dimensional Black-Scholes models (c.f. (Bučková et al. 2017) and references therein).*

4.2. Results of Numerical Approximation of Option Prices

In this part, we present results of computation of European style call options for the exponential utility function with a risk parameter $\gamma > 0$ and linear utility function.

The model and numerical parameters used can be found in Table 2. We used the exponential mesh $S_k = K \ln(x_k)$ where $\{x_k, k = 1, \cdots, N_S\}$ is an equidistant mesh of the interval $[K/2, 2K]$. The plot of the call option price as a function of the underlying asset price S is shown in Figure 1, for the buyer call option prices, i.e., $\delta = -1$. We used Matlab framework for computation of the solution on 3GHz Intel single Core machine. For numerical discretization parameters shown in Table 2 The computational time was 8.1 sec per one policy iteration. According to Remark 1 the overall complexity is of the order $O(N_\alpha \times N_\beta \times N_S \times p_{max})$.

Table 2. Model and numerical parameters of the numerical solution.

Model Parameters	Value	Num Params	Value	Num Params	Value
Strike price K	50	N_α	6	N_S	100
Transaction costs θ	0.01	L_α^-	0.2	S^+	100
Volatility σ	0.3	L_α^+	0.6	$\lambda_B = \lambda_C$	10
Risk-free rate r	0.05	N_β	6	N	10
Drift μ	0.1	L_β^-	-100	T	1
Risk-aversion γ	0.1	L_β^+	100		

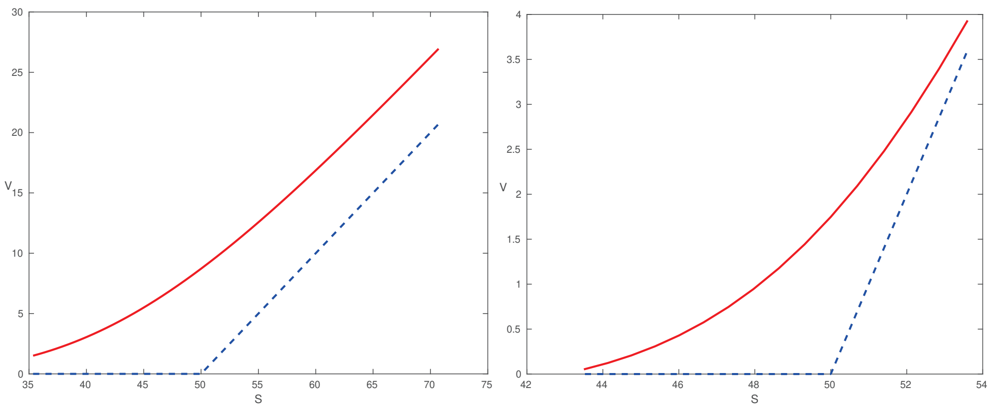


Figure 1. Linear (**left**) and exponential (**right**) utility indifference call option buyer price as a function of the underlying asset price S with the wealth function parameters $\alpha = 0.467, \beta = 33.3$.

5. Conclusions

In this paper we investigated and analyzed the system of Hamilton-Jacobi-Bellman equations arising in pricing financial derivatives. We followed the utility indifference option pricing model in which the option price is constructed as a difference of the certainty equivalents to the value functions solving the system of HJB equations. We analyzed solutions to the transformed nonlinear partial differential equation involving a possibly non-constant risk aversion function. Useful bounds on the option price were obtained using parabolic comparison principle. We also proposed a finite difference numerical discretization scheme. Various computational examples were also presented.

Author Contributions: The authors contributed equally to this research. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external support.

Acknowledgments: The authors gratefully acknowledge the contribution of the Slovak Research and Development Agency under the project APVV-20-0311.

Conflicts of Interest: The authors declare no conflict of interest.

References

Barles, Guy, and Halil Mete Soner. 1998. Option pricing with transaction costs and a nonlinear Black-Scholes equation. *Finance and Stochastics* 2: 369–97. [CrossRef]

Bučková, Zuzana, Matthias Ehrhardt, Michael Günther, and Pedro Pólvora. 2017. Alternating direction explicit methods for linear, nonlinear and multi-dimensional Black-Scholes models. In *Novel Methods in Computational Finance*. Cham: Springer, pp. 333–71.

Cantarutti, Nicola, Joao Guerra, Manuel Guerra, and Maria do Rosário Grossinho. 2017. Indifference pricing in a market with transaction costs and jumps. In *Novel Methods in Computational Finance*. Cham: Springer, pp. 31–46. [CrossRef]

Carmona, René, and Erhan Çinlar. 2009. *Indifference Pricing: Theory and Applications*. Princeton: Princeton University Press.

Chernogorova, Tatiana P., Miglena N. Koleva, and Radoslav L. Valkov. 2018. A two-grid penalty method for American options. *Journal of Computational and Applied Mathematics* 37: 2381–98. [CrossRef]

Clevenhaus, Anna, Matthias Ehrhardt, Michael Günther, and Daniel Ševčovič. 2020. Pricing American Options with a Non-Constant Penalty Parameter. *Journal of Risk and Financial Management* 13: 124. [CrossRef]

Davis, Mark H. A., Vassilios G. Panas, and Thaleia Zariphopoulou. 1993. European option pricing with transaction costs. *SIAM Journal on Control and Optimization* 31: 470–93. [CrossRef]

Dempster, Michael A. H., and Stanley R. Pliska, eds. 1997. *Mathematics of Derivative Securities*. Volume 15 of Publications of the Newton Institute. Cambridge: Cambridge University Press.

Dong, Yan, and Xiaoping Lu. 2021. Utility-indifference pricing of European options with proportional transaction costs. *Journal of Computational and Applied Mathematics* 397: 12. [CrossRef]

Hodges, Stewart D., and Anthony Neuberger. 1989. Optimal replication of contingent claims under transaction costs. *Review of Futures Markets* 8: 222–39.

Kallsen, Jan, and Johannes Muhle-Karbe. 2015. Option pricing and hedging with small transaction costs. *Mathematical Finance* 25: 702–23. [CrossRef]

Kilianová, Soňa, and Daniel Ševčovič. 2018. Expected utility maximization and conditional value-at-risk deviation-based Sharpe ratio in dynamic stochastic portfolio optimization. *Kybernetika (Prague)* 54: 1167–83.

Leland, Hayne E. 1985. Option pricing and replication with transactions costs. *The Journal of Finance* 40: 1283–301. [CrossRef]

Lesmana, Donny Citra, and Song Wang. 2015. Penalty approach to a nonlinear obstacle problem governing American put option valuation under transaction costs. *Applied Mathematics and Computation* 251: 318–30. [CrossRef]

Li, Wen, and Song Wang. 2009. Penalty approach to the hjb equation arising in european stock option pricing with proportional transaction costs. *Journal of Optimization Theory and Applications* 143: 279–93. [CrossRef]

Li, Wen, and Song Wang. 2014. A numerical method for pricing european options with proportional transaction costs. *Journal of Global Optimization* 60: 59–78. [CrossRef]

Meyer, John, and David Needham. 2014. Extended weak maximum principles for parabolic partial differential inequalities on unbounded domains. *Proceedings of the Royal Society A Mathematical Physical and Engineering Sciences* 478: 20140079. [CrossRef]

Monoyios, Michael. 2004. Option pricing with transaction costs using a Markov chain approximation. *Journal of Economic Dynamics and Control* 28: 889–913. [CrossRef]

Perrakis, Stylianos, and Jean Lefoll. 2000. Option pricing and replication with transaction costs and dividends. *Journal of Economic Dynamics and Control* 24: 1527–61. [CrossRef]

Post, Thierry, Yi Fang, and Martin Kopa. 2015. Linear tests for DARA stochastic dominance. *Management Science* 61: 1615–29. [CrossRef]

Ševčovič, Daniel. 2017. Nonlinear parabolic equations arising in mathematical finance. In *Novel Methods in Computational Finance*. Cham: Springer, pp. 3–15.

Ševčovič, Daniel, and Magdaléna Žitňanská. 2016. Analysis of the nonlinear option pricing model under variable transaction costs. *Asia-Pacific Financial Markets* 23: 153–74. [CrossRef]

Journal of
Risk and Financial Management

Article

Market Liquidity and Its Dimensions: Linking the Liquidity Dimensions to Sentiment Analysis through Microblogging Data

Francisco Guijarro [1,*], Ismael Moya-Clemente [2] and Jawad Saleemi [3]

[1] Research Institute for Pure and Applied Mathematics, Universitat Politècnica de València, 46022 València, Spain
[2] Faculty of Business Administration and Management, Universitat Politècnica de València, 46022 València, Spain; imoya@esp.upv.es
[3] Business School, Universitat Politècnica de València, 46022 València, Spain; j.saleemi@yahoo.com
* Correspondence: fraguima@upvnet.upv.es; Tel.: +34-963877000

Abstract: Market liquidity has an immediate impact on the execution of transactions in financial markets. Informed counterparty risk is often priced into market liquidity. This study investigates whether microblogging data, as a non-financial information tool, is priced along with market liquidity dimensions. The analysis is based on the Australian Securities Exchange (ASX), and from the results, we conclude that microblogging content in pessimistic periods has a higher impact on liquidity and its dimensions. On a daily basis, pessimistic investor sentiments lead to higher trading costs, illiquidity, a larger price dispersion and a lower trading volume.

Keywords: microblogging data; data mining; investor sentiments; asset pricing; market liquidity; liquidity dimensions

Citation: Guijarro, Francisco, Ismael Moya-Clemente, and Jawad Saleemi. 2021. Market Liquidity and Its Dimensions: Linking the Liquidity Dimensions to Sentiment Analysis through Microblogging Data. *Journal of Risk and Financial Management* 14: 394. https://doi.org/10.3390/jrfm14090394

Academic Editor: George Halkos

Received: 10 July 2021
Accepted: 20 August 2021
Published: 24 August 2021

1. Introduction

This work investigates whether microblogging data, as a source of information, can explain liquidity dimensions.

In the behavioral finance literature, emotion-driven market participants with stochastic predictions are gaining a considerable amount of interest. Recent research often quantified surveys, message boards (e.g., ragingbull.com, accessed on 19 August 2021) or financial news to construct sentiment indicators for modeling stock market behavior. Researchers are also exploring microblogging data for use in both modeling and predicting stock market behavior (Zhang et al. 2011). Moreover, microblogging sentiment indicators may be more economically meaningful than traditional sources of financial information (Oliveira et al. 2017).

The participation of companies in microblogging platforms can contribute to the development of valuable knowledge among investors (Prokofieva 2015), and increase the opportunity for significant returns (Bank et al. 2019). Market liquidity is often reported to be priced into asset returns (Saleemi 2020). An abundance of studies can be found that examine microblogging data for financial market prediction. However, there is still room to explore the impact of microblogging content on various liquidity dimensions.

The novelty of our work lies in the methodological contribution compared to related works. This study links the different dimensions of market liquidity with sentiment analysis using content from the popular social media platform Twitter.

In contrast to previous studies, investor sentiment tools are applied to uncover their role in the liquidity dimensions of microblogging content. This research fills a gap in the behavioral finance literature, and helps us to understand the impact on informed counterparty liquidity in a broader sense.

Liquidity, or its risk, is an active area of research as it imposes immediate consequences on the financial transaction (Guijarro et al. 2019). Market liquidity can be explained by its dimensions, which include transaction execution cost, trading quantity, immediacy of transaction execution and asset price dispersion (Le and Gregoriou 2020). Trading is considered illiquid (Gorton and Metrick 2010), and it is assumed that asymmetric information risk should be priced into liquidity (Saleemi 2020). Microblogging platforms allow market participants to exchange financial information on a real-time basis. To our knowledge, this is the first paper to study whether microblogging content, as an indicator of investor sentiments, is priced in the various dimensions of liquidity.

As microblogging content is gaining considerable attention in the behavioral finance literature, the aim of this research is therefore to explore whether liquidity dimensions can be significantly explained by microblogging sentiment indicators. As there is no previous literature on how investor sentiments may affect the different dimensions of liquidity, we do not hypothesize what the sign of the relationship between the two variables should be. Our paper aims to be the first empirical approach to the study of this problem. The results may have potential implications for both researchers and traders in terms of quantifying microblogging content-based sentiments with regards to market liquidity dimensions.

The rest of the paper is structured as follows. The literature is reviewed in Section 2. The procedure used to build the model and the data set is explained in Section 3. Section 4 discusses the findings of the research. Finally, Section 5 highlights the main results of the research.

2. Literature Review

The proliferation of behavior finance literature is attributable to the authoritative role of various sources of information on investor sentiments. Among the diversified structure of social networks, it may be of great interest to identify the most valued opinion providers. Microblogging platforms, in particular Twitter, allow participants to exchange potential content about financial markets on a real-time basis (Oliveira et al. 2017). Investor sentiment can be linked to systematic risk (Lee et al. 2002).

Investor sentiment determines asset price levels and therefore needs to be taken into account in the asset pricing model (Brown and Cliff 2005). Aggregate opinion has a significant impact on financial assets, the valuations of which are extremely subjective and difficult to arbitrage (Baker and Wurgler 2006). Moreover, financial assets without media coverage earn higher returns (Fang and Peress 2009), while monetary policy decisions in bear market periods have a greater impact on financial assets (Kurov 2010).

Incoming news significantly influences stock returns, volatility and trading volumes (Groß-Klußmann and Hautsch 2011). Microblogging content has some predictive power on returns, market-adjusted returns and future directional stock price movements (Oh and Sheng 2011). Twitter is a potential indicator of how the financial market will behave the next day (Zhang et al. 2011), while investor sentiments extracted from Twitter comments can predict asset price movements a few days in advance (Smailović et al. 2013).

Microblogging content can have greater effects on stock market performance than conventional media (Yu et al. 2013). Media investment interest plays a crucial role in reducing the information asymmetry, which in turn can stabilize the market, protect investors and improve corporate governance (Wei et al. 2014). In that sense, microblogging data can be a reliable source of stock-related news (Sprenger et al. 2014).

According to Walker (2016), the media can drive market behavior. In addition, companies' activity on Twitter can reduce the expected negative reactions in the market (Mazboudi and Khalil 2017). According to Li et al. (2018), users' attention to Twitter can better reflect stock trends. Aggregate opinion on Twitter is relevant for predicting a company's forthcoming quarterly earnings (Bartov et al. 2018), although Twitter content is less effective in determining market liquidity and trading cost (Guijarro et al. 2019).

Market liquidity and its related issues comprise one of the dominant strands of the asset pricing literature. With respect to the concept of information effects, the informed trader drives market liquidity (Glosten and Milgrom 1985). Immediacy, tightness, depth, breadth and resilience are the five key characteristics of a liquid market, according to Sarr and Lybek (2002). Market liquidity can be determined by trading cost, trading quantity, trading speed and price dispersion (Le and Gregoriou 2020). It follows then that informed trading risk must be priced in the liquidity (Saleemi 2020).

Liquidity is considered as a time-varying risk factor (Hasbrouck and Seppi 2001), as well as a crucial attribute of capital assets (Amihud and Mendelson 1991). The financial asset whose return is more sensitive to liquidity shocks has a higher expected return (Le and Gregoriou 2020). More recently, it has been found that returns are very sensitive to liquidity shocks in environments of high uncertainty, such as the current COVID-19 crisis (Saleemi 2021).

Market frictions are the costs associated with the execution of a transaction, which directly affect liquidity. Their impact has been shown to be time-varying (DeGennaro and Robotti 2007). Transaction costs can be divided into two major elements: the explicit cost and the implicit cost. The explicit cost is identifiable before the transaction takes place. However, the implicit cost is less observable and represents a large fraction of the total cost of the transaction. The bid–ask spread is a key point for the quantification of transaction costs, as it captures almost all the costs associated with the execution of the transaction (Sarr and Lybek 2002).

Since the late 1960s, the bid–ask spread has been extensively investigated in the asset pricing literature (Gregoriou 2013). Market-makers enable continuous trading by matching buy and sell orders. Liquidity providers facilitate the immediacy of trade execution by accepting the risk of holding inventory. Investors tend to reduce their risk exposure to future price uncertainty. In this context, liquidity providers impose a cost on the seller, i.e., a higher spread. The higher the volatility of asset prices, the higher the spread will be set by liquidity providers (Ho and Stoll 1981).

Another stream in the field links asymmetric information to the size of the spread. Information-sensitive stocks are illiquid. In the case of informed trading, there is a potential risk of loss for the uninformed party. Therefore, liquidity providers tend to increase the spread as compensation for this potential loss (Easley and O'Hara 2004). Another component of the spread is the order processing cost (Roll 1984). In case the order processing cost is higher, liquidity suppliers will buy an asset at the lowest bid price with the expectation of reselling it at the highest ask price.

Another interesting result is that the bid–ask spread is closely related to trading volume. The higher the cost of trading, the lower the amount of trading (Easley and O'Hara 1992). A small spread translates into a larger amount of trading, as the number of active trading participants causes the spread to become narrower. There are also causal effects on the spread of the amount of trading. A small trading volume reduces the size of the spread, which in turn adds liquidity to the market and improves price accuracy (Sarkissian 2016). According to Le and Gregoriou (2020), there is a strong relationship between higher trading volume and higher spread due to asymmetric information effects.

3. Materials and Methods

Our paper investigates whether informed trading based on microblogging content influences liquidity dimensions. To do so, we extracted investor sentiments from the popular social network Twitter, collecting different measures of each liquidity dimension in order to investigate the relationship between microblogging content and liquidity dimensions. Studies of asset pricing introduce several measures that capture one or more dimensions of market liquidity. This paper focuses on a small number of proxies for each dimension of liquidity, namely the bid–ask spread and liquidity based on the volume of price impact.

Depending on the frequency of the data, liquidity indicators are modeled in two ways: high-frequency data and low-frequency data. High-frequency measures estimate liquidity and its dimensions from intraday financial transactions. In contrast, the construction of low-frequency proxies is based on the daily characteristics of a security, such as the opening, high, low and closing prices (OHLC prices), as well as the volume traded. Unlike high-frequency data, low-frequency data are computationally less intensive and widely accessible to the markets. In this research, the analysis is based on low-frequency data from the Australian Securities Exchange (ASX), and was run over the period 3 January 2020 to 2 June 2021.

Among the measures of liquidity, the literature devotes much attention to the bid–ask spread. The spread captures the immediacy and cost of transactions. A large spread reflects a liquidity provider's unwillingness to accept an inventory position without imposing a higher cost on the seller. Most recently, Saleemi (2020) proposed a model of the cost-based market liquidity (CBML) measure, i.e., the bid–ask spread. The CBML model estimates the possible presence of an informed trader in the financial market. Based on the general foundations of the asset pricing literature, CBML is developed from Equation (1):

$$CBML_t = \sqrt{[(S_{t-1}) - (v_t^s)]^2} \tag{1}$$

where S_{t-1} is the ratio between the price range and the closing price on day $t-1$. This value is estimated by Equation (2):

$$S_{t-1} = \frac{high_{t-1} - low_{t-1}}{close_{t-1}} \tag{2}$$

where $high_{t-1}$ indicates the highest price on day $t-1$; low_{t-1} refers to the lowest price of day $t-1$; and $close_{t-1}$ is the closing price on day $t-1$. In the next trading session, the CBML method estimates the effects of asymmetric information on asset prices. v_t^s computes the ratio between the range price of an informed trader and the closing price on day t, as per Equation (3):

$$v_t^s = \frac{v_t^{ask} - v_t^{bid}}{close_t} \tag{3}$$

Assuming risk neutrality in the next trading session, the asset is valued at:

$$\eta_t = (high_t + low_t)/2 \tag{4}$$

where η_t is the mean of high and low prices on day t. If we consider the same probability of an informed trader, the estimated ask value for which the seller would redeem his position is assumed to be conditional on a trade such as:

$$v_t^{ask} = ask_t\pi + \eta_t\pi \tag{5}$$

where the estimated bid value for which the buyer would accept the inventory position is assumed conditional on a trade such as:

$$v_t^{bid} = bid_t\pi + \eta_t\pi \tag{6}$$

The liquidity model based on the impact of price on volume mainly estimates the level of liquidity by the dispersion of the asset price and its trading quantity. The Martin Liquidity Index (MLI) estimates the link between price changes and trading volume. The MLI model assumes that price dispersion influences trading volume and, as a result, impacts market liquidity. The higher the MLI value, the greater the price dispersion relative to the quantity traded. Hence, higher price dispersion leads to lower market liquidity. The analytical expression of the MLI for period t is given in Equation (7):

$$MLI_t = \sum\nolimits_{t=1}^{T} \frac{(close_t - close_{t-1})^2}{ln(vol_t)} \tag{7}$$

where vol_t is the quantity traded of the asset on day t. The model explains the price impact in terms of the effect that a traded unit has on the price. The illiquid asset requires less trading to move prices compared to the liquid asset. Note that our research only aims to estimate the influence of investor sentiments on stock market liquidity, therefore it is not necessary to deflate the price series. Such a deflation would make sense in a hypothetical case where one would want to analyze the profitability of an investment, as in the case where an analysis of investor sentiments could be used to derive a stock market investment strategy.

The R programming language was used to collect tweets from the ASX during the period from 3 January 2020 to 2 June 2021, using the libraries "ROAuth", "twitteR" and "rtweet". The study emphasizes pre-processing the unstructured text of the tweets. This process was carried out using the "NLP" and "tm" libraries, which allowed the original data to be cleaned and structured appropriately for further processing. Sentiment analysis tools were applied to convert intraday tweets into structured and valuable content. Tweets were structured by removing punctuation symbols, stop words and trailing spaces. In addition, the text was converted into lower case for the analysis of the microblogging financial conversation. For ethical reasons, market participants have been anonymized. For each tweet, the financial information was quantified in either a bullish (positive) or bearish (negative) score. Neutral opinions were not taken into account in the analysis. As the number of tweets posted on a single day is very large, the sentiment values for day t were aggregated for the analysis. This process was carried out through the "syuzhet" and "lubridate" libraries.

The basic sentiment indicators, i.e., the negative ratio (NR) and positive ratio (PR), were used as attributes according to Equations (8) and (9):

$$NR_t = \frac{Bear_t}{Bull_t} \tag{8}$$

$$PR_t = \frac{Bull_t}{Bear_t} \tag{9}$$

where $Bear_t$ is the aggregated bearish value on day t; and $Bull_t$ indicates the accumulated bullish value on day t.

First, we considered investor sentiment indicators as explanatory variables and liquidity dimensions as response variables, with both variables expressed in daily values. Next, the multiple linear regression model in Equation (10) was used to estimate the impact of investor sentiments on liquidity dimensions:

$$LD_t = \alpha + \beta_1 NR_t + \beta_2 PR_t + \epsilon_t \tag{10}$$

where LD_t refers to each measure of the liquidity dimension on day t; NR_t reflects the aggregated pessimistic sentiments on day t; PR_t indicates the aggregate optimistic sentiments on day t; and ϵ_t is the error term.

The following experiment is based on a multivariate forecasting algorithm, the vector autoregression (VAR) model. In this case, variables are modeled as a linear combination of their own lags and the past values of other variables. The Schwarz criterion (SC), also known as Bayesian information criterion, is applied to select the optimal lags. To estimate the impact of lags, the VAR model is structured through Equations (11)–(13):

$$L_t = \alpha_L + \beta_{11} L_{t-1} + \beta_{12} L_{t-2} + \gamma_{11} N_{t-1} + \gamma_{12} N_{t-2} + \varnothing_{11} P_{t-1} + \varnothing_{12} P_{t-2} + \epsilon_{L,t} \tag{11}$$

$$N_t = \alpha_N + \beta_{21} L_{t-1} + \beta_{22} L_{t-2} + \gamma_{21} N_{t-1} + \gamma_{22} N_{t-2} + \varnothing_{21} P_{t-1} + \varnothing_{22} P_{t-2} + \epsilon_{N,t} \tag{12}$$

$$P_t = \alpha_P + \beta_{31} L_{t-1} + \beta_{32} L_{t-2} + \gamma_{31} N_{t-1} + \gamma_{32} N_{t-2} + \varnothing_{31} P_{t-1} + \varnothing_{32} P_{t-2} + \epsilon_{P,t} \tag{13}$$

where L_t denotes each liquidity dimension on day t; L_{t-1} (L_{t-2}) refers to the lag value of each liquidity dimension on day $t-1$ ($t-2$); N_{t-1} (N_{t-2}) reflects the pessimistic sentiments on day $t-1$ ($t-2$); P_{t-1} (P_{t-2}) refers to the optimistic sentiments on day $t-1$ ($t-2$); $\epsilon_{L,t}$ is the white noise variable; N_t refers to the negative sentiments on day t; $\epsilon_{N,t}$ is the white noise variable; P_t refers to the positive sentiments on day t; and $\epsilon_{P,t}$ is another white noise variable.

In the following, we represent this model in a matrix notation:

$$\begin{bmatrix} L_t \\ N_t \\ P_t \end{bmatrix} = \begin{bmatrix} \alpha_L \\ \alpha_N \\ \alpha_P \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} \begin{bmatrix} L_{t-1} \\ L_{t-2} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \end{bmatrix} \begin{bmatrix} N_{t-1} \\ N_{t-2} \end{bmatrix} + \begin{bmatrix} \varnothing_{11} & \varnothing_{12} \\ \varnothing_{21} & \varnothing_{22} \\ \varnothing_{31} & \varnothing_{32} \end{bmatrix} \begin{bmatrix} P_{t-1} \\ P_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{L,t} \\ \epsilon_{N,t} \\ \epsilon_{P,t} \end{bmatrix} \quad (14)$$

Equation (14) is further elaborated as:

$$\begin{aligned} LS_t &= \begin{bmatrix} L_t \\ N_t \\ P_t \end{bmatrix}, \quad A = \begin{bmatrix} \alpha_L \\ \alpha_N \\ \alpha_P \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix}, \quad L_t = \begin{bmatrix} L_{t-1} \\ L_{t-2} \end{bmatrix}, \\ \gamma &= \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \end{bmatrix}, \quad N_t = \begin{bmatrix} N_{t-1} \\ N_{t-2} \end{bmatrix}, \quad \varnothing = \begin{bmatrix} \varnothing_{11} & \varnothing_{12} \\ \varnothing_{21} & \varnothing_{22} \\ \varnothing_{31} & \varnothing_{32} \end{bmatrix}, \\ P_t &= \begin{bmatrix} P_{t-1} \\ P_{t-2} \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \epsilon_{L,t} \\ \epsilon_{N,t} \\ \epsilon_{P,t} \end{bmatrix} \end{aligned} \quad (15)$$

Finally, we can rewrite the VAR model as Equation (16):

$$LS_t = A + \beta L_t + \gamma N_t + \varnothing P_t + \epsilon_t \quad (16)$$

4. Results and Discussion

The descriptive statistics of the data sample are shown in Table 1. It is noted that the variables are positively skewed with fat-tailed numerical distribution. Positive skewness of the data sample indicates a right-skewed distribution, with values to the right of mean. The fat-tailed numerical distribution, or higher kurtosis, indicates extreme values in the corresponding data set. The measures applied are based on distinct theoretical assumptions, which may influence the measurement of liquidity. The measures of the liquidity dimensions are plotted in Figure 1, where it is found that they are not constant, but vary over time.

Table 1. Descriptive statistics (daily basis).

Variables	Min	Median	Mean	Max	SD	S	K
CBML	0.0000522	0.008503	0.0110089	0.0869665	0.01043	3.3169	19.4480
MLI	0.000007	0.026949	0.120687	4.408370	0.35033	7.7186	80.1896
NR	0.1961	0.4984	0.5219	1.6038	0.16152	2.0675	11.3873
PR	0.6235	2.0065	2.0712	5.1	0.57910	1.1954	7.2555

Note: Cost-based market liquidity: CBML; Martin Liquidity Index: MLI; negative ratio: NR; positive ratio: PR; standard deviation: SD; skewness: S; kurtosis: K.

The microblogging sentiment indicators are depicted in Figure 2. It is also noted that investor sentiment indicators are not constant and change over time. It is worthwhile examining whether the Twitter feeds can influence the market liquidity dimensions. In this context, the sentiment analysis tools were applied to extract valuable content from unstructured Twitter feeds and the multivariate methods were applied to disentangle the various aspects involved. In our work, we aimed to analyze the impact of microblogging content on liquidity dimensions using multiple linear regression analysis.

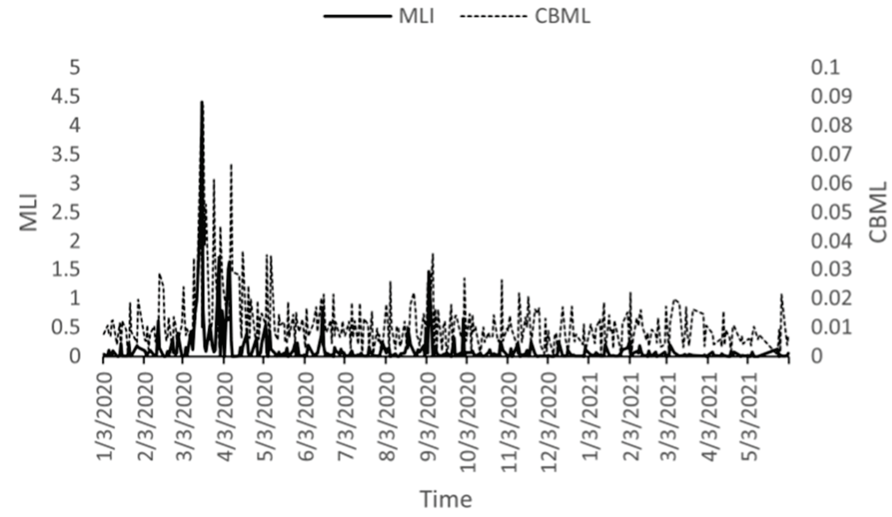


Figure 1. Time-varying market liquidity dimensions.

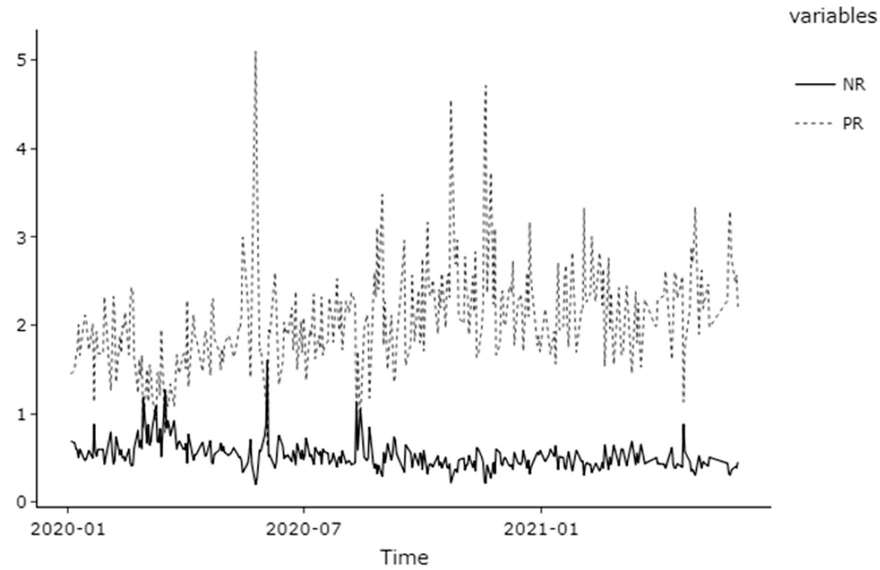


Figure 2. Time-varying investor sentiment indicators.

Table 2 presents the estimated regression values, where the investor sentiment indicators are the independent variables and each measure of the corresponding liquidity dimension acts as the dependent variable. On a daily basis, it is observed that the NR sentiment indicator is positive and significantly associated with the CBML measure. This implies that an increase in pessimistic investor sentiments leads to a higher spread. The higher spread illustrates the liquidity provider's unwillingness to accept the financial position without imposing a higher cost on the seller. A higher cost in pessimistic periods affects the speed of the transactions and therefore reduces liquidity for the ASX. Since the size of the spread is crucial in determining liquidity and its associated cost, a larger spread indicates illiquidity and a higher cost of trading in the Australian market during pessimistic periods. In contrast, the size of spread is not significantly explained by the optimistic sentiment measure, positive ratio. Therefore, optimistic sentiments based on microblogging content do not play a significant role in reducing the spread size in the Australian market.

Table 2. Regression analysis results.

Variables		Estimate	*p*-Value
	Intercept	−0.006133	0.4174
CBML (a)	NR	0.021341	0.0028 **
	PR	0.002899	0.1434
	Intercept	−0.46598	0.062395
MLI (b)	NR	0.81774	0.000533 ***
	PR	0.07718	0.237351

Note: (a) adjusted R-squared: 0.03674; F-statistic: 7.235; *p*-value: 0.0008427; (b) adjusted R-squared: 0.06915; F-statistic: 13.15; *p*-value: 0.000; significance codes: '***' < 0.001; '**' < 0.01.

The following experiment was conducted to analyze whether financial microblogging content can explain the dispersion of asset price and trading quantity. We found that pessimistic sentiments are positively and significantly associated with price impact volume-based liquidity. This indicates that a pessimistic bias in investor sentiments leads to a higher MLI value. The higher MLI value illustrates the greater price dispersion of the ASX relative to its trading volume. Therefore, investors would need a smaller amount of trades in the ASX to move its prices in the pessimistic periods. A higher MLI value, or higher price dispersion, illustrates the lack of liquidity in the ASX market. However, the optimistic mood of investors is not significantly associated with price impact volume-based liquidity.

Table 3 presents the VAR coefficients for the past time series of spreads and investor sentiments. Market liquidity and its associated cost are not significantly explained by the lagged coefficients of investor sentiments. Cost-based market liquidity is reported to be significantly correlated with its own past time series. Meanwhile, pessimistic investor sentiments are not significantly explained by the lagged coefficients of cost-based market liquidity and optimistic sentiments. Investors' optimistic sentiments are not significantly correlated with the past time series of cost-based market liquidity and pessimistic sentiments.

Table 3. Estimation of VAR coefficients and significance test values, CBML model.

Variables		Estimate	*p*-Value
CBML (a)	$\beta_{11,CBML}$	−0.8168	0.000 ***
	$\gamma_{11,N}$	−0.01081	0.0843
	$\varnothing_{11,P}$	−0.002544	0.1453
	$\beta_{12,CBML}$	−0.3748	0.000 ***
	$\gamma_{12,N}$	0.006916	0.2730
	$\varnothing_{12,P}$	0.0009216	0.5985
	α_{CBML}	−0.00001418	0.9797
N (b)	$\beta_{21,CBML}$	−0.095473	0.90380
	$\gamma_{21,N}$	−0.677182	0.000 ***
	$\varnothing_{21,P}$	−0.030498	0.25019
	$\beta_{22,CBML}$	−0.530606	0.49794
	$\gamma_{22,N}$	−0.279544	0.00375 **
	$\varnothing_{22,P}$	−0.019775	0.45731
	α_N	−0.001133	0.89382
P (c)	$\beta_{31,CBML}$	−1.855197	0.5173
	$\gamma_{31,N}$	0.149640	0.6639
	$\varnothing_{31,P}$	−0.463831	0.000 ***
	$\beta_{32,CBML}$	0.810466	0.7752
	$\gamma_{32,N}$	−0.054357	0.8757
	$\varnothing_{32,P}$	−0.195826	0.0429 *
	α_P	0.003898	0.8992

Note: (a) adjusted R-squared: 0.4645; F-statistic: 47.84; *p*-value: 0.000; ARCH test: 0.000; JB test: 0.000; (b) adjusted R-squared: 0.2608; F-statistic: 20.05; *p*-value: 0.000; ARCH test: 0.000; JB test: 0.000; (c) Adjusted R-squared: 0.1971; F-statistic: 14.25; *p*-value: 0.000; ARCH test: 0.000; JB test: 0.000; significance codes: '***' < 0.001; '**' < 0.01; '*' < 0.05.

Moreover, the results find that investors' sentiments are significantly associated with their own past time series. The Jarque–Bera (JB) test, the autoregressive conditional heteroscedastic (ARCH) test and the forecast error variance decomposition (FEVD) test are estimated. The JB test indicates that the residuals are not normally distributed. The ARCH test shows that the variables suffer from the ARCH effects. Figure 3 reveals that cost-based market liquidity and pessimistic investor sentiments are strongly influenced by their own variance shocks. Investors' optimistic sentiments are influenced by their own exogenous shocks and negative sentiments variance shocks.

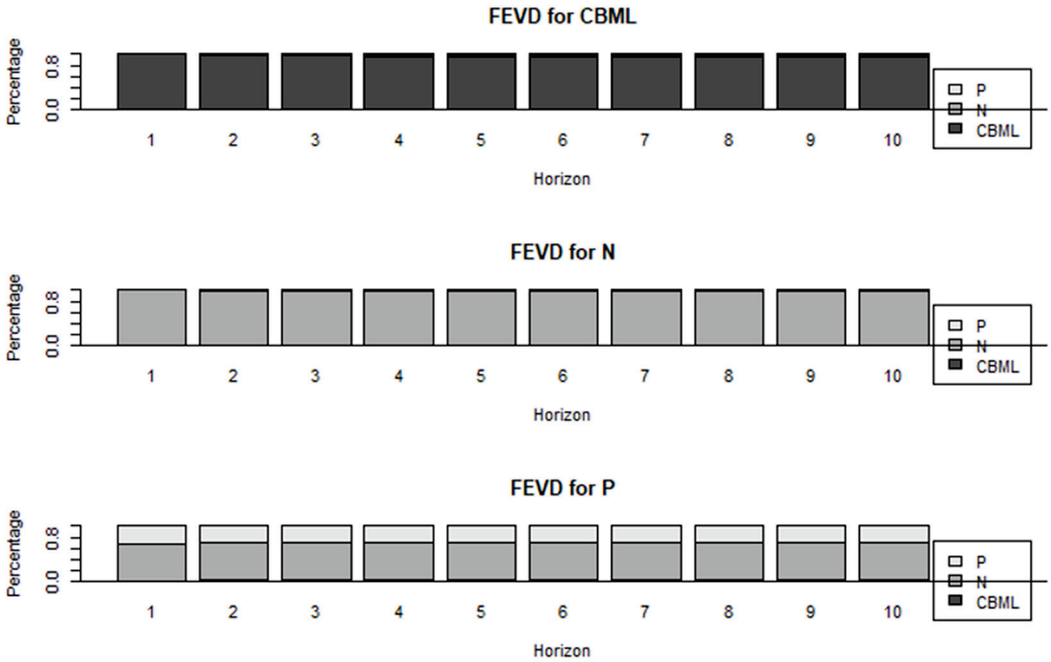


Figure 3. FEVD analysis for CBML and investor sentiments.

Based on the VAR model estimation results, the dispersion of the asset price and its trading quantity are not significantly explained by the lagged coefficients of the investor optimistic sentiments (Table 4). However, it is observed that price impact volume-based liquidity is significantly associated with its own past time series and the lag_{t-1} of pessimistic investor sentiments. Likewise, pessimistic investor sentiments are not significantly explained by the past time series of price impact volume-based liquidity and optimistic investor sentiments. Optimistic investor sentiments are not significantly associated with the lagged coefficients of price impact volume-based liquidity and pessimistic investor sentiments. Moreover, investors' sentiments are significantly explained by their own past time series. The JB test shows that the residuals are not normally distributed. The ARCH test reports that the variables suffer from the ARCH effects. Figure 4 illustrates that price impact volume-based liquidity and pessimistic investor sentiments are strongly influenced by their own exogenous shocks. Finally, investors' optimistic sentiments are influenced by their own variance shocks and exogenous negative sentiment shocks.

Table 4. Estimation of VAR coefficients and significance test values, MLI model.

Variables		Estimate	*p*-Value
MLI (a)	$\beta_{11,\text{MLI}}$	−0.3323240	0.000 ***
	$\gamma_{11,N}$	0.4591262	0.0268 *
	$\varnothing_{11,P}$	0.0901849	0.1156
	$\beta_{12,\text{MLI}}$	−0.2531608	0.000 ***
	$\gamma_{12,N}$	0.0268085	0.8976
	$\varnothing_{12,P}$	0.0373390	0.5145
	α_{MLI}	−0.0001461	0.9936
N (b)	$\beta_{21,\text{MLI}}$	−0.013135	0.60219
	$\gamma_{21,N}$	−0.668877	0.000 ***
	$\varnothing_{21,P}$	−0.029397	0.26646
	$\beta_{22,\text{MLI}}$	0.016084	0.51670
	$\gamma_{22,N}$	−0.276186	0.00437 **
	$\varnothing_{22,P}$	−0.020703	0.43411
	α_N	−0.001120	0.89491
P (c)	$\beta_{31,\text{MLI}}$	0.065824	0.4712
	$\gamma_{31,N}$	0.099416	0.7738
	$\varnothing_{31,P}$	−0.474563	0.000 ***
	$\beta_{32,\text{MLI}}$	−0.079112	0.3790
	$\gamma_{32,N}$	−0.047393	0.8920
	$\varnothing_{32,P}$	−0.189236	0.0491 *
	α_P	0.003904	0.8989

Note: (a) adjusted R-squared: 0.141; F-statistic: 9.862; *p*-value: 0.000; ARCH test: 0.000; JB test: 0.000; (b) adjusted R-squared: 0.2616; F-statistic: 20.13; *p*-value: 0.000; ARCH test: 0.000; JB test: 0.000; (c) adjusted R-squared: 0.1987; F-statistic: 14.39; *p*-value: 0.000; ARCH test: 0.000; JB test: 0.000; significance codes: '***' < 0.001; '**' < 0.01; '*' < 0.05.

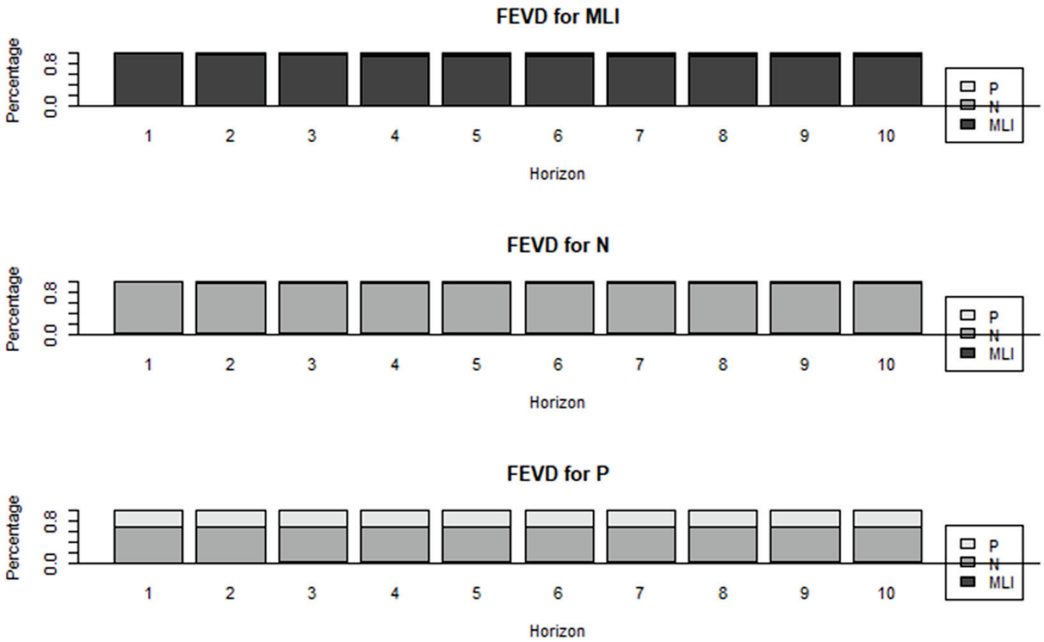


Figure 4. FEVD analysis for MLI and investor sentiments.

5. Conclusions

This research analyzed whether investor sentiments based on microblogging content influences the dimensions of market liquidity. Using time series and multivariate statistical models, the influence of investor sentiments on the liquidity of the Australian Securities Exchange was studied. To estimate investor sentiments, posts on the popular social network platform Twitter were analyzed and different liquidity measures were applied to estimate the relationship between microblogging content and liquidity dimensions. We found that investor sentiments in pessimistic periods were significantly associated with higher trading cost, illiquidity, higher price dispersion and lower trading volume. However, cost-based market liquidity and price impact volume-based liquidity were not significantly explained by optimistic investor sentiments.

From the multivariate model approach, market liquidity and its associated cost were not significantly associated with the past time series of pessimistic and optimistic investor sentiments. In contrast, price impact volume-based liquidity was found to be positive and was significantly explained by lagged pessimistic investor sentiments. Likewise, a significant relationship was found between market liquidity dimensions and their own past time series. Finally, market liquidity dimensions were discovered to be strongly influenced by their own variance shocks.

This research has important implications in terms of revealing the relationship between microblogging content and the various dimensions of liquidity that previous studies have ignored. This quantification of investor sentiments based on microblogging content may be useful for liquidity risk management and portfolio construction. Although the study fills a gap in the behavioral finance literature, the geographical dataset that was employed is a limiting element of the study. As this study covers the Australian market, the results may not be generalizable to other markets. The analysis therefore encourages other researchers to uncover the impact of microblogging content on liquidity dimensions at both the industry and the company level. This would undoubtedly provide insight into the authoritative role of microblogging content on liquidity dimensions more broadly.

Author Contributions: Conceptualization, F.G., I.M.-C. and J.S.; methodology, F.G., I.M.-C. and J.S.; software, F.G. and J.S.; validation, J.S.; formal analysis, I.M.-C.; investigation, F.G.; resources, J.S.; data curation, J.S.; writing—original draft preparation, J.S.; writing—review and editing, F.G. and I.M.-C.; visualization, J.S.; supervision, F.G. and I.M.-C.; project administration, F.G., I.M.-C. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: Two anonymous reviewers provided thoughtful and detailed comments that greatly improved the final version of this article. We would like to thank the reviewers for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

Amihud, Yakov, and Haim Mendelson. 1991. Liquidity, maturity, and the yields on U.S. treasury securities. *The Journal of Finance* 46: 1411–25. [CrossRef]

Baker, Malcolm, and Jeffrey Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61: 1645–80. [CrossRef]

Bank, Semra, Evrim E. Yazar, and Ugur Sivri. 2019. Can social media marketing lead to abnormal portfolio returns? *European Research on Management and Business Economics* 25: 54–62. [CrossRef]

Bartov, Eli, Lucile Faurel, and Partha S. Mohanram. 2018. Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review* 93: 25–27. [CrossRef]

Brown, Gregory W., and Michael T. Cliff. 2005. Investor sentiment and asset valuation. *The Journal of Business* 78: 405–40. [CrossRef]

DeGennaro, Ramon P., and Cesare Robotti. 2007. Financial Market Frictions. *Economic Review* 92: 1–16.

Easley, David, and Maureen O'Hara. 1992. Time and the process of security price adjustment. *The Journal of Finance* 47: 577–605. [CrossRef]

Easley, David, and Maureen O'Hara. 2004. Information and the cost of capital. *The Journal of Finance* 59: 1553–83. [CrossRef]

Fang, Lily, and Joel Peress. 2009. Media coverage and the cross-section of stock returns. *The Journal of Finance* 64: 2023–52. [CrossRef]

Glosten, Lawrence R., and Paul R. Milgrom. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14: 71–100. [CrossRef]

Gorton, Gary, and Andrew Metrick. 2010. Haircuts. *Federal Reserve Bank St Louis Review* 92: 507–20. [CrossRef]

Gregoriou, Andros. 2013. Earnings announcements and the components of the bid-ask spread: Evidence from the London stock exchange. *Journal of Economic Studies* 40: 112–26. [CrossRef]

Groß-Klußmann, Axel, and Nikolaus Hautsch. 2011. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance* 18: 321–40. [CrossRef]

Guijarro, Francisco, Ismael Moya-Clemente, and Jawad Saleemi. 2019. Liquidity Risk and Investors' Mood: Linking the Financial Market Liquidity to Sentiment Analysis through Twitter in the S&P500 Index. *Sustainability* 11: 7048. [CrossRef]

Hasbrouck, Joel, and Duane J. Seppi. 2001. Common Factors in Prices, Order Flows, and Liquidity. *Journal of Financial Economics* 59: 383–411. [CrossRef]

Ho, Thomas, and Hans R. Stoll. 1981. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics* 9: 47–73. [CrossRef]

Kurov, Alexander. 2010. Investor sentiment and the stock market's reaction to monetary policy. *Journal of Banking and Finance* 34: 139–49. [CrossRef]

Le, Huong, and Andros Gregoriou. 2020. How do you capture liquidity? A review of the literature on Low-frequency stock liquidity. *Journal of Economic Surveys* 34: 1170–86. [CrossRef]

Lee, Wayne Y., Christine X. Jiang, and Daniel C. Indro. 2002. Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance* 26: 2277–99. [CrossRef]

Li, Qing, Yan Chen, Jun Wang, Yuanzhu Chen, and Hsinchun Chen. 2018. Web media and stock markets: A survey and future directions from a big data perspective. *IEEE Transactions on Knowledge and Data Engineering* 30: 381–99. [CrossRef]

Mazboudi, Mohamad, and Samer Khalil. 2017. The attenuation effect of social media: Evidence from acquisitions by large firms. *Journal of Financial Stability* 28: 115–24. [CrossRef]

Oh, Chong, and Olivia Sheng. 2011. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. Paper presented at the International Conference on Information Systems, Shanghai, China, December 4–7.

Oliveira, Nuno, Paulo Cortez, and Nelson Areal. 2017. The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications* 73: 125–44. [CrossRef]

Prokofieva, Maria. 2015. Twitter-based dissemination of corporate disclosure and the intervening effects of firms' visibility: Evidence from Australian-listed companies. *Journal of Information Systems* 29: 107–36. [CrossRef]

Roll, Richard. 1984. A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance* 39: 1127–39. [CrossRef]

Saleemi, Jawad. 2020. An estimation of cost-based market liquidity from daily high, low and close prices. *Finance, Markets and Valuation* 6: 1–11. [CrossRef]

Saleemi, Jawad. 2021. COVID-19 and liquidity risk, exploring the relationship dynamics between liquidity cost and stock market returns. *National Accounting Review* 3: 218–236. [CrossRef]

Sarkissian, Jack. 2016. Option pricing under quantum theory of securities price formation. *SSRN Electronic Journal*. [CrossRef]

Sarr, Abdourahmane, and Tonny Lybek. 2002. Measuring liquidity in financial markets. *International Monetary Fund* 2: 1–64. [CrossRef]

Smailović, Jasmina, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. 2013. Predictive sentiment analysis of Tweets: A stock market application. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*. Berlin/Heidelberg: Springer. [CrossRef]

Sprenger, Timm O., Andranik Tumasjan, Philipp G. Sandner, and Isabell M. Welpe. 2014. Tweets and trades: The information content of stock microblogs. *European Financial Management* 20: 926–57. [CrossRef]

Walker, Clive B. 2016. The direction of media influence: Real-estate news and the stock market. *Journal of Behavioral and Experimental Finance* 10: 20–31. [CrossRef]

Wei, Cen, Li Shihao, and Tong Naqiong. 2014. The Influence of Investor Attention on the Stock Return and Risk: An Empirical Study Based on the "Easy Interactive" Platform Data of Shenzhen Stock Exchange. *Securities Market Herald* 7: 40–47.

Yu, Yang, Wenjing Duan, and Qing Cao. 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems* 55: 919–26. [CrossRef]

Zhang, Xue, Hauke Fuehres, and Peter A. Gloor. 2011. Predicting stock market indicators through Twitter "i hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences* 26: 55–62. [CrossRef]

Journal of
Risk and Financial Management

Article

Risk Perception, Accounting, and Resilience in Public Sector Organizations: A Case Study Analysis

Mouhcine Tallaki and **Enrico Bracci** *

Department of Economics and Management, University of Ferrara, 44121 Ferrara, Italy; tllmcn@unife.it
* Correspondence: enrico.bracci@unife.it

Abstract: There are various factors that can affect an organization's ability to overcome a crisis and the uncertainties that arise thereafter. Little is known about the process of organizational resilience and the factors that can help or prevent it. In this paper, we analyzed how public sector organizations build resilience/traits of risks awareness, and in doing that, we derived some elements that could affect the process of resilience. In particular, drawing on the conceptual framework proposed by Mallak we analyzed an in-depth case study in a public sector organization (PSO) identifying some contextual dimensions implicated in the process of building resilience. In our analysis, we identified two main elements that affect resilience: Risk perception and the use of accounting. Results shown how risk perception is perceived as a trigger, while accounting is considered as an enforcer in the process of building resilience capacity. The results also show the way accounting is implicated in the management of austerity programs and supporting the creation of a resilient public sector organization. In our case, the risk has become an opportunity for change. In the face of these budget cuts, management began refocusing the company's mission from infrastructure maintenance to providing services with a market-based logic.

Keywords: resilience capacity; cut-back management; crisis

Citation: Tallaki, Mouhcine, and Enrico Bracci. 2021. Risk Perception, Accounting, and Resilience in Public Sector Organizations: A Case Study Analysis. *Journal of Risk and Financial Management* 14: 4. https://doi.org/10.3390/jrfm14010004

Received: 3 December 2020
Accepted: 21 December 2020
Published: 24 December 2020

1. Introduction

The concept of resilience is holistic and complex as it is developed from several perspectives, i.e., organizational, information technology, industrial relation, engineering, business strategy, culture, organizational learning (Andersson et al. 2019). Organizational resilience is defined as the ability to build traits of risk awareness, preference for cooperation agility, and improvisation (Andersson et al. 2019). Organizational resilience is the ability to effectively absorb, develop specific responses to, and engage in transformative activities to capitalize on disruptive surprises that potentially threaten organization survival (Lengnick-Hall et al. 2011). Mainstream literature views resilience in terms of the ability to return to the previous state of perceived 'normality' (engineering perspective) or as the capacity for recovery from crisis by following the sequence of rescue, restoration of damaged infrastructure and then the rebuilding of markets (ecological perspective) (Barbera et al. 2017; Davoudi et al. 2012). According to the ecological perspective organizations could not only persist, but also adapt its structure under the new conditions, and this process implies gradual change as a new business concept may emerge (Dahles and Susilowati 2015). The ecological perspective recognizes the existence of multiple equilibria, and the possibility of systems to flip into alternative stability domains (Davoudi et al. 2012).

The concept of organizational resilience is based, therefore, on the notion that individuals, or a system, can withstand or recover its prior shape after shocks. This concept is described sometimes as a trait, a characteristic, or an outcome of the organization that faced adverse conditions (Carden et al. 2018). However, environmental conditions like financial crises or other external shocks are not sufficient to describe resilience in public sector organizations. In the public sector, various studies have explored governmental responses to the crisis but fewer studies deal with the managerial consequences of such phenomena for public organizations (Barbera et al. 2017; Tallaki and Bracci 2020). Austerity measures in public sector have focused the role of resilience in the public sector (Barbera et al. 2017; Leslie and Canwell 2010). Nowadays, the main challenges for the public sector is to deliver services in an age of austerity (Bracci et al. 2015; Leslie and Canwell 2010). Boin and Lodge (2016) argue that the time has come for public administration scholars to incorporate crisis and disaster management into the main research agendas of the field. De Aquino and Cardoso (2017) analyze the pattern presented by four Brazilian municipalities at the beginning of a serious revenue downturn, and the authors show that the cases analyzed do not have an anticipatory capacity or long-term strategic planning. The same applied in Greece where it was shown that the municipalities did not demonstrate significant anticipatory capabilities (Cohen and Hlepas 2017). The resilience pattern could be different between municipalities, its depending on how the shock has been perceived (Du Boys 2017) or other characteristics like organizational features and capacities, as well as the characteristics of the external environment (Barbera et al. 2017, 2020). In fact, in other studies, municipalities appeared to be a common set of anticipatory and coping capacities (Jones 2017).

Barbera et al. (2017) call for further studies about the internal and external determinant of resilience in public sector organizations, and in particular the role of accounting in supporting the development of resilience capacity. More recently, Barbera et al. (2020) argued on the need to further investigate how accounting can help public sector organization to change and adapt to crisis and risks. Moving from this gap in the literature, we analyzed how accounting technologies designed and used by public sector organizations can trigger the development of resilience/traits of risks awareness. In doing so, we draw on the conceptual framework defined by (Mallak 1998) and applied it in a case study conducted in a public sector organization (PSO). Results show how risk perception is perceived as a trigger, while accounting is considered as enforcer in the process of resilience. The results also shown the way accounting is implicated in the management of austerity programs and supporting the creation of a resilient organization. In our case, the risk has become an opportunity for change. In the face of these budget cuts, management began refocusing the company's mission from infrastructure maintenance to providing services with a market-based logic. Accounting played different roles in the construction of reality (Burchell et al. 1980; Hines 1988) and in managing uncertainty (Chong and Chong 1997). Accounting provided certain language and discourse that constitute a way of establishing legitimacy in highly ambiguous situations (Nahapiet 1988). We argue that accounting has a role in making public sector organizations more resilient, with relevant policy and managerial implications. To the best of our knowledge, this paper is one of the first attempting to investigate the relationship between accounting technologies, risk awareness, and the development of resilient capacities.

The remainder of the paper is structured as follows: Section 2 describes how to build resilience and the framework used. Section 3 reports the methodology. Section 4 describes the case study and the results emerging. The paper ends with a discussion and some final reflections to highlight the contributions of the paper as well as its limitations.

2. How to Build Resilience: Theoretical Framework

The resilience concept is multifaceted and depends on a variety of dimensions and characteristic (Linnenluecke 2017). In the process of containing external risks, i.e., resilience, organizations may activate three mechanisms (Powley 2009): Alteration of the formal relational structures and opening a temporal space for organization members to form and renew relationships (liminal suspension); awareness of community members of the needs of others (compassionate witnessing); and activation of relational networks across organizational and functional boundaries (relational redundancy). In the same way, organizations can absorb external risks by undergoing first-order change and single-loop learning (labelled resiliency), or they can adopt new practices through second-order change and double-loop learning (labelled retention) (Meyer 1982). Accordingly, resilience can be planned (anticipatory) and adaptive (coping with). Planned resilience involves the use of existing, predetermined planning capabilities. Adaptive resilience emerges as response to a crisis or other emergent situations (Lee et al. 2013).

Various internal and external factors could therefore support public sector organizations to be resilient, also affecting the professional identity of managers and employees (Barbera et al. 2017). Identifying the determinants of resilience or those internal and external factors makes possible to understand how to make an organization more performant in terms of resilience. This is because, as highlighted by Fraser et al. (1999, p. 136), "*If we can understand what helps some people to function well in the context of high adversity, we may be able to incorporate this knowledge into new practice strategies*". Accordingly, organizations need to understand their resilience strengths and weaknesses and must be able to evaluate the effectiveness of resilience strategies (Lee et al. 2013).

Carden et al. (2018) specify that a flexible workforce, not only in skills and job tasks, but also in mindsets could help organizations to survive. Biggs (2011) and Coutu (2002) highlight the importance of human capital, which takes in some of the features of innovation and collaboration, in developing organizational resilience. Meyer (1982) concludes that resilience is influenced by strategy and slacks in resources, ideologies, and by organizational structures. Sutcliffe and Vogus (2003) argue that resilience in organizations depends on broader information processing, loosening of control, and utilization of slack capabilities. Hamel and Välikangas (2003) suggest that innovation is the key for resilience. Another example of resilient organizations was presented by Gittell et al. (2006), whereby the authors highlight the importance of viable business model that allows financial reserves to build resilience in organizations.

In order to address our research question, we draw on Mallak (1998), who defined seven drivers of resilience, namely: Perceive experiences constructively, perform positive adaptive behavior, ensure external resources, expand decision-making boundaries, practice bricolage, develop tolerance for uncertainly, and build virtual rule system (Table 1). *Positive perception of experiences* is essential to the resilient individual because the individual is more likely to be able to solve the problem when he perceives experiences positively. In this case, individual takes an active approach, i.e., *positive adaptive behavior*, toward solving problems in the workplace.

Table 1. Summary of resilience dimensions.

Principle	Putting the Principle to Work
Perceive experiences constructively	Even if the experience causes pain, find the positive angle and move forward.
Perform positive adaptive behaviors	Perceive change as opportunity, not danger. Allow responses to adapt to the needs of the situation, rather than execute ineffective "programmed" responses.
Ensure adequate external resources	Ensure access to adequate resources to allow positive adaptive response to approach a wide variety of possible events.
Expand decision making boundaries	Provide greater decision making authority to support positive adaptive response and the use of resources to achieve the objective.
Practice bricolage	Develop the ability to create solutions on the spot using materials on hand.
Develop tolerance for uncertainty	Develop the ability to make decisions with less than the desired amount of information.
Build virtual role systems	In a team, individuals have a shared understanding of the team's mission and can fill in wherever needed to ensure smooth functioning of the team.

Source: Mallak (1998).

Adequate external resources include resources of advice, information, finances, emotional support, and practical help, and this principle works in concert with *expanded decision-making* boundaries to maximize the potential for positive adaptive responses. *Bricolage* is the ability to fashion a solution on the spot using materials on hand; to do that individuals needs expanded decision-making authority and access to resources. Regarding *develop tolerance for uncertainty*, the latter refers to situations where the amount of information we need is greater than the information we have. Individuals with this capacity could make good decisions under conditions of uncertainty. *Virtual role* is identified as an advanced form of work team relationships. In this paper, we focused on the framework of Mallak (1998) to analyze, in particular, the role of accounting in supporting public sector organizations to develop risk awareness and become resilient.

3. Research Methodology

The paper adopts a qualitative approach with a longitudinal case study. In doing so, we also respond to the call for more qualitative analyses (Vaivio 2008) by developing a case study about PSO. We used a mix of primary and secondary sources. We interviewed managers and employees involved in the process under study. We did total of 18 interviews, the average duration of the interviews is equal to 1.44 h (Table 2). All interviews were taped and transcribed. The transcripts were sent to the interviewees to validate and confirm the adherence of the content with their thought and perception.

Table 2. Interviewees by position/duration.

Interviewees' Role	Duration (In Hour)	Number of Interviews	Number of Interviewees
Managing director	3.0 h	2	1
CFO	3.0 h	2	1
Management control manager	1.5 h	1	1
Human resources manager	1.2 h	1	1
Manager of strategical investment	5.8 h	4	4
Maintenance manager	3.0 h	2	2
Maintenance workers	5.5 h	4	4
Quality manager	1.5 h	1	1
Purchasing Manager	1.5 h	1	1
Total	**26.0 h**	**18**	**16**

Thus, we analyzed and codified the interviews, highlighting the keywords that related to resilience and the way the company reacted to a financial crisis. One of the critical elements in qualitative data analysis is the systematic coding of interviews (Miles et al. 2014) as it attributes interpreted meaning to each datum that contains references to specific categories of information (Saldaña 2013). Given the research question and the explorative approach adopted, the code list was an inductive task, based on what respondents said (Carey et al. 1996).

The interviews were analyzed separately by the authors to reduce subjectivity, and results were compared and discussed to reach a common understanding of the meaning of the interviewees' responses. We also took detailed notes of several informal interviews. These informal interviews allowed us to go back to the original interviews to specify, clarify and ask for additional evidence. We also collected data through direct observation and participation in meetings. We first entered the site in 2014 and started the research in 2015, which allowed us to follow changes over time. We attended various board and management meetings. Furthermore, we had access to all internal and external documentation (minutes, budgets, guidelines, procedures, public documents concerning the sector's data, and the financial situation of the owner).

4. Analysis of Results

The subject of the case study is a PSO that deals with ordinary and extraordinary maintenance of regional government infrastructure and is 100% controlled by the regional government. The company employs some 250 people and has a budget around €19 million (in 2018). Its main source of income is an annual transfer made by the regional government for infrastructure maintenance. Over the last few years, public transfers have fallen by around 35%, or some €5.5 million. The cuts were incremental but the financial performance remained in line with the historical trend. This is because the company has managed to compensate the cut in resources with the development of a new business line.

4.1. Risk Perception

Over the years, the company has managed to overcome the crisis by developing new business opportunities. The perception of risk, as the interviewees pointed out, following the cut of funds, was the trigger that led to a change in the business model. In fact, financial risk intensively affects business environment. Risk perception is linked to uncertainty. The latter is considered as risk when it becomes on object of management (Power 2007). Risk could be distinguished on the basis of the source of risk. Black (2005) reported that internal risks arise within the organization, external risks, instead, arise from the environment in which the organizations operate. The reduction of funds, considered as the source of risk, generated uncertainties regarding the future of the company, as the interviewees pointed out. Various operators said they are worried about the company's future.

Renn (2012) reported that organizations could deal with three different approach to govern risk: Technocratic risk management approach that focuses on risk minimization; decisionistic risk management that extends the technocratic approach by opening debate to risk evaluation; and transparent risk governance considers the engagement with the nature, perceptions, and contested benefits of risk in complex situations. In the public sector, risk is usually related to technocratic/bureaucratic approach of risk (Brown and Osborne 2013; Power 2007). In our case, the risk has become an opportunity for change. In the face of these budget cuts, management began refocusing the company's mission from infrastructure maintenance to providing services with a market-based logic.

4.2. Building Resilience

In the following sections, we analyze the resilience process according to the various dimensions of Mallak (1998).

✓ *Perceive experiences constructively*

According to Mallak (1998), an individual is more likely to be able to solve his problem when he perceives experiences positively. Even if the experience causes pain, organizations have to find the positive angle and move forward. This positive perception of change was noted by researchers both in meetings held between various managers and in interviews conducted. Regarding these changes one of manager of strategical investment said:

> *I saw the company change the type of mentality and approach to some problems or some processes that were not previously checked or held in particular consideration, this is positive for me . . . . the change in mentality allows to go in parallel with market changes.*

This orientation to change and the positive perception of changes was also noted by the human resources manager who said:

> *The desire for learning of the staff seems positive. I think it's a good thing. It also serves us to bring together people from different business realities in order to homogenize. This was done in particular for maintenance personnel of electrical installations and station managers who had different operating modes and sometimes did not understand each other.*

The company had a strong predisposition to change and consequently it perceived experiences constructively. With the cutting of resources, the company found itself faced with a significant financial risk. The proposed solution, and then introduced by the management, concerned the reconversion of the company's mission, that is, from an infrastructure maintenance to a production company with a more market logic. The first two years the cut was not significant and this allowed the company to experiment with the new logic. The adherence to the reconversion, by the staff, was voluntary. The staff involved in this conversion process seem satisfied, as pointed out one by of the maintenance operators: *For me it is a challenge, but the company never leaves you alone.* To motivate the staff involved, an additional production award was given. The accompaniment, the training, and the motivation have had positive results because the rate of adhesion to the conversion process has increased through time.

To verify the sustainability of the new business idea the PSO developed a business plan. Various meetings in order to establish the operational objectives and in order to establish the production standards have been carried out. In addition, benchmarking was also done to understand the productivity of similar companies. All of this has been translated into various simulations that have led the company to set productivity goals. In a first phase, the company started with high-level uncertainty of both the objectives and the actions, because there were no clear management and operational objectives, the choice made to convert the staff without having an idea about productivity and therefore on the consequence of the action taken created uncertainty regarding actions. This emerged in the various meetings, in which the researchers also took part, where all responsible were skeptical of the future. However, there were no alternatives, as pointed out by the general manager in various occasions.

✓ *Perform positive adaptive behaviors*

The second principle of resilience defined by Mallak (1998) as the ability of organizations to respond in a positive and adaptive fashion to a wide range of events. Organizations that perform adaptive behaviors take an active approach toward solving problem in workplace. This principle is connected to the first principle of resilience. In fact, organizations that perceive constructively experiences adopt positive adaptive behaviors. This resilience principle focuses on adaptive as opposed to programmed responses. Over the years, the company has shown to have a mentality oriented towards adaptation and change. One of the respondents specified that:

> *In recent years we are transformed into a modern company. I mean a company with market orientation. A company that has active processes instead of an administrative bandwagon. Along with this, there was a staff rejuvenation process, a very high training process on which a lot was invested. Also, because here there were patchy skills, that is, there were those who knew a lot and those who knew very little. And those who knew did not share information with others.*

He added:

> *The offices were deserted and everything was improvised. Instead, we started hiring young people, people who have a certain kind of vision and preparation. Now there has been a significant change in the management approach.*

The company in recent years reviewed all the processes in order to improve their functioning and to reduce inefficiencies in time and costs. The revision of the process has allowed the company to adapt and direct the behavior of the operators in order to obtain the necessary information. The quality manager said that:

> *We have transformed what were historical procedures not even described in formalized procedures; now we no longer use procedures related to the experience of those who managed but linked to an idea that was translated into a procedure regardless of who was responsible.*

In this process, the organizational chart was also revised and new employees were identified as responsible, most of them young or recently employed. This has allowed the company to obtain a logic of responsibility that aimed to direct the behavior of the various operators. The choice of new younger managers has made it possible to overcome some resistance, in particular among some older ones, as one manager of strategical investment emphasized.

Even at an individual level, the operators try to be adaptable to any problem and to find unplanned solutions. In the event that, for some reason, the material is missing, the maintenance operators do not block the work and try to recover the material from some other abandoned plant. Sometimes even at the last minute the operators change the work program because there is a delay in the delivery of the materials by the suppliers trying to do another activity that does not require the material or for which the material is available. The change in the type of activity was made, for example, even during rainy days. There are activities that must be done outside the company and others that can be done within the company, in the case of bad weather conditions even at the last moment in order to not block and slow down the work operators change program and try to carry out activities within the company. This adaptive behavior to the conditions of the context was also noticed during the working meetings. In one case, for example, of a design error, the project manager identified the error and also highlighted the solution without addressing to external designers. These adaptive behaviors on an individual level have been made in particular to avoid blocking work and respecting company objectives.

✓ *Ensure adequate external resources*

According to Mallak (1998) external resource encompasses resources such as advice, information, finances, emotional support, and practical help. In these years, the company has succeeded in guaranteeing the necessary resources for both the maintenance and the renewal of the railway network. With the cuts in resources, the company was able to convert part of the maintenance staff to carry out the renovation of the railway network. This activity was previously outsourced. The training and the accompaniment with external consultants has allowed the company to have the necessary skills in order to reconvert the staff. This was also possible thanks to the acquisition of the know-how of other companies in the sector (through agreements) or the involvement of external consultants with extensive experience in the sector. This ability to attract and guarantee external resources in terms of advice, information, emotional support, and practical help was accompanied by the adoption of an accounting system that allowed the production of the data necessary for the purpose of formulating to formulate advice, information, emotional support, and practical help. Accounting contributed to formulate advice, information, emotional support, and practical help, and therefore to reduce the uncertainties, i.e., to undertake actions that are no longer inspired but calculated and based on various simulations. In this sense, accounting played a role of "ammunition machine" providing the general manager or other manager with the information needed to support a specific decision.

✓ *Expand decision-making boundaries*

Another principle of resilience is the expansion of decision-making boundaries. According to Mallak (1998) resilient individuals need the ability and authority to make decisions on the spot in a variety of situations. The formalized procedures introduced by the company helped to establish the responsibilities and the flow of authorization with respect to the various activities. In addition, the company monitors the results monthly and every month at least two meetings are held to discuss operational problems. This made it possible to have a shared timeline decision-making system and to expand decision-making boundaries. The presence of a director with managerial and technical skills helped to strengthen decision-making autonomy. As underlined by one manager of strategical investment:

> *The general director is always present, technically prepared, more oriented to share decisions making and to give the necessary autonomy; He is good manager for this company and other large private companies because he has a mentality that he evaluates you based on what you are doing. Precisely for this reason, I find myself very well.*

Furthermore, the sharing of operational objectives, as well as the almost daily support with external consultants, have allowed all operators to be involved and participate in the various decisions making. Their involvement in the decision-making phase allowed the company to make them aware regarding the economic impact. Their involvement has also allowed the reduction of uncertainties with respect to the consequences of the choices made. One of the operators said that:

> *In my opinion, the new mode of objectives definition has brought much more order. Before, I was not aware of the company objectives. It was a negotiation among executives.*

He added:

> *Now we are strictly interconnected . . . everyone is involved...sharing goals was one of the causes that accelerate this process.*

This favored a sense of inter-dependency, dialogue, mutual understanding, and trust. In fact, involving people along the decision-making process allows to sensitize people and to accelerate the decision-making process.

✓ *Practice bricolage*

Bricolage is the ability to fashion a solution on the spot using materials on hand, it refers to the practice of creating order out of whatever materials were available at the time (Mallak 1998). From the analysis of the procedures followed by the operators it has been noticed that in some cases, some operators introduced different tools or procedures that aimed to control or to improve activities. As pointed out by one of the maintenance operators:

> *Instead of using rag and brush, losing time, I sprayed food degreaser and in the meantime I waited I could do voltage measurement tests. This avoided waste of time and then cleaning the system improves the functionality of the system, other teams did not use food degreaser and therefore lost more time.*

The objective of these tools, according to the operators, is to find immediate and inexpensive solutions to some operational or control aspects and to respect the times of work. Another aspect related to this aspect of resilience is for example the reuse and recycling of maintenance material. Once a plant is decommissioned, the operators do not throw away all the material, all the material that can be recovered is put into storage. In these years, for example, with the cutting of resources and personnel reconversion, the company has continued to recover and use the material from the decommissioned plants, this on the one hand has guaranteed financial resources for company and on the one hand has allowed the operation as that material has long delivery times. As pointed out, the objective of using the material recovered from abandoned plants is to speed up the work, solving the practical problems, and therefore to achieve the objectives shared at the company level.

✓ *Develop tolerance for uncertainty*

Uncertainty refers to situations where the amount of information we need is greater than the information we have, and resilient individual has the capacity to make good decisions under these conditions (Mallak 1998). Until a few years ago, the company had public transfers that guaranteed both the ordinary maintenance and the modernization of the network. The uncertainties were almost null because the necessary resources were guaranteed by public transfers. With the cutting of financial resources, the company has started living in an uncertain environment. Currently, most of the revenue (37%) is no longer guaranteed by transfers but depends on the work of the staff. The director says:

> *It is a challenge that we must face, our survival depends on what we can do; Certainly we do not have the tools to evaluate what will happen in the future but better to prove that stay still waiting the end.*

At the company level, from management to operators, the uncertainty of the future has become the order of the day. The planning and monitoring of the results for the purpose of survival has entered everyone's mindset. The use of accounting, as highlighted by various respondents, helped to reduce and manage uncertainty. To verify the sustainability of the new business idea, the PSO developed business plan. To do that, the company carried out an in-depth analysis with respect to the needs and availability of the skills able to accompany the company. This allowed the company to understand the availability of the administrative and technical skills needed to carry out the work and therefore the feasibility of the operation. In this phase, the business plan contributed to rationalize the process and thus reduce uncertainty. Considering the high level of uncertainty, the decision was made in some way by inspiration. Various simulations made allowed first of all to rationalize the process and secondly to legitimize the choice towards the public partner. Then, the business plan gave an appearance of feasibility of the new activity. Referring to the business plan, the management control manager stated that:

> *We knew that objectives of the business plan were challenging . . . The business plan made us believe we could make it.*

At the same time, the budget role and function changed significantly, from a traditional public sector approach, where the budget is commitment based, with a high degree of certainty in particular for the revenues, to a managerial approach with uncertainty of revenues. As such, thee accounting system, on the one hand, has rationalized and legitimized the process of resilience.

✓ *Build virtual role systems*

The virtual role system refers to an advanced form of work team relationships. Resilient individual can operate in a virtual role systems environment; when many people in the organization are effective in this systems, the organization becomes more resilient. The virtual role system provides a work environment where the team can continue in the absence of one or more members (Mallak 1998). The review of the various processes and of the formalization of the various procedures allowed the creation of a coordinated system among the various company structures. The introduction of a new ERP helped to unify both the procedures and the language between the operators. The review of the organization chart, the assignment of responsibilities, and the implementation of a performance-based incentive system contributed to involve all stakeholders in the change process. Moreover, the training, also technical for the administrators, allowed the staff to have a broader vision with respect to the company. Having written and formalized the various activities and roles and having introduced ERP that gives a cross-sectional view of the whole company, the roles have become interchangeable and consequently the team can continue in the absence of one or more members. The various activities were mapped, described, and standard working times were defined. In the mapping of the various operational activities, the company has given an overall view to the various operators regarding the work to be done helping to complete the virtual system. This virtual system was also concretized at the operational level, getting the operators to alternate with respect to the activities. With this logic, the company has created workers with a broad vision and has been able to create a virtual system that has served in subsequent years.

5. Discussion

Barbera et al. (2017) highlighted that organizations could respond by self-regulation, by constrained or reactive adaptation, or by contented or powerless fatalism. Renn (2012) reported that organizations could deal with three different approaches to govern ris: Technocratic risk management, decisionistic risk management, and transparent risk governance. In our case, the perception of risk has become an opportunity for change. In the face of these budget cuts, management began refocusing the company's mission from infrastructure maintenance to providing services with a market-based logic. In the case analyzed, the management adopted the transparent risk approach (Renn 2012), and in doing that, they adopted the reactive approach of resilience (Barbera et al. 2017) by generating benefits from the situation.

The case portrays a vivid picture of the role of accounting in supporting the development of resilient capacities. In particular, the adoption of tools like a business plan, budget, and other accounting technologies creates a particular realm of calculation with which judgements and decisions were made. In this sense, accounting rationalized and legitimized the process of resilience. Accounting contributed to formulate advice, information, emotional support, and practical help and therefore to reduce the uncertainties. Accordingly, accounting has given the company the basis for decisions ((Hopwood 1990, 1992; Miller and O'Leary 2019). Decision making and judgement requires ability to interpret the context.

Accounting practices offered a way of making things real, constructing seemingly objective and neutral records for abstract and complex (Hopwood 1990, 1992; Miller and O'Leary 2019), and this helped the case-organization to develop resilient capacities. It also helped to encourage interaction between various company areas and allowed to create a common language and favor the interaction between participants who do not usually operate in the same context (Lai et al. 2014). This is because accounting can influence perceptions, change language, and infuse dialogue (Hopwood 1990). In fact, the way in which the objectives were defined and checked made it possible to create communication channels between various operators. This certainly helped create interconnections and common language between the various structures and contributed to make the company resilient.

Another role of accounting in the resilience process has concerned the reduction of uncertainty to developing tolerance for uncertainty. The information perceived to be important by decision-makers are related to perceived environmental uncertainty (Gordon and Narayanan 1984). Accounting helped the management to reduce uncertainty and to improve decision making (Gul and Chia 1994). Our study shows how the development of resilient capacity goes through the management risk awareness supported by the existing and newly designed accounting systems. As observed above, in the case-organization the business plan, as an example of newly developed accounting tool, supported the formalization of risks and the management of the opportunities of the changing environment to anticipate and cope with the external shock.

6. Conclusions

In this paper, we investigate the process of building resilience capacities. We highlighted how risk perception could be the trigger for this process, and the role of accounting. Organizations facing financial crisis could respond differently. In particular, drawing on Mallak (1998), we investigated the linkages between risk perceptions, accounting, and the emergence of resilience capacities.

The paper contributions are two-fold. Firstly, our analysis reveals that Mallak's framework is adequate in order to understand how resilience is built, and can therefore represent a point of reference for future studies. Second, the paper is one of the first attempting to show how accounting is implicated in making public sector organization resilient and its relationship with the risk awareness. During financial shock, public sector organizations face high level of uncertainties, both in terms of objectives to be pursued and the means through which achieve them. As the case shows, accounting in such contexts can support the organizational resilience by providing the legitimation of choices, thanks to the rational representation of uncertain objects. Furthermore, accounting support adaptive behavior by providing knowledge and ready to use answers. During financial shocks, the role of accounting need to be understood in a dynamic perspective. Whereas at the beginning of the financial shock, accounting plays a role as rationalization machine, during the implementation process of the solutions to cope or exploit the crisis, the accounting technologies' role change, supporting the organization more as an answer machine (Burchell et al. 1980). This conclusion is supportive of Mouritsen and Kreiner (2016) view that accounting support decisions not by solidifying decision but with a continuous process of decision "from causality to effectuation and from solutions to generation of alternatives" (p. 21). As such, a resilient organization is supported by accounting assuming multifaceted role not for the search of a new equilibrium but to continuously adapt to the uncertain internal and external environment (Barbera et al. 2020).

The results have both policy and managerial implications. From a policy perspectives, austerity and cut-back policy measures need to be planned supporting public sector organizations to develop resilient capacities. As we have shown in the case study, budget cut-back triggered organizational innovation and change, mobilizing the resilient capacities available. From a managerial perspective, public managers need to pay attention to the way accounting technologies are used throughout the organization. Accounting can play several roles, but that will be dependent on the way accounting is designed and developed. Future studies may be interested in studying the role of accounting versus other dimensions of resilience and understanding how accounting can affect performance in a resilience process.

This study is not without limitations. First and foremost, we based our analysis on one in-depth case study, therefore the results cannot be generalized in different contexts. Second, we observed the case analysis during a discrete point of time. As the development of resilient capacities takes time, a more longitudinal study would help in understanding more clearly the dynamics of resilience. Future studies can address the above limitations by extending the depth and length of analysis to corroborate, extend and amend our findings.

Author Contributions: Conceptualization, E.B. and M.T.; methodology, E.B.; formal analysis, M.T.; investigation, E.B. and M.T.; resources, E.B.; writing—original draft preparation, M.T.; writing—review and editing, E.B.; supervision, E.B.; project administration, E.B.; funding acquisition, E.B. The paper's sections can be attributed as follows: Sections 1 and 3–5 to E.B., Sections 2–4 and 6 to M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

Andersson, Thomas, Mikael Cäker, Stefan Tengblad, and Mikael Wickelgren. 2019. Building Traits for Organizational Resilience through Balancing Organizational Structures. *Scandinavian Journal of Management* 35: 36–45. [CrossRef]

De Aquino, André C. B., and Ricardo Lopes Cardoso. 2017. Financial Resilience in Brazilian Municipalities. In *Governmental Financial Resilience*. Bingley: Emerald Publishing Limited, pp. 53–71. [CrossRef]

Barbera, Carmela, Enrico Guarini, and Ileana Steccolini. 2020. How Do Governments Cope with Austerity? The Roles of Accounting in Shaping Governmental Financial Resilience. *Accounting, Auditing & Accountability Journal*. ahead-of-print. [CrossRef]

Barbera, Carmela, Martin Jones, Sanja Korac, Iris Saliterer, and Ileana Steccolini. 2017. Governmental Financial Resilience under Austerity in Austria, England and Italy: How Do Local Governments Cope with Financial Shocks? *Public Administration* 95: 670–97. [CrossRef]

Biggs, Duan. 2011. Understanding Resilience in a Vulnerable Industry: The Case of Reef Tourism in Australia. *Ecology and Society* 16: art30. [CrossRef]

Black, Julia. 2005. The Emergence of Risk-Based Regulation and the New Public Risk Management in the United Kingdom. *Public Law*, 512–48.

Boin, Arjen, and Martin Lodge. 2016. Designing Resilient Institutions for Transboundary Crisis Management: A Time for Public Administration. *Public Administration* 94: 289–98. [CrossRef]

Du Boys, Celine. 2017. Resilience Patterns of French Municipalities: A Case Study. In *Governmental Financial Resilience: International Perspectives on How Local Governments Face Austerity*. Bingley: Emerald Publishing Limited, pp. 93–113. [CrossRef]

Bracci, Enrico, Christopher Humphrey, Jodie Moll, and Ileana Steccolini. 2015. Public Sector Accounting, Accountability and Austerity: More than Balancing the Books? *Accounting, Auditing & Accountability Journal* 28: 878–908. [CrossRef]

Brown, Louise, and Stephen P. Osborne. 2013. Risk and Innovation. *Public Management Review* 15: 186–208. [CrossRef]

Burchell, Stuart, Colin Clubb, Anthony Hopwood, John Hughes, and Janine Nahapiet. 1980. The Roles of Accounting in Organizations and Society. *Accounting, Organizations and Society* 5: 5–27. [CrossRef]

Carden, Lila L., Tiffany Maldonado, and Raphael O. Boyd. 2018. Organizational Resilience: A Look at McDonald's in the Fast Food Industry. *Organizational Dynamics* 47: 25–31. [CrossRef]

Carey, James W., Mark Morgan, and Margaret J Oxtoby. 1996. Intercoder Agreement in Analysis of Responses to Open-Ended Interview Questions: Examples from Tuberculosis Research. *CAM Journal* 8: 1–5. [CrossRef]

Chong, Vincent K., and Kar Ming Chong. 1997. Strategic Choices, Environmental Uncertainty and SBU Performance: A Note on the Intervening Role of Management Accounting Systems. *Accounting and Business Research* 27: 268–76. [CrossRef]

Cohen, Sandra, and Nikolaos Hlepas. 2017. Financial Resilience of Greek Local Governments. In *Governmental Financial Resilience*. Bingley: Emerald Publishing Limited, pp. 135–52. [CrossRef]

Coutu, Diane L. 2002. The Anxiety of Learning. *IEEE Engineering Management Review*. [CrossRef]

Dahles, Heidi, and Titi Prabawa Susilowati. 2015. Business Resilience in Times of Growth and Crisis. *Annals of Tourism Research* 51: 34–50. [CrossRef]

Davoudi, Simin, Keith Shaw, L. Jamila Haider, Allyson E. Quinlan, Garry D. Peterson, Cathy Wilkinson, Hartmut Fünfgeld, Darryn McEvoy, Libby Porter, and Simin Davoudi. 2012. Resilience: A Bridging Concept or a Dead End? 'Reframing' Resilience: Challenges for Planning Theory and Practice Interacting Traps: Resilience Assessment of a Pasture Management System in Northern Afghanistan Urban Resilience: What Does It Mean in Planni. *Planning Theory & Practice* 13: 299–333. [CrossRef]

Fraser, Mark W, Jack M Richman, and Maeda J Galinsky. 1999. Risk, Protection, and Resilience: Toward a Conceptual Framework for Social W. *Social Work Research* 23: 131–43. [CrossRef]

Gittell, Jody Hoffer, Kim Cameron, Sandy Lim, and Victor Rivas. 2006. Relationships, Layoffs, and Organizational Resilience. *The Journal of Applied Behavioral Science* 42: 300–329. [CrossRef]

Gordon, Lawrence A., and V.K. Narayanan. 1984. Management Accounting Systems, Perceived Environmental Uncertainty and Organization Structure: An Empirical Investigation. *Accounting, Organizations and Society* 9: 33–47. [CrossRef]

Gul, Ferdinand A., and Yew Ming Chia. 1994. The Effects of Management Accounting Systems, Perceived Environmental Uncertainty and Decentralization on Managerial Performance: A Test of Three-Way Interaction. *Accounting, Organizations and Society* 19: 413–26. [CrossRef]

Hamel, Gary, and Liisa Välikangas. 2003. The Quest for Resilience. *Harvard Business Review* 81: 52–63, 131.

Hines, Ruth D. 1988. Financial Accounting: In Communicating Reality, We Construct Reality. *Accounting, Organizations and Society* 13: 251–61. [CrossRef]

Hopwood, Anthony G. 1992. Accounting Calculation and the Shifting Sphere of the Economic. *European Accounting Review* 1: 125–43. [CrossRef]

Hopwood, Anthony G. 1990. Accounting and Organisation Change. *Accounting, Auditing & Accountability Journal* 3: 09513579010145073. [CrossRef]

Jones, Martin. 2017. English Resilience in the Face of Austerity. In *Governmental Financial Resilience*. Bingley: Emerald Publishing Limited, pp. 73–91. [CrossRef]

Lai, Alessandro, Giulia Leoni, and Riccardo Stacchezzini. 2014. The Socializing Effects of Accounting in Flood Recovery. *Critical Perspectives on Accounting* 25: 579–603. [CrossRef]

Lee, Amy V., John Vargo, and Erica Seville. 2013. Developing a Tool to Measure and Compare Organizations' Resilience. *Natural Hazards Review* 14: 29–41. [CrossRef]

Lengnick-Hall, Cynthia A., Tammy E. Beck, and Mark L. Lengnick-Hall. 2011. Developing a Capacity for Organizational Resilience through Strategic Human Resource Management. *Human Resource Management Review* 21: 243–55. [CrossRef]

Leslie, Keith, and Adam Canwell. 2010. Leadership at All Levels: Leading Public Sector Organisations in an Age of Austerity The Leadership Challenge for the Public Sector. *European Management Journal* 28: 297–305. [CrossRef]

Linnenluecke, Martina K. 2017. Resilience in Business and Management Research: A Review of Influential Publications and a Research Agenda. *International Journal of Management Reviews* 19: 4–30. [CrossRef]

Mallak, Larry. 1998. Putting Organizational Resilience to Work. *Industrial Management* 40: 8.

Meyer, A. D. 1982. Adapting to Environmental Jolts. *Administrative Science Quarterly* 27: 515–37. [CrossRef]

Miles, M., A. M. Huberman, and J. Saldana. 2014. *Qualatative Data Analysis*, 3rd ed. Thousand Oaks: SAGE Publications, Ltd.

Miller, Peter, and Ted O'Leary. 2019. Accounting and the Construction of the Governable Person. *Management Control Theory* 411–41. [CrossRef]

Mouritsen, Jan, and Kristian Kreiner. 2016. Accounting, Decisions and Promises. *Accounting, Organizations and Society* 49: 21–31. [CrossRef]

Nahapiet, Janine. 1988. The Rhetoric and Reality of an Accounting Change: A Study of Resource Allocation. *Accounting, Organizations and Society* 13: 333–58. [CrossRef]

Power, Michael. 2007. *Organized Uncertainty: Designing a World of Risk Management*. New York: Oxford University Press.

Powley, Edward H. 2009. Reclaiming Resilience and Safety: Resilience Activation in the Critical Period of Crisis. Edited by Nick Turner and Garry C. Gray. *Human Relations* 62: 1289–1326. [CrossRef]

Renn, Ortwin. 2012. *Risk Governance: Coping with Uncertainty in a Complex World*. London: Routledge. [CrossRef]

Saldaña, Johnny. 2013. *The Coding Manual for Qualitative Researchers*, 2nd ed. Thousand Oaks: SAGE Publications Inc. [CrossRef]

Sutcliffe, Kathleen M, and Timothy J Vogus. 2003. Organizing for Resilience. In *Positive Organizational Scholarship: Foundations of a New Discipline*. San Francisco: Berrett-Koehler, pp. 94–110.

Tallaki, Mouhcine, and Enrico Bracci. 2020. Resilience Capacities and Management Control Systems in Public Sector Organisations. *Journal of Accounting and Organizational Change*. Forthcoming. [CrossRef]

Vaivio, Juhani. 2008. Qualitative Management Accounting Research: Rationale, Pitfalls and Potential. *Qualitative Research in Accounting & Management* 5: 64–86. [CrossRef]

Journal of
Risk and Financial Management

MDPI

Article

CoCDaR and mCoCDaR: New Approach for Measurement of Systemic Risk Contributions

Rui Ding * and Stan Uryasev

Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA; stanislav.uryasev@stonybrook.edu
* Correspondence: rui.ding.1@stonybrook.edu

Received: 2 October 2020; Accepted: 1 November 2020; Published: 3 November 2020

Abstract: Systemic risk is the risk that the distress of one or more institutions trigger a collapse of the entire financial system. We extend CoVaR (value-at-risk conditioned on an institution) and CoCVaR (conditional value-at-risk conditioned on an institution) systemic risk contribution measures and propose a new CoCDaR (conditional drawdown-at-risk conditioned on an institution) measure based on drawdowns. This new measure accounts for consecutive negative returns of a security, while CoVaR and CoCVaR combine together negative returns from different time periods. For instance, ten 2% consecutive losses resulting in 20% drawdown will be noticed by CoCDaR, while CoVaR and CoCVaR are not sensitive to relatively small one period losses. The proposed measure provides insights for systemic risks under extreme stresses related to drawdowns. CoCDaR and its multivariate version, mCoCDaR, estimate an impact on big cumulative losses of the entire financial system caused by an individual firm's distress. It can be used for ranking individual systemic risk contributions of financial institutions (banks). CoCDaR and mCoCDaR are computed with CVaR regression of drawdowns. Moreover, mCoCDaR can be used to estimate drawdowns of a security as a function of some other factors. For instance, we show how to perform fund drawdown style classification depending on drawdowns of indices. Case study results, data, and codes are posted on the web.

Keywords: systemic risk; conditional value-at-risk; CVaR; CVaR regression; drawdown; conditional drawdown-at-risk; fund style classification

---

1. Introduction

Systemic risk is the risk that the distress of one or more institutions triggers a collapse of the entire financial system. The CoVaR measure for systemic risk contributions was first proposed by Adrian and Brunnermeier (2008). This measure is the value-at-risk (VaR) of the financial system conditional on an institution (bank) being in financial distress. The systemic risk contribution of an institution is defined as a difference of VaR conditioning on the institution being under distress and being in its normal state. Huang and Uryasev (2017) replaced VaR by conditional value-at-risk (CVaR) and proposed the CoCVaR measure. CVaR has superior mathematical properties as compared to VaR; see, for instance, Rockafellar and Uryasev (2002). CVaR takes into account losses in the distribution tail, while VaR is not sensitive to outcomes in the tail.

Similar to Huang and Uryasev (2017), this paper is based on CVaR, but returns are replaced by drawdowns. The relevant risk measure is called conditional drawdown-at-risk (CDaR). By applying the CoCVaR approach to drawdowns, we defined CoCDaR. Therefore, CoCDaR is CDaR of the financial system conditioned on an institution in distress measured by drawdown. The intuition behind CDaR instead of VaR or CVaR is that these two measures do not take into account consecutive losses. As a result, small consecutive losses resulting in a large cumulative loss are not picked up by VaR or CVaR. Drawdown, which is capturing cumulative losses, is popular in active portfolio management. The idea behind CoCDaR is that large drawdowns of financial institutions have a strong effect on the system as a whole. Hence, by conditioning on large drawdowns of institutions we can analyze systemic risk contributions (compared to effect of one-period negative returns).

We further extended CoCDaR with multiple regression framework and developed so-called mCoCDaR. This measure allows for multiple institutions being in distress, while CoVaR, CoCVaR and CoCDaR assume that only one institution is in distress and others are in normal states. Similar to mCoCDaR, we considered a multiple regression version of CoCVaR, called mCoCVaR. Therefore, mCoCVaR and mCoCDaR account for multiple marginal risk contributions of institutions and are well-defined Shapley values. This approach is motivated by the idea of identifying a risk contribution of each institution that is independent of contributions of other institutions. The estimation of CoCDaR and mCoCDaR was performed with CVaR regression developed in Rockafellar et al. (2014) and Golodnikov et al. (2019). The CVaR regression in CoCDaR uses drawdowns, while CoCVaR uses returns.

The mCoCDaR framework was also illustrated with fund style classification by using drawdowns instead of returns. This approach extends Bassett and Chen (2001), which used quantile regressions of fund returns depending on returns of indices. In addition, we have considered portfolio optimization formulations with CoCVaR and CoCDaR objectives and risk constraints.

CoCDaR and mCoCDaR approaches were demonstrated with a case study for the 10 largest USA banks. Furthermore, we have performed drawdown style classification of the Magellan fund using four stock indices. CVaR regression was implemented with Portfolio Safeguard (PSG) developed by AORDA (http://aorda.com). Case studies results and codes are posted on the web for verification purposes.

2. Methodology

2.1. Drawdown Definition

Suppose $r_1, \dots, r_T$ are the rates of return of a risky instrument coming from a distribution of return random variable X. Let ξ_t be the cumulative rate of return of the instrument for time t, which can be either uncompounded and defined by $\xi_t = \sum_{k=1}^{t} r_k$ or compounded and defined by $\xi_t = \prod_{k=1}^{t}(1 + r_k) - 1$. Further analysis in this section holds for either definition of the cumulative return, however, for the sake of tractability of optimization problems, ξ_t is defined as uncompounded cumulative rate of return.

The drawdown of the instrument at time t with τ-window is defined as follows (see Chekhlov et al. (2005); Zabarankin et al. (2014)),

$$y_t = \max_{t_\tau \le k \le t} \xi_k - \xi_t, \qquad t_\tau = \max\{t - \tau, 1\}, \qquad t = 1, \dots, T, \quad \tau = 1, \dots, T. \tag{1}$$

At time t, the drawdown is the loss of the instrument, since a peak of ξ_t that occurs within the τ-window $[t_\tau, t]$ ($t_\tau = 1$ for $t \le \tau$ and $t_\tau = t - \tau$ for $t > \tau$). If at time t, the cumulative rate of return ξ_t is the highest on $[t_\tau, t]$, then $y_t = 0$. The drawdown is always nonnegative and is often referred to as underwater curve. It is zero for all time moments only if returns are nonnegative for all period. See Figure 1 for the illustration of the drawdown definition (the figure is borrowed from Zabarankin et al. (2014)).

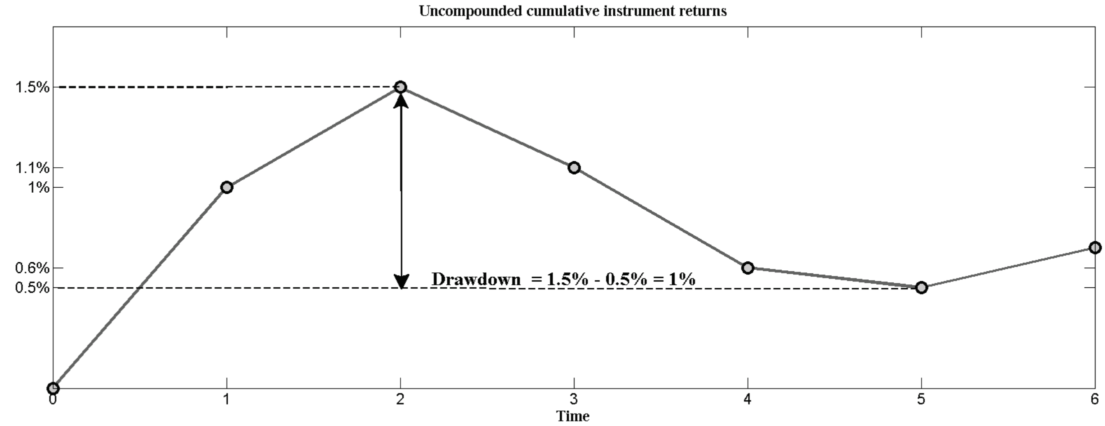


Figure 1. Drawdown example: the solid line is the uncompounded cumulative rate of return, which at time t is the sum of rates of return over periods $1, \ldots, t$. Here, $\tau = 6$. For $t = 5$, $\xi_5 = 0.5\%$, whereas the maximum of ξ_t over time moments preceding $t = 5$ occurs at $t = 2$ with $\xi_2 = 1.5\%$. Consequently, $y_5 = 1.5\% - 0.5\% = 1\%$. The instrument maximum drawdown over time period $[0, 6]$ occurs at $t = 5$.

2.2. CoCDaR Definitions

Conditional value-at-risk (CVaR) of a random variable X (see Rockafellar and Uryasev (2000, 2002)), can be defined as follows,

$$CVaR_\alpha(X) = \min_C \{C + \frac{1}{1-\alpha} E[(X-C)^+]\} ,$$

where $A^+ = \max\{0, A\}$. The α-conditional drawdown is an expectation over the worst $1-\alpha$ drawdowns occurring in the considered horizon. We can look at $\{y_t\}_{1 \le t \le T}$ as a nonlinear transformation of observations from the random return variable X and denote the random variable for drawdowns by Y. The conditional drawdown-at-risk (CDaR) for X is defined as CVaR of Y:

$$CDaR_\alpha(X) = CVaR_\alpha(Y) .$$

Let X^{sys} denote return of a financial system and let returns of financial institutions $i = 1, \ldots, I$ be denoted as X^i. Given a sample path of data $\{x_t^{sys}, x_t^1, \ldots, x_t^I\}_{1 \le t \le T}$, we can obtain the drawdown observations for the financial system as well as all the institutions and denote them by $\{y_t^{sys}, d_t^1, \ldots, d_t^I\}_{1 \le t \le T}$. Let $Y^{sys}, D^1, ..., D^I$ denote random variables associated with these observations.

Similar to Huang and Uryasev (2017), we can define CoCDaR as:

$$CoCDaR_\alpha^{sys|i} = CDaR_\alpha(X^{sys}|X^i, M_1, ..., M_n) = CVaR_\alpha(Y^{sys}|D^i, M_1, ..., M_n) .$$

Here $M_1, ..., M_n$ are state factor variables which we define in the next section. They are the lagged system variables used in Huang and Uryasev (2017), but transformed to provide more explanatory powers in drawdown regression (considered later on).

$CoCDaR_\alpha^{sys|i}$ gives a measure of CDaR of a system (index) conditioning on the drawdown level of an individual institution (stock) i, along with some other state variables. By using drawdown instead of return, we are looking specifically at the impact of individual institution drawdown to the entire financial system drawdown as a measure of systemic risk contribution (that takes into account consecutive distress periods). This intuition will be further developed in Section 2.6.

2.3. State Variables

Lagged state variables $M_{1,t-1}, ..., M_{n,t-1}$ used in the regression in the following section were introduced by Adrian and Brunnermeier (2008):

(1) VIX = The Chicago Board Options Exchange Volatility Index;

(2) Liquidity Spread = A short-term liquidity spread, defined as the difference between the three-month repurchase agreement rate and the three-month treasury bill rate;

(3) 3M Treasury Change = The change in the three-month T-bill rate;

(4) Term Spread Change = The change in the slope of the yield curve, measured by the yield spread between the ten-year treasury rate and the three-month bill rate;

(5) Credit Spread Change = The change in the credit spread between Baa-rated bonds and the treasury rate;

(6) Equity Returns = The equity market return from S&P 500 Index;

(7) Real Estate Excess Return = The real estate sector return in excess of the market return.

2.4. Estimation of CoCDaR

Consider the following regression similar to Adrian and Brunnermeier (2008) and Huang and Uryasev (2017),

$$Y_t^{sys} \sim \beta_0 + \beta_1 D_t^i + \omega_1 M_{t-1,1} + ... + \omega_n M_{t-1,n} \ .$$

We define the residual random variable as:

$$L = Y^{sys} - (\beta_0 + \beta_1 D^i + \omega_1 M_1 + ... + \omega_n M_n) \ .$$

This regression problem uses a single institution's drawdown and lagged state variables as factors to model the drawdown of the financial system. We have T observations of the system drawdowns, drawdowns of institution i, and the state factors. We next perform a CVaR regression with the above model and find CVaR of the system's drawdown conditioned on drawdowns of institution i. Here, the state factors are cumulative changes of each fundamental factor in the period of the current drawdown of the financial system. In particular, VIX and liquidity spread are given in numbers so we calculate the time lagged difference of them in the current period of system drawdown. The other state factors are given in percentage changes, so we calculate their cumulative changes in the current period of system drawdown.

For each time step t, we consider the system cumulative returns ξ_t^{sys} and find the historic peak time (used in drawdown definition), denoted by

$$\nu(t) = \underset{t_\tau \le s \le t}{\arg\max} \ \xi_s^{sys} \ . \tag{2}$$

Let the original state variable values (numeric or cumulative changes in percentage) be denoted by m_t. The transformed state variables for the CoCDaR regression are hence defined for each $j = 1, \ldots, n$:

$$M_{t-1,j} = m_{t-1,j} - m_{\nu(t),j} \ .$$

The estimate of the α-CVaR of Y^{sys} can be obtained by minimizing the CVaR (superquantile) error from Rockafellar et al. (2014):

$$\mathcal{E}_\alpha^{CVaR}(L) = \frac{1}{1-\alpha} \int_0^1 CVaR_\gamma^+(L)\, d\gamma - E[L] \ . \tag{3}$$

Golodnikov et al. (2019) proved that minimization of error (3) for CVaR regression can be reduced to the minimization of the Rockafellar error (convex and liner programming formulations are in Appendix A, Golodnikov et al. (2019)). The Rockafellar error belongs to the mixed quantile quadrangle, as defined by Rockafellar and Uryasev (2013). For given confidence levels $\alpha_k \in (0,1)$ and weights $\lambda_k > 0$, $k = 1, \ldots, K$ such that $\sum\limits_{k=1}^{K} \lambda_k$=1, the Rockafellar error equals:

$$\mathcal{E}^{ROC}(L) = \min_{C_1,\dots,C_K} \left\{ \sum_{k=1}^{K} \lambda_k \mathcal{E}^{KB}_{\alpha_k}(L - C_k) \mid \sum_{k=1}^{K} \lambda_k C_k = 0 \right\} , \quad (4)$$

where the rescaled Koenker–Bassett (KB) error equals:

$$\mathcal{E}^{KB}_{\alpha}(L) = E\left[\frac{\alpha}{1-\alpha} L^{+} + (-L)^{+} \right] . \quad (5)$$

Koenker and Bassett (1978) suggested estimating a conditional quantile by minimizing error (5). Since CVaR is an integral of quantile (VaR), then it is not surprising that CVaR can be estimated with Rockafellar error (4) which is a weighted average of KB-errors. The Rockafellar error is quite a complicated function: it is a minimum of a convex nonsmooth function with respect to variables $C_1, \dots, C_K$ with a linear constraint. However, since this error is a convex piece-wise linear function, it can be minimized very efficiently; see for instance results of numerical experiments in Golodnikov et al. (2019). The resulting coefficients will provide an estimate of the α-CVaR of the dependent variable conditioned on the independent variables.

Denote by $\hat{\beta}^{\alpha}_0, \hat{\beta}^{\alpha}_1, \hat{\omega}^{\alpha}_1, ..., \hat{\omega}^{\alpha}_n$ the regression coefficients obtained by minimizing the Rockafellar error (4). CoCVaR of the system's drawdown, which is CoCDaR of the system, is estimated by:

$$CoCDaR^{sys}_{t,\alpha} = \hat{\beta}^{\alpha}_0 + \hat{\beta}^{\alpha}_1 D^i_t + \hat{\omega}^{\alpha}_1 M_{t-1,1} + ... + \hat{\omega}^{\alpha}_n M_{t-1,n} .$$

This regression estimation is done for every institution, $i = 1, \dots, I$.

2.5. Institutional Drawdown-at-Risk

To calculate system CoCDaR at some risk level conditioned on institution i being in drawdown distress, we need to set an institutional distress level D^i_t.

α-value-at-risk (VaR), which is also α-quantile, of a random loss variable L is defined as:

$$VaR_{\alpha}(L) = \inf\{x : F_L(x) \geq \alpha\} .$$

We define α-drawdown-at-risk (α-DaR) of an institution i as the α-quantile (VaR) of the drawdown loss random variable D^i corresponding to its return random variable X^i, where $\alpha \in [0, 1]$,

$$DaR_{\alpha}(X^i) = VaR_{\alpha}(D^i) .$$

Similar to Huang and Uryasev (2017), we can use quantile regression for estimation of α-DaR:

$$D^i_t \sim \gamma^i_0 + \gamma^i_1 M^i_{t-1,1} + ... + \gamma^i_n M^i_{t-1,n} .$$

Here, the state factors $M^i_1, ..., M^i_n$ are defined differently compared to the CoCDaR regression. They are the same fundamental factor changes but calculated in the current period of each institution drawdown. Define $\nu_i(t) = argmax_{t_\tau \leq s \leq t}\, \xi^i_s$ for each institution $i = 1, \dots, I$, where ξ^i_t are the cumulative returns observations. The transformed state variables for the DaR regression are hence defined for each $j = 1, \dots, n$:

$$M^i_{t-1,j} = m_{t-1,j} - m_{\nu_i(t),j} .$$

Let the residual term be denoted as:

$$G^i = D^i - (\gamma^i_0 + \gamma^i_1 M^i_1 + ... + \gamma^i_n M^i_n) .$$

By minimizing KB-error, $\mathcal{E}^{KB}_{\alpha}(G^i)$, we find coefficients $\hat{\gamma}^i_0, ..., \hat{\gamma}^i_n$ and estimate the α-quantile of D^i_t:

$$DaR^i_{t,\alpha} = \hat{\gamma}^i_0 + \hat{\gamma}^i_1 M^i_{t-1,1} + ... + \hat{\gamma}^i_n M^i_{t-1,n} .$$

2.6. Sytemic Risk Contribution

We have defined the CoCDaR measure and suggested an estimation procedure with CVaR regression. Next we show how to use this measure for systemic risk contribution measurement. We follow definitions from Huang and Uryasev (2017) and define:

$$X_t^{sys} = 100 \ln \frac{I_t}{I_{t-1}} \ ,$$

as the system's return variable which is the log return of the index value, such as the Dow Jones Index. Similarly, the i-th financial institution log return X_t^i is defined as:

$$X_t^i = 100 \ln \frac{P_t^i}{P_{t-1}^i} \ ,$$

where P_t^i is the closing price of institution i at time t.

Using the definitions in previous sections, we get the drawdown observations Y_t^{sys}, D_t^i for the financial system and an institution i. We also have state factors, $M_{t-1,1}, ..., M_{t-1,n}$, for every time moment t is in the considered horizon.

We first perform the quantile regression in Section 2.5 to estimate $DaR_{t,\alpha'}^i$ for all t for two particular levels: $\alpha'_1 = 0.9$ and $\alpha'_2 = 0.5$. The level $\alpha'_1 = 0.9$ corresponds to the distress level of the institution in terms of its drawdown and $\alpha'_2 = 0.5$ corresponds to the median (normal) state of the institution.

Next we perform the CVaR regression from Section 2.4 and obtain an estimate of the α-CoCDaR of the financial system conditioned on the drawdown level of institution i and state factors. Here α is different from α' used in the previous quantile regression. For every time step t, we calculate:

$$CoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,\alpha'}^i} = \hat{\beta}_0^\alpha + \hat{\beta}_1^\alpha DaR_{t,\alpha'}^i + \hat{\omega}_1^\alpha M_{t-1,1} + ... + \hat{\omega}_n^\alpha M_{t-1,n} \ .$$

By choosing $\alpha'_1 = 0.9$ and $\alpha'_2 = 0.5$ for the DaR level for an individual institution and selecting a separate risk level α for system CoCDaR, we obtain:

$$\Delta CoCDaR_{t,\alpha}^{sys|i} = CoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.9}^i} - CoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.5}^i} \ .$$

This difference is defined as the systemic drawdown risk contribution of institution i to the financial system at the selected risk level α. More concretely, it calculates the difference in conditional drawdown-at-risk values of the financial system given that the drawdown level of institution i is at its distress level or its normal level as a measure of systemic drawdown risk contribution.

2.7. mCoCDaR Definition

Using the same set of state factors and extending the idea of CoCDaR as a measure of systemic risk contribution, we propose a more comprehensive measure called multiple-CoCDaR, which measures the conditional drawdown-at-risk of the financial system conditioned on the distress levels of all I institutions being considered. The idea is an extension of the CoCDaR approach defined above by combining it with a generalization of the multiple-CoVaR method defined in Bernardi et al. (2013) and Bernardi and Petrella (2014). In their paper, a similar approach was developed that defines conditional tail risk of a system/institution conditioned on the distress level of multiple institutions at the same time. A similar approach was also seen in Cao (2013). Different from their methods, our approach uses a simple multiple regression formulation. In the multiple regression framework, we can measure risk contribution of an institution by taking the difference between CoCDaR values of the system under different drawdown levels of that institution alone, while holding other institutions' drawdown values fixed at their normal levels. We define mCoCDaR as:

$$mCoCDaR_{\alpha}^{sys|1,...,I} = CDaR_{\alpha}(X^{sys}|X^1,...,X^I,M_1,...,M_n) = CVaR_{\alpha}(Y^{sys}|D^1,...,D^I,M_1,...,M_n) \ .$$

2.8. Estimation of mCoCDaR

Consider the following regression using the same set of state factors as in CoCDaR regression,

$$Y_t^{sys} \sim \beta_0 + \beta_1 D_t^1 + ... + \beta_I D_t^I + \omega_1 M_{t-1,1} + ... + \omega_n M_{t-1,n} \ ,$$

with residual:

$$L = Y^{sys} - (\beta_0 + \beta_1 D^1 + ... + \beta_I D^I + \omega_1 M_1 + ... + \omega_n M_n) \ .$$

This regression problem uses I institutions drawdowns and lagged state variables as factors to model the drawdown of the financial system. We have T observations for the system drawdown random variable, the I institutions' drawdown random variables, and the state factors' random variables. We next perform a CVaR regression of the above model to find CVaR of the system drawdown conditioned on all I institution drawdown. Denote by $\hat{\beta}_0^{\alpha}, \hat{\beta}_1^{\alpha}, ..., \hat{\beta}_I^{\alpha}, \hat{\omega}_1^{\alpha}, ..., \hat{\omega}_n^{\alpha}$ coefficients obtained by minimizing Rockafellar error (4). These coefficients allow one to compute the α-CDaR of the financial system conditioned on drawdowns of all the institutions and state factors.

The multiple-CoCVaR of the system drawdown, which is equivalent to the multiple-CoCDaR of the financial system, is estimated by:

$$mCoCDaR_{t,\alpha}^{sys} = \hat{\beta}_0^{\alpha} + \hat{\beta}_1^{\alpha} D_t^1 + ... + \hat{\beta}_I^{\alpha} D_t^I + \hat{\omega}_1^{\alpha} M_{t-1,1} + ... + \hat{\omega}_n^{\alpha} M_{t-1,n} \ .$$

This procedure applies one regression problem using all institutions' drawdown observations to obtain coefficient estimates. The institutional DaRs are calculated exactly the same way as in Section 2.5.

2.9. Sytemic Risk Contribution using mCoCDaR

We have defined mCoCDaR measure and the estimation procedure with CVaR regression. We use this measure for systemic risk contribution measurement, following the definitions in Section 2.4. The drawdown observations are denoted by $Y_t^{sys}, D_t^1, ..., D_t^I$ for the financial system and all I institutions respectively. We also have lagged state variables $M_{t-1,1}, ..., M_{t-1,n}$ for every time moment t.

We first perform the quantile regression in Section 2.5 to estimate $DaR_{t,\alpha'}^i$ for all t and for all i for two particular levels: $\alpha'_{i,1} = 0.9$ and $\alpha'_{i,2} = 0.5$. Level $\alpha'_{i,1} = 0.9$ corresponds to the distress level of the i-th institution in terms of its drawdowns and $\alpha'_{i,2} = 0.5$ corresponds to its median (normal).

Next we perform the CVaR regression from Section 2.8 and estimate the financial system's conditional drawdown-at-risk conditioned on the drawdown levels of all I institutions and state factors. For every time step t, we calculate,

$$mCoCDaR_{t,\alpha}^{sys|D_t^1 = DaR_{t,\alpha'_1}^1,...,D_t^I = DaR_{t,\alpha'_I}^I} = \hat{\beta}_0^{\alpha} + \hat{\beta}_1^{\alpha} DaR_{t,\alpha'_1}^1 + ... + \hat{\beta}_I^{\alpha} DaR_{t,\alpha'_I}^I + \hat{\omega}_1^{\alpha} M_{t-1,1} + ... + \hat{\omega}_n^{\alpha} M_{t-1,n} \ .$$

Now, to analyze the effect of a single institution i on the financial system, we compute the mCoCDaR values based on $\alpha'_{i,1} = 0.9$ and $\alpha'_{i,2} = 0.5$, while holding $\alpha'_{-i} = 0.5$ fixed where $-i$ means all the institutions other than i, and calculate the difference in mCoCDaR,

$$\Delta mCoCDaR_{t,\alpha}^{sys|i} = mCoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.9}^i, D_t^{-i} = DaR_{t,0.5}^{-i}} - mCoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.5}^i, D_t^{-i} = DaR_{t,0.5}^{-i}} \ .$$

This difference is the incremental/marginal systemic drawdown risk contribution of the distress of institution i to the financial system, while other institutions are at their normal states.

We can switch back to the original return observations instead of the drawdown observations and perform the regression procedure in Sections 2.7 and 2.8. This way we get another measure for systemic risk contribution which we call mCoCVaR. It measures the incremental/marginal conditional value-at-risk of the financial system's returns conditioned on one institution's return being in distress while all the other institutions are in their normal states.

2.10. Advantages of mCoCDaR and mCoCVaR

As we have seen in the previous section, the multiple version of the systemic risk conditional estimation provides a more general framework to analyze the effect on the financial system posed by a particular institution's distress, or perhaps multiple financial institutions' joint distress. It is based on the idea that during periods of financial instability, several institutions may experience financial distress at the same time, so their risk contributions can be highly correlated. Switching from the CoCVaR and CoCDaR to their multiple regression versions helps to mitigate these dependencies on risk contribution measures.

With mCoCDaR, we can measure the contribution to the financial system's conditional drawdown-at-risk conditioned on the drawdown levels of two institutions i, j as follows,

$$\Delta mCoCDaR_{t,\alpha}^{sys|i,j} = mCoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.9}^i, D_t^j = DaR_{t,0.9}^j, D_t^{-i,j} = DaR_{t,0.5}^{-i,j}} - mCoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.5}^i, D_t^j = DaR_{t,0.5}^j, D_t^{-i,j} = DaR_{t,0.5}^{-i,j}} .$$

There is a lot of flexibility on the risk levels to choose for this type of analysis, which means the DaR level for the two institutions in distress can be set differently. This approach considers the joint impact of two institutions without distinguishing their respective contributions, which is not included in the original framework without using multiple regression. The flexibility given by mCoCDaR, and similarly mCoCVaR, does not come at additional computation costs. In fact, by combining all institutions in one regression problem, we save computational time.

Another advantage of the multiple-CoCVaR and multiple-CoCDaR are their consistency as risk distribution measures. Bernardi et al. (2013) noticed that the original $\Delta CoVaR^{sys|i}$ is not a desirable risk distribution measure, because summing up $\Delta CoVaR^{sys|i}$ for all institutions i does not generally equal their overall effect on the system. This issue is addressed in Bernardi et al. (2013) and Bernardi and Petrella (2014) via the Shapley value, which transforms the calculated contribution using $\Delta Multiple - CoVaR$ to a Shapley value for each institution so that their contribution adds up to the joint contribution of all institutions together on the system. The Shapley value methodology was originally proposed to measure shared utility or cost among participants of a cooperative game.

We observe that the individual risk contribution calculated with $\Delta mCoCVaR$ or $\Delta mCoCDaR$ does not have this drawback. For instance, for mCoCDaR:

$$mCoCDaR_{t,\alpha}^{sys|D_t^1 = DaR_{t,\alpha'_1}^1, \ldots, D_t^I = DaR_{t,\alpha'_I}^I} = \hat{\beta}_0^\alpha + \hat{\beta}_1^\alpha DaR_{t,\alpha'_1}^1 + \ldots + \hat{\beta}_I^\alpha DaR_{t,\alpha'_I}^I + \hat{\omega}_1^\alpha M_{t-1,1} + \ldots + \hat{\omega}_n^\alpha M_{t-1,n} .$$

Once we have estimated the coefficients via CVaR regression, we can calculate the individual contribution of institution i entering stress level 0.9 as:

$$\begin{aligned}\Delta mCoCDaR_{t,\alpha}^{sys|i} &= mCoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.9}^i, D_t^{-i} = DaR_{t,0.5}^{-i}} - mCoCDaR_{t,\alpha}^{sys|D_t^i = DaR_{t,0.5}^i, D_t^{-i} = DaR_{t,0.5}^{-i}} \\ &= \hat{\beta}_i^\alpha (DaR_{t,0.9}^i - DaR_{t,0.5}^i) \equiv V_{sys}(i) .\end{aligned}$$

The total contribution of all financial institutions distress on the systemic risk is:

$$\Delta mCoCDaR_{t,\alpha}^{sys|1,\ldots,I}$$

$$= mCoCDaR_{t,\alpha}^{sys|D_t^1 = DaR_{t,0.9}^1, \ldots, D_t^I = DaR_{t,0.9}^I} - mCoCDaR_{t,\alpha}^{sys|D_t^1 = DaR_{t,0.5}^1, \ldots, D_t^I = DaR_{t,0.5}^I}$$

$$= \sum_{i=1}^{I} \hat{\beta}_i^{\alpha} (DaR_{t,0.9}^{i} - DaR_{t,0.5}^{i}) \; = \; \sum_{i=1}^{I} V_{sys}(i) \; .$$

A similar statement is valid for $mCoCVaR$. The entire systemic risk is exactly distributed to its institutional components. $\Delta mCoCVaR$ and $\Delta mCoCDaR$ are both Shapley value functions, denoted by $V_{sys}(i)$ for contributor i, such that they satisfy the following desirable mathematical properties as outlined in Bernardi et al. (2013). Let $\mathcal{S}$ be a set of I institutions:

(1) Efficiency: $\sum_{i=1}^{I} V_{sys}(i) = V_{sys}(\mathcal{S})$. This axiom states that the total risk is distributed to participants.

(2) Symmetry: For $i \neq j$ such that $V_{sys}(H \cup i) = V_{sys}(H \cup j), \forall H$ such that $i, j \notin H$, then $V_{sys}(i) = V_{sys}(j)$. This axiom states that the contribution measure is permutation invariant and fair for all contributors.

(3) Dummy axiom: $V_{sys}(H \cup i) = V_{sys}(i), \forall i \in H$ and $H \supseteq \mathcal{S}$. This means, if the risk of institution j is independent of all other institutions, then its risk contribution to the system should be its own risk. Generally, in CoCVaR and CoCDaR approaches (also CoVaR), the risks are not orthogonal among institutions. Hence, their ranking should differ from those provided by a Shapley value measure such as $mCoCVaR$ and $mCoCDaR$.

(4) Linearity (additivity): If $i, j \in H, i \neq j$ are two institutions, where $V_{sys}(i) \neq V_{sys}(j)$, let $w_i > 0, w_j > 0, k = w_i i + w_j j$, then new combined risk contributions equal the weighted average of individual risk contributions: $V_{sys}(k) = w_i V_{sys}(i) + w_j V_{sys}(j)$.

(5) Zero player: If $i \notin S, V_{sys}(i) = 0$. A null player receives zero risk contribution.

2.11. mCoCDaR Versus mCoCVaR

We do not claim that one of the considered risk measures is better for analyzing systemic risk contributions than the other one. CoCVaR is concerned with the conditional risk in terms of the returns' tail behavior, while CoCDaR is concerned with the conditional risk in terms of the drawdowns. These measures have a nonlinear relationship embedded in their definitions.

When a financial system's large drawdowns are significantly correlated with large drawdowns of some particular institutions, it can be hypothesized that the CoCDaR measure will provide a more reasonable estimate of the risk contributions and therefore give a more reasonable ranking of the systemic risk contribution of each institution. This can be generalized to comparing mCoCVaR and mCoCDaR measures which are proposed in this work. Another intuition for using drawdown based approaches is that drawdown measures a psychological effect from a consistent distress in stock returns.

3. Case Studies

This case study uses data from CoCVaR paper Huang and Uryasev (2017), which is posted at this link [1]. Codes, data, calculation results for this case study are posted at this link [2].

We first computed the drawdowns from the return data and transformed the state factors corresponding to each regression problem. Next, we proceeded to the quantile regressions on institutional drawdowns and the CVaR regression on the system's drawdowns. The CVaR regression is implemented using Portfolio Safeguard (PSG)[3] in MATLAB environment. PSG includes efficiently implemented (precoded) Koenker–Bassett, Rockafellar and CVaR errors.

1 http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial_engineering/case-study-cocvar-approach-risk-contribution-measurement
2 http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial_engineering/case-study-cocdar-approach-systemic-risk-contribution-measurement
3 Portfolio Safeguard (PSG) is a product of American Optimal Decisions: http://aorda.com

3.1. Financial Institutions

We consider the ten largest publicly traded banks in the USA as of 31 December 2014:

1. JP Morgan Chase & Company (JPM)
2. Bank of America (BAC)
3. Citigroup Inc (C)
4. Wells Fargo & Company (WFC)
5. The Bank of New York Mellon Corporation (BK)
6. US Bancorp (USB)
7. Capital One Financial Corporation (COF)
8. PNC Financial Services Group Inc (PNC)
9. State Street Corporation (STT)
10. The BB&T Corporation (BBT)

Data period is from 18 February 2000 to 30 January 2015. Closing prices are downloaded from Yahoo Finance for the Dow Jones US Financial Index and the financial institutions.

3.2. CoCVaR Calculation Results

In addition to numerical results presented in the following sections for new measures, we calculated systemic risk contributions based on the CoCVaR method. We reproduced the case study described in Section 3.3.4 of Huang and Uryasev (2017) with corrected input data (corrected the wrong sign in return data of financial instruments). We considered negative returns (losses) for each bank and the index. We averaged each bank's contributions across time and ranked them accordingly, where larger values correspond to stronger contributions to system's CoCVaR (the units of all reported values are 100%):

1. WFC: 0.03608
2. BBT: 0.03210
3. PNC: 0.03089
4. JPM: 0.03077
5. BAC: 0.03063
6. STT: 0.02905
7. COF: 0.02776
8. BK : 0.02740
9. USB: 0.02341
10. C : 0.00187

3.3. CoCDaR Calculation Results

The drawdown-at-risk values of each institution at two different risk levels, $\alpha'_1 = 0.9$ for distress level and $\alpha'_2 = 0.5$ for normal level, are computed using the quantile regression defined in Section 2.5. The CoCDaR values of the system at a specific risk level $\alpha = 0.9$ conditioned on each institution's DaR being at a distress level and a normal level are respectively computed based on the CVaR regression in Section 2.4. Following Section 2.6, the difference in CoCDaR values is taken and this results in a time series of $\Delta CoCDaR^{sys|i}_{t,\alpha}$ for each institution i and for each observation time t.

We observe that the quantile regression for DaR using the state variables as regressors yields different behaviors for different institutions. Responses are different for the state variables: some are positive while others are negative. This is true for both the distress level and the normal level. The pseudo R^2 metrics for these quantile regressions are generally between 0.5 to 0.7, which indicates a descent level of explanatory power as compared with using just the original state factors. The observation is consistent with that made in Huang and Uryasev (2017). The CoCDaR and DaR calculation results are posted in the CoCDaR case study[2], see Problems 1 and 3. For the ten CVaR regressions of index drawdowns on the state variables and respective institution drawdowns, we observe that the coefficients for each factor typically have the same sign (with a few exceptions). The pseudo R^2 for CoCDaR regressions are all above 0.8.

We averaged each bank's contributions to CoCDaR across time and ranked the ten banks accordingly, where larger values correspond to stronger contributions to system's CoCDaR:

1. WFC: 0.27695 [1]
2. BAC: 0.22285 [5]
3. BBT: 0.07073 [2]
4. COF: 0.06107 [7]
5. USB: 0.05502 [9]
6. BK: 0.03564 [8]
7. PNC: 0.02558 [3]
8. JPM: 0.02306 [4]
9. STT: 0.00242 [6]
10. C: –0.01390 [10]

The number in brackets is the ranking based on $\Delta CoCVaR$ in Section 3.2. Results show that only Citigroup Inc. has negative CoCDaR contribution to the index on average, hinting that its drawdowns could have a negative correlation with index drawdowns. All other institutions are contributing positively to the system's conditional drawdown-at-risk.

In particular, PNC Financial Services Group Inc (PNC) was ranked third by $\Delta CoCVaR$ but ranked seventh by $\Delta CoCDaR$. On the other hand, Bank of America (BAC) was ranked fifth by $\Delta CoCVaR$ but ranked second by $\Delta CoCDaR$. Clearly, these two approaches provide different perspectives.

3.4. mCoCVaR Calculation Results

This section demonstrates the performance of suggested mCoCVaR, which is the multiple version of the CoCVaR approach developed in Huang and Uryasev (2017). We begin by performing the mCoCVaR analysis of the ten financial institutions in one CVaR regression. The pseudo R^2 for mCoCVaR regression is 0.76. The value-at-risk for normal and distress states are calculated for every institution respectively using quantile regressions on the original state variables. The procedure for VaR calculation is described in Huang and Uryasev (2017), Sections 2.3 and 3.3.2. By holding all other institutions' return values to their VaR values in a normal state (which corresponds to the median) and looking at the differences resulting from changing one particular institution's return value to its VaR value in a distress state, we obtain a time series of $\Delta mCoCVaR^{sys|i}_{t,\alpha}$ for each institution and for each observation time t. We averaged each bank's contributions to mCoCVaR across time and ranked the ten banks accordingly, where larger values correspond to stronger contributions to system's mCoCVaR:

1. BBT: 0.00813 [2]
2. BAC: 0.00721 [5]
3. BK: 0.00619 [8]
4. JPM: 0.00598 [4]
5. PNC: 0.00494 [3]
6. STT: 0.00463 [8]
7. WFC: 0.00306 [1]
8. COF: 0.00266 [7]
9. USB: 0.00149 [9]
10. C: 0.00062 [10]

The number in the bracket is the ranking according to $\Delta CoCVaR$ in Section 3.2. The results based on $\Delta mCoCVaR$ are similar to those based on $\Delta CoCVaR$, but there are some significant differences. For instance, WFC, originally ranked the highest, dropped to the seventh place in this new ranking. This might have been caused by its returns having a high correlation to returns of other institutions, for example The BB&T Corporation (BBT). This effect is neglected in the previous CoCVaR method, but in our multiple regression setting, by explicitly fixing the other institutions' returns to their respective normal states, we are analyzing the marginal impact of WFC's distress. Hence, the drop in ranking may indicate that WFC is not a key systemic risk contributor in the sense that its risk contributions are dependent on the high risk contributions of other institutions. Clearly, CoCVaR and mCoCVaR provide different perspectives regarding the ranking of financial institutions' risk contributions.

3.5. mCoCDaR Results

The drawdown-at-risk of each institution at two different risk levels, $\alpha'_1 = 0.9$ for distress level and $\alpha'_2 = 0.5$ for normal level, are computed using the quantile regression defined in Section 2.5; this step is identical to the first step performed in Section 3.3. The mCoCDaR values of the system at a specific risk level $\alpha = 0.9$ conditioned on each institution's DaR being at a distress level and a normal level are computed respectively based on the CVaR regression with multiple institutions as specified in Section 2.8. Following Section 2.9, the difference in mCoCDaR values is taken and this results in a time series of $\Delta mCoCDaR_{t,\alpha}^{sys|i}$ for each institution i and for each observation time t.

Since we are using the same quantile estimates for DaR, we obtained the same observations as that in Section 3.3. For the CVaR regression of the drawdowns of the index on the state variables and the institution drawdowns, we observe that some institutions' regression coefficients are positive in the CVaR regression, while others are negative. The pseudo R^2 for mCoCDaR regression is 0.9. The mCoCDaR and DaR results are posted in the CoCDaR case study[2], see Problems 2 and 3.

We averaged each bank's contributions to mCoCDaR across time and ranked the ten banks accordingly, where larger values correspond to stronger contributions to system's mCoCDaR:

1. BAC: 0.20572 [2]
2. BBT: 0.02964 [3]
3. USB: 0.02485 [5]
4. STT: 0.02011 [9]
5. COF: 0.01749 [4]
6. BK: 0.01548 [6]
7. C: 0.00434 [10]
8. PNC: –0.01353 [7]
9. JPM: –0.01353 [8]
10. WFC: –0.06106 [1]

The number in the bracket is the ranking according to $\Delta CoCDaR$. $\Delta mCoCDaR$ and $\Delta CoCDaR$ rankings are mostly similar, yet have some interesting differences as well. While WFC is ranked highest by $\Delta CoCDaR$, it is ranked last by $\Delta mCoCDaR$. This observation coincides with what we saw in Section 3.4, indicating the high correlation that WFC might have with other top risk contributors such as BB&T and BAC. Furthermore, while STT is ranked second last by $\Delta CoCDaR$, it is ranked fourth by $\Delta mCoCDaR$.

3.6. Comparative Summary of the Proposed Methods

Table 1 provides a complete summary of the rankings of the ten banks with the four risk measures.

Compared with CoCVaR, CoCDaR takes into account drawdowns and focuses on consecutive losses. Using drawdowns is particularly insightful because drawdowns identify cumulative losses (negative cumulative returns), hence the dependencies between institutions and the system in "good" times are ignored. Dependencies in "bad" times are captured, which is important for risk analysis. We observe that CoCVaR and CoCDaR may provide very different rankings. For instance, USB with mCoCDaR and CoCDaR are ranked 3 and 5, accordingly (i.e., BAC is a top contributor), but with mCoCVaR and CoCVaR it is ranked 9 (i.e., close to bottom contributor). Even more surprisingly, JPM is ranked 9 and 8 with mCoCDaR and CoCDaR, but ranked 4 with mCoCVaR and CoCVaR.

mCoCVaR and mCoCDaR approaches add further insights to CoCVaR and CoCDaR, since they employ a multiple regression that marginalizes the systemic risk contributions of individual institutions. Running the multiple regression instead of individual ones enables us to look at institutions' contributions in a unified way, since their fraction contributions sum up to one.

Risk contributions based on CoVaR and CoCVaR measures, as a function of time, demonstrate a similar pattern for different institutions, see Huang and Uryasev (2017). This is probably because the methodology is based on separate regression for each institution. On the other hand, mCoCDaR results (plotted below) show that the time series of mCoCDaR risk contributions exhibit quite different patterns compared to CoVaR and CoCVaR, and compared across different institutions. With multiple regression, marginal risk contributions of each institution change significantly over time.

Table 1. Systemic Risk Contribution Ranking Summary.

	mCoCDaR	CoCDaR	mCoCVaR	CoCVaR
JPM	9	8	4	4
BAC	1	2	2	5
C	7	10	10	10
WFC	10	1	7	1
STT	4	9	6	6
PNC	8	7	5	3
USB	3	5	9	9
COF	5	4	8	7
BK	6	6	3	8
BTT	2	3	1	2

Furthermore, we plot time dependent drawdowns and mCoCDaR contributions; see Figures 2–11. Each institution graph on the left plots its drawdown curve in blue versus the orange curve showing drawdowns of the Dow Jones index in the same time period, both based on cumulative uncompounded returns on a weekly basis. Every graph on the right plots fraction contribution to the total systemic risk from an individual bank. This fraction is obtained by normalizing individual contributions measured by $\Delta mCoCDaR$ described in Section 3.5. Normalization is done by dividing individual contributions by the total contribution from the ten banks. By construction, the normalized contributions sum up to one for each time step. As a result of applying the mCoCDaR regression setting, we observe that individual contributions significantly vary over time as well as across institutions. Moreover, risk contributions may have different signs. For instance, JPM and WFC always have negative contributions (see, Figures 2 and 5). Citigroup starts with negative contributions and moves to contributing positively (see, Figure 4), while the others always have positive contributions.

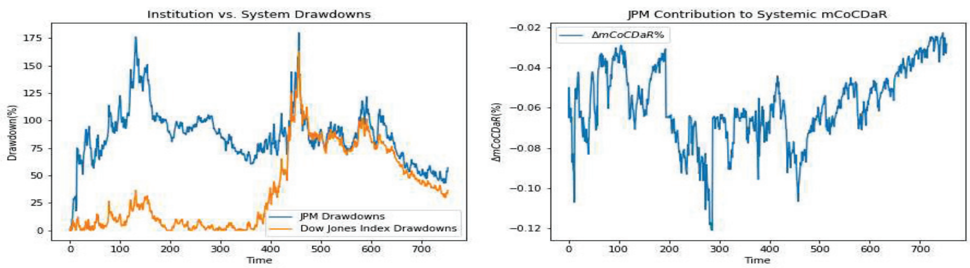


Figure 2. JP Morgan Chase & Company.

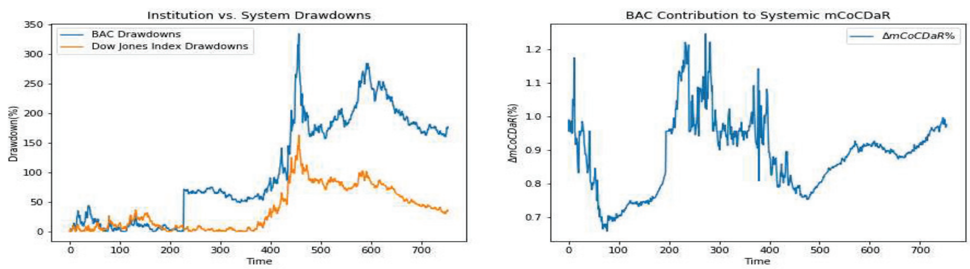


Figure 3. Bank of America.

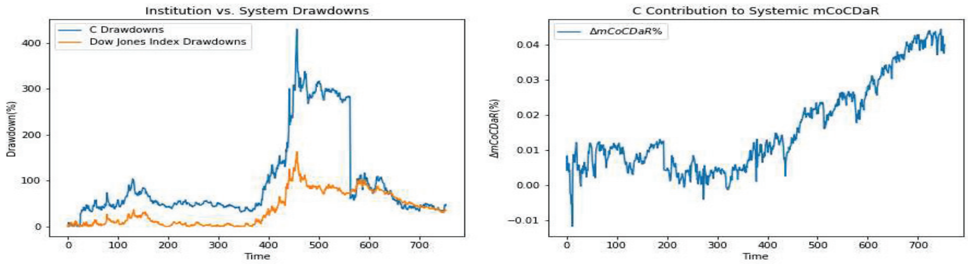


Figure 4. Citigroup Inc.

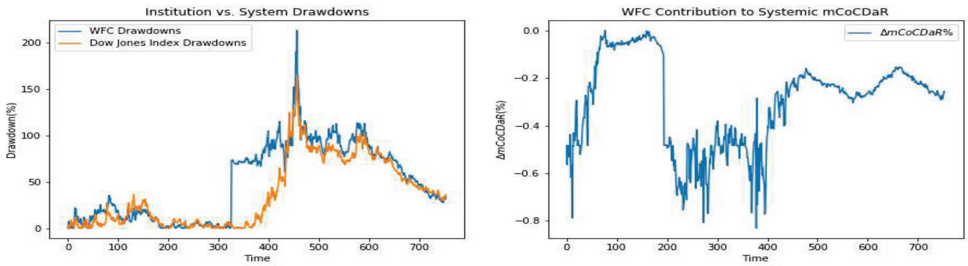


Figure 5. Wells Fargo & Company.

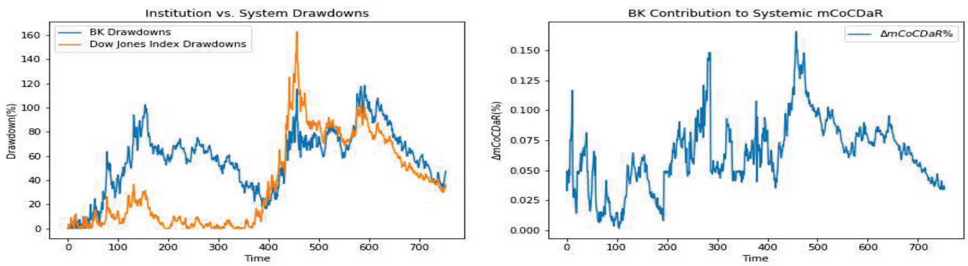


Figure 6. The Bank of New York Mellon Corporation.

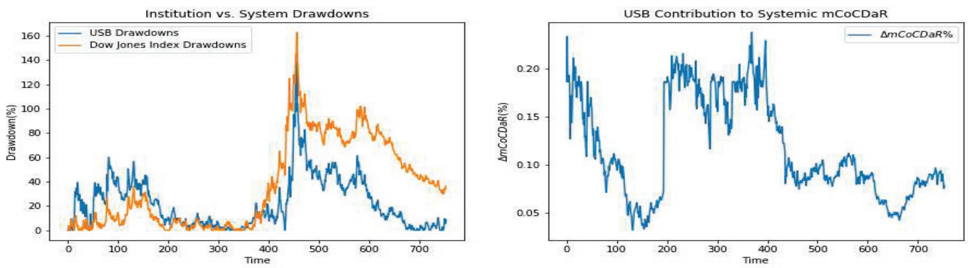


Figure 7. US Bancorp.

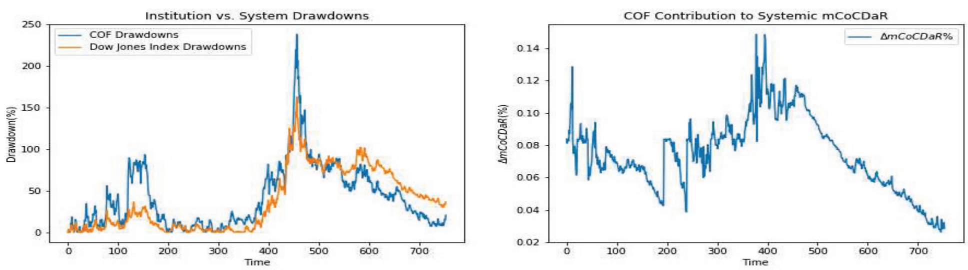


Figure 8. Capital One Financial Corporation.

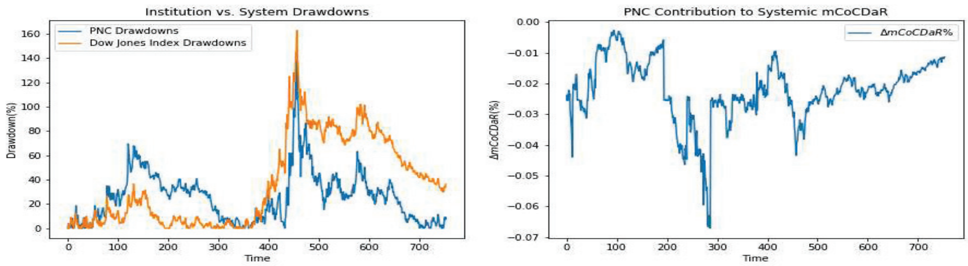


Figure 9. PNC Financial Services Group Inc.

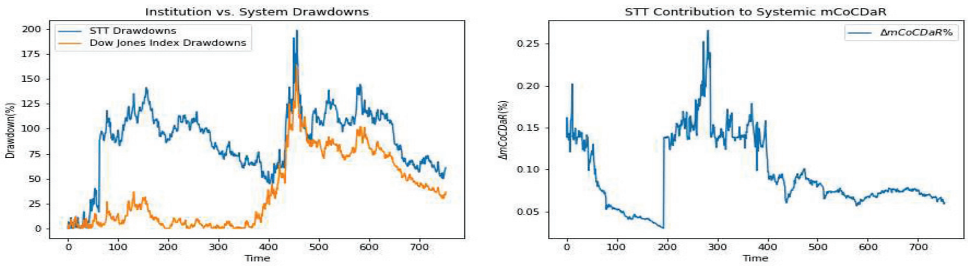


Figure 10. State Street Corporation.

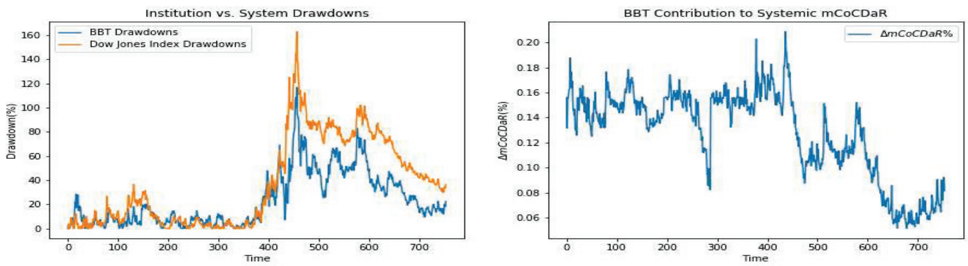


Figure 11. The BB&T Corporation.

4. mCoCDaR Application to Style Classification

This section extends the approach to hedge fund style classification. We show how to estimate CDaR as a function of drawdowns of several market indices. Style classification is a well studied topic approached by Sharpe (1992) and Carhart (1997) with a standard regression (for returns of instruments). Furthermore, it was extended by Bassett and Chen (2001) using quantile regression. Here, we demonstrate results with the mCoCDaR method. This classification explains fund drawdowns, as a function of drawdowns of several market indices (as factors). Codes, data, and results for this case study are posted at this link[4].

Similar to Bassett and Chen (2001), we investigated dependence of drawdowns of the Magellan fund (fund) from four indices: Russell 1000 value index (rlv), Russell 1000 growth index (rlg), Russell 2000 value index(ruj), and Russell 2000 growth index (ruo). These indices correspond to four equity classes: large value stocks, large growth stocks, small value stocks, and small growth stocks.

[4] http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial_engineering/case-study-style-classification-with-mcocdar-regression/

We used a dataset from a quantile regression style classification case study posted at[5]. Golodnikov et al. (2019) considered the same dataset for testing CVaR regression, which is posted at this link[6]. The dataset contains 1264 weekly return observations for the fund and indices.

We calculated drawdowns for the fund and the four indices using weekly returns in the considered time period. CVaR regression of drawdowns is done by minimizing CVaR2 error in PSG[3], as follows,

$$D_{fund,t} = \beta_0 + \beta_1 D_{rlv,t} + \beta_2 D_{rlg,t} + \beta_3 D_{ruj,t} + \beta_4 D_{ruo,t} \ ,$$

where $D_{i,t}$ for $i = rlv,\ rlg,\ ruj,\ ruo$ are (uncompounded) drawdowns of index i at time t and $D_{fund,t}$ are (uncompounded) drawdowns of the fund. See definition of drawdowns of a financial instrument in Section 2.1.

For 0.9-CVaR regression the pseudo-R square equals 0.91 and the estimated coefficients are:

$$\hat{\beta}_0 = 0.3713,\ \hat{\beta}_1 = 0.4621,\ \hat{\beta}_2 = 0.5493,\ \hat{\beta}_3 = -0.0171,\ \hat{\beta}_4 = -0.0591 \ .$$

We considered also 0.0-CVaR regression, which estimates mean and corresponds to an ordinary least squares regression. Pseudo-R square equals 0.91 and estimated coefficients are:

$$\hat{\beta}_0 = -0.2733,\ \hat{\beta}_1 = 0.4891,\ \hat{\beta}_2 = 0.5150,\ \hat{\beta}_3 = -0.0618,\ \hat{\beta}_4 = -0.0003 \ .$$

Regression coefficients show that both large and average drawdowns of the Magellan fund are mostly explained by drawdowns in large value stocks index (coefficient $\hat{\beta}_1$) and large growth stocks index (coefficient $\hat{\beta}_2$). The fund exhibits roughly 50–50% mix of these two classes of stocks in the sense of drawdown behavior.

Furthermore, we compared these results with previous studies, which used CoVaR- and CoCVaR-based measures. The CoVaR approach based on quantile regression[5] (see Problem 1 in the link) gives the following coefficient estimate:

$$\hat{\beta}_0 = -0.0089,\ \hat{\beta}_1 = 0.4602,\ \hat{\beta}_2 = 0.5176,\ \hat{\beta}_3 = -0.0156,\ \hat{\beta}_4 = 0.0001 \ .$$

and the CoCVaR approach based on CVaR regression[6] (see Problem 1, α=0.9 in the link) gives the following estimate:

$$\hat{\beta}_0 = 0.0105,\ \hat{\beta}_1 = 0.6058,\ \hat{\beta}_2 = 0.4721,\ \hat{\beta}_3 = -0.0778,\ \hat{\beta}_4 = -0.0071 \ .$$

We observe that this particular dataset considered regressions of a similar style with around a 50–50% mix of two stock indices.

5. On Portfolio Optimization with mCoCDaR and mCoCVaR

Previous sections defined and tested mCoCDaR and mCoCVaR multiple regression versions for systemic risk measurement. It should be considered that risk measures can be used for other purposes. For instance, we can build a portfolio minimizing CoCVaR or CoCDaR, conditioned on the distress level of several market indices (or factors), under the constraint that the expected return meets some target. Similar problems were studied in Kurosaki and Kim (2013a, 2013b) with CoAVaR and CoVaR measures for conditional risk. Here, we present portfolio optimization problems using mCoCVaR and mCoCDaR risk measures:

$$\min_{\vec{w}_t} \quad mCoCVaR_{\alpha,t}^{\vec{w}_t|f_t^1,\ldots,f_t^K} \quad s.t. \ \sum_{i=1}^{I} w_t^i r_t^i = r^\star, \ \sum_{i=1}^{I} w_t^i = 1$$

---

5 http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial_engineering/style-classification-with-quantile-regression/

6 http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial_engineering/on-implementation-of-cvar-regression/

$$\min_{\vec{w}_t} \quad mCoCDaR_{\alpha,t}^{\vec{w}_t|f_t^1,\ldots,f_t^K} \quad s.t. \; \sum_{i=1}^{I} w_t^i r_t^i = r^\star, \; \sum_{i=1}^{I} w_t^i = 1$$

where K is the number of market index factors, $f_t^1, \ldots, f_t^K$ are risk levels at time t of K factors (market indices), $\vec{w}_t$ is vector of portfolio weights for I stocks, r_t^i is return of stock i at time t, and $r^\star$ is a target return. Systemic risk-driven portfolio selection problems were also studied in Capponi and Rubtsov (2019), where they considered portfolio optimization given a systemic event. Detailed analysis of these portfolio optimization problems is beyond the scope of this paper. We have included a short description to show that considered risk measures can be used in various areas of finance.

6. Conclusions

This paper proposed a new systemic risk measure, CoCDaR, which is based on conditional drawdown-at-risk and inspired by the CoCVaR approach from Huang and Uryasev (2017). We further extended the approach to mCoCDaR, which calculates conditional drawdown-at-risk of the financial system conditioned on all the institutions' drawdown distress levels. These measures can rank institutions according to their incremental (marginal) contributions to the systemic risk of the system, conditional on other institutions' distress levels. The multiple regression setting is applied to the CoCVaR measure from Huang and Uryasev (2017) and resulted in so-called mCoCVaR. Since mCoCDaR and mCoCVaR are based on multiple regression, they have the flexibility to measure joint contributions of multiple institutions. These measures are also well-defined Shapley value functions with desirable mathematical properties for a risk contribution measure. After normalization, individual risk contributions sum up to one. These advantages do not come at any additional computational cost.

CoCDaR and mCoCDaR measures are based on drawdowns (path dependent cumulative losses). These two measures capture the impact of an institution's drawdowns on the financial system's drawdowns, which is particularly suitable for market crash situations. They are useful for determining which institution may lead to a bigger crash in the market in terms of large drawdown events.

We performed a case study for the three proposed methods, CoCDaR, mCoCVaR, and mCoCDaR, using data from the ten largest banks and the Dow Jones Index, along with some state factors. The case study with codes and data are posted on the web. We have also reproduced the case study for CoCVaR measure from Huang and Uryasev (2017), with corrected signs in the returns data. We compared the ranking of institutions based on contributions to system's CoCDaR, mCoCDaR, mCoCVaR, and CoCVaR. The difference in applying CVaR- and CDaR-based measures is observed from quite different rankings of institutions. Multiple regression identifies key drivers in systemic risk because effects are marginalized. We compared time dependent curves of risk contributions for mCoCDaR and CoCVaR. Risk contributions based on CoCVaR are quite similar across institutions, while those based on mCoCDaR have very different patterns. These different patterns are implied by both the use of drawdowns and the use of multiple regressions.

Other applications of the proposed method include fund style classifications based on mCoCDaR regression. We have conducted a case study analyzing drawdowns of the Magellan fund as a function of drawdowns of four market indices. We have posted this case study to the web. The suggested methodology may also be used in various other areas of finance. In particular, we have stated portfolio selection problems with mCoCVaR or mCoCDaR objectives and constraints on expected returns.

Author Contributions: R.D. and S.U. stated the problem; S.U. obtained the data and provided the software; R.D. processed the data, wrote the programs and obtained the results; R.D. and S.U. analyzed the results; R.D. prepared the first draft manuscript; R.D. and S.U. revised the text and the conclusions. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Adrian, Tobias, and Markus K. Brunnermeier. 2008. *CoVaR*. Federal Reserve Bank of New York Staff Report No. 348. St. Louis: Federal Reserve Bank.

Bassett, Gilbert W., and Hsiu-Lang Chen. 2001. Portfolio style: Return-based attribution using quantile regression. *Empirical Economics* 26: 293–305. [CrossRef]

Bernardi, Mauro, Antonello Maruotti, and Lea Petrella. 2013. Multivariate Markov-Switching Models and Tail Risk Interdependence. Available online: https://arxiv.org/pdf/1312.6407.pdf (accessed on 3 November 2020).

Bernardi, Mauro, and Lea Petrella. 2014. Interconnected risk contributions: A heavy-tail approach to analyze U.S. financial sectors. *Journal of Risk and Financial Management* 8: 198–226. [CrossRef]

Cao, Zhili. 2013. *Multi–CoVaR and Shapley Value: A Systemic Risk Measure*. Paris: Bank de France Working Paper.

Capponi, Agostino, and Alexey Rubtsov. 2019. Systemic Risk Driven Portfolio Selection. May 25. Available online: https://ssrn.com/abstract=3394471orhttp://dx.doi.org/10.2139/ssrn.3394471 (accessed on 3 November 2020).

Carhart, Mark M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52: 57–82. [CrossRef]

Chekhlov, Alexei, Stan Uryasev, and Michael Zabarankin. 2005. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance* 8: 13–58. [CrossRef]

Golodnikov, Aexander, Viktor Kuzmenko, and Stan Uryasev. 2019. CVaR regression based on the relation between CVaR and mixed-quantile quadrangles. *Journal of Risk and Financial Management* 12: 107. [CrossRef]

Huang, Wei-Qiang, and Stan Uryasev. 2017. The CoCVaR approach: Systemic risk contribution measurement. *The Journal of Risk* 20: 75–93. [CrossRef]

Koenker, Roger, and Gilbert Bassett. 1978. Regression quantile. *Econometrica* 46: 33–50. [CrossRef]

Kurosaki, Tetsuo, and Young Shin Kim. 2013a. Systemic risk measurement in the global banking stock market with time series analysis and CoVaR. *Investment Management and Financial Innovations* 10: 184–96.

Kurosaki, Tetsuo, and Young Shin Kim. 2013b. Mean-CoAVaR optimization for global banking portfolios. *Investment Management and Financial Innovations* 10: 15–20.

Rockafellar, R. Tyrell, Johannes O. Royset, and S.I. Miranda. 2014. Superquantile regression with applications to buffered reliability, uncertainty quantification and conditional value-at-risk. *European Journal of Operations Research* 234: 140–54. [CrossRef]

Rockafellar, R. Tyrell, and Stan Uryasev. 2000. Optimization of conditional value-at-risk. *The Journal of Risk* 2: 21–42. [CrossRef]

Rockafellar, R. Tyrell, and Stan Uryasev. 2002. Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance* 26: 1443–71. [CrossRef]

Rockafellar, R. Tyrell, and Stan Uryasev. 2013. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* 18: 33–53. [CrossRef]

Sharpe, William F. 1992. Asset allocation: Management style and performance measurement. *Journal of Portfolio Management (Winter)* 18: 7–19. [CrossRef]

Zabarankin Michael, Pavlikov Konstantin, and Stan Uryasev. 2014. Capital asset pricing model (CAPM) with drawdown Measure. *European Journal of Operational Research* 234: 508–17. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of
Risk and Financial Management

Article

The Effects of ERM Adoption on European Insurance Firms Performance and Risks

Doureige J. Jurdi * and Sam M. AlGhnaimat

Department of Economics and Finance, La Trobe Business School, La Trobe University, Melbourne, VIC 3086, Australia; s.al-ghnaimat@latrobe.edu.au
* Correspondence: d.jurdi@latrobe.edu.au

Abstract: We investigate the effects of adopting enterprise risk management (ERM) on the performance and risks of European publicly listed insurance firms. Using a dataset for 24 years, we report new results which show that ERM adopters realize significant ERM premiums after controlling for other covariates and endogeneity. Several firm characteristics such as size, opacity, and the choice of external monitoring agents such as auditors are significant determinants of adopting ERM. We fill a gap in the literature by assessing the impact of adopting ERM on firm risks and report new findings for our sample, which show that ERM adopters effectively reduce firm total and systematic risks and, to a greater extent, idiosyncratic risk. Firm-level variables such as size, leverage, dividend payments events, and diversification impact firm total risk. Insurers use corporate events such as dividend payments to signal information about reducing risk. Industry and international diversification reduce firm total risk and idiosyncratic risk, respectively.

Keywords: enterprise risk management; firm characteristics; firm performance; firm risk; insurance firms

JEL Classification: G20; G22; G30; G32

Citation: Jurdi, Doureige J., and Sam M. AlGhnaimat. 2021. The Effects of ERM Adoption on European Insurance Firms Performance and Risks. *Journal of Risk and Financial Management* 14: 554. https://doi.org/10.3390/jrfm14110554

Academic Editor: David Allen

Received: 9 August 2021
Accepted: 12 November 2021
Published: 16 November 2021

1. Introduction

Enterprise risk management (ERM) provides a holistic approach for identifying, evaluating, managing, and mitigating risks at the enterprise level. The Committee of Sponsoring Organizations of the Treadway Commission (COSO) defined ERM as: "process, effected by an entity's board of directors, management and other personnel, applied in a strategy setting and across the enterprise, designed to identify potential events that may affect the entity, and manage risk to be within its risk appetite, to provide reasonable assurance regarding the achievement of entity objectives" (COSO 2004). Eckles et al. (2014) argue that compliance with the Sarbanes Oxley Act, which aims at enhancing corporate governance, has been a catalyst for firms to build their ERM infrastructure. Nocco and Stulz (2006) claim that ERM creates value for the firm and strengthens its ability to execute its strategic plan. Bohnert et al. (2019) indicate that ERM activities in insurance firms are relevant to implementing Solvency II requirements, especially Pillar 2, and these activities enhance firm value. Hoyt and Liebenberg (2011) find that ERM adoption improves firm value, whereas Beasley et al. (2008) find an insignificant market reaction to appointing a Chief Risk Officer. Anton and Nucu (2020) provide a review of various studies on ERM programs applied in various sectors and document mixed evidence on the effect of ERM adoption on firm performance. Other studies find that ERM reduces the cost of capital (Berry-Stölzle and Xu 2018), stock return volatility (Eckles et al. 2014), and firm solvency levels (Nguyen and Vo 2020).

This study identifies the determinants of ERM in European publicly listed insurance firms and evaluates the effects of ERM adoption on market-based performance and risks of these firms.[1] We add new empirical evidence to existing findings on the relationship between ERM adoption and insurance firm performance using a 24-year long dataset. Existing studies often use short periods, from 1 to 8 years, in the empirical investigation (see Hoyt and Liebenberg 2008; Lechner and Gatzert 2018; Bohnert et al. 2019); an extended dataset is necessary for our analysis, considering that the effects of ERM adoption may require time to be realized in a constantly changing risk environment. In addition, we contribute to the existing literature by assessing the impact of ERM adoption on publicly listed European insurance firms' market-based measures risk and consider firm total risk, systematic risk, and idiosyncratic risk. While it might be reasonable to expect that ERM adoption may reduce the levels of firm risk, implementing adequate risk controls is a consequence of risk assessments and requires time. In addition, the emergence of new risks in the financial services industry and the complexity of the risk management environment pose significant challenges to the effectiveness of ERM programs and their running costs. Furthermore, like banks, insurance firms are subject to moral hazards when they reinsure allowable exposures, which may induce excessive risk-taking. At the same time, one may question whether ERM adoption contributes to risk reduction in the insurance sector, which is highly regulated and often subject to scrutiny by various regulators.

Therefore, we aim to provide new insights to researchers and practitioners by investigating the impacts of ERM adoption on firm risks. Our analysis begins with estimating a logistic regression to identify the determinants of ERM in our sample. Results show that the likelihood of adopting an ERM program is associated with firm size, opacity, and being audited by one of the four global audit firms. We estimate a full maximum likelihood treatment effects model to assess the value premium and identify factors that influence performance. The model addresses the potential of endogeneity bias arising from firm factors that simultaneously affect ERM choice and performance (see Hoyt and Liebenberg 2011). Our results reveal an ERM value premium of 54.3% and show that firm size and trailing profitability are positively associated with firm performance. Then, we consider the effects of ERM adoption on firm risks since performance improvement in insurance firms might also be a byproduct of minimizing risks. Estimates from a full maximum likelihood treatment effects model show that ERM adoption reduces firm total risk by 28%. Firm size and leverage are positively related to total risk due to the potential of facing higher cash flow volatility and the likelihood of experiencing financial distress, respectively. In addition, we find that firms rely on dividend payment events as a signaling mechanism to convey information about their ability to reduce total risk and achieve a similar effect by reaping benefits from industry diversification. Results also show that European insurance firms in our sample benefit significantly from adopting ERM by reducing systematic and idiosyncratic risks.

The remainder of the paper is organized as follows. Section 2 provides a summary of the relevant literature and the hypotheses development. Section 3 explains the methodology used in the empirical analysis. Section 4 describes the data. Section 5 reports the main results, and Section 6 concludes.

2. Literature Review and Hypotheses Development

The literature documents mixed evidence on the value of ERM implementation for firms across various sectors. Using a sample of 120 U.S. companies, from various industries, between the years 1992 and 2003, Beasley et al. (2008) find that the announcement of appointing senior risk executives does not impact firm cumulative abnormal returns in both financial and non-financial firms. However, the market reacts positively to implementing ERM programs as it reduces agency costs and financial distress. Tahir and Razali (2011) conduct a similar study on a sample of 528 Malaysian firms from ten different industries, other than the financial services industry, during 2007 and fail to find a significant relationship between ERM and firm value measured by Tobin's Q. Farrell and Gallagher (2015) analyze the implications of ERM maturity score on firm valuation using data for 225 firms from various industries from 2006 to 2011, most of which operate in the U.S. The authors find that firms with mature ERM implementation have high ERM attribute scores and attract a valuation premium of around 25%. Larger firms achieve better ERM attribute scores than smaller firms; for example, for adopting an ERM-based approach, ERM process management, and others. These scores decrease for internationally diversified firms due to difficulties in maintaining ERM practices across different countries. Lechner and Gatzert (2018) identify firm characteristics that impact implementing an ERM system using listed German firms, from various industries, between 2009 and 2013. The authors consider 2013 in their analysis and find that the firm's size is positively related to ERM implementation, whereas financial leverage is negatively related. Firms with international operations and those operating in the banking, insurance, and energy industries impact ERM implementation positively. The authors show that ERM activities increase shareholders' value (as proxied by Tobin's Q) by 41.6%. Using a sample of 162 Spanish-listed firms from various industries, excluding firms from the financial and insurance sectors, from 2012 to 2015, González et al. (2020) conclude that ERM adoption is not associated with changes in performance or reducing the probability of firm bankruptcy. However, the authors find that the appointment of a chief risk officer, although it reduces firm performance, may also improve its financial health.

In the insurance industry, Hoyt and Liebenberg (2008) study the extent to which insurance firms in the U.S. have implemented ERM programs between 2000 and 2005 and evaluate the value of these programs. The authors use a treatment effects model that simultaneously examines ERM determinants and value for 125 firms. They show that ERM programs are positively related to firm size and the pressure from institutional owners and negatively related to leverage and reinsurance. Additionally, they find that ERM adopting firms are valued 16.7% higher than non-ERM firms after controlling for value determinants and endogeneity bias. Three years later, Hoyt and Liebenberg (2011) conduct a similar study on 117 US insurance firms between the years 1998–2005 to find that, on average, ERM increases firm value by approximately 20%. McShane et al. (2011) utilize the S&P ERM insurance rating to study the value relevance of ERM on a sample of 82 insurance companies for which the S&P has issued an ERM rating. The S&P ERM insurance rating captures all aspects of the risk management program and reflects its extent of implementation. Using data from 2008, the authors estimate a regression model to investigate the relationship between firm value and the degree to which insurers have implemented ERM. They find a positive relationship between the level of risk management and shareholder value in categories that capture increasing levels of traditional risk management (TRM) but no additional increase in value when shifting from TRM to ERM.[2] Bohnert et al. (2019) investigate the factors that affect ERM adoption on performance using 42 European insurance firms between 2007 and 2015. Their results show that the firm's size is positively related to ERM implementation, while leverage and return volatility are negatively related to ERM. The authors also find that insurance firms with high-quality risk management ratings have a higher Tobin's Q of 6.5% than firms with lower risk management ratings.

Given the evidence cited above, we hypothesize that adopting ERM improves firm value and performance. This hypothesis is based on two assumptions; first, we assume that the benefits of adopting ERM exceed costs, and second, adopting ERM would contribute to improving the quality of risk governance (Bohnert et al. 2019) and would not be limited just to a response to regulatory or monitoring agents' pressure. Pagach and Warr (2011) indicate little or no benefit from ERM if adopted purely in response to regulatory pressures.

Nguyen and Vo (2020) investigate the relationship between ERM adoption and solvency for 101 publicly listed insurance firms in the European Union from 2007 to 2013. The authors find that ERM adoption adversely affects firm solvency. They also find that ERM adoption is determined by various factors, including firm size, performance, business type/line, international operations, and demand. Eckles et al. (2014) examine the implications of ERM adoption on U.S. insurance firm risk-taking behavior using data from 1990 to 2008 and report a reduction in the firm's stock return volatility. The authors find that adopting ERM lowers the marginal costs of reducing risks and improves the operating profits per unit of risk. ERM adoption may lead to effective risk governance. Baxter et al. (2013) find, using a sample of banks and insurance companies from 2006 to 2008 selected from the S&P Ratings Direct database, that ERM quality is associated with better governance. The authors argue that market participants react favorably to ERM quality announcements. They also find that higher-risk companies have lower quality ERM.

We view ERM adoption as a progressive step for improving firm risk governance and exploiting natural hedges, and therefore, we hypothesize that ERM adoption reduces insurance firm risks. We focus on market-based risk measures constructed from stock returns or estimated using standard asset pricing models and consider firm total risk, systematic and idiosyncratic risks. We contend that risk reduction could be achieved if the benefits from ERM activities exceed costs. ERM adopters would then exploit natural hedges and eliminate risk duplications, reducing risk management costs while improving performance (Hoyt and Liebenberg 2011). In addition, ERM facilitates risk combinations which would reduce risks if correlations were less than one. Eckles et al. (2014) demonstrate based on results from modern portfolio theory that moving away from a silo-based risk management approach to ERM results in improved loss allocations across firm segments leading to a reduction in risk per dollar spent. In addition, Hann et al. (2013) argue that coinsurance enables diversified firms to transfer resources across business segments to avoid countercyclical deadweight costs; therefore, reducing systematic risk. Berry-Stölzle and Xu (2018) link the adoption of ERM to reducing the cost of capital by reducing a firm's systematic risk. However, reducing systematic risk depends on the extent of coinsurance among diversified business units (Hann et al. 2013). Lastly, we hypothesize that adopting ERM reduces firm idiosyncratic risk. Stulz (1996) and Stulz (2003) argue that risk management activities create value to shareholders in the presence of agency costs and market imperfections. These activities mitigate firm-specific lower tail events. Alternatively, risk management activities could be costly when firms do not face tail risk events, and therefore, these activities may not be value-creating. In addition, if investors can diversify away idiosyncratic risk (Markowitz 1952), then investing in risk management activities may not be desirable.

3. Methodology

We begin with estimating a logistic regression to quantify the association between factors likely to affect the decision to adopt or engage in an ERM program and the likelihood of adopting the program.

$$P(ERM|x) = \frac{\exp(x\beta)}{1+\exp(x\beta)} \tag{1}$$

where $P(ERM)$ is the probability of adopting an ERM program by an insurance firm. The vector of explanatory variables and the constant is denoted by x, and β is a parameter vector. We use the method of maximum likelihood to estimate the log form specification of Equation (1) where the dependent variable ERM = 1 for firms adopting an ERM program and 0 otherwise. Following the literature, we consider the following determinants of ERM:

Firm size: Hoyt and Liebenberg (2011) and Lechner and Gatzert (2018) show that firm size plays a significant role in determining ERM implementation. Implementing ERM requires substantial resources to acquire the relevant technology, appointing key personnel, and funding related administrative costs, which may discourage small firms with limited resources (Beasley et al. 2005). Allayannis and Weston (2001) report a negative relationship between size and firm value. We define insurance firm size as the natural logarithm of the book value of assets.

Financial leverage: Golshan and Rasid (2012) argue that firms who adopt ERM programs may reduce their debt ratio to lower the probability of financial distress. Conversely, firms with a holistic risk management framework may decide to increase leverage as they would be better positioned to manage risks associated with higher debt levels (Hoyt and Liebenberg 2011; Pagach and Warr 2011). We define financial leverage as the total debt to total assets ratio.

Capital opacity: Pagach and Warr (2011) argue that opaque assets are associated with information asymmetry between the firm and investors, typically resulting in undervaluing the firms' assets. Adopting ERM addresses this problem as it facilitates reflecting the firm's risk profile more and its financial strength more accurately to outside stakeholders (Liebenberg and Hoyt 2003). We define capital opacity as the ratio of total intangible assets to the book value of assets.

Financial slack: Pagach and Warr (2011) argue that ERM users are more inclined to increase the level of financial slack to reduce the probability of financial distress. Conversely, ERM users have a lower likelihood of experiencing financial distress due to improved risk management, requiring less liquidity. Several studies report a positive relationship between financial slack and ERM adoption (Pagach and Warr 2011; Berry-Stölzle and Xu 2018). We proxy for financial slack by the ratio of cash and marketable securities to total assets.

Big Three (rating): Lechner and Gatzert (2018) explain that market participants react positively to firms that receive a high credit rating, especially if that rating is from one of the Big Three rating agencies (Standard and Poor's, Moody's, or Fitch). As a result, companies are more likely to implement ERM to signal their commitment to improved management strategies and higher transparency, which ultimately will be positively reflected in their ratings (Hoyt and Liebenberg 2011). We define the big three rating variable as a binary variable that takes a value of 1 if the insurance firm is rated by anyone of the big three rating agencies and zero otherwise.

Big Four (audit): Companies that appoint high-quality auditors are more likely to have better risk management systems because they will be reflected in their published audit reports (Beasley et al. 2005; Lechner and Gatzert 2018). We define big four as a binary variable that takes a value of 1 if the insurance firm is audited by any one of the big four audit firms (Deloitte, PricewaterhouseCoopers, KPMG, and Ernst & Young) and zero otherwise.

Industrial diversification: Firms operating in more than one segment are less likely to implement ERM due to realizing diversification benefits (Pagach and Warr 2011). However, as the number of business units increases, the complexity and range of risks facing a firm may also increase, leading to an increasing need for a holistic risk management system. Lechner and Gatzert (2018) find that industrial diversification is positively related to ERM. We define industrial diversification as a binary variable that takes a value of 1 if the insurance firm operates in more than one insurance segment and zero otherwise.

International diversification: Internationally diversified firms generally face more risks than domestic firms and must comply with different regulatory requirements across different countries (Hoyt and Liebenberg 2011; Lechner and Gatzert 2018). We define international diversification as a binary variable that takes a value of 1 if the insurance firm has international operations and zero otherwise.

One potential disadvantage of the logistic regression is that it ignores the selectivity bias that occurs when the choice of adopting ERM programs is endogenous. To address a potential endogeneity bias arising from factors jointly affecting the decision to adopt ERM and firm performance, we estimate a full maximum likelihood treatment effects model following Hoyt and Liebenberg (2011).

The treatment effects model estimates the effect of the ERM variable, which is considered as an endogenous binary treatment variable on an observed continuous variable such as performance (or risk) conditional on other determinants of performance.

The model requires estimating a performance regression and selection equations. The performance equation is

$$y_{it} = ERM_{it}\delta + x'_{it}\beta + \varepsilon_{it} \tag{2}$$

where y_{it} denotes a performance or a risk variable, ERM_{it} is a binary variable that takes a value of 1 if firm i was subject to ERM treatment during year t and zero otherwise. x'_{it} contains a set of explanatory variables. The decision to adopt ERM is modeled as an outcome of an unobserved latent variable ERM^*, which is a linear function of ω'_{it}, a vector that contains a set of explanatory variables affecting the decision to adopt ERM,

$$ERM^*_{it} = \omega'_{it}\gamma + \mu_{it} \tag{3}$$

where ERM_{it} is a binary variable equal to 1 if $ERM^* > 0$; that is if an insurance firm adopts ERM and 0 otherwise. The error terms ε_{it} and μ_{it} are assumed to follow a bivariate normal distribution with a zero mean vector and covariance matrix $\begin{bmatrix} \sigma_p & \rho \end{bmatrix}'$. Equations (2) and (3) are estimated simultaneously using the method of maximum likelihood.

The primary independent variable of interest is ERM adoption (*ERM*). We identify whether a firm adopts or implements an ERM program using a method suggested by Hoyt and Liebenberg (2011). This method involves searching for particular words, terms, or phrases that identify risk management practices linked to ERM.[3] This variable takes a value of zero for all years before the first ERM event recorded and one for the year of the event onwards. We use the variables explained earlier as determinants of ERM.

With respect to performance equation, we use Tobin's Q as a proxy of firm value (Bohnert et al. 2019; Hoyt and Liebenberg 2008, 2011; Lechner and Gatzert 2018; Li et al. 2014; Lin et al. 2012),

$$Tobin's\ Q = \frac{Market\ value\ of\ equity + Book\ value\ of\ liabilities + Preferred\ Stock}{Book\ Value\ of\ Total\ Assets}$$

Tobin's Q indicates that a firm creates (destroys) value by utilizing assets efficiently when it is larger (less) than 1. The ratio is free from management discretion and is a prospective performance measure that reflects future expectations rather than history (Lindenberg and Ross 1981; Hoyt and Liebenberg 2011). To estimate Equation (2), we use the following explanatory variables.

Firm size: McShane and Cox (2009) find that larger life-health insurers enjoy an enhanced value resulting from economies of scale, greater market power, and lower insolvency risk than smaller firms. On the other hand, Allayannis and Weston (2001) explain that large firms may face agency problems that deteriorate firm's value. Hoyt and Liebenberg (2011) and Lechner and Gatzert (2018) report a positive relationship between size and value.

Financial leverage: The relationship between capital structure and firm value is ambiguous in the literature (Lechner and Gatzert 2018). Relying more on debt can increase firm value by reducing the free cash flow that could have been invested in sub-optimal projects by self-interested managers (Hoyt and Liebenberg 2011; Jensen 1986). Furthermore, an increased debt level allows for tax savings (Tahir and Razali 2011). On the other hand, an increased debt level may lead to financial distress, and in some cases, bankruptcy (Hoyt and Liebenberg 2011).

Profitability: Allayannis and Weston (2001) explain that profitable firms tend to be overvalued compared to non-profitable companies. We control for differences in the firms' profitability in the sample by adding the return on assets as a control variable in the models used (Hoyt and Liebenberg 2011; McShane et al. 2011).

Revenue growth: Previous studies such as Smith and Watts (1992) and Allayannis and Weston (2001) emphasize the importance of controlling for possible influence from varying investment opportunities in firms. Since firm value is primarily dependent on cash flows from future investments, companies with higher sales are more likely to be able to invest in positive net present value projects, thus enhancing firm value. On the other hand, the uncertainty associated with future growth increases the information asymmetry in the capital market, leading to increasing debt costs, thus negatively impacting value. We define revenue growth as the percentage change in revenues from year to year.

Dividends: Paying dividends may signal to investors that the company has exhausted its growth capabilities, negatively impacting firm value (Allayannis and Weston 2001). Conversely, paying dividends reduces the free cash flow that may have been used for managers' self-interest rather than maximizing value (Hoyt and Liebenberg 2011), typically implying that dividends may also positively impact firm value. Further, paying dividends is sometimes perceived as a signal of strong financial health, which also increases firm value. We define dividends as a binary variable that takes a value of 1 if the insurance firm pays dividends in a financial year and zero otherwise.

Industrial diversification: A higher level of diversification allows for risk reduction opportunities by realizing interdependencies between various types of risks. Moreover, diversified firms may observe performance enhancement caused by the achievement of economies of scope in addition to greater access to internal capital (Ai et al. 2018). Conversely, diversification increases the complexity of coordination among conglomerates, which may result in the loss of information between them, potential duplication of management activities, and an increase in agency costs (Ai et al. 2018; Hoyt and Liebenberg 2011; Lechner and Gatzert 2018).

International diversification: The same argument for industrial diversification applies equally to international diversification. The benefits from scope economies and the ability to reduce risk will be positively reflected in the firm's value (Hoyt and Liebenberg 2011). Conversely, the decrease in management efficiency and the inability to align the interest of shareholders and insider managers may lead to a reduction in shareholder value instead (Denis et al. 2002).

Besides performance, we estimate the treatment effects model to test whether adopting ERM reduces the risk in insurance firms. We consider three measures of risk as left-hand side variables in Equation (2). These measures are total risk, systematic risk, and idiosyncratic risk (Chen et al. 2006; Pathan 2009). Total risk is calculated as the natural logarithm of the standard deviation of daily stock returns for each fiscal year. Systematic risk is measured by the regression coefficient of insurance firm return on the relevant market index excess return. Idiosyncratic risk is computed annually as the standard deviation of residuals obtained from regressing daily return observations on the Fama-French European three factors.

The determinants of ERM, performance, and risk are summarized in the following functional forms:

$$ERM_{it} = \mathrm{f}(Leverage,\ Opacity,\ Size\ Slack,\ Big\ Three,\ Big\ Four,\ Ind\ Div,\ \ Int\ Div)_{it},$$

$$y_{it} = \mathrm{f}(ERM,\ Size,\ Leverage,\ ROA,\ Revenue\ growth,\ Dividend,\ \ Ind\ Div,\ \ Int\ Div)_{it}.$$

4. Data

We obtain data for European insurance firms from the Bloomberg and Factiva databases. European insurance firms were identified using the global industry classification system (GICS) provided via the Bloomberg database. The data set is from 1995 to 2018, and it contains 80 firms headquartered in 20 different European countries.[4]

Figure 1 shows that the proportion of ERM engagement increased over time by insurance firms in our sample, reaching 90% in 2018. We also hand collect data for some variables from audited annual reports and use the Kenneth French data library. Table 1 provides variable definitions and reports the sources of data.

Descriptive Statistics

Table 2 reports the univariate statistics for the entire sample. ERM years account for 64.1% of the total firm years available in the data, indicating that most firms in the sample have implemented ERM at some point during the sample period. We test for differences in the mean and median between ERM adopting and non-adopting firms using the *t-test* with unequal variances and the *Wilcoxon rank-sum (R.S.) test*. These tests show statistically significant differences in Tobin's Q, total risk, and idiosyncratic risk between non-ERM adopting and adopting firms (*ERM = 0 – ERM = 1*) and across most variables used in the analysis.

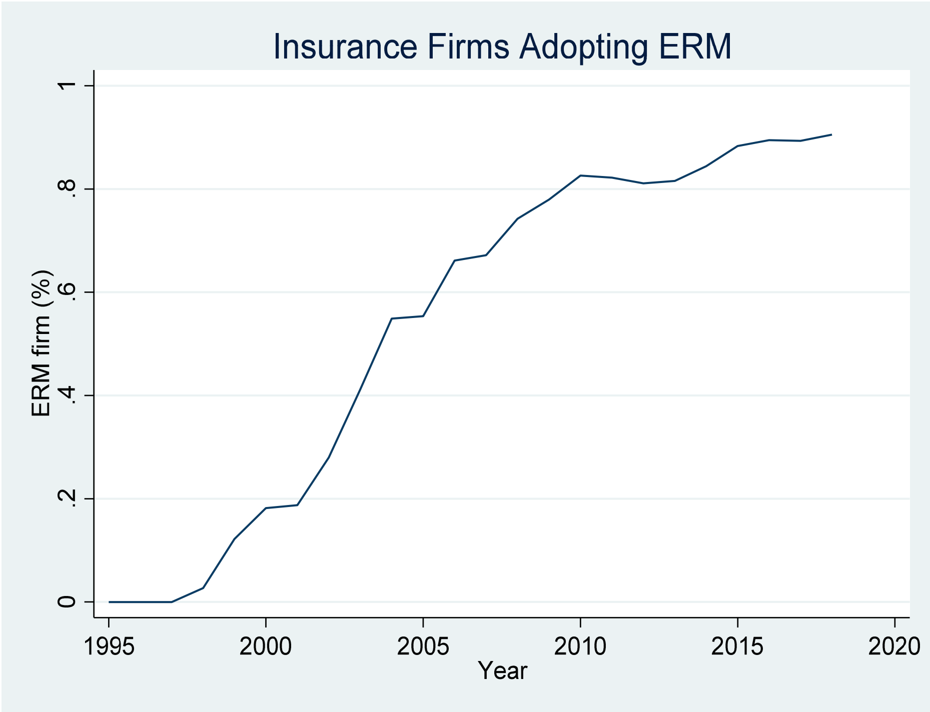


Figure 1. ERM adopting firms.

Table 1. Variable description.

Variable	*Definition*	*Source*
ERM	ERM = 1, Otherwise = 0.	Annual reports, Bloomberg, and Factiva.
Tobin's Q	(Market Cap + Total Liabilities + Preferred Equity + Minority Interest)/Total Assets.	Bloomberg
Total risk	The annualized standard deviation of the relative price change for the 360 most recent trading days closing price.	Bloomberg
Systematic risk	Beta is the percent change in the stock price given a 1% change in the market index.	Bloomberg
Idiosyncratic risk	The annual standard deviations of residuals are estimated using the daily Fama-French European 3-factor model.	Kenneth French library
Dividends	1 = for years the company has paid dividends, 0 = otherwise.	Bloomberg
ERM Leverage	Total Liabilities/Total Assets.	Bloomberg
Q Leverage	Total Liabilities/Market Value of Equity.	Bloomberg
Opacity	Total Intangible Assets/Total Assets.	Bloomberg
Revenue Growth	$(Revenue_t - Revenue_{t-1})/Revenue_{t-1}$.	Bloomberg
ROA	Trailing 12 Month Net Income/Book Value of Assets.	Bloomberg
Size	Natural Logarithm of Total Assets.	Bloomberg
Slack	Cash and Marketable Securities/Total Assets.	Bloomberg
Big three (Ratings)	1 = for years the company is rated by one of the big three rating agencies, 0 = otherwise.	Annual Reports
Big four (Audit)	1 = for years the company is audited by one of the big four audit firms, 0 = otherwise.	Annual reports and Bloomberg
Industry Div.	1 = if the firm operates in more than 1 segment, 0 otherwise.	Annual reports and Bloomberg
International Div.	1 = if firm has international operations, 0 otherwise.	Annual reports and Bloomberg

Table 2. Descriptive statistics.

Variable	N	Mean	Median	Std. Dev.	Min	Max	t-Test	R.S. Test
ERM	1369	0.641	1.000	0.480	0.000	1.000	-	-
Tobin's Q	1369	1.182	1.030	0.504	0.685	7.665	0.154 ***	0.050 ***
Total risk	1369	0.325	0.270	0.173	0.060	1.625	0.046 ***	0.050 ***
Systematic risk	1366	0.008	0.010	0.128	−2.769	1.832	−0.002	0.000
Idiosyncratic risk	1369	0.938	0.780	0.590	0.096	4.024	0.305 ***	0.190 ***
Dividends	1207	0.848	1.000	0.360	0.000	1.000	−0.111 ***	0.000 ***
ERM Leverage	1369	0.846	0.900	0.148	0.247	0.994	−0.068 ***	−0.040 ***
Q Leverage	1369	13.154	7.740	20.163	0.078	147.219	−3.207 ***	−3.730 ***
Opacity	1369	0.035	0.010	0.070	0.000	0.584	-0.006	−0.010 ***
Revenue growth	1369	0.177	0.050	0.881	−0.710	8.098	0.207 ***	0.050 ***
ROA	1369	0.019	0.010	0.048	−0.357	0.431	0.009 ***	0.000 ***
Size	1369	9.530	10.310	2.815	−0.984	13.875	−2.700 ***	−2.110 ***
Slack	1369	0.061	0.030	0.105	0.000	0.769	0.043 ***	0.000
Big three (Ratings)	1369	0.522	1.000	0.500	0.000	1.000	−0.314 ***	−1.000 ***
Big four (Audit)	1369	0.888	1.000	0.316	0.000	1.000	−0.221 ***	0.000 ***
Industry Div.	1369	0.970	1.000	0.171	0.000	1.000	−0.017 *	0.000 *
International Div.	1369	0.768	1.000	0.422	0.000	1.000	−0.264 ***	0.000 ***

Note: ***, **, * denote the statistical significance at the 1%, 5%, 10% levels, respectively.

On average, ERM adopters in our sample have lower Tobin's Q, total risk, and idiosyncratic risk than non-adopters. While differences in the results of these univariate tests are pronounced, these results do not account for other factors that impact performance or risk. Moreover, estimates show that 52.2% of firms with an ERM program are rated by one of the three biggest rating agencies, and 88.8% are audited by one of the big four auditing firms. In addition, approximately 97% of the firms provide insurance products across various industries, and 76.8% operate internationally across different markets.

Table 3 presents Pearson's correlation matrix for all the variables. Overall, these correlations are weak to moderate. We computed the variance inflation factors for the independent variables and found no evidence of multicollinearity in the data.[5] Tobin's Q, total risk, and idiosyncratic risk negatively correlate with ERM, whereas the correlation between ERM and systematic risk is positive. The correlation between firm-level variables and ERM shows mixed signs. However, there is a positive correlation between variables that proxy for external monitoring agents and diversification with ERM. These correlations are pairwise, and therefore their signs do not account for the impact of other variables.

Table 3. Pearson Correlation Coefficients.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
1. ERM	1.000																
2. Tobin's Q	−0.147	1.000															
3. Total risk	−0.129	−0.078	1.000														
4. Systematic risk	0.008	−0.005	0.003	1.000													
5. Idiosyncratic risk	−0.248	0.159	0.157	0.005	1.000												
6. Dividends	0.141	−0.016	−0.197	−0.021	−0.313	1.000											
7. ERM Leverage	0.221	−0.472	0.002	−0.002	−0.322	0.059	1.000										
8. Q Leverage	0.076	−0.176	0.152	−0.083	−0.005	−0.090	0.376	1.000									
9. Opacity	0.038	0.319	−0.103	−0.004	0.012	0.046	−0.335	−0.193	1.000								
10. Revenue growth	−0.113	0.033	0.167	0.004	0.135	−0.082	−0.145	−0.020	0.030	1.000							
11. ROA	−0.093	0.566	−0.155	−0.014	0.099	0.115	−0.586	−0.243	0.200	0.026	1.000						
12. Size	0.460	−0.421	−0.176	0.005	−0.496	0.253	0.568	0.163	−0.215	−0.137	−0.261	1.000					
13. Slack	−0.198	0.517	0.069	0.001	0.264	−0.067	−0.517	−0.189	0.241	0.055	0.383	−0.470	1.000				
14. Big three (Ratings)	0.302	−0.216	−0.079	−0.029	−0.162	0.184	0.243	0.162	−0.050	−0.080	−0.140	0.524	−0.188	1.000			
15. Big four (Audit)	0.336	−0.212	−0.147	−0.004	−0.420	0.294	0.356	0.105	−0.080	−0.091	−0.164	0.560	−0.290	0.220	1.000		
16. Industry Div.	0.047	−0.043	−0.139	−0.002	−0.224	0.137	−0.057	−0.002	−0.020	−0.112	0.075	0.188	−0.061	0.175	0.155	1.000	
17. International Div.	0.300	−0.256	−0.093	0.030	−0.441	0.303	0.325	0.065	−0.023	−0.115	−0.175	0.476	−0.341	0.276	0.336	0.066	1.000

5. Main Results

We begin with estimating a logistic regression to examine the determinants of ERM for insurance firms in our sample and report the results in Table 4. These results show that the coefficient of leverage is statistically significant and negative, indicating that firms adopting ERM reduce their leverage to reduce the likelihood of default.

Table 4. Logistic Regression ERM determinants.

Variable	***ERM***
ERM Leverage	−5.636 **
	(2.527)
Opacity	12.452 ***
	(4.284)
Size	2.308 ***
	(0.275)
Slack	−1.478
	(3.334)
Big three (Ratings)	2.646 ***
	(0.604)
Big four (Audit)	1.371 **
	(0.595)
Industry Div.	−0.302
	(1.111)
International Div.	0.862
	(0.920)
Constant	−19.584 ***
	(2.979)
Number of Obs.	1369
Number of Clusters (Firms)	80
Wald test (Chi-squared)	126.33 ***

Note: ***, **, * denote the statistical significance at the 1%, 5%, 10% significance levels; robust standard errors are reported below parameter estimates in parentheses.

Firms with more opaque assets are more likely to adopt ERM. Implementing ERM increases the firms' ability to better assess their own risk profile, which may reduce information asymmetry between the firm and its stakeholders. This is likely when firms attempt to improve risk management to reduce the likelihood of financial distress. As expected, larger firms are more likely to engage in ERM. These firms have more resources to adopt an enterprise-wide risk management program to mitigate risks and address continuously evolving directives that regulate the insurance industry and the continuously changing risk environment. Consistent with Hoyt and Liebenberg (2011), firms rated by one of the big three rating agencies are more likely to engage in ERM. The results also show that hiring one of the big four auditing firms positively correlates with ERM. Beasley et al. (2005) explain that firms that employ high-quality auditors are more likely to commit to an integrated risk management framework, provided that it will be reflected in annual reviews. ERM engagement signals the firm's commitment to improving risk management systems to external monitoring agents, which would be reflected positively in their financial review. Lastly, the Wald test indicates that the coefficients are simultaneously different from zero.

Next, we estimate a full maximum likelihood treatment effects model that addresses the endogeneity bias, explained in Section 3. The results are reported in Table 5. The most important result is that ERM adopters are valued 53.4% higher than non-adopters.

Regarding the control variables, results show a negative relationship between size and performance. Larger insurers have greater bureaucracies and may face more complex agency problems (Lechner and Gatzert 2018), negatively impacting firm value. Consistent with Hoyt and Liebenberg (2011), ROA is positively associated with Tobin's Q.

Table 5. Full maximum likelihood treatment effects model—Tobin's Q.

Variable	***Tobin's Q***	***ERM***
ERM	0.534 ***	
	(0.101)	
Dividends	−0.014	
	(0.054)	
ERM Leverage		−0.221
		(1.036)
Q Leverage	0.00004	
	(0.001)	
Opacity		3.778 ***
		(1.185)
Revenue growth	−0.014	
	(0.014)	
ROA	5.768 ***	
	(1.315)	
Size	−0.084 ***	0.170 ***
	(0.019)	(0.049)
Slack		0.706
		(0.855)
Big three (Ratings)		0.209
		(0.128)
Big four (Audit)		0.850 ***
		(0.288)
Industry Div.	−0.145	−0.406
	(0.205)	(0.631)
International Div.	−0.105	0.215
	(0.082)	(0.229)
Constant	1.756 ***	−1.902 *
	(0.243)	(0.995)
Number of Obs.	1207	
Number of Clusters (Firms)	77	
Wald test (Chi-squared)	58.60 ***	
Wald test of indep. equations	17.23 ***	

Note: ***, **, * denote statistical significance at the 1%, 5%, 10% significance levels; robust standard errors are reported below parameter estimates in parentheses.

Concerning the ERM equation, the results show that firm size, opacity, and being audited by one of the big four audit firms remain significant determinants of ERM adoption. The Wald test for independent equations rejects the null hypothesis that residuals from Equations (2) and (3) are uncorrelated and support their joint estimation.

Next, we report results using a full maximum likelihood treatment effects model to assess the relationship between adopting ERM and the total risk measure in Table 6.

These results show that ERM engagement is effective in reducing a firm's total risk by 28%. Dividend payment events are negatively associated with total risk. This result is consistent with the risk reduction hypothesis (Grullon et al. 2002; Von Eije et al. 2014; Tripathy et al. 2021), suggesting that firms use these events to convey information to market participants about the reduction in their riskiness. Leverage, however, amplifies total risk. Evidence in the literature indicates that taking on leverage increases the probability of financial distress (Pagach and Warr 2011). In addition, although weakly significant, size is positively associated with total risk. Larger firms engage in complex resource administration and management to enable ERM (Beasley et al. 2005; Nocco and Stulz 2006). These firms face significant uncertainties (Altuntas et al. 2011) and, more importantly, experience volatile cash flows (Pagach and Warr 2011), which in turn will be reflected in stock price volatility. The results also show that insurers benefit from industry diversification. Industry diversification may yield better outcomes from economies of scope, with higher diversification leading to a decrease in operational and financial risks (Pagach and Warr 2011). However, international diversification increases firm total risk, possibly due to a higher number and complexity of risks and the need to comply with different national regulatory requirements (Hoyt and Liebenberg 2011). Reeb et al. (1998) argue that firms with foreign sales have higher systematic risk. These firms tend to have higher cash flow volatilities from international activities that are not offset adequately by diversification benefits. The Wald test for independent equations rejects the null hypothesis that the residuals from Equations (2) and (3) are uncorrelated and supports their joint estimation.

Table 6. Full maximum likelihood treatment effects model—Total Risk.

Variable	***Total Risk***	***ERM***
ERM	−0.280 ***	
	(0.033)	
Dividends	−0.051 **	
	(0.020)	
ERM Leverage		−0.782
		(0.554)
Q Leverage	0.001 ***	
	(0.0004)	
Opacity		2.243 **
		(1.090)
Revenue growth	0.013	
	(0.010)	
ROA	−0.224	
	(0.200)	
Size	0.009 *	0.184 ***
	(0.005)	(0.045)
Slack		−0.794
		(0.736)
Big three (Ratings)		0.135
		(0.116)
Big four (Audit)		0.609 **
		(0.277)
Industry Div.	−0.133 **	−0.461
	(0.067)	(0.359)
International Div.	0.057 *	0.173
	(0.031)	(0.202)
Constant	0.543 ***	−1.005 *
	(0.057)	(0.606)
Number of Obs.	1207	
Number of Clusters (Firms)	77	
Wald test (Chi-squared)	179.19 ***	
Wald test of indep. equations	42.44 ***	

Note: ***, **, * denote statistical significance at the 1%, 5%, 10% significance levels; robust standard errors are reported below parameter estimates in parentheses.

We extend the analysis to quantify the effect of ERM adoption on systematic and idiosyncratic risks and report the results in Table 7. The results reported in Panel A show that ERM engagement reduces systematic risk by 16.5%. Similar to the results from Table 6, we find a positive association between size and systematic risk. However, leverage is negatively related to systematic risk. Previous studies have not investigated the relationship between leverage and systematic risk directly; however, Berry-Stölzle and Xu (2018) suggest that ERM adoption reduces a firm's cost of capital, thus decreasing systematic risk.

Results reported in Panel B of Table 7 show that ERM adoption effectively reduces exposure to idiosyncratic risk by 74.3%. Dividend payment events are negatively associated with idiosyncratic risk, which is consistent with the signaling argument discussed above. The results also show that international diversification reduces idiosyncratic risk, which is consistent with rational expectations. Firms may face a higher number of risks and more complex risks (Hoyt and Liebenberg 2011); however, compliance with various national regulations may improve their ability to manage their risk. Opacity and size are significant determinants of ERM engagement in the estimates of the treatment effects model for both systematic and idiosyncratic risk, whereas audit appears to be a significant determinant in the latter. The Wald test supports the joint estimation of the risk and ERM equations.

Table 7. Full maximum likelihood treatment effects model—Systematic and idiosyncratic risks.

	Panel A		*Panel B*	
Variable	***Systematic Risk***	***ERM***	***Idiosyncratic Risk***	***ERM***
ERM	−0.165 **		−0.743 ***	
	(0.069)		(0.151)	
Dividends	−0.010		−0.203 ***	
	(0.009)		(0.076)	
ERM Leverage		−0.390		0.924
		(0.560)		(1.119)
Q Leverage	−0.0006 **		0.003	
	(0.0003)		(0.002)	
Opacity		1.724 **		4.384 ***
		(0.849)		(1.375)
Revenue growth	0.0004		−0.024	
	(0.001)		(0.022)	
ROA	−0.086		−0.230	
	(0.073)		(1.416)	
Size	0.010 **	0.149 ***	−0.016	0.128 ***
	(0.005)	(0.038)	(0.022)	(0.045)
Slack		−0.602		−1.091
		(0.670)		(0.760)
Big three (Ratings)		0.225		−0.061
		(0.146)		(0.170)
Big four (Audit)		0.362		1.006 ***
		(0.243)		(0.260)
Industry Div.	−0.011	−0.419	−0.385	−0.405
	(0.029)	(0.384)	(0.323)	(0.337)
International Div.	0.030	0.191	−0.241 **	0.207
	(0.018)	(0.212)	(0.122)	(0.242)
Constant	0.030	−0.849	2.289 ***	−2.304 **
	(0.030)	(0.607)	(0.402)	(1.021)
Number of Obs.	1207		1207	
Number of Clusters (Firms)	77		77	
Wald test (Chi-squared)	12.15		60.77 ***	
Wald test of indep. equations	4.95 **		28.72 ***	

Note: ***, **, * denote statistical significance at the 1%, 5%, 10% significance levels; robust standard errors are reported below parameter estimates in parentheses.

Overall, we find that ERM adoption leads to value creation for insurance firms in our sample, which is realized by improved performance and reduced risks as indicated by the market-based variables used in our analysis.

6. Conclusions

This paper examines the impact of adopting ERM on European publicly listed insurance firms' performance and risks using a 24-year long dataset. We identify firms that adopt ERM using a text search method proposed by Hoyt and Liebenberg (2011). Using logistic regression, we find that the determinants of ERM adoption in our sample include firm financial variables such as leverage, opacity, and size and variables that identify leading monitoring agents who provide credit ratings and auditing services.

Our assessment of the impact of ERM adoption on firm performance and risks relies on using the full maximum likelihood treatment effects model, which addresses the endogeneity bias arising from variables affecting ERM adoption and performance (or risks) simultaneously. We summarize our contributions as follows: First, we report a new result that identifies, for our sample, a significant value premium for ERM adopters.

Second, we show that adopting ERM reduces firm total risk significantly. This result holds for systematic and idiosyncratic risks. These findings are new, and they complement existing results in relevant ERM studies that have considered other risks such as default risk, earnings volatility, and solvency (Altuntas et al. 2011; Berry-Stölzle and Xu 2018; Pagach and Warr 2011; Nguyen and Vo 2020).

The reported effects of ERM adoption by publicly listed insurance firms in our sample have important implications for policymakers and regulators concerned about the financial health and stability of the insurance industry. Adopting ERM programs by insurance firms provides another corporate governance mechanism that improves firm value and reduces risks. Our findings lead us to recommend the adoption of effective ERM programs which may facilitate the application of Solvency 2 standards and, in particular, selected provisions set for Pillar 2, which focus in part on risk governance and setting up processes that enable the firm's own risk and solvency assessment (OSRA). In addition, ERM adoption may add value to the role of monitoring agents, and this appeals to investors in insurance firms as our results have shown that these firms have realized, on average, significant value premiums and reduced risks (total risk, systematic, and idiosyncratic risks).

We note, however, that we have only considered publicly listed insurance firms due to data limitations. Several firms operating in the insurance industry in Europe are not listed on stock exchanges. For example, Nguyen and Vo (2020) indicate that many insurance firms in France are mutual corporations that are not publicly listed. Hence, considering non-listed insurance firms together with listed firms provides a more comprehensive assessment for the insurance industry and forms the basis of a potential extension for this paper.

Author Contributions: Conceptualization, D.J.J.; methodology, D.J.J.; software, D.J.J. and S.M.A.; validation, D.J.J.; formal analysis, D.J.J.; investigation, D.J.J.; data curation, S.M.A.; writing—original draft preparation, D.J.J. and S.M.A.; writing—review and editing, D.J.J.; supervision, D.J.J.; project administration, D.J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research has not received funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset was obtained from subscription databases identified in the Data Section of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 The dataset covers a long period (1995–2018) relative to other studies. During this period several directives and regulations have been issued across European countries and have impacted risk management systems in European insurance firms. Pradier and Chneiweiss (2017) document various directives and regulations that have impacted the evolution of the insurance industry in Europe since the 1970s. More recently, the European Parliament and the Council agreed that the Solvency II Directive (including the amendments introduced by Omnibus II) should apply as of 1 January 2016.

2 McShane et al. (2011) discuss the differences between TRM and ERM. The former is silo-based and uncoordinated focusing mainly on financial risk, whereas the latter is broader and includes operational and strategic risks besides financial risks.

3 The terms we search for include: 'chief risk officer', 'risk management director', 'enterprise risk management', 'holistic risk management', 'strategic risk management', 'risk committee', 'risk' and various other synonyms including: 'enterprise-wide risk management', 'integrated risk management', 'firm-wide risk management', 'group risk management', 'comprehensive risk management', 'risk coordinator', 'risk manager', 'COSO', 'ORSA' and 'integrated framework' and other.

4 Insurance firms in this sample are headquartered in Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, the United Kingdom, Greece, Iceland, Ireland, Isle of Man, Italy, Luxembourg, Malta, Netherlands, Norway, Spain, Sweden, and Switzerland. Firms are included in the dataset if they have four years or more of data.

5 Estimates of the variance inflation factor for independent variables are lower than 10.

References

Ai, Jing, Vickie Bajtelsmit, and Tianyang Wang. 2018. The Combined Effect of Enterprise Risk Management and Diversification on Property and Casualty Insurer Performance. *The Journal of Risk and Insurance* 85: 513–43. [CrossRef]

Allayannis, George, and James P. Weston. 2001. The Use of Foreign Currency Derivatives and Firm Market Value. *The Review of Financial Studies* 14: 243–76. [CrossRef]

Altuntas, Muhammed, Thomas R. Berry-Stoelzle, and Robert E. Hoyt. 2011. Implementation of Enterprise Risk Management: Evidence from the German Property-Liability Insurance Industry. *The Geneva Papers on Risk and Insurance. Issues and Practice* 36: 414–39. [CrossRef]

Anton, Sorin Gabriel, and Anca Elena Afloarei Nucu. 2020. Enterprise Risk Management: A Literature Review and Agenda for Future Research. *Journal of Risk and Financial Management* 13: 281. [CrossRef]

Baxter, Ryan, Jean C. Bedard, Rani Hoitash, and Ari Yezegel. 2013. Enterprise Risk Management Program Quality: Determinants, Value Relevance, and the Financial Crisis. *Contemporary Accounting Research* 30: 1264–95. [CrossRef]

Beasley, Mark S., Richard Clune, and Dana R. Hermanson. 2005. Enterprise Risk Management: An Empirical Analysis of Factors Associated with the Extent of Implementation. *Journal of Accounting and Public Policy* 24: 521–31. [CrossRef]

Beasley, Mark, Don Pagach, and Richard Warr. 2008. Information Conveyed in Hiring Announcements of Senior Executives Overseeing Enterprise-Wide Risk Management Processes. *Journal of Accounting, Auditing & Finance* 23: 311–32. [CrossRef]

Berry-Stölzle, Thomas R., and Jianren Xu. 2018. Enterprise Risk Management and the Cost of Capital. *The Journal of Risk and Insurance* 85: 159–201. [CrossRef]

Bohnert, Alexander, Nadine Gatzert, Robert E. Hoyt, and Philipp Lechner. 2019. The Drivers and Value of Enterprise Risk Management: Evidence from Erm Ratings. *The European Journal of Finance* 25: 234–55. [CrossRef]

Chen, Carl R., Thomas L. Steiner, and Ann Marie Whyte. 2006. Does Stock Option-Based Executive Compensation Induce Risk-Taking? An Analysis of the Banking Industry. *Journal of Banking & Finance* 30: 915–45. [CrossRef]

COSO (Committee of Sponsoring Organizations of the Treadway Commission). 2004. Enterprise Risk Management—Integrated Framework. Available online: www.coso.org (accessed on 1 August 2021).

Denis, David J., Diane K. Denis, and Keven Yost. 2002. Global Diversification, Industrial Diversification, and Firm Value. *The Journal of Finance (New York)* 57: 1951–79. [CrossRef]

Eckles, David L., Robert E. Hoyt, and Steve M. Miller. 2014. The Impact of Enterprise Risk Management on the Marginal Cost of Reducing Risk: Evidence from the Insurance Industry. *Journal of Banking & Finance* 43: 247–61. [CrossRef]

Farrell, Mark, and Ronan Gallagher. 2015. The Valuation Implications of Enterprise Risk Management Maturity. *Journal of Risk and Insurance* 82: 625–57. [CrossRef]

Golshan, Nargess, and Siti Zaleha Abdul Rasid. 2012. Determinants of Enterprise Risk Management Adoption: An Empirical Analysis of Malaysian Public Listed Firms. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 6: 242–49.

González, Luís Otero, Pablo Durán Santomil, and Aracely Tamayo Herrera. 2020. The Effect of Enterprise Risk Management on the Risk and the Performance of Spanish Listed Companies. *European Research on Management and Business Economics* 26: 111–20. Available online: https://www.sciencedirect.com/science/article/pii/S2444883420303028 (accessed on 1 September 2020).

Grullon, Gustavo, Roni Michaely, and Bhaskaran Swaminathan. 2002. Are Dividend Changes a Sign of Firm Maturity? *The Journal of Business (Chicago, Ill.)* 75: 387–424. [CrossRef]

Hann, Rebecca N., Maria Ogneva, and Oguzhan Ozbas. 2013. Corporate Diversification and the Cost of Capital. *The Journal of Finance* 68: 1961–99. [CrossRef]

Hoyt, Robert E., and Andre P. Liebenberg. 2008. The Value of Enterprise Risk Management: Evidence from U.S. Insurance Industry. Available online: https://www.soa.org/globalassets/assets/files/resources/essays-monographs/2008-erm-symposium/mono-2008-m-as08-1-hoyt-abstract.pdf (accessed on 5 March 2021).

Hoyt, Robert E., and Andre P. Liebenberg. 2011. The Value of Enterprise Risk Management. *Journal of Risk and Insurance* 78: 795–822. [CrossRef]

Jensen, Michael C. 1986. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review* 76: 323–29.

Lechner, Philipp, and Nadine Gatzert. 2018. Determinants and Value of Enterprise Risk Management: Empirical Evidence from Germany. *The European Journal of Finance* 24: 867–87. [CrossRef]

Li, Qiuying, Yue Wu, Udechukwu Ojiako, Alasdair Marshall, and Maxwell Chipulu. 2014. Enterprise Risk Management and Firm Value within China's Insurance Industry. *Acta Commercii* 14: e1–e10. [CrossRef]

Liebenberg, André P., and Robert E. Hoyt. 2003. The Determinants of Enterprise Risk Management: Evidence from the Appointment of Chief Risk Officers. *Risk Management and Insurance Review* 6: 37–52. [CrossRef]

Lin, Yijia, Min-Ming Wen, and Jifeng Yu. 2012. Enterprise Risk Management: Strategic Antecedents, Risk Integration, and Performance. *North American Actuarial Journal* 16: 1–28. [CrossRef]

Lindenberg, Eric B., and Stephen A. Ross. 1981. Tobin's Q Ratio and Industrial Organization. *The Journal of Business (Chicago, Ill.)* 54: 1–32. [CrossRef]

Markowitz, Harry. 1952. Portfolio Selection. *Journal of Finance* 7: 77–91. [CrossRef]

McShane, Michael K., and Larry A. Cox. 2009. Issuance Decisions and Strategic Focus: The Case of Long-Term Care Insurance. *The Journal of Risk and Insurance* 76: 87–108. [CrossRef]

McShane, Michael K., Anil Nair, and Elzotbek Rustambekov. 2011. Does Enterprise Risk Management Increase Firm Value? *Journal of Accounting, Auditing & Finance* 26: 641–58. [CrossRef]

Nguyen, Duc Khuong, and Dinh-Tri Vo. 2020. Enterprise Risk Management and Solvency: The Case of the Listed Eu Insurers. *Journal of Business Research* 113: 360–69. [CrossRef]

Nocco, Brian W., and René M. Stulz. 2006. Enterprise Risk Management: Theory and Practice. *Journal of Applied Corporate Finance* 18: 8–20. [CrossRef]

Pagach, Donald, and Richard Warr. 2011. The Characteristics of Firms That Hire Chief Risk Officers. *The Journal of Risk and Insurance* 78: 185–211. [CrossRef]

Pathan, Shams. 2009. Strong Boards, Ceo Power and Bank Risk-Taking. *Journal of Banking & Finance* 33: 1340–50. [CrossRef]

Pradier, Pierre-Charles, and Arnaud Chneiweiss. 2017. The Evolution of Insurance Regulation in the Eu since 2005. In *Financial Regulation in the Eu: From Resilience to Growth*. Edited by Raphaël Douady, Clément Goulet and Pierre-Charles Pradier. Cham: Springer International Publishing, pp. 199–251.

Reeb, David M., Chuck C. Y. Kwok, and Hyungkee Young Baek. 1998. Systematic Risk of the Multinational Corporation. *Journal of International Business Studies* 29: 263–79. [CrossRef]

Smith, Clifford W., and Ross L. Watts. 1992. The Investment Opportunity Set and Corporate Financing, Dividend, and Compensation Policies. *Journal of Financial Economics* 32: 263–92. [CrossRef]

Stulz, René M. 2003. *Rethinking Risk Management. The Revolution in Corporate Finance*, 4th ed. Hoboken: Blackwell Publishing, pp. 367–84.

Stulz, René M. 1996. Rethinking Risk Management. *Journal of Applied Corporate Finance* 9: 8–25. [CrossRef]

Tahir, Izah Mohd, and Ahmad Rizal Razali. 2011. The Relationship between Enterprise Risk Management and Firm Value: Evidence from Malaysian Public Listed Companies. *International Journal of Economics and Management Sciences* 1: 32–41.

Tripathy, Niranjan, Da Wu, and Yi Zheng. 2021. Dividends and Financial Health: Evidence from U.S. Bank Holding Companies. *Journal of Corporate Finance* 66: 101808. [CrossRef]

Von Eije, Henk, Abhinav Goyal, and Cal B. Muckley. 2014. Does the Information Content of Payout Initiations and Omissions Influence Firm Risks? *Journal of Econometrics* 183: 222–29. [CrossRef]

Journal of
Risk and Financial Management

Article

The Implications of Derisking: The Case of Malta, a Small EU State

Simon Grima [1,*], Peter J. Baldacchino [2], Jeremy Mercieca Abela [2] and Jonathan V. Spiteri [1]

1 Department of Insurance, Faculty of Economics, Management and Accountancy, University of Malta, MSD 2080 Msida, Malta; jonathan.v.spiteri@um.edu.mt
2 Department of Accountancy, Faculty of Economics, Management and Accountancy, University of Malta, MSD 2080 Msida, Malta; peter.j.baldacchino@um.edu.mt (P.J.B.); jeremy.mercieca-abela.15@um.edu.mt (J.M.A.)
* Correspondence: simon.grima@um.edu.mt

Received: 13 August 2020; Accepted: 11 September 2020; Published: 18 September 2020

Abstract: In this study, we explore the emerging derisking phenomenon by identifying and analysing the main factors that are affected by, and the implications of, the derisking process by focusing on the key drivers and implications of derisking specific to Malta. To do this, we carried out 32 interviews with individuals who have a good or excellent level of expertise in derisking and administered a survey, completed by 296 participants who were filtered to ensure their level of expertise, resulting in 285 valid participant surveys. In total, between the interviews and the survey, we had 317 valid participants. Findings showed that to maximise the effectiveness of derisking, one needs to find the right balance of adequately managing risks without extinguishing business needs. This implies a need for the regulations to be balanced and proportionate. This study is a relevant contributor to future derisking to be conducted in Malta and serves as a benchmark for further studies. Moreover, this research project accentuates the need for increased awareness, knowledge and expertise of derisking in Malta. Consequently, the provision of education to professionals is important so that such professionals are able to keep abreast with all the latest developments regarding derisking and AML/CFT (antimoney laundering and combatting the financing of terrorism).

Keywords: derisking; drivers and implications of derisking; risk management; antimoney laundering (AML); combatting the financing of terrorism (CFT); financial services; Malta; small EU state; proportionality

1. Introduction

The implementation of a derisking strategy is a recent trend; many banks and financial institutions worldwide are undertaking derisking to simplify their business models, minimise their risk exposures and to comply with the increasing AML/CFT (antimoney laundering and combatting the financing of terrorism) regulations. Other key driving factors, which have led to many banks around the world opting for derisking, include a steady increase in costs necessary to ensure compliance with AML/CFT regulations, rising fines and penalties imposed for noncompliance and the lower profitability of certain customer bases.

Derisking is a process which involves the closure of the bank accounts of, and/or the termination of relationships with, clients categorised as being high risk after considering the risk appetite of the organisation (Durner and Shetret 2015). Nisar (2016) provides some examples of these types of clients, which can include money service businesses (MSBs), correspondent banks, embassies, international charities, nonprofit organisations (NPOs) and multinational corporations (MNCs). Derisking can be carried out in two ways: (1) Wholesale derisking involves the derisking of an entire category of customers (known as a customer base) or a particular sector. Examples of sectors which are often subject to wholesale derisking are MSBs, NPOs, charities and the defence sector (Durner and Shetret 2015). Alternatively, (2) derisking on a case-by-case basis is when a bank first considers the attributes of an individual customer and the relationship the bank has with such a customer, and then carries out a customer risk assessment. Based on all the information obtained, the bank finally decides on whether or not to carry out derisking (Artingstall et al. 2016).

Moreover, as part of their derisking strategy, banks and other financial institutions may also decide to limit the range of financial services they are willing to provide to these types of clients (Haley 2017). However, although it is thought that by engaging in derisking and terminating relationships with the so-classified high-risk clients, banks and other financial institutions are able to minimise their risk exposures, this has significant negative repercussions on clients and such clients will end up bankless, ultimately leading to increased financial exclusion (Durner and Shetret 2015).

Haley (2017) mentions three main forms of derisking:

1. When banks and other financial institutions decide to close the bank accounts of certain individuals or firms or when banks and other financial institutions opt not to open bank accounts for these individuals or firms in the first place. This form of derisking is also known as debanking. The costs arising from this type of derisking are largely borne by the individuals. Furthermore, these costs are usually only limited to restricting financial integration and do not represent a systemic risk. This holds true assuming that domestic banks are able to provide services for transactions that are not cross-border in nature.
2. When banks and other financial institutions restrict or completely withdraw services from money service businesses (MSBs; Haley 2018). MSBs are nonbank institutions that offer a variety of financial services at a much lower cost when compared to traditional banking institutions (Durner and Shetret 2015). Some of the services that MSBs provide include the transmission of money, currency exchange and the checking of cheques (Durner and Shetret 2015). One of the most noteworthy types of MSBs is money transfer organisations (MTOs). These are defined as financial companies, typically not banks, that are involved in the crossborder transfer of funds (IMF 2008, 2012). When it comes to the transfer of funds, MTOs can either use their own internal systems or access "crossborder banking networks" (IMF 2008, p. 9). MTOs provide services which are vital when it comes to the efficient transferring of remittances worldwide. Therefore, if MTOs are derisked, this implies that other informal channels for the transferring of funds will emerge. These informal channels may open pathways for money laundering and financing of terrorism, thus posing a significant risk to efforts being made regarding AML/CFT (antimoney laundering/combatting the financing of terrorism; Haley 2017).
3. The loss of correspondent banking relationships (CBRs; Haley 2018). These are an arrangement under which one bank (correspondent) holds deposits owned by other banks (respondents) and provides payment and other services to those respondent banks (CPMI 2003, p. 16). Essentially, in a CBR, domestic banks hold accounts in large international banks, and, in turn, such large international banks will provide access to a global payment and settlement system. As a result, if CBRs are severed, domestic banks will be unable to process payments and provide services to their clients. This represents a potential systemic risk and could lead to an inefficient global payment system (Haley 2017).

2. Rationale for the Study

Over the past few years, the derisking practice has also gained popularity in Malta, and, in fact, there are quite a few domestic and international banks operating in Malta that have engaged in, or are in the process of engaging in, derisking.

Given that at present, very limited research regarding the implications of derisking vis-à-vis the Maltese scenario has been carried out, this implies the need for such a study. This study fills such a gap by identifying and subsequently laying out the analysis of the factors that are affected by, and more importantly, the drivers and implications of, the derisking process.

Although we herein highlight the main drivers of the derisking process in a small EU state and island, as suggested by various researchers such as Bezzina et al. (2014), Briguglio (1995), and King (1993), given the jurisdiction's size and nature, the results can be used as small-scale laboratory test findings for more complex politics, regulations and policies of larger countries. Therefore, the findings of this study are of significance to international policymakers, regulators and governments.

3. The Objective and Research Questions of the Study

The researcher has one primary objective: to identify and analyse the implications that are affected by the derisking process. Moreover, underpinning this objective are two research questions:

RQ1: What are the main implication variables of a derisking strategy within the context of Malta?

RQ2: What are the key effects of derisking with regards to the Maltese scenario?

In doing this, we also check whether the drivers differ with the different demographics.

The aim is to understand what the implications variables are, which should be looked at when carrying out a derisking strategy, that is, a checklist to help risk managers develop a risk matrix and measure the implications of derisking as one single qualitative or quantitative value. This can also serve as a tool for policymakers.

4. Why Was Malta Chosen as the Focal Point of This Study?

As noted above, over the years, many leading and well-known researchers such as King (1993), Briguglio (1995), Baldacchino (2006), Bezzina et al. (2014), and Xuereb et al. (2019) have made use of small island states like Malta in order to understand the complex financial environments, dynamics, politics, policies and regulations of larger countries (Xuereb et al. 2019). Therefore, small island states are utilised as laboratories on a smaller scale (Xuereb et al. 2019; Pirotta 2001).

Moreover, the principle of proportionality is of utmost importance to ensure that smaller states still have a voice. Proportionality means that the EU should only take the action that is necessary to achieve its aims. This principle is set out in Treaty 5 of the EU under Article 5. In reality, when it comes to the EU's decision-making process, it is usually the larger states that have more prominence and are able to influence the outcome of final decisions. Therefore, in practice, the concept of proportionality is not always adhered to and is not effectively managed (Xuereb et al. 2019). Adopting an approach which is proportional involves taking into account the size, nature, systemic importance, complexity and risk profile of the stakeholders involved and the services being offered. Therefore, it is important that regulations are tailored according to such requirements. This leads to lower costs related to compliance and a decreased regulatory burden for smaller and less complex organisations that might otherwise experience an undue reduction in their competitiveness. Consequently, this study is of significant importance as it gives smaller states the opportunity to have a much-needed voice and allows further understanding to be obtained regarding the implications of derisking vis-à-vis smaller jurisdictions (Xuereb et al. 2019; Lautenschläger 2017; Grima et al. 2016).

5. Literature

5.1. The International De-Risking Landscape

5.1.1. Traditional Risk Management

The traditional risk management approach used reactive models, defined as a managerial process, an administrative process or a decision-making process of planning, organizing, leading, and controlling the organization's activities to lower the adverse effects of any accidental losses that a firm might incur. It was more of a reporting and monitoring function.

It is a risk-based approach, which places responsibility on various departments and business units and the focus is on pure risk (hazard risk where the consequences may or may not be losses) and refers to individual risks as though they do not interact with each other. Therefore, this approach does not align with the risk management needs of an organisation since risks should be treated as a whole (taking into consideration the direct and indirect effects). The result is, therefore, unsatisfactory due to different types of risks being managed and treated independently. This fragmented approach to risk cannot work within an aggregated approach to risk throughout an organisation (Head 1973).

The complexity, volatility and unpredictability of the current economic and social environment are a constant reminder of the many risks we need to face to reach objectives. The traditional approach is, therefore, no longer sufficient, and it is necessary for today's companies to invest significant resources to identify, measure and manage risks and ensure longevity. The risk management approach has a major role in an organisation's ability to avoid, reduce and turn risks into opportunities.

5.1.2. The Context of Derisking

Over the past years, regulatory scrutiny in the global financial services sector has significantly increased, especially regarding financial crime and AML/CFT regulation (Badour et al. 2017). The reaction of banks and other financial institutions to this new regulatory environment was to strengthen and invest more in their compliance and risk management functions (Badour et al. 2017). As costs associated with regulation and compliance continued to increase, some banks and other financial institutions started carrying out cost–benefit analyses with regards to certain activities and operations they carry out. Such banks and other financial institutions found that in some cases, it is more beneficial to stop servicing certain customer segments and/or to exit specific product lines (Badour et al. 2017).

In the current global landscape which comprises different actors concerned with derisking, there are many of such actors (e.g., the G20) that view this phenomenon as being an issue that is related to financial stability (HSC and ECNL 2018). This is because derisking has a considerable impact on CBRs, which may, consequently, negatively impact trade financing and economic development. Other actors such as the World Bank believe that derisking is broader in scope and may even affect financial integrity and inclusion (HSC and ECNL 2018). It is important that banks and other financial institutions do not only aim to safeguard their margins but also act as "good global citizens [by] providing a public service" (HSC and ECNL 2018, p. 12).

On the other hand, according to regulatory framework architects, derisking is not the result of stricter AML/CFT regulations, but principally, the result of banks and other financial institutions not properly applying and/or misinterpreting the requirements related to AML/CFT. At the same time, such banks and financial institutions complain that clean money is disappearing into banking channels that are unregulated (HSC and ECNL 2018).

5.1.3. The Political Landscape

Long-established political concerns with regards to financial crime and terrorism collide with an increased demand for regulation that creates "legal, political and operational imperatives that result in derisking" (HSC and ECNL 2018, p. 13).

The pressure of ever-increasing international policy responses when it comes to financial crime and terrorism has transformed how the global financial system works and operates (HSC and ECNL 2018). In the 1980s, international architecture regarding the surveillance and control of customers was introduced. Following this was the establishment of enforcement mechanisms, which mainly comprise of economic and financial sanctions. As a result, financial services providers are now also being held responsible for monitoring activities that are carried out by their customers (HSC and ECNL 2018).

The Financial Action Task Force (FATF) is one of the most important and powerful international organisations with respect to derisking. The FATF was established in 1989 by the G7 countries with the primary aim of developing a framework to combat money laundering that could be applied on an international level. Upon its inception, this organisation imposed obligations on three levels (HSC and ECNL 2018):

1. Firstly, financial services providers had to start conducting due diligence and identifying customers engaging in illicit activities.
2. On a domestic level, states had to start investigating suspicious transactions and financial crime.
3. Finally, on an international level, cross-border cooperation with respect to financial crime investigations and the freezing of assets of individuals involved in such financial crimes were encouraged.

Following the September 11 attacks in 2001, the mandate of the FATF was broadened in scope to include the combatting of the financing of terrorism so as to stop money from reaching terrorist organisations. Here, the ultimate aim is that of increasing peace and security on a global scale (HSC and ECNL 2018). The FATF also has the power to ensure that countries implement their recommendations. Countries that do not comply are blacklisted from the global financial system until they achieve full compliance once again (HSC and ECNL 2018).

The logic behind the transformation that the FATF underwent over the years can also be applied to the AML/CFT framework, which focuses on areas of failure and does not take into account other concerns such as human rights and accountability (HSC and ECNL 2018). The unintended spill-offs of this approach, such as derisking, are oftentimes tolerated and not questioned. It is only when derisking leads to funds being pushed into shadow banking channels that it starts becoming a concern for many prominent stakeholders (HSC and ECNL 2018).

5.1.4. The Regulatory Landscape

To minimise the risk of noncompliance, the banking and financial services sectors have opted to internalise AML/CFT regulations and are regulated both statutorily and on a "self-regulation" basis. Individually, banks have also developed numerous internal mechanisms to ensure compliance and mitigate risk (HSC and ECNL 2018).

The role of central banks is to make sure that AML/CFT regulations that have been incorporated into the banking regulatory framework are actually applied in practice (HSC and ECNL 2018). "Breaches of sanctions regimes or failures to conduct proper due diligence" (HSC and ECNL 2018, p. 16) can lead to the imposition of fines and even the threat of criminal prosecution. In this regard, the biggest driver of derisking is not the cost of financial penalties, but the enforcement actions' reputational costs if the banks involved are allegedly found to be colluding with terrorist groups or organised criminals (HSC and ECNL 2018). Consequently, banks are revising their risk appetites and implementing more widespread derisking (HSC and ECNL 2018).

The internal compliance mechanisms that banks are introducing to avoid the risks and fines for not being compliant are usually developed by individual third-party financial services providers (HSC and ECNL 2018). Therefore, determining whether individual clients or entire business segments are perceived as being high-risk is a commercial decision based on the algorithmic decision-making of a third-party actor and not on an impartial and objective risk assessment (HSC and ECNL 2018). The industry that provides such compliance services has thus normalised the derisking process (HSC and ECNL 2018).

When it comes to derisking, it is important that relevant stakeholders take into consideration not only certain risk management tools that are provided by the industry (HSC and ECNL 2018), but also the "political economy of the compliance sector" (HSC and ECNL 2018, p. 16).

Derisking is not always a rational process based on sound cost–benefit analyses. It can be "the result of (highly politicised) processes predicated on the creation of 'suspect communities' to maintain an adequate supply of risk to be managed" (HSC and ECNL 2018, p. 17). This is not an appropriate baseline for the prevention of financial crime and terrorism and leads to statutory regulation not being applied on a proportionate basis (HSC and ECNL 2018).

5.1.5. The Relationship between Money Laundering, Corruption and Bribery

Money laundering and corruption are two serious offences which are very much linked to one another. The prevalence of money laundering in any particular country is a clear indicator that corruption is also prevalent and vice versa (Mugarura 2016). The ideal environment in which money laundering and corruption thrive is one which is characterised by a lack of adequate oversight by regulatory authorities, bad governance and weak laws and regulations (Mugarura 2016).

One very common type of corruption is bribery. The underlying objective of bribery is to influence the opinions and/or behaviour of certain individuals such as public officials. Bribes often entail gifts and rewards; however, they are not always monetary in value. One example of a nonmonetary bribe is hospitality. The prevalence of bribery within a particular country leads to both short-term and long-term economic and reputational damage (Clark 2019; Consiglio and Grima 2012).

Since money laundering and corruption are intertwined, it is important that countries not only introduce anticorruption policies but also adopt robust AML strategies and make sure that such strategies are effectively implemented and enforced. Proper enforcement of AML strategies helps to hold corrupt public officials accountable for their actions (Mugarura 2016).

The manifestation of money and corruption hinders the ability of individual countries to pursue development goals on a national level and also facilitates crimes such as human, drug and small arms trafficking, terrorism financing, as well as prostitution (Mugarura 2016).

Corruption and bribery are often mistaken for one another or used interchangeably since their meanings are very similar. However, to prevent, manage and mitigate the problems they cause, it is important to understand the differences between them.

Corruption is defined as any actions by a person in a high position to take an unfair advantage, usually for personal gain. This is the abuse by persons with entrusted power for private gain, which can be of any size, commercial or political.

Bribery, on the other hand, is a type of corruption. An example is offering gifts to individuals so as to influence their opinions or behaviour (Consiglio and Grima 2012).

5.1.6. A Sustainable Model for AML in the United Nations Development Goals

In 2015, all the member states of the United Nations (UN) adopted 17 sustainable development goals (SDGs) that focus on tackling a number of prominent issues faced by all countries around the world. Since money laundering destabilises domestic economies, it needs to be scaled down so that such SDGs can then be achieved (Dobrowolski and Sulkowski 2019).

According to Dobrowolski and Sulkowski (2019), money laundering "distorts the efficient and effective outcomes of economic activities and serves as a catalyst for other types of crime, including corruption, fraud, drug trafficking, and terrorism" (p. 1). Therefore, money laundering can have a crippling effect on economies worldwide and leads to distorted international finances. Furthermore, money laundering significantly decreases integrity, in turn leading to reduced transparency and accountability. Ultimately, this results in higher levels of public mistrust, adversely impacting global markets and stifling innovation and development (Dobrowolski and Sulkowski 2019).

Supreme audit institutions (SAIs) have the power and capacity to audit, investigate and evaluate AML systems and antimoney laundering organisations (AMLOs) to assess whether such systems and organisations are effectively fulfilling their obligations and tasks. In addition, SAIs have the responsibility of providing decision-makers and governments with accurate and reliable information, which allows them to establish AML systems which are efficient and effective. Therefore, SAIs promote good governance. Notwithstanding, the capacity of SAIs is sometimes limited due to a lack of standards and guidelines regarding the auditing of AML systems and organisations (Dobrowolski and Sulkowski 2019).

To combat money laundering in an effective manner, Dobrowolski and Sulkowski (2019) proposed the use of the MAMA (multiple antimoney laundering activity) model as a universal framework of AML and an assessment tool for auditors. This model is beneficial because it can assist all types of organisations from all countries around the world. Moreover, this model allows different organisations to compare their different AML systems, meaning that benchmarks can be established (Dobrowolski and Sulkowski 2019). The MAMA model consists of 8 criteria (Dobrowolski and Sulkowski 2019):

- Five of these criteria are "enablers" that focus on the core functions of AMLOs—(1) leadership, (2) people, (3) policy and strategy, (4) partnership and resources, and (5) processes.
- The other three criteria are "results" that focus on the achievements of AMLOs—(1) employee results, (2) institutional stakeholders results, and (3) society results.

5.2. The Drivers of Derisking

When analysing the driving factors of the derisking process, it is imperative to take into account recent developments in the spheres of banking and regulation (Artingstall et al. 2016). The approach being adopted regarding risk is changing as the priority for banks and other financial institutions is to decrease their costs. Furthermore, banks and financial institutions aim to reduce risk-weighted assets and slim down their business in order to lower their risk profiles (Artingstall et al. 2016).

The derisking process is driven by an amalgamation of factors such as the ever-increasing AML/CFT regulations, higher compliance costs, more stringent and demanding prudential requirements, increased risk management controls and low profitability of certain business areas (Langthaler and Niño 2017).

To facilitate comprehension, these aforementioned drivers of derisking and other factors are going to be grouped under specific subheadings and explained in more detail in the following pages.

5.3. Decreasing Risk Appetites

In response to the 2008 global financial crisis, many banks started scaling back their risk appetites and adopting a risk-based approach (RBA) to AML/CFT. By adopting an RBA to AML/CFT, banks reduced the extent to which the services they provide to customers can be abused for the purpose of financial crime. This is done by such banks discontinuing relationships that they believe present too high a risk of this abuse (Artingstall et al. 2016). The reduction in the risk appetites of banks and other financial institutions has led to more widespread derisking of clients that do not fall within such decreased risk appetites.

The exiting of relationships with customers perceived as presenting the highest risk helps minimise counterparty risk. Consequently, banks can improve their risk profile and ensure compliance with the RBA to AML/CFT regulations, which many regulators advocate in favour of (Badour et al. 2017). Furthermore, it is beneficial for banks to have strong practices associated with risk management in place as this leads to reduced regulatory risk and a lower likelihood of regulatory action (Badour et al. 2017).

It is paramount that any organisation determines its level of risk appetite as this ensures that effective decisions are taken when determining strategies to manage risk by, for example, transferring or accepting part of or the whole risk and determining which tools to utilise when doing this. ISO 31000 defines risk management as being "the amount and type of risk that an organisation is willing to pursue or retain" (ISO 2009). Such risk is pursued and retained so that the organisation's strategic objectives can be achieved (Alix et al. 2015; Bromiley et al. 2015; Lamanda and Voneki 2015; as cited in Zhang 2016). In the context of derisking, banks and other financial institutions usually derisk the clients that are deemed to be too high-risk (in the light of the appetite of the shareholders, transmitted through the board of directors) and fall outside of their risk appetite. Moreover, the implementation of proper risk mitigation strategies decreases the adverse effects of risks that an organisation is exposed to.

Risk appetite is sometimes also defined as that amount of financial crime residual risk that a bank or financial institution is willing to have onboard and the view of how efficiently and effectively it can manage such risks (Artingstall et al. 2016). Risk management involves two complementary types of mitigation—enhanced due diligence and continuous monitoring. Beyond such mitigation strategies, interventions will then involve some form of derisking (Artingstall et al. 2016).

RBAs with respect to financial crime identify and analyse the risk that is associated with factors that include the type of business involved, the sector, the occupation, jurisdiction and political risk, distribution channels and the products or services required by the customer (Artingstall et al. 2016). These customer risk ratings allow banks to assign scores to specific customers and then determine whether the financial crime risks involved are acceptable and fall within the particular bank's predetermined parameters. This all depends on the financial crime risk appetite that the bank has (Artingstall et al. 2016).

5.4. Increasing AML/CFT and Regulatory Compliance

5.4.1. AML/CFT Compliance

With AML/CFT regulations always increasing and becoming ever more complex, this has resulted in higher costs related to compliance (Haley 2017). A survey on global antimoney laundering carried out by KPMG in 2014 found that 78% of respondents, representing the most prominent global banks, stated that their banks reported a substantial increase in AML compliance, with 22% of such respondents revealing that their banks experienced a 50% increase from the years 2011 to 2014 (KPMG 2014). In 2014, HSBC spent $800 million on risk management and compliance, which was about a $200 million increase from the previous year (Arnold and Hughes 2014).

The main factor which has led to rising compliance costs is the imposition of enhanced customer due diligence (CDD) requirements (Haley 2018). These require banks to follow Know Your Customer (KYC) guidelines, which mandate that banks need to identify and verify their customers and also the ultimate beneficial owners of their legal entity customers (Langthaler and Niño 2017). Additionally, banks must comprehend the nature and purpose of the relationships with their customers and conduct monitoring that is ongoing. In the case of clients and countries that are both perceived as being higher risk, KYC requirements and due diligence can be very costly and time-consuming (Langthaler and Niño 2017).

Rising AML/CFT compliance costs have a magnified effect on jurisdictions that have smaller and more restricted banking markets as such jurisdictions are not always able to spread the fixed cost component of these AML/CFT compliance costs over a large enough number of transactions (Haley 2017).

In addition, rising compliance costs may disincentive larger banks to uphold certain interbank relationships that provide additional cover or more transactional options. This negatively affects small and medium-sized banks, which may end up having to stop servicing some of their customers and other banks (Artingstall et al. 2016).

In their study entitled Drivers and Impacts of Derisking, Artingstall et al. (2016) analysed banks that have engaged in derisking and closed several of their customers' bank accounts. The banks investigated were not willing to share a portion of these increased compliance costs for bank accounts with their customers by using the customers' risk rating as a basis. If these compliance costs could be calculated for each individual customer, there might have been customers willing to pay a premium to maintain their bank accounts (Artingstall et al. 2016). This is essentially a form of market failure, where if costs and resources are allocated more efficiently against the actual money laundering and terrorism financing risks involved, banks would need to derisk less (Artingstall et al. 2016).

Badour et al. (2017) argue that the derisking process actually leads to lower compliance costs and allows banks to operate with a greater level of efficiency. Exiting relationships with high-risk customers that typically require considerable compliance resources allow such resources to be used elsewhere. By focusing on fewer product lines that do not involve high compliance costs, this will allow banks to operate more effectively and simultaneously, also possibly boosting their profitability (Badour et al. 2017).

5.4.2. Regulatory Compliance

According to Durner and Shetret (2015), regulatory approaches often diverge across "state, national and international jurisdictions" (p. 10). Such divergence is usually considered to be an invisible cost associated with globalisation and results in increased costs related to compliance (Durner and Shetret 2015).

Apart from the significant fines, agreed settlements of regulatory action may include the imposition of certain business restrictions as part of such regulatory settlement for the organisation under consideration. This is known as regulatory derisking (Artingstall et al. 2016).

An example of a regulatory settlement reached when certain AML/CFT requirements are breached is a deferred prosecution agreement (DPA; Artingstall et al. 2016). With regards to DPAs, banks agree to voluntarily abide by several conditions in exchange for the suspension of their criminal charges. The conditions, stipulated in a DPA, comprise of financial penalties and improvements that must be made with regards to particular compliance measures and cooperation with the applicable regulatory and law enforcement agencies (Durner and Shetret 2015).

Costs associated with compliance of a regulatory nature may ultimately end up being shifted to customers in the form of increased fees, reduced availability of services and restricted credit (Durner and Shetret 2015). Low-income individuals and businesses that have low profit margins may not be able to sustain these newly added costs, leading to services being discontinued and these individuals and businesses being financially excluded (Durner and Shetret 2015).

5.4.3. Rising Fines and Penalties

The imposition of significant fines and penalties on banks and financial institutions for noncompliance with and breaches of AML/CFT requirements is another main driver of derisking (Haley 2018). Fines and penalties may also be issued if there are deficiencies in AML/CFT compliance and if sanctions are violated (Durner and Shetret 2015). Table 1, depicted below, taken from Haley (2018, p. 5), shows some examples of fines and penalties that were imposed on international banks for noncompliance.

After the end of the 2008 global financial crisis, regulators worldwide started facing increasing pressures to hold organisations accountable for misconduct (English and Hammond 2014). Nevertheless, such regulators have stipulated that the imposition of substantial fines and enforcement actions is only for the most egregious and sustained offenders (Durner and Shetret 2015).

Table 1. Fines and penalties imposed for noncompliance.

Bank	US$ (Billions)	Date
HSBC	1.9	December 2012
J.P. Morgan Chase	1.7	January 2014
BNP Paribas	8.9	July 2014
Commerzbank	1.5	March 2015

Source: author's compilation.

5.4.4. Declining Customer Profitability

Low profitability of certain customer bases is another factor which drives derisking. Indeed, there are some cases where "it is simply not worth the hassle and costs" (Durner and Shetret 2015, p. 17) to engage with a particular customer base. If the profitability associated with a specific customer base does not compensate for the higher risk, the additional costs that must be incurred and the further AML/CFT controls that must be implemented, the customer base is derisked (Durner and Shetret 2015).

Customer profitability is also important when it comes to assessing CBRs. Nowadays, banks and other financial institutions engage in a practice known as Know your Customer's Customer (KYCC), which means that the responsibility of banks and other financial institutions is not only limited to their customers but also the customers of their customers (Durner and Shetret 2015).

Since KYCC represents an additional risk for banks and other financial institutions engaging in CBRs (especially in high-risk jurisdictions) and is considered as being a costly and rigorous process, such banks and other financial institutions often resort to derisking (Durner and Shetret 2015).

Given the fact that KYCC is a highly intensive and expensive process, banks and other financial institutions may decide that the risk involved is not worth the reward with regards to certain high-risk customer bases. If this is the case, the subsequent closure of the bank accounts of these clients represents a clear market failure, especially considering the financial inclusion benefits derived if these sectors continue being serviced (Durner and Shetret 2015).

5.4.5. Reputational Concerns

The consequences of noncompliance are not only limited to fines and penalties. Other types of enforcement actions, e.g., limitations, when it comes to the provision of certain services can also result in the organisation incurring high financial costs (English and Hammond 2014).

Certain enforcement actions can lead to a bank incurring reputational damage, and this can have devastating effects. As regulatory scrutiny continues to increase, so do the chances of banks being found guilty of noncompliance with AML/CFT procedures and sanctions. Often, there is an element of uncertainty and concern regarding the bank's ability to survive the enforcement action. This can have an adverse effect on the relationship that such a bank has with investors, meaning that the share price may be impacted (Durner and Shetret 2015).

Since banks provide crucial services to vulnerable communities, this implies that the derisking of such communities can have negative ramifications when it comes to public relations (Durner and Shetret 2015). Therefore, the continued servicing of these vulnerable communities can lead to banks obtaining "reputation returns" (Durner and Shetret 2015, p. 12).

5.4.6. Higher Capital and Liquidity Requirements

A key driving factor that also influences the derisking carried out by banks and other financial institutions is the imposition of increased capital requirements and liquidity thresholds following the 2008 global financial crisis (Haley 2017).

Such higher requirements and thresholds have created an environment in which it is harder for banks to maintain profitable customer relationships, with some banks resorting to deleveraging (Artingstall et al. 2016). Moreover, many banks worldwide have opted to undertake a strategic review of their main business and functions, which has led to such banks focusing more on businesses that are the most core (FCA 2016).

5.5. The Factors and Implications of Derisking

The main factors, implications, and effects of derisking, identified through various sources of the existing literature, are described below:

5.5.1. Shift of AML/CFT Risk and Shadow Banking Channels

When banks and other financial institutions engage in derisking, this results in a shift of AML/CFT risk. This is because if CBRs are terminated, affected clients will then have to start relying on smaller banks and credit institutions to obtain the financial services they require. The problem with this is that such smaller banks and credit institutions may not have the expertise and capacity necessary to service higher risk clients (Durner and Shetret 2015).

Some customers may have no other choice but to turn to shadow banking channels to address their financing needs (Durner and Shetret 2015). What makes these alternative channels an attractive proposition is their ease of access and low costs (Langthaler and Niño 2017). However, this does not always hold true—if these underground channels are the only way certain customers can obtain funds, this implies that such customers may have to incur higher costs to acquire finance (Durner and Shetret 2015). Moreover, these channels present a higher risk as they are not very regulated and monitored, suggesting that there is less transparency and accountability towards customers (Durner and Shetret 2015).

These aforementioned points create a regulatory paradox. This is because the introduction of stricter regulation and increased AML/CFT compliance requirements has resulted in the implementation of more derisking strategies by banks and other financial institutions, and operations are being pushed to markets that are not properly regulated and monitored (Langthaler and Niño 2017).

5.5.2. Financial and Socioeconomic Implications

The worst effects of the global phenomenon of derisking are felt the most by smaller countries that typically have low volumes of crossborder transactions. Countries that are also at a high level of risk are those that have a significant dependence on remittance payments (Haley 2018).

The derisking process can lead to reduced international trade finance. At times, international banks may decide to terminate certain CBRs with smaller foreign banks and close their correspondent bank accounts. This subsequently limits the access of such foreign banks to foreign currencies. Access to foreign currencies is vital because it is necessary for international trade to be conducted and it also facilitates international investment (Durner and Shetret 2015). Therefore, if a country's access to such currencies is lowered, this leads to the whole country being debanked, adversely impacting economic growth (Durner and Shetret 2015).

According to Langthaler and Niño (2017), tourism is one of the industries that suffers the most when a country's access to foreign currencies is restricted, especially in the case of Caribbean countries. If foreign exchange services become more difficult to access, tourism flows and related investments will decrease (Langthaler and Niño 2017).

Correspondent banks provide a variety of financial services that are paramount with respect to crossborder trade. Such services include import/export credit letters, contract guarantees and discounting. If banks and other financial institutions stop providing these services, this will have an effect on the balance of payments of many countries since small and medium-sized exporters will be less able to engage in trade (Langthaler and Niño 2017). The termination of services provided by banks has the largest negative impact on small and medium-sized firms in poor countries (IFC 2017).

5.5.3. Increased Financial Exclusion

The implementation of a derisking strategy not only increases financial exclusion but also negatively impacts financial inclusion (Durner and Shetret 2015) and puts the world's financial integrity at risk (Babe 2017). Since financial inclusion and financial integrity complement each other, it is paramount that goals related to financial inclusion are met so that risks are mitigated in an effective manner and, more importantly, financial crimes can be combatted (World Bank 2016).

If a country or existing banked population is derisked, this means that the access of such country or existing banked population to financial services becomes curtailed (Durner and Shetret 2015). Therefore, poverty will increase, and there will be greater inequality when it comes to income (Demirgüç-Kunt and Klapper 2013). In fact, according to authors such as McLean et al. (2018), Erbenova et al. (2016), Starnes et al. (2017), and MacDonald (2019), such derisking and, particularly, the severing of CBRs also negatively impact initiatives that are primarily aimed at alleviating poverty and inequality. This is because derisking disrupts remittance flows. Additionally, people who have access to savings and credit mechanisms are in a better position to cope with economic shocks and smoothen their income (Durner and Shetret 2015).

Durner and Shetret (2015) stated that the barriers of financial inclusion include "a lack of financial literacy, low income and erratic cash flows, and high transportation and opportunity costs" (p. 21). The lack of availability of financial services is particularly detrimental in the case of developing countries, especially if there are impoverished and marginalised communities involved. Without access to crucial financial services, people are not able to purchase essential goods, pay for education or medical care and remit funds abroad (Taylor 2016).

The disruptions in global remittance flows caused by derisking is an issue of a humanitarian concern (Durner and Shetret 2015). Somalia is an example of a country that is heavily reliant on remittances—over 40% of Somali people rely on such flows (Paul et al. 2015). A decrease in remittance flows has a major impact on the country and its most vulnerable communities (Orozco and Yansura 2013). Without such remittance flows, families would not have enough funds to pay for food, housing, education and health care (Gutale 2015).

Derisking is often perceived as being a market failure as it may adversely affect financial inclusion goals. The implementation of a derisking strategy may lead to affected clients having to rely on smaller financial institutions that do not always have sufficient resources and AML/CFT capacity to minimise the impact of risks involved. Additionally, in certain cases, higher risk clients may end up being completely pushed out of the financial sector (Nisar 2016).

5.6. Categories of Clients Affected by Derisking

5.6.1. Correspondent Banks

The effects of the loss of CBRs:

i. Trade—If the clients of banks find it difficult to send and receive foreign payments and maintain business relationships with foreign suppliers and customers, trade flows will decrease, meaning that imports and exports will also be lower than before (Haley 2018). As a result, revenues will also start to deteriorate, and businesses may find themselves unable to pay off their bank loan payments (Haley 2017).
ii. Banking—Weakened domestic banks may not be able to finance investment by granting loans to businesses (Haley 2017).
iii. Investment—If domestic banks are unable to effectively service their clients, foreign direct investment (FDI) is disincentivised. If investment expenditure starts decreasing, this may hinder a country's productivity and growth in the long-term (Haley 2017).
iv. Competition—Lower competition adversely affects growth and may lead to monopoly pricing, implying that social welfare can also be affected by a decrease in competition (Haley 2017).

In order to maintain its profitability, correspondent banking requires a high volume of transactions and operations (Langthaler and Niño 2017). Therefore, one of the reasons why correspondent activities are being retracted is due to this sector experiencing a decrease in returns (Langthaler and Niño 2017). Moreover, some correspondent banks are inefficiently grouping larger proportions of customers as being high risk to minimise risk and avoid potential AML/CFT scandals. This leads to such customers losing access to invaluable banking services which they require (Langthaler and Niño 2017, as affirmed by the World Bank 2015, p. 29).

Given the potential damage to reputation and balance sheet from any enforcement case, firms seek to avoid any counterparties or jurisdictions where there is uncertainty.

5.6.2. Money Service Businesses (MSBs) and Money Transfer Organisations (MTOs)

The withdrawal of correspondent banking results in derisking and directly affects MSBs, MTOs and other remittance institutions. With regards to MSBs, even if they are in full compliance with the regulations of the sending jurisdictions, transactions may still present a risk. This is not only because the jurisdictions receiving the funds sometimes have inadequate frameworks related to AML/CFT, but also due to the fact that their bordering jurisdictions may be subject to certain sanctions or have a conflict that is undergoing (Durner and Shetret 2015).

MSBs have loyal clients that regularly utilise their services. However, the amount associated with an average transaction is small. Regulatory frameworks are complex and varied, and there are additional operational and compliance burdens involved with processing a high volume of transactions. Thus, the extra costs associated with providing banking services to particular customer bases may not be offset even if there is a considerable volume of transactions (Durner and Shetret 2015).

Derisking of MTOs results in the loss of remittance services, and this can have harmful ramifications on the poor living in small countries that are not so developed when it comes to financial regulation. Examples include African, Latin American and Caribbean countries. This is because people in these countries are highly dependent on cash flows coming from abroad (Haley 2017).

This loss of income for individuals living in these countries has a negative impact on global efforts for poverty alleviation. Additionally, an increase in costs incurred to transfer money may lead to a decrease in export earnings. This results in lower investment and countries being unable to meet certain growth and development targets (Haley 2017).

5.6.3. Nonprofit Organisations (NPOs)

Derisking inhibits the ability of NPOs to keep on providing essential services in countries going through a humanitarian crisis, to affect aid and relief worldwide, and to campaign for changes within political and social spheres on a global scale (Goswami 2017; Durner and Shetret 2015). This especially applies to those NPOs that operate in zones experiencing conflict or instability (Durner and Shetret 2015). NPOs are perceived as being quite risky due to the nature of the work they conduct, such as transferring of money on an international level, or because the areas they operate from or both these reasons (Durner and Shetret 2015).

Although the derisking practice threatens the implementation of the mandates of humanitarian organisations and human right activists (Van Broekhoven et al. 2014), most prominent NPOs remain hesitant to speak out about the negative impacts of such practice. This is because they are afraid of even greater financial exclusion and suffering damage to their reputation (Keatinge 2014).

5.6.4. Other Types of Affected Clients

(i) Charities—Charitable organisations depend on banking facilities in order to collect and receive funds (Artingstall et al. 2016). These organisations often lack the knowledge and resources needed to satisfy questions that are related to due diligence (Badour et al. 2017). The derisking process has an amplified impact on smaller charities. This is because, following a derisking exercise, smaller charities would have to start operating on a cash basis, implying more costs and higher risks (Artingstall et al. 2016).

(ii) Politically Exposed Persons (PEPs)—Elected officials and other individuals who hold important and high profile public positions represent a higher risk due to their susceptibility to bribery and corruption. Embassies housing PEPs may also be significantly impacted by derisking (Badour et al. 2017).

(iii) Diplomats and Foreign Students—Members of the diplomatic service of a particular jurisdiction may be discriminated against when it comes to opening a bank account or accessing financial services due to the position they hold. Additionally, foreign students may find it difficult to open bank accounts in certain jurisdictions, and this is mainly the result of problems related to identification documents and costs that may be incurred in the verification and due diligence processes (Artingstall et al. 2016).

(iv) Defence Sector—Small and medium-sized enterprises (SMEs) specialising in the defence sector may sometimes find it difficult to obtain the required financing and access to financial services. Since several controversies related to this sector have transpired in the past, banks have started to exercise increased caution to avoid being conspicuous (Artingstall et al. 2016).

(v) Fintech—Customers resort to this novel industry due to their inability to access, or dissatisfaction with, more traditional banking services. This industry is often the target of derisking exercises because of the high risks involved, especially in the case of virtual currency operators, electronic money institutions (EMIs) and payment institutions (PIs) (Artingstall et al. 2016; Badour et al. 2017). EMIs and PIs are extremely important because they provide services to populations that are debanked (Artingstall et al. 2016).

5.7. The Unbanked, Vulnerable Communities, Minorities and Gender Issues

Unbanked populations in developing countries often struggle to obtain access to financial services because of barriers that constrain their financial inclusion, such as financial illiteracy, a low level of income, inconsistent cash flows and significant costs related to transportation (Durner and Shetret 2015). The rural poor living in developing countries have historically been financially excluded by banks due to the high costs incurred to operate from remote locations (Durner and Shetret 2015).

The communities that are most significantly affected by a lack of access to financial services are "rural, low-income, and minority communities, such as women and youth" (Durner and Shetret 2015, p. 22). Women are sometimes victims of procedures that are inherently discriminatory in nature. In fact, women sometimes face financial exclusion as finance is usually much harder to obtain when compared to men, inevitably hindering women's empowerment (Durner and Shetret 2015).

5.8. Technology-Based Solutions to the Derisking Problem

5.8.1. Blockchain and Distributed Ledger Technologies (DLTs)

These help to mitigate the negative effects associated with derisking as their implementation helps improve the data gathering and identification processes. This ultimately leads to reduced burdens associated with compliance and enhanced transparency (Babe 2017). Moreover, blockchain and DLTs result in lower regulatory and compliance costs such as KYC requirements (IFC 2017), reduced costs associated with the verification of customers, and increased transparency of transactions.

These technology-based solutions offer numerous benefits: they facilitate the sending and receiving of remittances, make trade finance easier for businesses to obtain and assist charities operating in certain areas of conflict (Neocapita 2017).

While blockchain and DLTs are able to adequately protect confidential information, the level of anonymity provided by such technologies enables bad actors to conceal their identities. Therefore, this means that individual payments may end up being difficult to track (IFC 2017).

5.8.2. Big Data, Machine Learning and Biometrics

- Big data consists of datasets that are high in volume, velocity and variety, meaning that unlike traditional datasets, the former necessitates the use of specialised systems and analytical techniques. One of the main advantages of big data is that it permits a wide variety of data types to be stored in the same place, implying that staff will now spend less time gathering information from different sources. Big data also increases the range and scope of information that can be used when it comes to KYC and suspicious transaction investigations (Woodsome and Ramachandran 2018).
- Machine learning is essentially a form of artificial intelligence that leads to computers becoming more efficient at performing certain tasks through repeated iterations. Such new technology can be used to help transform compliance functions of organisations and also results in more accurate monitoring of transactions. This is because machine learning identifies previously undetected illicit finance techniques and reduces false alerts (Woodsome and Ramachandran 2018).
- Biometrics are used to authenticate the identity of a person and control his/her accessibility to a particular system. By utilising biometrics, banks are able to identify and verify the identities of their customers and carry out due diligence in a proficient manner. This addresses gaps related to identification that are present in developing countries (Woodsome and Ramachandran 2018).

5.9. Derisking in the Context of Malta

5.9.1. The MONEYVAL Report of 2019

In 2019, MONEYVAL, a monitoring body of the Council of Europe (CoE) specialising in antimoney laundering, published a report that provided a brief overview of the AML/CFT measures Malta has implemented and commented on the robustness of the AML/CFT system the country has in place.

Moreover, this report analysed the extent to which Malta is compliant with the FATF recommendations (CoE 2019a).

To obtain a more comprehensive understanding of money laundering and financing of terrorism risks, Malta conducted a national risk assessment (NRA) exercise in 2013/2014. Following this NRA exercise, Maltese authorities were able to more extensively comprehend the risks and vulnerabilities in Malta's AML/CFT system, especially in the case of regulated sectors (CoE 2019a). However, MONEYVAL noted that there were factors, including predicate offences, legal persons and arrangements and financing of terrorism, which appeared to be insufficiently understood (CoE 2019b).

5.9.2. Derisking in Malta: A Local Case

The banking sector in Malta has been experiencing a downward trend for these past few years. Late in the year 2018, Pilatus Bank, which was considered Malta's most controversial bank, had to close down its operations after allegations of money laundering breaches emerged. In July 2019, Satabank was fined a sum of €3 million for having weak structures in place, which facilitated the carrying out of criminal activities such as money laundering (Vento 2019).

Lately, Malta's banking sector has increasingly suffered as foreign banks are withdrawing their correspondent bank services from Malta (Costa 2019, Costa 2019). This has mainly been the result of failures vis-à-vis antimoney laundering in Europe (Vento 2019).

In the year 2019, a core domestic bank, Bank of Valletta (BOV), embarked on a derisking exercise (Grech 2019). This resulted in BOV ceasing the provision of certain banking services to some of its customers and closing their bank accounts (Martin 2019). To ensure that it followed international best practices, BOV decided to undertake this derisking exercise alongside a KYC retail client remediation exercise (Grech 2019).

The decision to initiate such derisking was taken following discussions BOV had with the Malta Financial Services Authority (MFSA) and the European Central Bank (ECB) (Grech 2019). It was brought to light that the ECB had given BOV clear instructions to engage in derisking in order to establish a new risk profile that involves decreased risk (Martin 2019). Derisking was also carried out so as to allow BOV to strengthen its frameworks, systems, policies and procedures regarding risk, compliance, antifinancial crime and antimoney laundering (Costa 2019).

The customers affected by this derisking exercise mainly comprised of businesses deemed as being high risk, such as iGaming companies. These iGaming companies, including Malta Gaming Authority (MGA) licensees, had their bank accounts closed (Martin 2019). Other affected customers that also had their accounts closed consisted of gaming affiliate companies (Martin 2019), as well as individual investor programme (IIP) buyers who had opened a bank account in Malta to obtain a Maltese passport (Grech 2019).

Derisking was carried out with respect to both international corporate customers (ICC) and international personal banking (IPB) customers. ICCs typically consist of companies that are registered in Malta with the Malta Business Registry (MBR), where the majority of the company's business activity occurs outside of Malta, and the beneficial owner is not Maltese (Grech 2019). When it comes to IPB customers, there is a mix of clients, with some located outside EU jurisdictions. The accounts of ICCs and IPB customers that BOV determined as not having an economic nexus in Malta were closed (Grech 2019).

The implementation of this derisking strategy has made it difficult and cumbersome for foreign companies to set up their operations in Malta and open a bank account (Martin 2019). Additionally, given limited local options, a number of foreign clients had no other choice but to transfer their operations overseas or start making use of e-money institutions (Vento 2019).

6. Methodology

6.1. Tool Used to Collect Data

For the purpose of this study, a purpose-built semistructured survey was designed and administered to personnel in banks, financial institutions, accounting firms, regulatory authorities, as well as any other relevant parties and people dealing with these institutions.

Some of these surveys, which were built using a software application (Qualtrics XM) in order to provide an online link, were administered using social media and email; the rest were carried out over the phone or using other communication applications such as Skype.

The survey consisted of three sections. The first section consisted of five demographic questions asking for the "age", "gender", "locality", "capacity" and "level of expertise" of the participant. The next section consisted of six themes with three statements each (i.e., 18 statements in total), to which participants were asked to rate each statement using a 1 to 5 Likert scale, "1" being "Strongly Disagree" and "5" being "Strongly Agree". The themes were associated with certain drivers, factors and implications of derisking identified from existing literature (derived using the thematic approach, as suggested by Braun and Clarke 2006). These themes were "Financial", "Socioeconomic", "Financial Exclusion", "Antimoney Laundering/Combatting the Financing of Terrorism (AML/CFT) and Risk", "Compliance and Regulation" and "Technology-Based Solutions". The third section consisted of a set of seven open-ended questions related to the topic of derisking and its main drivers, factors, implications and effects, particularly with regards to the Maltese scenario (Saunders et al. 2009; Yin 2002; Yazan 2015; Stake 1995). Other questions dealt with

- The advantages and disadvantages of derisking;
- Potential solutions to the main derisking problems;
- The role of regulators and regulatory authorities vis-à-vis derisking;
- The Maltese sectors and industries that are the most negatively affected by derisking; and
- The prevalence of derisking in Malta in the years to come.

6.2. Sample Population

Since the population of possible candidates was not known, we used a mix of nonprobability purposive sampling by contacting persons on our social media and email, and they connected us with further possible participants (nonprobability snowballing sampling; Mack et al. 2005; Neuman 2005).

The initial sample of potential study participants consisted of 60 organisations employing individuals who possess a relatively high level of knowledge regarding the derisking practice, money laundering and terrorism financing. The organisations that were targeted included banks operating in Malta, the Central Bank of Malta (CBM), local regulatory authorities (the MFSA and the FIAU), the large accounting firms and other accounting and risk management firms.

To determine the participants' level of expertise, a 1 to 3 Likert scale was utilised, where 1 corresponded to "a fair level of expertise", 2 corresponded to "a good level of expertise" and 3 corresponded to "an excellent level of expertise". To ensure that the data obtained were reflective of expert knowledge, we filtered participants by level of expertise and only used data from participants who expressed a good or excellent level of expertise.

Between 15 November 2019 and 21 April 2020, we carried out a total of 32 interviews with individuals who have a good or excellent level of expertise in derisking. Although it seemed that additional interviews beyond this point did not provide new information or added value, we maintained an open mind and supplemented our data with the 296 completed online surveys received from participants during the same period in which we carried out the interviews. This ensured that our selection of participants was, as much as possible, free from any selection bias since the surveys were answered anonymously. These surveys were then filtered to ensure that the level of expertise criteria was met, and these resulted in 285 valid participants. In total, we had 317 participant surveys.

6.3. Analysis

Following the collection of data, such data was first transferred into Microsoft Excel and Microsoft Word, and, later, the quantitative data were loaded onto Statistical Package for the Social Sciences (SPSS version 21) software. Subsequently, the data were analysed using frequencies and by subjecting the data responses to the statements to exploratory factor analysis, which is a method of testing the theoretical understanding of the factors for the main drivers of derisking posed in the first research question (RQ1), and, as determined from the literature, no a-priori fixed number of factors was set. The final number of factors was determined by the data and the authors' interpretation of them (Hair et al. 1998).

Since the items used the ordinal scale of measurement, we used the median (Md) as a measure of central tendency and the interquartile range (IQR) as a measure of spread. Where a group of items could be grouped into a construct (or theme), we assessed the internal consistency reliability of the measures via Cronbach's alpha. After the items were combined into a single Likert scale, we computed the mean (M) as a measure of central tendency and the standard deviation (SD) as a measure of spread.

We then carried out multiple linear regression to determine how this measure varies with the demographic factors; we also used STATA application software applying White robust standard errors to account for any heteroskedasticity and analysed the answers to the last section using the thematic analysis approach, as suggested by Braun and Clarke (2006, p. 6), to answer RQ2.

7. Results and Discussion

7.1. Summary Statistics

As can be noted from the tables in Appendix B, the largest group of participants were aged between 30 and 39 (39.7%), most were men (68.5%), coming from a central location (54.9%), and users of financial institutions (67.5%), with a good level of expertise (59.6%). This shows that most of the participants were exposed to a similar culture and environment, and this could be the reason why the results of the multiple linear regression, noted and discussed below, are not significant.

Additionally, as noted in Appendix A, all participants are in agreement (mean between 3.4 and 4.19) with the statements posed in the survey/questionnaire, with the exception of Statement 13 (S13) wherein the participants strongly agreed. This further confirms our findings from the multiple linear regression carried out and explained below.

7.2. Exploratory Factor Analysis

For exploratory factor analysis, we used direct oblimin via principal components extraction with Kaiser normalization. The Kaiser–Meyer–Olkin (KMO) statistic, which is a measure of sampling adequacy for the appropriateness of applying factor analysis, fell within the acceptable range (above 0.6) with a value of 0.93. This further supported the continuance of factor analysis, and so the analysis proceeded.

Exploratory factor analysis loaded best as three factors and 22 statements, which, in combination, explained 60.68% of the variance of the perceived implications of derisking in Malta. Table 2 shows which statements are grouped under each of the three factors. Factor 1, which has now been termed "Research and Development Solutions", explained 46.687% of the variance and comprised five items that include the following statements:

S3 Derisking leads to an increase in costs incurred to transfer money. Therefore, export earnings decrease, resulting in a lower level of investment. Moreover, given that domestic banks will not be able to effectively service their clients, this leads to a lower level of foreign direct investment (FDI). As a result, a country's productivity and growth in the long-term may be negatively affected, rendering such a country unable to meet growth and development targets.

S5 The implementation of a derisking strategy increases financial exclusion and negatively impacts financial inclusion. If a country or existing banked population is derisked, this means that the access of such country or banked population to financial services becomes limited. When access to financial services is restricted, poverty and income inequality increase.

S16 In order to mitigate the negative effects associated with derisking, certain technology-based solutions can be implemented. These include blockchain, distributed ledger technologies (DLTs) and biometrics. The implementation of such solutions improves data gathering and identification processes, ultimately leading to reduced burdens associated with compliance and enhanced transparency.

S17 Blockchain and DLTs help minimise the negative impacts of derisking by facilitating the sending and receiving of remittances, making trade finance easier for businesses to obtain and assisting charities operating in certain conflict areas.

S18 While blockchain and DLTs are able to adequately protect confidential information, the level of anonymity such technologies allow makes it easy for bad actors to conceal their identities. Therefore, if this happens, individual payments become more difficult to track.

Table 2. Exploratory factor analysis [a].

	Factors		
	1	**2**	**3**
S17	0.857		
S16	0.803		
S18	0.791		
S7	0.719		
S6	0.546		
S14		0.812	
S1		0.697	
S5		0.675	
S3		0.564	
S13		0.519	
S15		0.496	
S12		0.404	
S8			−0.794
S10			−0.759
S9			−0.741
S4			−0.620
S11			−0.448
S2			−0.371

Extraction method: principal component analysis. Rotation method: oblimin with Kaiser normalization. [a] Rotation converged in 12 iterations. Source: authors' computations.

Factor 2, which has now been termed "Compliance and Regulatory", explained 7.621% of the total variance and comprised seven items that include the following statements:

S1 At times, banks and other financial institutions may opt to terminate certain correspondent banking relationships (CBRs) with smaller foreign banks and close their correspondent bank accounts, consequently limiting their access to foreign currencies. Therefore, the derisking process leads to a reduction in international trade finance.

S3 Low profitability of customer bases also brings about derisking. If the profitability of a particular customer base does not compensate for the higher risks and additional costs that must be incurred, such a customer base is often derisked. An example of these additional costs is Know Your Customers' Customers (KYCC), which means that banks and other financial institutions are not only responsible for their customers, but also for the customers of their customers.

S5 The derisking practice leads to lower imports and exports as the customers of banks engaging in derisking will not be able to send or receive foreign payments and maintain important business relationships with foreign suppliers and customers.

S12 In response to derisking, many customers are resorting to shadow banking channels and fintech to obtain required funds. These environments are unknown and much less regulated and monitored, implying that risk is higher as well. If such alternatives are the only source of finance available, this means that consumers may have to incur higher costs to address their financing needs.

S13 With AML/CFT regulations always increasing and becoming ever more complex, this has resulted in a rise in compliance costs. These increased compliance costs have a magnified effect on jurisdictions that have a smaller and more restricted banking market. This is because such jurisdictions are often unable to spread the fixed cost component of these AML/CFT compliance costs over a large enough number of transactions.

S14 Since the fines and penalties imposed for noncompliance with and breaches of AML/CFT requirements are considerable, many banks and other financial institutions have opted to implement a derisking strategy.

S15 The imposition of increased capital and liquidity requirements following the 2008 global financial crisis is also another factor which is influencing the derisking carried out by banks and other financial institutions.

Factor 3, which has now been termed "Direct and Indirect Effects", explained 6.373% of the total variance and comprised six items that include the following statements:

S2 Correspondent banks provide a variety of financial services that are paramount with respect to crossborder trade. Such services include import/export credit letters, contract guarantees and discounting. If banks and other financial institutions cease providing these services, small and medium-sized exporters will be less able to participate in trade. This especially applies to small and medium-sized firms in poorer countries.

S4 The withdrawal of CBRs and derisking of money transfer organisations (MTOs) result in the loss of remittance services. This can have negative ramifications for the poor living in small countries that are not so developed when it comes to financial regulation, e.g., African, Latin American and Caribbean countries.

S8 The lack of availability of financial services is particularly detrimental in the case of developing countries, especially if there are impoverished and marginalised communities involved. Without access to financial services, people would not be able to purchase essential goods, pay for education or medical care, and remit funds abroad.

S9 Derisking has a significant negative impact on nonprofit organisations (NPOs) as it inhibits their ability to provide crucial services in countries currently undergoing humanitarian crises, affect aid and relief, and campaign for changes in political and social spheres.

S10 Following the 2008 global financial crisis, banks and other financial institutions worldwide started to curtail their risk appetite and adopted a risk-based approach (RBA) with regards to AML/CFT. This reduction in the risk appetite has led to an increase in derisking.

S11 When banks and other financial institutions engage in derisking, this results in a shift in AML/CFT risk. This is because if CBRs are terminated, the affected clients will then have to start relying on smaller banks and credit institutions to obtain the financial services they require. The problem with this is that such smaller banks and credit institutions may not have the expertise and capacity necessary to service high-risk and poor populations.

The 18 statements can be used as an inventory checklist grouped under the 3 main themes "Research and Development Solutions", "Compliance and Regulatory" and "Direct and Indirect Effects" to understand the implications that a derisking exercise can bring to a financial firm. Moreover, one can use this checklist/inventory to develop a risk exposure matrix to arrive at a single qualitative rating and/or quantitative measure of the implication of derisking.

This is of specific interest to risk managers working in the financial services sector, who, as per regulatory requirements, would need to summarise and classify the risk rating of a firm, project and strategy to the board of directors, shareholders and the regulators. It is also of interest to policymakers who can develop policies based on such results.

7.3. Reliability Test

The Cronbach's alpha coefficients of this scale were between 0.83 and 0.86. Therefore, we can conclude that this scale is reliable as part of our statistical analysis (Table 3).

Table 3. Cronbach's alpha values (n = 317).

Factor	Item	Mean	Min–Max	Cronbach's Alpha
1	5	4.09	3.97–4.27	0.88
2	7	3.91	3.80–4.21	0.83
3	6	4.09	3.88–4.29	0.86

Source: authors' computations.

The computed "Derisking Model" for Malta shows a mean of 4.03 (SD = 0.67). All the factors (Factors 1, 2 and 3) produced means that were close to the computed model. This shows that participants from Malta, overall, believe that these are the drivers of derisking in Malta.

7.4. Multiple Linear Regression and ANOVA

The computed one-way analysis of variance (ANOVA) was used to show that there are no statistically significant differences between the means of the independent (unrelated) groups ($p > 0.01$; Table 4).

Table 4. ANOVA [a].

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.371	5	0.474	1.043	0.392 [b]
	Residual	141.380	311	0.455		
	Total	143.751	316			

[a] Dependent variable: derisking drivers. [b] Predictors (constant): level of expertise, age, provider–user, gender, and locality. Source: authors' computations.

Moreover, multiple regression analysis ($p > 0.01$) reveals that the perception of participants and interviewees does not change as an effect of the different demographics, i.e., (1) age, (2) gender, (3) locality, (4) whether they are a provider or a user of financial services, and (5) level of expertise (Table 5). The same result is obtained after following the use of White robust standard errors to account for heteroscedasticity (Table 6). This shows that all participants are aligned in their opinion and reasoning about the main drivers of a derisking strategy, and these opinions are not dependent on any demographic factors. This may be due to the fact that Malta is a small island and one of the small European Union jurisdictions with a population of approximately 480k in an area of 246 km squared, with one main national university, a single financial services regulator (Malta Financial Services Regulator (MFSA)) and a Financial Intelligence Analysis Unit (FIAU). Therefore, most of the participants have practiced, worked and studied in/under the same environment or are in contact at similar conferences.

Table 5. Coefficients [a].

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.868	0.313		12.341	0
	Age	0.008	0.045	0.011	0.169	0.866
	Gender	−0.102	0.095	−0.07	−1.074	0.284
	Locality	0.058	0.07	0.056	0.832	0.406
	Provider-User	−0.111	0.085	−0.077	−1.302	0.194
	Level of Expertise	0.128	0.088	0.093	1.451	0.148

[a] Dependent variable: derisking drivers. Source: authors' computations.

Table 6. Accounting for Heteroskedasticity.

Variable	Coefficient
Age	0.0075 (0.0401)
Gender	−0.1025 (0.0991)
Locality	0.0579 (0.062)
Provider-User	−0.1108 (0.0849)
Level of Expertise	0.1269 (0.084)
Constant	3.8713 *** (0.2937)
N	317
R-Squared	0.0165
F-Statistic	1.08

Notes: Dependent variable is the score denoting derisking drivers. Robust standard errors are reported in parentheses. *** denotes statistical significance at the 1% level; ** denotes statistical significance at the 5% level; * denotes statistical significance at the 10% level. Source: authors' computations.

7.5. Solutions to the Main Problems of Derisking (Participants Responses to the Interviews And/or Open-Ended Questions)

With respect to mitigating the negative effects of derisking, emerging technology such as blockchain and distributed ledger technology (DLT) were the two most common solutions provided by the participants (35). If coupled with a "robust regulatory environment and governance structure", blockchain and DLTs have the potential to reduce the costs and burdens of regulation and compliance and also increase the overall level of transparency. This is in agreement with the views of Babe (2017) and the IFC (2017).

Nevertheless, some participants (21%) emphasised the fact that blockchain and DLTs are still in their "infancy stage" and not widely accepted yet. In fact, they are by no means a "silver bullet" and there are still considerable limitations related to these technologies that need to be sorted out. Firstly, blockchain and DLTs depend a great deal on other technologies such as the internet, which is not always available in certain geographical areas, particularly in developing countries and other countries perceived as being high risk. In addition, these technologies lack scalability, and the speed of transactions using such technologies is considered as being slower compared to transactions being processed using existing technologies. The underlying concept of decentralisation conflicts with regulatory objectives as it makes it difficult to determine the controlling parties. Finally, there are also issues related to data protection.

One participant stated that while "blockchain and DLTs are tools that can be both beneficial and detrimental" since these technologies are designed to build a system of trust amongst a group of individuals, they can be abused if such technologies are used for the wrong reasons. A further argument brought up by another interviewee is that these technologies are usually considered as being ancillary tools and not essential tools for banks and other financial institutions. This is because of the high costs associated with implementing and monitoring the use of such technologies. One interviewee was completely against the use of blockchain and DLTs and argued that these technologies are not mature enough, and there are too many hurdles which need to be surpassed for them to become mainstream.

Alternative solutions to the problems of derisking suggested by a few interviewees (13) include legal entity identifiers (LEIs) and Know Your Customer (KYC) utilities. LEIs are based on International Standards Organisation (ISO) 17,442 and consist of a code that is made up of 20 alphanumeric characters. These identifiers allow legal entities participating in financial transactions to be clearly and uniquely identified, leading to enhanced transparency on a global scale. KYC utilities enable financial institutions to effectively and efficiently carry out their KYC procedures through the centralised collection, verification and sharing of customer information.

Another participant believed that a possible solution to mitigate the adverse impacts of derisking strategies implemented in Malta is to have a financial services sector that constitutes a large number of small players, each having a different and varied risk appetite, specialising in the management of customer relationships characterised with their own unique risk level. Naturally, the management of customers that present a higher risk attracts higher fees for the services provided. Blockchain, DLTs and fintech adopted in a regulated manner "are all key in this multiplayer landscape of financial services providers".

The introduction of stricter regulation and AML/CFT compliance requirements means that enhanced oversight by the ECB and Maltese regulatory authorities, such as the MFSA and the Financial Intelligence Analysis Unit (FIAU), is now more crucial than ever. All interviewees (32) advocated that regulators and regulatory authorities are responsible for issuing robust guidance to banks and financial institutions, not only in the case of Malta but also internationally. This thereby ensures that an appropriate RBA is applied and that controls are administered when there are high-risk customers involved. One interviewee noted that although oversight by local and international regulators and regulatory authorities has been increasing steadily, such oversight does not always take into consideration the principle of proportionality. The example given by this interviewee was that a Maltese bank and German bank might be treated in the same way, "notwithstanding the obvious difference in size and systemic importance". To ensure fairness and equality, it is, therefore, paramount that regulators and regulatory authorities apply this principle well.

Another interviewee expressed that regulators and regulatory authorities must ensure that meaningful derisking exercises are executed. This implies that the sufficient controls that allow the appropriate management of risks need to be in place so that residual risks are kept to a minimum. This interviewee then added that these controls should be continuously monitored "so that their effectiveness is maximised". It is also very important that regulators and regulatory authorities ascertain that the customer bases held by banks and other financial institutions are sustainable.

In the scenario of Malta, it seems to be that local regulatory authorities are the ones initiating and imposing derisking exercises themselves, rather than the Maltese banks and other financial institutions being the ones voluntarily engaging in derisking. One of the interviewees believed that this approach is being taken because such regulatory authorities want to make sure that local banks and other financial institutions understand their business models well and manage the risks involved. This leads to "an increase in the level of risk awareness across the general society".

One interviewee believed that after the MONEYVAL Report was issued in 2019, Maltese regulatory authorities shifted their focus and changed their role from being watchdogs to "blood hounds".

However, this interviewee feared that the introduction of more stringent requirements and higher fines and penalties will not solve the money laundering and terrorism financing issue with regards to the Maltese situation and contended that different strategies will need to be explored in the long-term.

Twelve interviewees explained that local regulatory authorities play an important role when it comes to safeguarding the integrity and stability of Malta's financial system. Their role also ensures that the recommendations of prominent regulatory bodies, such as the Basel Committee on Banking Supervision (BCBS) and the FATF, are adhered to as this "helps to greatly improve the country's international standing from a compliance perspective". Consequently, compliance costs for correspondent banks would decrease as such banks would start viewing Malta as presenting a lower risk, implying that less money and resources are required for the monitoring of relationships.

While strict supervisory mechanisms are a fundamental component that helps build and restore trust in banks and other financial institutions, it is of utmost importance that local and international regulators and regulatory authorities do not overregulate the market as this may force certain customer bases into less regulated channels. According to one interviewee, the "right balance which adequately manages risk, while not extinguishing business, needs to be struck". Another interviewee added that regulators and regulatory authorities "need to be careful so as to not overburden banks and financial institutions with too much bureaucracy to the point where it then becomes overkill".

7.6. Derisking within the Maltese Context

The NRA that was carried out by the Maltese authorities in 2013/2014 identified certain sectors and industries that were the most adversely impacted by derisking undertaken by the local banks and financial institutions. Such sectors and industries, together with other sectors that the general public has a negative perception of, include the iGaming sector, companies which provide services that are related to virtual financial assets and passport by investment scheme promoters.

Almost all of the interviewees (29) remarked that the majority of gaming companies in Malta are being derisked by local banks, even if such companies are licensed by the MGA. This is because of the high risk that these companies carry. This ties in with the literature; as part of their derisking strategy, Maltese banks sometimes decide to cease providing certain services to gaming companies and close their bank accounts. One interviewee stated that such derisking by banks and financial institutions is a product of the "extended regulatory scrutiny that Malta has been under due to being open to high-risk sectors with limited enforcement".

Companies that provide services in the area of virtual financial assets, such as cryptoassets and DLT assets, may find it difficult to find a bank that is willing to offer them traditional banking services. This is a result of the unknown risks that are associated with new and innovative technologies like cryptocurrencies and DLTs.

The Maltese banking and financial services sectors are also considerably impacted by the derisking process. Several interviewees (22) expressed that it is not easy for local banks and other financial institutions to establish CBRs with foreign banks and financial institutions, which leads to lower activity in business lines such as the "provision of payments in foreign currencies and trade finance". This, in turn, affects the business generation capabilities of companies that have a material proportion of their revenues generated from international trade. A lack of a proper structure in place that facilitates foreign currency transactions impedes business from being conducted.

Other Maltese sectors, industries and organisation types that interviewees (28) noted as being relatively high risk from a money laundering and terrorism financing perspective include fintech, tourism (incorporating hospitality and catering), shipping, property development, certified financial planners offering trustee and fiduciary services, company service providers, trading companies and MSBs.

7.7. Will Derisking Become More Prevalent in Malta in the Years to Come?

All the interviewees, except for one, claimed that in the near future, derisking in Malta will continue to be undertaken by banks and other financial institutions. First of all, signs in the global economy are indicating that a step back from globalisation levels achieved so far seems to be desirable. Consequently, it is possible that this will lead to further derisking, not only in Malta but also on an international level. When taking into consideration the local context, many interviewees (22) stated that the drivers of derisking are ever-increasing regulatory pressures and compliance requirements imposed by the ECB and Maltese authorities. As a result, these interviewees predicted that these drivers, coupled with all the recent developments in AML/CFT legislation, will result in the further exercise of the derisking practice in Malta.

According to one interviewee, it is inevitable that derisking strategies will continue being implemented in the short- to medium-term as pressure keeps on mounting for Maltese authorities and institutions to "come clean". However, in the long-term, this interviewee believed that business opportunities might open up for institutions that are smaller in size in the form of customers that will end up unbanked by larger institutions. If these smaller institutions obtain a comprehensive and thorough understanding of the inherent risks associated with banking these customer bases, they would be able to offer their services and make a decent return.

In line with the findings of the MONEYVAL Report issued in 2019, local banks and other financial institutions are expected to continue engaging in derisking so that existing AML/CFT deficiencies can be addressed and Malta's standing can improve. This sentiment was conveyed by five interviewees, who also added that continued and ongoing derisking by Maltese institutions will make it easier for them to maintain business relationships with foreign institutions and will decrease compliance risk.

Other interviewees (12) contended that in the years to come, derisking exercises in Malta will start being carried out not only by banks and other financial institutions but also by gaming, insurance, property development, asset management and wealth management companies.

Several interviewees also discussed derisking in Malta within the context of the recent COVID-19 outbreak. Currently, in the midst of this unprecedented pandemic, Maltese authorities are trying to relax and alleviate regulatory requirements so that development and growth are not stifled. One of the interviewees commented that "since everything is covered by a shroud of uncertainty", it is highly possible that regulatory requirements may be relaxed for a prolonged period of time. This is due to the fact that post-COVID 19, such relaxed requirements would serve a stimulus for banks and financial institutions to start regenerating the economy.

A final interesting argument brought forward by another interviewee is the fact that although derisking is unavoidable in the current environment, a point of saturation may eventually be reached, and this would probably end up stalling the entire sector. This would then be followed by a "new cycle of easing of regulations", with the aim of achieving a new equilibrium.

8. Conclusions

Derisking is a global phenomenon that has recently been gaining traction both locally and internationally. In the case of Malta, derisking is still an emerging practice that is not very well-established, which has a lot of potential for further growth and development. In fact, it is only during the past few years that Maltese banks and other financial institutions have started undertaking derisking exercises to better manage their risks, focus on customer bases that fall within their risk appetite and ensure compliance with regulatory and AML/CFT requirements.

There are many factors that are driving derisking within the Maltese scenario. Firstly, a considerable amount of EU directives and regulations, especially those regarding AML/CFT, have been introduced. Since the costs of compliance associated with such directives and regulations are relatively high, this has led to the increased prevalence of derisking both in Malta and the EU.

Locally, regulatory pressures from the ECB and local regulatory authorities on Maltese banks and other financial institutions have been significantly intensifying. Pressure has been steadily mounting for these local banks and financial institutions to re-evaluate their risk profiles and enhance their strategies related to AML/CFT. Consequently, such banks and financial institutions have no choice but to derisk certain riskier customer bases.

The MONEYVAL report, issued in 2019, is another factor leading to derisking in Malta; this report has had a two-fold effect. Firstly, the report calls on Maltese authorities to improve the AML/CFT measures the country has in place, and this has, in turn, led to the authorities compelling local banks and other financial institutions to decrease their risk appetite, resulting in higher derisking. Secondly, given that the report pinpoints several AML/CFT deficiencies present in Malta, foreign banks and financial institutions have started to exercise more caution and sometimes restrict the services provided to, or simply exit relationships with, certain local banks and financial institutions.

Reputational risk is also leading to increased derisking in Malta. The imposition of higher fines and penalties by regulators on banks and other financial institutions for noncompliance with AML/CFT requirements, and the adverse media brought about by such fines and penalties, is leading to a decline in the risk appetite of local banks and financial institutions. This ultimately results in the derisking of customer bases. Malta's current turbulent political climate and the many scandals and allegations of corruption that were recently brought to light have led to a deterioration in the country's reputation. As a result, foreign correspondent banks have initiated the severing of relationships with some Malta banks and other financial institutions because the significant AML/CFT compliance costs are not justified by the high risk and low volume of transactions.

Following the 2008 global financial crisis, capital and liquidity requirements have considerably increased. Banks now need to hold larger capital buffers so that they are able to decrease the risks of their portfolios. This has had a negative effect on the profitability of both local banks as well as banks within the EU and is leading to more widespread derisking. In fact, some relationships with higher risk customers that are not that profitable to maintain are being terminated.

With regards to the general implications and effects of the derisking process, the interviewees mentioned a shift in AML/CFT risk. After being derisked, customers have to resort to other smaller banks and financial institutions to obtain the financial services they require, even though such banks and financial institutions may lack the capacity needed to service such customers. Other implications that the interviewees highlighted were the use of shadow banking channels to acquire finance, the loss of legitimate business and the financial exclusion of certain customer bases.

The interviewees explained that derisking can be seen from two perspectives. From the perspective of banks and financial institutions engaging in derisking, adopting this practice lowers their risk exposures, reduces the risk of fines and penalties, decreases compliance costs and minimises reputational risk. From the perspective of financial markets, derisking safeguards the interests of lower-risk customers and enhances the integrity of the financial system.

Since Malta is one of the smallest countries in the EU, this implies that a robust banking sector is vital in order to attract top-level players to the island. The country's shift from a sales-oriented mentality to a compliance and regulatory mentality signifies that more derisking exercises need to be introduced. The implementation of derisking strategies also ensures that the reputational risk of Maltese banks and other financial institutions is kept to a minimum.

The undertaking of increased derisking in Malta should lead to less corruption and money being leaked from the economy due to money laundering, meaning that trust in the local banking and financial services sectors should also increase. By adjusting their risk appetites, local banks and other financial institutions can contribute to the rebuilding of Malta's reputation. Furthermore, by engaging in derisking, such banks and financial institutions will shrink and become more stable, implying that additional attention can be directed towards core customers.

Apart from its benefits, derisking also has its drawbacks. Derisking exercises may result in a deceleration of economic activity, and this has an impact on the whole socioeconomic system, leading to a decrease in GDP, employment levels and economic growth. Subsequently, Maltese institutions and organisations may have to downsize their operations, and some employees may end up redundant.

When asked about whether the advantages of derisking outweigh the disadvantages, especially in relation to the Maltese scenario, the responses of interviewees were mixed. Some interviewees argued that proper implementation of derisking achieves more pros than cons as it restores the balance between profitability and having customers that are of high quality. In the short-term, the derisking practice leads to a decrease in the profits of banks and other financial institutions, but in the long-term, profits would be more sustainable.

Other interviewees stated that although the adoption of derisking can result in lower compliance costs and higher efficiency, a lack of proper understanding and capacity can result in the loss of legitimate business, defeating the purpose of this exercise. Additionally, the many drawbacks of the derisking process, such as declining CBRs, the use of unregulated banking channels and financial exclusion, can render derisking unsustainable. The solution provided to maximise the pros and minimise the cons of derisking is to find the right balance and mechanism. Regulation should also be balanced and proportionate.

When it comes to the Maltese scenario, given all the recent AML/CFT developments and regulatory pressures by the ECB and local authorities, derisking is on the rise and expected to continue increasing in the near future. Maltese regulatory authorities have the important role of not only providing oversight and supervision but also of facilitating the undertaking of meaningful derisking. This ensures that local banks and financial institutions are able to effectively manage their risks and have customer bases which are sustainable.

In conclusion, the objective of this study, i.e., to identify and analyse the factors affected by and the implications of the derisking process, has been achieved, as clearly shown in the summary above. The research questions that examine the drivers and implications of derisking in the case of Malta have also been answered and exhaustively analysed and discussed throughout this study.

In light of the recent scandals and corruption allegations, the Maltese authorities need to, now more than ever, strengthen the country's fight with regards to money laundering and terrorism financing. One way that this can be achieved is by the imposition of derisking exercises on other sectors of the economy.

The derisking approach should be standardised across the board to ensure a level playing field for all stakeholders. If all Maltese institutions implementing a derisking strategy adopt certain common criteria and standards, this will lead to enhanced transparency and less discriminatory practices applied. Standardisation of the derisking process ensures that the derisking that is conducted is more meaningful and only the customer bases that pose the greatest risk are derisked. This standardised approach should result in more fairness and consistency, therefore maximising the effectiveness that can be derived from undertaking derisking.

Furthermore, the Maltese regulatory authorities should be provided with adequate human and financial resources as this allows them to have the capacity required to carry out enhanced oversight and risk-based supervision based on the size, complexity and risk profile of Malta's private sector. This would also permit such authorities to introduce more practical AML/CFT measures and carry out more investigations and prosecutions. Moreover, by having sufficient resources, the authorities would be able to monitor those Maltese institutions undergoing derisking and ensure that the process is being properly and equitably conducted. Law enforcement authorities should, similarly, have the resources necessary to pursue high-level cases of money laundering, bribery and corruption.

More education and workshops regarding this practice should also be provided to professionals working within the fields of banking, financial services and accountancy. Moreover, it is important that such professionals keep up-to-date with the latest changes and developments when it comes to directives and regulations concerning AML/CFT, as well as derisking in general.

Author Contributions: J.M.A. and S.G. contributed towards the conceptualization, methodology, software, validation, formal analysis, investigation, and resources. J.M.A. took care of the data curation, and writing—original draft preparation. S.G. carried out the writing—review and editing, visualization, supervision, and project administration. P.J.B. carried out the final editing, analysis, and review. J.V.S. carried out the final editing, analysis,

and review. "The Implications of Derisking: The Case of Malta, a Small EU State". All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This paper is based on an unpublished thesis as the final part for the MA in Accountancy at the Accountancy Department, University of Malta, by J.M.A. (2019). "The Implications of Derisking: The Case of Malta" was supervised by Simon Grima—herein the main authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Summary statistics.

	N	Minimum	Maximum	Mean	Std. Deviation
S1	317	1.00	5.00	3.8013	0.96206
S2	317	1.00	5.00	4.1956	0.88924
S3	317	1.00	5.00	3.9685	0.98996
S4	317	1.00	5.00	4.1041	0.92361
S5	317	1.00	5.00	3.8549	1.05441
S6	317	1.00	5.00	4.2650	0.92391
S7	317	1.00	5.00	3.9811	1.00613
S8	317	1.00	5.00	4.2902	0.88106
S9	317	1.00	5.00	4.0410	0.96204
S10	317	1.00	5.00	3.8801	1.03643
S11	317	1.00	5.00	4.0000	0.97760
S12	317	1.00	5.00	3.8170	1.01786
S13	317	1.00	5.00	4.2082	1.05281
S14	317	1.00	5.00	3.8801	1.06356
S15	317	1.00	5.00	3.8549	1.01465
S16	317	1.00	5.00	3.9653	1.08005
S17	317	1.00	5.00	4.1230	1.01601
S18	317	1.00	5.00	4.1199	1.06356
Valid N (listwise)	317				

Source: authors' compilation.

Appendix B

Table A2. Age.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	21–29	22	6.9	6.9	6.9
	30–39	126	39.7	39.7	46.7
	40–49	67	21.1	21.1	67.8
	50–59	102	32.2	32.2	100.0
	Total	317	100.0	100.0	

Table A3. Gender.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	217	68.5	68.5	68.5
	Female	100	31.5	31.5	100.0
	Total	317	100.0	100.0	

Table A4. Locality.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	North	99	31.2	31.2	31.2
	Central	174	54.9	54.9	86.1
	South	44	13.9	13.9	100.0
	Total	317	100.0	100.0	

Table A5. Provider–user.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	User of Financial Institutions	214	67.5	67.5	67.5
	Financial Services Provider	103	32.5	32.5	100.0
	Total	317	100.0	100.0	

Table A6. Level of expertise.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Good	189	59.6	59.6	59.6
	Excellent	128	40.4	40.4	100.0
	Total	317	100.0	100.0	

References

Alix, Michael, Shyam Venkat, Zubin Mogul, Som-lok Leung, Marcia Banks, and Juliane Saary-Littman. 2015. Risk appetite frameworks: Insights into evolving global practices. *Global Credit Review* 5: 1–17. [CrossRef]

Arnold, Martin, and Jennifer Hughes. 2014. Strategies shift as regulators renew scrutiny of bank compliance. *Financial Times*. August 20. Available online: https://www.ft.com/content/5279ef02-289d-11e4-8bda-00144feabdc0 (accessed on 7 February 2020).

Artingstall, David, Nick Dove, John Howell, and Michael Levi. 2016. Drivers & Impacts of Derisking. Available online: http://orca.cf.ac.uk/104017/1/drivers-impacts-of-derisking.pdf (accessed on 11 February 2019).

Babe, Ann. 2017. The Downsides of De-Risking. Available online: https://innovationsjournal.net/thedownsides-of-de-risking-fdc079ed8d56 (accessed on 4 February 2020).

Badour, Ana, Claire Gowdy, and Arie Van Wijngaarden. 2017. De-Risking in Financial Services: Perspectives and Implications. Available online: https://www.mccarthy.ca/en/insights/blogs/snipits/de-risking-financial-services-perspectives-and-implications#page=1 (accessed on 13 February 2020).

Baldacchino, Godfrey. 2006. Islands, Island Studies. *Island Studies Journal* 1: 3–18.

Bezzina, Frank, Simon Grima, and Josephine Mamo. 2014. Risk management practices adopted by financial firms in Malta. *Managerial Finance* 40: 587–612. [CrossRef]

Braun, Virginia, and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3: 77–101. [CrossRef]

Briguglio, Lino. 1995. Small island developing states and their economic vulnerabilities. *World Development* 23: 1615–32. [CrossRef]

Bromiley, Philip, Michael McShane, Anil Nair, and Elzotbek Rustambekov. 2015. Enterprise Risk Management: Review, Critique, and Research Directions. *Long Range Planning* 48: 265–76. [CrossRef]

Clark, Daniel. 2019. What Is the Difference between Corruption and Bribery? Available online: https://www.engageinlearning.com/faq/compliance/preventing-bribery-in-business/what-is-the-difference-between-corruption-and-bribery/ (accessed on 23 February 2020).

Council of Europe (CoE). 2019a. Malta—Fifth Round Mutual Evaluation Report: Executive Summary. Available online: https://rm.coe.int/summary-moneyval-2019-5-5th-roundmermalta/16809737c2 (accessed on 27 February 2020).

Council of Europe (CoE). 2019b. Malta should Step up Its Efforts to Investigate and Prosecute Money Laundering as well as to Strengthen Its Supervisory System. Available online: https://search.coe.int/directorate_of_communications/Pages/result_details.aspx?ObjectId=09000016809732a4 (accessed on 27 February 2020).

Consiglio, John, and Simon Grima, eds. 2012. *A Law Based Financial Services Architecture*. London: Pearson.

Costa, Massimo. 2019. APS chief warns country's reputation could trigger correspondent banks'de-risking. *Business Today*. June 21. Available online: https://www.businesstoday.com.mt/banking-finance/bankingfinance/220/aps_chief_warns_countrys_reputation_could_trigger_correspondent_banks_derisking#.XrVaO2gzZPb (accessed on 13 February 2020).

Committee on Payments and Market Infrastructures (CPMI). 2003. A Glossary of Terms Used in Payments and Settlement Systems. Available online: https://www.bis.org/cpmi/publ/d00b.htm?m=3%7C16%7C266 (accessed on 3 March 2019).

Demirgüç-Kunt, Asli, and Leora Klapper. 2013. Measuring Financial Inclusion: Explaining Variation in Use of Financial Services across and within Countries. Available online: https://www.brookings.edu/wp-content/uploads/2016/07/2013a_klapper.pdf (accessed on 20 February 2020).

Dobrowolski, Zbyslaw, and Lukasz Sulkowski. 2019. Implementing a Sustainable Model for Anti-Money Laundering in the United Nations Development Goals. *Sustainability* 12: 244. [CrossRef]

Durner, Tracey, and Liat Shetret. 2015. Understanding Bank Derisking and its Effects on Financial Inclusion. Available online: https://oxfamilibrary.openrepository.com/bitstream/handle/10546/582310/rr-bank-de-risking-181115-en.pdf?sequence=4&isAllowed=y (accessed on 7 February 2019).

English, Stacey, and Susannah Hammond. 2014. The Rising Costs of Non-Compliance from the End of a Career to the End of a Firm. Available online: https://blogs.thomsonreuters.com/answerson/rising-costs-non-compliance/ (accessed on 28 February 2020).

Erbenova, Michaela, Yan Liu, Nadim Kyriakos-Saad, Alejandro Lopez-Mejia, Giancarlo Gasha, Emmanuel Mathias, Mohamed Norat, Francisca Fernando, and Yasmin Almeida. 2016. *The Withdrawal of Correspondent Banking Relationships: A Case for Policy Action*. IMF Staff Discussion Note. Washington, DC: International Monetary Fund.

Financial Conduct Authority (FCA). 2016. FCA Research into the Issue of de-Risking. Available online: https://www.fca.org.uk/news/news-stories/fca-research-issue-de-risking (accessed on 3 February 2020).

Goswami, Sangeeta. 2017. Blog: Derisking and Civil Society: Drivers, Impact and Possible Solutions. Available online: https://dochas.ie/blog/blog-derisking-and-civil-society-drivers-impact-and-possible-solutions (accessed on 21 February 2020).

Grech, Helena. 2019. BOV Says De-risking Follows 'International Best Practices'. *Malta Chamber*. November 9. Available online: https://www.maltachamber.org.mt/en/bov-says-de-risking-follows-international-best-practices (accessed on 13 February 2020).

Grima, Simon, Inna Romanova, Frank Bezzina, and Frank Chetcuti Dimech. 2016. Alternative investment fund managers directive and its impact on Malta's financial service industry. *International Journal in Economics and Business Administration* 4: 70–85. [CrossRef]

Gutale, Ali. 2015. Life after losing remittances: Somalis share their stories. *The Guardian*. June 18. Available online: https://www.theguardian.com/global-development-professionals-network/2015/jun/18/life-after-losing-remittances-somalis-share-their-stories-somalia (accessed on 4 March 2020).

Hair, Joseph, Ronald Tatham, Rolf Anderson, and William Black. 1998. *Multivariate Data Analysis*, 5th ed. Upper Saddle River: Prentice-Hall.

Haley, James. 2017. De-risking: Effects, Drivers and Mitigation. Available online: https://www.cigionline.org/sites/default/files/documents/Paper%20no.137web.pdf (accessed on 11 February 2019).

Haley, James. 2018. De-Risking of Correspondent Banking Relationships: Threats, Challenges and Opportunities. Available online: https://www.wilsoncenter.org/sites/default/files/media/documents/article/2018_haley_report-edits-2-2018-final.pdf (accessed on 11 February 2019).

Head, George. 1973. Teaching Risk Management as a Decision Process. A Review of the New I.I.A. Program in Risk Management. *The Journal of Risk and Insurance* 40: 149–51. [CrossRef]

Human Security Collective (HSC), and European Centre For Not-For-Profit Law (ECNL). 2018. At the Intersection of Security and Regulation: Understanding the Drivers of 'De-Risking' and the Impact on Civil Society Organizations. Available online: http://ecnl.org/wp-content/uploads/2018/05/Understanding-the-Drivers-of-De-Risking-and-the-Impact-on-Civil-Society-Organizations_1.pdf (accessed on 12 February 2019).

International Finance Corporation (IFC). 2017. Can Blockchain Technology Address De-Risking in Emerging Markets? Available online: https://www.ifc.org/wps/wcm/connect/f0defcac-ebac-487e-912a-8d8a7a8b07ff/EMCompass+Note+38+Derisking+and+Blockchain+5-24+FINAL.pdf?MOD=AJPERES&CVID=lMWVNUO#:~{}:text=Blockchain%20technology%20can%20help%20with,%E2%80%9CKnow%20Your%20Customer%E2%80%9D%20requirements.&text=Blockchain%20has%20the%20potential%20to%20improve%20this%20situation (accessed on 4 February 2020).

International Monetary Fund (IMF). 2008. International Transactions in Remittances: Guide for Compilers and Users. Available online: https://www.imf.org/external/np/sta/bop/2008/rcg/pdf/guide.pdf (accessed on 3 February 2020).

International Monetary Fund (IMF). 2012. Malta: 2012 Article IV Consultation—Staff Report; and Public Information Notice on the Executive Board Discussion. Available online: https://www.imf.org/external/pubs/ft/scr/2012/cr12105.pdf (accessed on 3 February 2020).

International Organisation for Standardization (ISO). 2009. ISO 31000: ISO/Guide 73:2009(en.)—Risk Management—Vocabulary. Available online: https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en (accessed on 18 February 2020).

Keatinge, Tom. 2014. "Counter-Terrorist Regulation Restricts Charity Banking Worldwide ... "—Ncharitable Behaviour. Available online: https://www.demos.co.uk/files/DEMOSuncharitablebehaviourREPORT.pdf?1419986873 (accessed on 5 March 2020).

King, Russell. 1993. The geographical fascination of islands. In *The Development Process in Small Island States*. Edited by Douglas Lockhart, David Drakakis-Smith and John Schembri. London: Routledge, pp. 13–37.

KPMG. 2014. Global Anti-Money Laundering Survey 2014. Available online: https://assets.kpmg/content/dam/kpmg/pdf/2015/03/global-anti-money-laundering-survey-latest.pdf (accessed on 6 March 2020).

Lamanda, Gabriella, and Zsuzsanna Tamásné Voneki. 2015. Hungry for risk: A risk appetite framework for operational risks. *Public Finance Quarterly* 2: 212–25.

Langthaler, Jan, and Jonas López Niño. 2017. An Overview on De-Risking: Drivers, Effects and Solutions. Available online: http://www.asbasupervision.com/es/bibl/i-publicaciones-asba/i-2-otros-reportes/1598-an-overview-on-de-risking-drivers-effects-and-solutions/file (accessed on 11 February 2019).

Lautenschläger, Sabine. 2017. Is Small Beautiful? Supervision, Regulation and the Size of Banks. Available online: https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp171014.en.html (accessed on 3 February 2020).

MacDonald, Scott. 2019. Is there a 'New Normal' for De-Risking in the Caribbean? KYC360, October 22. Available online: https://www.riskscreen.com/kyc360/news/is-there-a-new-normal-for-de-risking-in-the-caribbean/ (accessed on 28 August 2020).

Mack, Natasha, Cynthia Woodsong, Kathleen MacQueen, Greg Guest, and Emily Namey. 2005. *Qualitative Research Methods: A Data Collector's Field Guide*. North Carolina: Family Health International.

Martin, Ivan. 2019. Bank of Valletta closes accounts of 'high risk' clients. *Times of Malta*. October 5. Available online: https://timesofmalta.com/articles/view/bank-of-valletta-closes-accounts-of-high-risk-clients.739819 (accessed on 13 February 2020).

McLean, Sheldon, Ydahlia Metzgen, Ranjit Singh, and Nyasha Skerrette. 2018. Economic Impact of de-risking on the Caribbean: Case Studies of Antigua and Barbuda, Belize and Saitn Kitts and Nevis. Available online: https://repositorio.cepal.org/bitstream/handle/11362/43310/S1701290_en.pdf?sequence=1&isAllowed=y (accessed on 16 February 2020).

Mugarura, Norman. 2016. Uncoupling the relationship between corruption and money laundering crimes. *Journal of Financial Regulation and Compliance* 24: 74–89. [CrossRef]

Neocapita. 2017. Blockchain Use Cases in International Development. Available online: https://medium.com/@neocapita/blockchain-use-cases-in-international-development-c02f940e960 (accessed on 20 February 2020).

Neuman, Lawrence. 2005. *Social Research Methods: Qualitative and Quantitative Approaches*, 6th ed. London: Pearson Allyn and Bacon.

Nisar, Ahsan. 2016. Bank De-risking and its Impact on MNCs. *Pakistan & Gulf Economist* 35: 10–12.

Orozco, Manuel, and Julia Yansura. 2013. Keeping the lifeline open: Remittances and Markets in Somalia. Available online: https://oxfamilibrary.openrepository.com/bitstream/handle/10546/297149/bp-keeping-lifeline-open-somalia-remittances-310713-en.pdf (accessed on 6 February 2020).

Paul, Scott, Anne-Marie Schryer-Roy, Ben Murphy, and Ed Pomfret. 2015. Hanging by a Thread: The Ongoing Threat to Somalia's Remittance Lifeline. Available online: https://oxfamilibrary.openrepository.com/bitstream/handle/10546/344616/bnhanging-by-thread-somaliaremittances-190215-en.pdf?sequence=14 (accessed on 5 February 2020).

Pirotta, Godfrey. 2001. A New Creation or an Image and Likeliness? The Maltese Experience of Establishing Local Government in a Centralised Micro-State. *Public Organization Review: A Global Journal* 1: 245–60. [CrossRef]

Saunders, Mark, Philip Lewis, and Adrian Thornhill. 2009. *Research Methods for Business Students*, 5th ed. Essex and London: Financial Times Prentice Hall.

Stake, Robert. 1995. *The Art of Case Study Research*. Thousand Oaks: SAGE Publications.

Starnes, Susan, Michael Kurdyl, Arun Prakash, Ariane Volk, and Shengnan Wang. 2017. De-risking and other Challenges in the Emerging market Financial Sector. *International Finance Corporation* 20: 2018.

Taylor, Paul. 2016. How Banks Can Avoid the De-Risking Trap. Available online: https://www.americanbanker.com/opinion/how-banks-can-avoid-the-de-risking-trap (accessed on 6 February 2020).

Van Broekhoven, Lia, Vanja Skoric, and Ben Hayes. 2014. De-Risking and Non-Profits: How Do You Solve a Problem that No-One Wants to Take Responsibility for? Available online: https://www.opendemocracy.net/en/de-risking-and-non-profits-how-do-you-solve-problem-that-n/ (accessed on 27 February 2020).

Vento. 2019. BOV's de-Risking Scheme Paves the Way for the Fall of Malta's Banking Sector. *MaltaWinds*. August 6. Available online: http://maltawinds.com/2019/08/06/bovs-de-risking-scheme-paves-the-way-for-the-fall-of-maltas-banking-sector (accessed on 13 February 2020).

Woodsome, Jim, and Vijaya Ramachandran. 2018. Fixing AML: Can New Technology Help Address the De-risking Dilemma? Available online: https://www.cgdev.org/reader/fixing-aml-can-new-technology-help-address-de-risking-dilemma?page=0 (accessed on 24 February 2020).

World Bank. 2015. Withdrawal from Correspondent Banking: Where, Why, and What to Do about It. Available online: http://documents.worldbank.org/curated/en/113021467990964789/pdf/101098-revised-PUBLIC-CBR-Report-November-2015.pdf (accessed on 19 February 2020).

World Bank. 2016. De-risking in the Financial Sector. Available online: https://www.worldbank.org/en/topic/financialsector/brief/de-risking-in-the-financial-sector (accessed on 4 March 2020).

Xuereb, Kieran, Simon Grima, Frank Bezzina, Andre Farrugia, and Pierpaolo Marano. 2019. The Impact of the General Data Protection Regulation on the Financial Services' Industry of Small European States. *International Journal of Economics and Business Administration* 7: 243–66. [CrossRef]

Yazan, Bedrettin. 2015. Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report* 20: 134–52.

Yin, Robert. 2002. *Case Study Research: Design and Methods*, 3rd ed. Thousand Oaks: SAGE Publications.

Zhang, Xiang. 2016. Factors that Shape an Organisation's Risk Appetite: Insights from the International Hotel Industry. Available online: https://pdfs.semanticscholar.org/e8a0/3a6d62000aa374fc9444f512bb7bee7c093e.pdf (accessed on 18 February 2020).

Journal of
Risk and Financial Management

Article

Plato as a Game Theorist towards an International Trade Policy

George E. Halkos *, Nicholas C. Kyriazis and Emmanouil M. L. Economou

Department of Economics, University of Thessaly, 38333 Volos, Greece; nkyr@uth.gr (N.C.K.); emmoikon@uth.gr (E.M.L.E.)
* Correspondence: halkos@econ.uth.gr; Tel.: +30-2421-074-664

Abstract: In the beginning of the second book of his *Politeia* (Republic) Plato in passage 2.358e–359a–c raises the issue of the administration of justice as a means of motivating people to behave fairly regarding their relationships and when cooperating with each other because, at the end, this is mutually beneficial for all of them. We argue that this particular passage could be seen as a part of a wider process of evolution and development of the institutions of the ancient Athenian economy during the Classical period (508–322 BCE) and could be interpreted through modern theoretical concepts, and more particularly, game theory. Plato argued that there are two players, each with two identical strategies, to treat the other justly or unjustly. In the beginning, each player chooses the "unjust" strategy, trying to cheat the other. In this context, which could be seen as a prisoner's dilemma situation, both end with the worst possible outcome, that is, deceiving each other and this has severe financial consequences for both of them. Realizing this, in a repeated game situation, with increasing information on the outcome and on each other, they choose the "just" strategy so achieving the best outcome and transforming the game in a cooperative one. We analyze this, formulating a dynamic game which is related to international commercial transactions, after explaining how such a situation could really arise in Classical Athens. We argue that this is the optimal scenario for both parties because it minimizes the risk of deceiving each other and creates harmony while performing financial transactions.

Keywords: Plato; fairness; commercial transactions; international trade; game theory

JEL Classification: B11; C71; C73; D81; F33; K20; N23

Citation: Halkos, George E., Nicholas C. Kyriazis, and Emmanouil M. L. Economou. 2021. Plato as a Game Theorist towards an International Trade Policy. *Journal of Risk and Financial Management* 14: 115. https://doi.org/10.3390/jrfm14030115

Academic Editor: Shigeyuki Hamori

Received: 21 February 2021
Accepted: 6 March 2021
Published: 10 March 2021

1. Introduction

During recent years, scholars have researched various aspects of the ancient Athenian economy of the Classical period (508–323 BC), showing its modern character and institutional setup in many areas, such as banking, insurance and other financial services.

Bresson (2016a, 2016b), O'Halloran (2018), Economou and Kyriazis (2019) and Bitros et al. (2020) among other authors, have analyzed the market-type of economic institutions that ensured the smooth functioning of the Athenian market while Bitros and Karayiannis (2008) have analyzed the related issue of entrepreneurship in Athens under free market economic principles. Stewart et al. (2020) focused on market specialization and division of labor in Classical Athens. Economou and Kyriazis (2017, 2019) focused on the importance of creation of an institutional setup which effectively protected private property and commercial contracts through the establishment of courts. These courts known as *Heliaia*, provided justice rapidly through highly capable and experienced jurors. Such commercial cases, known as *dikai emporikai* were judged within a month, so that justice was provided rapidly among litigants (Cohen 1992). The speed of litigation is of high importance so as to reduce significantly transaction costs (Coase 1960), and protection of commercial transactions through the law so as to become credible is a pivotal intertemporal axiom (Hodgson 2015).

Bresson (2016a, 2016b), Harris and Lewis (2016), O'Halloran (2018), Bitros et al. (2020), and Economou et al. (2021) analyze the international commercial orientation of the Athenian economy. During 478–404 BCE Athens benefited by international commerce, mainly with its 316 allied states (according to Figueira 1998, p. 52) through the Delian League and this led its economy to achieve economic growth, at least for some specific years. Thus, Athens had actually managed to integrate a large part of the Eastern Mediterranean region into a unified area of economic cooperation and security through its mighty fleet. In relation to this Halkos and Kyriazis (2010) and Bitros et al. (2020) analyzed the development of a prosperous "shipping industry" that finally lead the port of Athens, Piraeus, to became, according to Cohen (1992, p. 141) an international entrepôt of antiquity. In the next section we further elaborate on the recent literature regarding the Athenian economy during the Classical period.

In this paper we argue that by trying to raise the issue of justice in human affairs, Plato, who as a citizen of Athens experienced the practice all these economic development of his times, tried to interpret the different outcomes according to whether a party deal is conducted (i) in an unfair manner ("unjust"), or (ii) in a fair manner ("just"). For Plato, the best way to ensure a mutually prosperous situation for two parties is to behave in a "just" way since this decreases the risks of wrongdoing between two parties which is important for both of them to be satisfied through their cooperation. We believe that Plato developed some ideas through which, in a wider sense, he could be considered as a pioneer "game theorist" ever as we purport to show in the next sections and this aspect of his writing has not been analyzed before. To show this, we make use of a game theoretical analysis by trying to interpret the commercial diastasis of Plato's relative passage as a two-stage prisoners' dilemma. This game has a sub-game perfect outcome for every outcome of the first stage, while we have an exclusive Nash equilibrium for the second stage game. Actually, to ensure perfect information, we include the state and the inspection mechanisms avoiding a sub-game perfect game with players' strategies representing a Nash equilibrium output in every sub-game. Instead, and under the threat of inspection mechanisms, cooperation seems to be the way to limit risk in commercial transactions.

With this paper, we contribute to the international bibliography which argues that fair and honest behavior while performing commercial transactions is a key intertemporal axiom for reducing financial risks. We argue that such a behavior further sets the conditions for trade to flourish and the economy to grow in the long run, providing that the functioning of the economy is supported by a series of efficient institutions and auditing mechanisms which ensure the protection of private property, sound money, and the development of successful financial institutions such as banking, insurance, etc. Characteristic historical cases of nations which played a key economic and geopolitical role in their time, such as Medieval Venice, Early Modern Europe's England, and the United Provinces (Dutch Republic), Great Britain, and currently the USA (Ferguson 2008; Puga and Trefler 2014; Economou and Kyriazis 2017), verify this financial axiom and with this paper we argue that we contribute to the relative discussion which should also include evidence from Athens during the Classical period and the time of Plato and later on.

The paper is structured as follows: Section 2 further analyzes in brief the institutional setup of the Athenian economy by arguing that it was export-oriented and favorable to international commerce. We do this in order to show that Plato has expressed his relevant views in an economic environment that favored, in practice, the formulation of a number of such views. We focus on two very important institutions, the *agoranomoi* and the *dokimastai*, and we argue that a key reason for the success of the internationally-oriented Athenian economy was the usage of sound money. In Section 3 we provide the relative passage of Plato in a verbatim way, followed by explanations and a further theoretical analysis. In Section 4, we present a fictional case regarding the commercial diastasis of Plato's passage by describing a transaction between two trading parties, followed by the formulation of a two-stage repeated prisoner's dilemma setup with complete but imperfect information.

Equally important is that such kind of transactions were the norm in the city-state of Athens during Plato's era.

We argue that in the beginning, each player chooses the "unjust" strategy, trying to cheat the other and this could be seen as a prisoner's dilemma situation, which ends in the worst possible outcome. Realizing this, in a repeated game situation, both parties choose the "just" strategy by adding auditing mechanisms in the game and they manage to achieve the best outcome by transforming the game into a cooperative one which maximizes the outcome for both.

Put it differently, being inspired by Plato's views, we argue that the success of commercial cooperation between two partiers can be achieved when auditing institutional mechanisms of minimizing the risks of performing financial transactions do exist and are functional in practice. Section 5 summarizes our findings and concludes by providing some further research directions inspired by the overall analysis.

2. The Athenian Economy during the 4th Century BCE Period

As it was already mentioned in the introduction, the success of the Athenians should be attributed to the superior package of (the free market type of) economic institutions they developed, even if in a primitive form in relation to today. Bresson (2016a, 2016b), Harris and Lewis (2016), O'Halloran (2018), Bitros et al. (2020), Economou et al. (2021), among others, have analyzed the Athenian "international" trade policy through the establishment of a unified "economic area" for interstate commercial transactions between the member city-states of the Delian League while Economou et al. (2021) add that this area should be also seen as an ad hoc monetary union. According to Cohen (1992, p. 141) in the port of Piraeus, the products imported could be sold on site and directly exported to new end-user markets. The port of Piraeus had become an international entrepôt of antiquity, as was Alexandria during the Hellenistic period (323–146 BCE), Amsterdam during the 17th century AD and among others, Hong Kong, Shanghai, and Rotterdam today. Bresson (2016a, p. 414) writes:

> "In the Classical period, Athens was obviously the greatest market in the Mediterranean world, the one that attracted all the merchants who wanted to sell a large cargo rapidly at a good price."

In accordance to the above, Cohen (1992) and Bitros et al. (2020), among others, have analyzed the Athenian internationally-oriented banking system. In particular, Cohen (1992) in his seminal book *Athenian Economy and Society: A Banking Perspective* analyzed exhaustively the sophisticated way banks functioned in Classical Athens. Banks' main services were: (i) the exchange of coins and foreign currencies, (ii) the acceptance of deposits, (iii) carrying out payments on behalf of their customers, (iv) providing loans to various business operations, including bottomry loans in shipping and even financing of consumer credit, (v) providing sureties, negotiated claims, and offered guarantees and personal advice to important customers, (vi) accepting documents and valuables for safekeeping but they did not act in the capacity of pawnbrokers, and (vii) facilitating export-import activities by settling payments among importers from and exporters to merchants abroad.

Cohen (1992, p. 141), Acton (2014, pp. 252–53), and Bitros et al. (2020), among others, further argue that banks also offered insurance on these maritime loans to ship-owners and merchants so as to reduce any possible negative financial consequences related to the great risks that were involved in transport journeys over sea lanes. Thus, insurance went hand in hand with the process of issuing loans with interest, so as to further promote international trade activities by minimizing the risks when performing international financial transactions.

Even primitive versions of joint stock companies were developed for various purposes such as financing maritime commerce through banks with more than one owner-shareholder, or the exploitation of the silver mines at Laurion (Cohen 1992, p. 153; Bitros et al. 2020, p. 24). Here again, these companies were the result of the attempt of the Athenian bankers or merchants to limit any negative financial consequences while performing trade; in other words, to minimize the risk of providing loans through an international trade environment. Halkos and Kyriazis (2010) further analyzed the process of "industrialization" that made Athens the first economy in history where "services" and "industrial" (handicrafts) sectors contributed more to GDP than agriculture.

To this point it has also to be acknowledged that the literature and the references that we provide herein consist only a small part of the international bibliography in the last 30 years that revisits the older, though influential, views of Finley (see, e.g., Finley 1973, 1983) and his followers who argued that the Athenian (and by extension ancient Greek) economy was characterized by primitive structures and organization and that it was "embedded" on social, political, and religious aspects, meaning that it was not actually a market economy. Among many others Tridimas (2019, p. 12) acknowledges that, during the 1970s and 1980s, Finley's view was the established orthodoxy but, in light of new evidence and advances in economic theory, "formalist" accounts gradually won ground. He adds that "formalism" maintains that economic actors are essentially the same in every place and in every period, so that we can treat both ancients and modern societies' citizens as seeking to maximize profits and utility as posited by the rational choice model. It is also necessary to be mentioned at this point that for space limitation purposes we intentionally choose to provide only selected references regarding some of the most recent works available in the bibliography that are related to the specific issues we discuss in this paper.

Under such a context, below we focus only on two key importance institutions which are linked with what we call as Plato's theory of justice and the commercial diastasis of his views which we further elaborate through a game theoretical analysis. These were the *agoranomoi* and the *dokimastai*. The first was a group of 10 public magistrates, elected by the Athenian Assembly of citizens under direct democracy procedures, responsible to ensure that the prices of goods in the market were not excessive. These actions were important so as to combat profiteering (Elliot 2018; Bitros et al. 2020). The *agoranomoi* also had to ensure that the quality of the products met their specifications. If this was not the case, they put fines to the sellers. They further solved disputes among merchants or between buyers and sellers in the market. This intervention could result in transactions without fear of violence or intimidation according to Harris and Lewis (2016, p. 30). The second and equally important institution were the so-called *dokimastai* (*dokimastes* in singular) which are related to the introduction of reliable currencies for the exchange of goods and services and the impressive increase of the volume of commerce between Athens and its allies (Economou et al. 2021).

Figueira (1998), van Alfen (2011), and Bitros et al. (2020), among others, argued that the so-called Athenian silver drachma became the universal coin in the Eastern Mediterranean because of its purity of silver content, which minimized the risks of performing reliable financial-commercial transactions between Athens and its allies or every other state. The most famous of these series of drachmae issued by the Athenian state silver mint, known as *Argurocopeion*, was the Athenian *tetradrachm* (four drachmae) which rapidly became the internationally-accepted leading currency throughout the Mediterranean (Scheme 1).

Scheme 1. An Athenian *tetradrachm*. Source: http://www.wildwinds.com/coins/greece/attica/athens/sg2526.jpg (accessed on 2 January 2021).

The reliability of these coins was safeguarded through the introduction by the Athenian Assembly of so-called *Monetary Law of Nicophon* (being the lawgiver who proposed the law). Under its provisions, all "good" (meaning of correct silver content) foreign coins and all "good" imitations of the Athenian drachma issued in mints outside Athens could circulate in the Athenian economy together with Athenian drachmae. Thus, foreign traders, etc., could use the coins of their preference for their transactions without having to change them into Athenian providing that they were not counterfeits.

In cases of doubt about the "purity" of the content of the foreign coins (and as a state guarantee against fraud) the office of the "testers" (known as *dokimastai*) was introduced. The "testers", being themselves public slaves who acted as state officials, had an office, a bench in the Agora (the marketplace) of Athens and at the harbor of Piraeus. If one of the private contracting parties had doubts regarding the purity of the foreign coins, he could bring them to the "tester" who examined their purity. If found impure (e.g., in the case of fraud) the coins were confiscated, then cut and next, became sacred property of the treasury-temple of the Mother of the Gods. If found to be authentic imitations of the Athenian drachma, (meaning that they were of the proper silver content) they were returned to their owners, who then had the right to use them legally in order to perform commercial transactions in the market (Figueira 1998; Engen 2005). Having the above in mind (Ober 2008, pp. 222–25), Bitros et al. (2020) and Economou et al. (2021) argue that the Law of Nicophon reduced transaction costs and generated, by its existence and provisions, trust between the parties.

As a general comment, both the *agoranomoi* and the *dokimastai* were institutions introduced by the Athenian state so as to limit to the minimum any kind of risk regarding the reliability of performing financial transactions in the Athenian market.

3. Plato's Game Theory of Justice: The Concept

Except the above brief description regarding the commercial orientation of the Athenian economy, what is also important is that by analyzing the ancient evidence in books, speeches, extant decrees, etc., it becomes clear that Greek thinkers/philosophers had anticipated many theoretical aspects of the working of the economy, such as the importance of efficient and fair administration of public economics (Xenophon Memorabilia, 3.4.7-12), the crucial issue of property and property rights protection (Aristotle, Politics, 1262b. 23–25, 1263a, 1278a; Demosthenes Against Timocrates, 149), the issue of the specialization of professions (Lysias Against Eratosthenes, 19), and the importance and the benefits of commercial exchange in order to establish viable international markets Pseudo-Xenophon (Constitution of the Athenians, 2.11-12; Plato Republic 2.370e–2.371a, 2.373d–e; Isocrates Panegyricus, 4.42).

Furthermore, the issues of division of labor and labor productivity were discussed by Xenophon (Cyropaedia, VII, 2.5), almost 21 centuries before Adam Smith raises this issue in Book I. 1.17–18 of his *Wealth of Nations*. In addition, the market mechanism, and the adjustment of prices of goods through the forces of demand and supply were also discussed by Xenophon (Ways and Means, IV). The benefits of free circulation of good currencies and the need to withdraw counterfeited coins from the market as a means of minimizing the risks when performing commercial transactions were discussed by Aristophanes (Frogs, 718–733). Even what we call today "marginalist" thinking was, to some extent, discussed by Plato (Republic 2.369b–e, 370–372, 373b–c).

We have to acknowledge that the evidence we provide here regarding the economic thinking of ancient Greek philosophers needs more clarification, however, this exceeds the scope of our paper. Detailed analysis regarding the ancient Greek philosophical thinking regarding economic matters are provided, among others, by Amemiya (2004), Baloglou (2012), and Stewart et al. (2020).

It is also to be clarified that the above philosophical views regarding economic institutions were not just the outcome of philosophical thinking or logical deductions of the philosophers regarding the theoretical foundations of economic principles but, more importantly, it was the description of how the Athenian economy of the time worked in practice on a daily basis. What we mean is that when Plato wrote *The Republic* in 375 BCE he was experiencing the daily working of the functioning of the institutions as described in Section 2 in practice. We believe that through such an economic environment Plato developed some ideas through which, in a wider sense, he could be considered as the first "game theorist" ever, and this aspect of his writing has not been analyzed before.

Our main aim in this paper is to focus and analyze in more detail Plato's 2.358e–359a–c passage. In *The Republic* (2.373d) Plato writes that two neighboring city states try to permanently acquire things (land in this case) in competition with one another, thus "undertaking the marginal point of their needs". Then, in *The Republic* (2.358e–359a–c) he writes the following (We follow the translation of the ancient Greek text as is provided by the Perseus Digital Library of the University of Tufts):

> "By nature, they say, to commit injustice is a good and to suffer it is an evil, but that the excess of evil in being wronged is greater than the excess of good in doing wrong. So that when men do wrong and are wronged by one another and taste of both, those who lack the power to avoid the one and take the other determine that it is for their profit to make a compact with one another neither to commit nor to suffer injustice; and that this is the beginning of legislation and covenants between men, and that they name the commandment of the law the lawful and the just, and that this is the genesis and essential nature of justice—a compromise between the best, which is to do wrong with impunity, and the worst, which is to be wronged and be impotent to get one's revenge. Justice, they tell us, being mid-way between the two, is accepted and approved, not as a real good, but as a thing honored in the lack of vigor to do injustice, since anyone who had the power to do it and was in reality 'a man' would never make a compact with anybody either to wrong nor to be wronged; for he would be mad. The nature, then, of justice is this and such as this, Socrates, and such are the conditions in which it originates, according to the theory. If we grant to each, the just and the unjust, license and power to do whatever he pleases, and then accompany them in imagination and see whither his desire will conduct each. We should then catch the just man in the very act of resorting to the same conduct as the unjust man because of the self-advantage which every creature by its nature pursues as a good, while by the convention of law it is forcibly diverted to paying honor to equality . . . "

Plato proposes here an underlying utility approach to justice. Justice is not "loved" per se, but because, if applied equally, increases utility in that it prohibits wrongdoing, which, for the person receiving it, reduces his utility. The "biggest good", for a person, would be if he could do wrong without punishment, which would increase his potential utility. The "biggest bad" is if he receives wrong without being able to retaliate, so that justice equilibrates or is an intermediation between those two extremes, balancing positive and negative potential effects of the two persons.

Further, it becomes clear that Plato, known as a moral philosopher, here perceives persons as "pure" utility maximizers. A person does not refrain from doing wrong due to his morals, etc., but because this could potentially harm his utility if the other person retaliates, or if the law prohibits it and he faces consequences if caught. Thus, in 2.359c Plato calls for establishing an environment of commonly acceptable values as rules of behavior regarding justice since this will be the way to ensure equality between individuals, the people themselves, in the society. In actuality, Plato calls for common protocols of behavior which minimize the risk of fraud and ensure the establishment of a prosperous financial environment for international trade.

Such protocols of behavior are related to the set of institutions as mentioned above in Sections 1 and 2 in brief. However, the success of the international economic-commercial networks that Athens introduced with her allies, especially in the 4th century BCE, should be attributed to some further institutions that were introduced, such as *proxeny* (a system of developing diplomatic relationships between states), *asylia* (exemption from reprisals from one city-state to another), *isopoliteia* (political equality), *enktesis* (rewarding citizenship and property rights to own land to foreigners), and *symbola* (interstate monetary union and bilateral commercial-diplomatic agreements (Harris and Lewis 2016; Economou 2020).

The connection between the rule of law and equality as a prerequisite for establishing a prosperous society, meaning a fair and just society, is also verified by modern literature (see among others, Bingham (2011) and Gowder (2016)). Epistemic disciplines in social sciences, such as the New Institutional Economics school of thought, consider justice and the rule of law mainly front the institutional point of view: Equality in justice is seen as a highly important prerequisite which guarantees responsible governance and contributes to economic growth (Acemoglu and Robinson 2013; Hodgson 2015). Plato's call for commonly accepted rules of behavior that are related to justice is also verified by modern scholars such as Greif (2006, p. 30), who concludes at the same wavelength that institutions are a system of rules, beliefs, patterns of behavior, and organizations that together create a regularity of social behavior.

In the following section we present a scenario with fictional characters regarding how Plato's theory of justice functioned in practice, through a game theoretical analysis. It is worth to be mentioned that the following game is not just a fictional scenario but a way to bring to life how the financial institutions of *agoranomoi* and the *dokimastai* functioned in practice in Classical Athens in the time of Plato. A description on how the institution of the *dokimastai* further functioned is provided by various ancient authors such as Aristotle (*Athenian Constitution* 51.2.).

4. Plato's Game Theory of Justice. A Fictional Scenario Based on Real Life Practices

As it has already been argued, in Plato's time, Piraeus, Athens twin city and harbor, was the Mediterranean's entrepot. Merchants from all over the world could find in Piraeus all goods (even rare ones like silk and spices) without having to travel to their sources.

4.1. A Fictional Case: Apollonius; A Merchant from Emporium

A merchant from distant Emporium (today's Ampurias in Southeastern Spain) Apollonius, arrives in Athens in order to buy Athenian jewelry, reputed for its artistic quality and craftsmanship. There, he meets the Athenian trader and jewelry merchant Heraiphon. They agree on the transaction, the jewelry to be bought, and the price to be paid.

Apollonius suggests paying in his own currency, Emporium drachmae, instead of changing in Athenian drachmae, in order to avoid the banking fee, something legally possible under Nicophon's Decree which permits the free circulation of all goods, thus reducing transaction costs. Heraiphon agrees and asks about the silver content of Emporium drachmae, Apollonius answering that it is exactly the same as Athenian drachmae (which would give as an exchange rate of one to one and the same price in both currencies). Still, Heraiphon has some doubts about the silver content of the Emporium drachma (with which he has never traded up till now) and decides to bring them in front of the official tester (dokimastes), provided by the city under Nicophon's Decree in order to be sure that the coins are indeed pure. The tester takes one at random, tests it, and discovers that Apollonius tried to cheat: His drachmae are not of pure silver content, but silver plated. The tester, as the Decree provides, confiscates all of the coins, and Apollonius suffers a substantial financial loss.

Angry, Apollonius wishes to retaliate and calls an *agoranomos*, one of those officials of the city being responsible for the good functioning of the market, protection against quality fraud, profiteering, etc. He accuses Heraiphon that he tried to cheat him selling him jewelry, alleged to be of pure gold, while in fact they are gold plated. The *agoranomos* tests the jewelry and vindicates Apollonius. Heraiphon tried to cheat him offering him gold plated jewelry instead of pure gold. Now, Heraiphon has to pay a fine, which corresponds to a financial loss. This is exactly a situation portrayed by Plato in the first part of the game: Both try to cheat, are found out, and in the end are worse off.

After some years, Apollonius comes back to Athens and meets Heraiphon. In view of their previous experience, negative for both, this time they are both honest and conclude their exchange. This is the second situation described by Plato; they agree "neither to do wrong". What the Athenian institutional setup and, in particular, Nicophon's Law provided for, was to transform asymmetrical information to symmetrical. Thus, market exchanges were facilitated, since information gathering is part of the transaction cost. Symmetrical information lowers transaction costs. By contrast, the existence of asymmetric information is well known and has been analyzed by many prominent economists such as Akerlof (1970) as one of the main problems of the functioning of markets.

The Athenian institutional setup achieved two particular aims: it gave symmetrical information on the quality-purity of the different currencies, and information on the character of the market participants, if they were honest or dishonest. In the first period of the exchange (or game) the two participants have asymmetric or even non-existent information on both issues. The Athenian does not know if the foreign currency is good, and the citizen from Ampurias does not know if the jewelry is original, and both participants do not know each other's character. The element of trust is not present. The office of the tester provides the solution. The tester gives his verdict on the quality of the currency, so that both acquire symmetrical information. Both know then if the currency is good or not. Further, they acquire symmetrical information on each other's character: They know if they are honest or not. Building upon this acquired information, in the next period they can modify their strategy, thus breaking out of the prisoner's dilemma and transforming the game into a cooperative one.

Cooperative games are based on trust and shared symmetrical information. A vast bibliography of modern literature in economics and organizational theory verifies the view that trust is a lubricant enabling organizations and societies to achieve Pareto superior outcomes (i.e., Arrow 1974; Fukuyama 1995; Kramer 1999). For example, Adler (2001) argues that trust itself is a coordination mechanism that lowers transaction costs between firms. By contrast, if uncertainty increases this has financial consequences which may jeopardize the smooth functioning of an economy (Halkos and Zisiadou 2020). Nowadays, trust and reliability of commercial transactions through digital money is ensured through consumer protection laws and through specific practical norms and procedures regarding using machines for money transferring, such as ATMs, e-banking through authorized web-pages, money transfers from account to account, etc. In Classical Athens, this was ensured by auditing mechanisms such as the *agoranomoi* and the *dokimastai*.

4.2. The Game

We set up the non-zero sum prisoners' dilemma in the normal form representation with each player simultaneously selecting a possible strategy. If we denote the strategy space of the set of possible strategies for player *i* as Ω_i then ω_i is an arbitrary element of this strategy space with $\omega_i \epsilon \Omega_i$, that is each element is part of the strategy space Ω_i.

If $(\omega_1, \omega_2, \ldots, \omega_n)$ is the combination of strategies one for each player and w_i the payoff function then:

$$w_i(\omega_1, \omega_2, \ldots, \omega_n)$$

and the game may be expressed as:

$$G = \{\Omega_1, \Omega_1, \ldots, \Omega_1; \omega_1, \omega_2, \ldots, \omega_n\} \tag{1}$$

In our game the players choose their strategies simultaneously having complete information as if they will/may have the help of the authorities.

In the normal-form game (1) if ω_i' and ω_i'' are elements of the strategy space and feasible strategies for player *i* then strategy ω_i' will be strictly dominated by ω_i'' if:

$$w_i(\omega_1, \omega_2, \ldots, \omega_{i-1}, \omega_i', \omega_{i+1}, \ldots, \omega_n) < w_i(\omega_1, \omega_2, \ldots, \omega_{i-1}, \omega_i'', \omega_{i+1}, \ldots, \omega_n)$$

If we consider a game of complete (due to the feasibility of using the appropriate state inspection mechanisms) but with imperfect information as both players select their strategies simultaneously, player 1 selects action ω_1 from feasible set Ω_1 and player 2 observing ω_1 selects action ω_2 from feasible set Ω_2 with payoffs u1(ω_1, ω_2) and u2(ω_1, ω_2). In a two-stage game of complete but imperfect information players 1 and 2 select at the same time actions ω_1 and ω_2 from feasible sets Ω_1 and Ω_2, respectively. Players 1 and 2 observe the first stage (ω_1, ω_2) and select at the same time ω_3 and ω_4 from feasible sets Ω_3 and Ω_4 respectively with payoffs:

$$u_i(\omega_1, \omega_2, \omega_3, \omega_4) \qquad \text{for } i = 1,2$$

For each feasible outcome of the first stage (ω_1, ω_2) we assume the second stage of the players has a unique Nash equilibrium as:

$$\omega_3^*(\omega_1, \omega_2),\ \omega_4^*(\omega_1, \omega_2)$$

Expecting these payoffs for players 1 and 2 in the first stage then if (ω_1^*, ω_2^*) is the unique Nash equilibrium outcome of this simultaneous game then $(\omega_1^*, \omega_2^*, \omega_3^*(\omega_1^*, \omega_2^*),$ $\omega_4^*(\omega_1^*, \omega_2^*))$ is the sub-game perfect outcome of this two-stage game. If player 1 does not play α_1^* in the first stage player 2 may think about the expectation (assumption) that player 1 will play $\omega_1^*(\omega_1, \omega_2)$ in the second stage.

We can create our analysis regarding Plato's passage 2.373d through a two-stage prisoners' dilemma and for each possible outcome of the first stage of the game (α_1,α_2) the second stage prisoners' dilemma gives a unique Nash equilibrium represented as $(\omega_3^*(\omega_1, \omega_2), \omega_4^*(\omega_1, \omega_2))$.

In Table 1 we present the game payoffs matrix described in the fictional case as analyzed above. In this first stage of the game both players trust each other and there is no need for state intervention through its auditing mechanisms (the *agoranomoi* and the *dokimastai*). The fact that there is no state auditing intervention increases the risk of both parties to be cheated and, thus, suffer inferior financial outcomes.

Table 1. First stage payoffs with players trusting each other—no need for state intervention.

Strategies		
Heraiphon \ Apollonius	Unjust (dishonest)	Just (honest)
Unjust (dishonest)	(2, 2)	(15, 5)
Just (honest)	(5, 15)	(10, 10)

In the first case described by Plato, (both "unjust") the outcome is negative for both and this is akin to a prisoner's dilemma situation. Both Apollonius and Heraiphon face a loss. Apollonius has been deceived since he accepts jewelry of inferior quality while Heraiphon accepts fraudulent money. Based on the procedures as described by Bitros et al. (2020) and Economou et al. (2021) regarding currency issuing and the role of the Athenian silver mint we assume that at least both Apollonius and Heraiphon do not face a total loss (0,0). The jewelry that Apollonius accepts has at least a small value, while, on the other hand, Heraiphon can apply to the silver mint, and ask from the staff to melt the fraudulent coins, then receiving back a small percentage of silver bullion, and then asking the staff to produce (unadulterated) silver coins for him under a small *seignorage* cost in favor of the state. Thus, we assume that the two players achieve a very small gain, that is (2, 2). On their part, the pairs (5, 15) and (15, 5) denote that, in both cases, one of the two players is cheated while the other player benefits more that the legal gain due to deceiving the other party. Finally, in this game both players gain when, from the beginning, they wish to play fair (10, 10).

If the game is repeated (as Plato presumes) both players "learn", by incorporating their past experience and, thus, choose the alternative "just" strategy, which maximizes the outcome for both. It is to their benefit to agree and behave in a "just" way. The game is transformed into a cooperation game. We symbolize "just" as "J" and "unjust" as "U".

The pure strategy payoffs are (2, 2) (5, 15) (15, 5), and (10, 10). Other possible payoffs include the pairs (x,x) for (2 < x < 10) which come from weighted averages of (2, 2) and (10, 10) an the pairs (y, z) for y + z = 15 and 5 < y < 15 as weighted averages of (5, 15) and (15,5). The other pairs on the interior of the shaded region are weighted averages of more than two pure strategy payoffs. These are shown in Figure 1 inspired from Gibbons (1992).

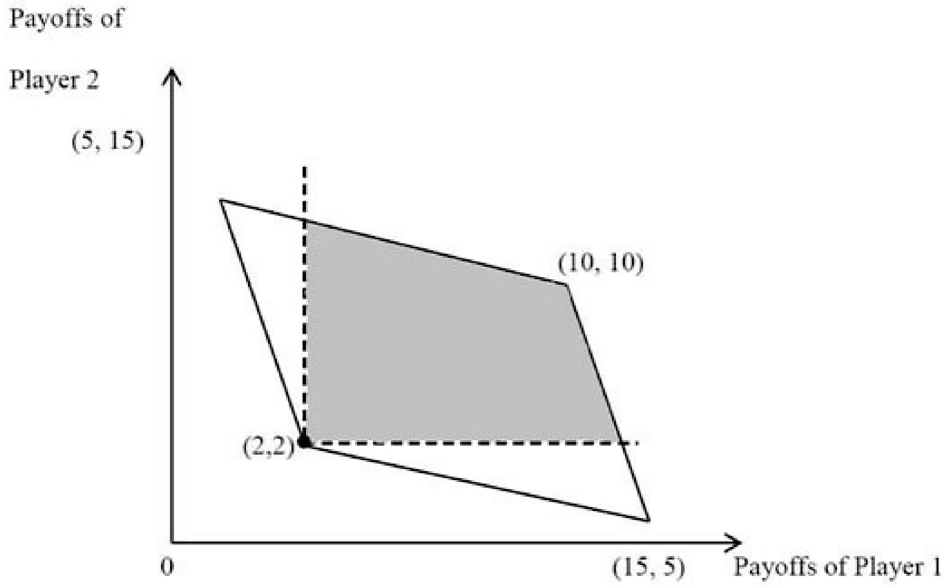


Figure 1. Graphical presentation of potential payoffs.

If we allow for exchange rate differentiation other than 1:1 (Athenian drachmae and foreign drachmae from Ampurias) any point in the shaded area in Figure 1 can be attained as the average payoff in a sub-game perfect Nash equilibrium of the repeated game but with a discount rate close to 1:1. To have the payoffs for different time horizons available we recall than and the sequence becomes (Fudenberg and Tirole 1993):

$$\frac{1-e}{1-e^{T-1}}\sum_{t=0}^{T} e^t [G_1(\omega)^t] \quad (2)$$

with the use of the average discounted payoff e. If the game is played once only, the cooperation is strongly dominant and the unique equilibrium for both players is just. If the game is repeated a finite number of times sub-game perfection requires both players to defect in the last period and backward induction implies that the unique sub-game period equilibrium is for both players to be just in every time period.

If the game is played infinitely then "both just every period" remains a sub-game perfect equilibrium. If the horizon is infinite and $e > \frac{1}{2}$ then they may cooperate in the first period and continue to cooperate so long as no player ever deviates from the just:

$$(1-e)(1+e+\ldots+e^{t-1}+2e^t+0+\ldots) = 1-e^t(2e-1) \quad (3)$$

which is less than 1 as $e > \frac{1}{2}$. With these strategies there are two classes of sub-games, class A with no player being just and class B with just i having occurred.

In the game theory setup the notions of dominance, iterated dominance, as well as Nash equilibrium stem from the hypothesis that players predict their opponents' moves by "introspection and deduction" relying on their knowledge of their opponents' rational payoffs and the simultaneous knowledge that each player knows that the other player is familiar with these matters. In this prisoners' dilemma setup it can be observed that if the game G presents a unique Nash equilibrium we may expect for any finite time horizon the repeated game G(T) to take a unique sub-game perfect result as an outcome the Nash equilibrium in each stage G. More specifically, if the game is played twice and we observe the first stage outcome before then there is a sub-game perfect outcome in this repeated game if (U,U) is played in the first stage. The two-stage prisoners' dilemma is a game with a sub-game perfect outcome as for each possible outcome of the first stage there is a unique Nash equilibrium for the second-stage game.

Allowing for state intervention leads to cooperation as the way to reduce risk in commercial transactions. That is, now the whole situation changes if we add the government and the inspection mechanisms, that is, the *agoranomoi* and the *dokimastai* as active participants in the market, who can impose fines to both parties, Heraiphon and Apollonius.

In this case, we hypothesize that both Heraiphon and Apollonius do not trust each other and before finishing their transaction they both wish to apply to the state auditing mechanisms, the *agoranomos* and the *dokimastes*. Then we have multiple Nash equilibrium situations that do not belong to this class of games. The best outcome for both Heraiphon and Apollonius will be (10, 10). They both choose to play 'just', but they still wish to turn to state auditing mechanisms, the *agoranomos* and the *dokimastes*, for an even more formal control of money and jewelry, perhaps because both of them are extremely law-abiding! It should also be noted that resorting to the two auditing mechanisms as mentioned above does not incur any costs for both Heraiphon and Apollonius. This is because a state fine can be imposed only to those who try to cheat. In other words, since no one tried to cheat the state's payoff in this case is (0, 0).

Thus, players expect that (10, 10) will be the second stage outcome if the first stage outcome was (10, 10) having now (1, 1) as a payoff if any of the other first stage outcomes takes place under complete information put forward by the inspection mechanisms. These inspection mechanisms may avoid other strategies, like trigger or tit-for-tat, in an infinitely repeated prisoners' dilemma game leading to sub-game perfect Nash equilibriums in every sub-game.

In Table 2 the pair $(-1, -1)$ indicates that the *agoranomos* finds that the jewelry of Hairephon is of lesser value, and orders their confiscation. Additionally, counterfeit money is confiscated by the *dokimastes*. The two merchants lose everything, and, in addition, a fine is imposed on both of them by the *agoranomos* because they tried to cheat each other, which could also cause malfunction on the Athenian market in a wider sense. Thus, here there is a total loss for both of them. The state on its part receives (1, 1) because it can impose fines to both of them, because they tried to cheat each other and this behavior may cause malfunctions or erode the prestige of the Athenian market.

Table 2. Payoffs of the second stage—both players do not trust each other.

		Strategies		
		Apollonius		
	Apollonius / Heraiphon	Unjust (dishonest)	Just (honest)	State (inspection mechanisms)
Haraiphon	Unjust (dishonest)	(−1, −1)	(−1, 8)	(1, 0)
	Just (honest)	(8, −1)	(10, 10)	(0, 0)
	State (inspection mechanisms)	(0, 1)	(0, 0)	(1, 1)

Similarly, the pair (−1, 8) means that Heraiphon tries to cheat while Apollonius tries to play justly. Thus, before finalizing their transaction, Apollonius demands that the jewelry of Heraiphon is inspected by an *agoranomos*. The later inspects the jewelry of Heraiphon, he finds that Heraiphon tried to cheat and, thus, he confiscates the jewelry of Heraiphon and also imposes upon him a fine. On their part, Apollonius does not earn 10 because he wished the transaction to be done, which it was not. Thus, although he keeps his pure (not fraudulent) coins, he scores 8 and not 10 because his utility would have been maximized if the transaction had been done, a situation that did not materialize in the end. In this case the state also earns a small profit because it imposes a fine to Hairephon because he tried to cheat Apollonius. By contrast, the state earns nothing from Apollonius since it cannot put a fine on his sound money and pure silver coins. Thus, the pair for the state in this (1, 0).

On the other hand, the pair (8, −1) denotes the inverse case of (−1, 8)., and under the same logic the state's payoff is only (0, 1) since it can only receive a small profit by putting a fine (only) to Apollonius.

5. Conclusions

In this paper we approached the issue of justice in human affairs as an important parameter regarding the establishment of an institutional environment, which is beneficial to international trade. Our analysis links the issue of financial risk through a game theoretical approach regarding the famous philosopher Plato's views on trust between parties and especially while performing international commercial transactions.

Since fair and honest behavior is a key intertemporal axiom for performing efficient commercial transactions, our motivation, among others in this paper, was to trace how deeply this financial behavior goes back in the course of history. We found that elements of such an intertemporal value system for reducing risk in financial transactions dates back to the time of Plato in Athens during the Classical times.

To prove this, we chose to elaborate a famous passage from Plato's Republic by formulating a repeated game of "just" and "not just" while performing financial transactions. Essentially Plato argued that the absence of regulatory mechanisms of justice in a society causes unproductive outcomes. The opposite also applies. When two parties decide to cooperate with each other by playing fairly and this is also backed up by strong institutions, that is, efficient auditing mechanisms by the state so as to apply justice fairly, as this will be to the benefit of both parties.

We formulated a relative conceptual framework through a game theoretical analysis seeing Plato as "game theorist". We argued that, in the beginning, each player chooses the "unjust" strategy, trying to cheat the other and could be seen as a prisoner's dilemma situation, which ends to the worst possible outcome. Realizing this, in a repeated game situation, both parties choose the "just" strategy so achieving the best outcome and transforming the game into a cooperative one that maximizes the outcome for both, that is, it is to their benefit to agree and behave in a "just" way and their cooperation is transformed into a cooperation game.

We have also found that the whole situation changes if we add the government and the inspection mechanisms. In a repeated game and having the state inspection mechanisms (S_1,S_2) ensures the trust outcome of (J_1,J_2) with cooperation achieved in a complete information repeated game. The outcome of trust between the two players (J_1,J_2) is Pareto superior as both players are better off compared to unjust ones (U_1,U_2) but the situation with this strategy is unstable as each player has an incentive to unjust if he considers the other player will play just. In case both players play justly and do not cheat then the outcome (J_1,J_2) will be chosen. This is ensured with the inspection mechanisms that eventually lead to the cooperative outcome.

Our results contribute to the current knowledge regarding the world economy competitive arena by conveying the intertemporal axiom that what is important so as to ensure the establishment of an effective system of commercial transactions at a supranational level is to create an environment of robust and functional financial institutions where trading parties behave "just". This is a necessary, but not a sufficient, condition, though. Such a financial system, in order to be effective requires to be backed by effective auditing mechanisms and, equally important, bonds of trust between trading parties and the state institutions should be forged and be mutually reinforcing. As argued in the introduction, these prerequisites were also among the key reasons of success in later historical cases, such as Venice, the United Provinces, England, Great Britain, and the USA nowadays.

Further research directions could perhaps include focusing on the common protocols of behavior that systems, such as monetary unions, develop so as to achieve the "just" behavior among trading parties at a supranational level. As already argued, some authors consider Athens and her Delian League allies as an ad hoc monetary union. Later monetary unions include, among others, the Achaean League as one of the most advanced federations in the Hellenistic period (Economou 2020) and, in later times, the United Provinces, a federal state which also established a monetary union, the USA, the German Zollverein of 1834, the Latin Monetary Union of 1865, the Austro-Hungarian Monetary Union (AHMU) of 1867, the Scandinavian Monetary Union of 1873 (Ryan and Loughlin 2018) and, of course, the current Eurozone of the European Union (EMU) established in 1992. Further research could, for example, include a comparative analysis of the cases mentioned above with the case of Classical Athens and the Delian League as a coherent whole in terms of a monetary union institutional setting. Such research could also include commercial and customs unions, such as the Hanseatic League of 1356, the European Free Trade Association (EFTA) of 1960, or even the recent Eurasian Economic Union (EAEU) established in 2015, among others. Another possible future research direction would be to compare the scientific impact and the methods we used in this paper with the results of other studies with similar hypotheses that possibly exist. Due to space limitations we promise further research in the future in this direction.

Of course, there are also some limitations regarding the research we performed with this paper. We have no cliometric data available for the case of Classical Athens and, generally, for every ancient economy, thus, in this case we can only rely on the findings of history by using them properly through interpretative tools from disciplines such as financial history, historical (and international) political economy, and the history of economic thought as our paper does.

As a final comment, we hope that the approach we follow in this paper, which combines (i) issues of risk while performing financial transactions, (ii) game theory as an interpretative tool for financial transactions, (iii) financial history, will stimulate future interest in the academic community on related issues.

Author Contributions: Conceptualization: E.M.L.E., N.C.K.; methodology: G.E.H., E.M.L.E., N.C.K.; validation: G.E.H., E.M.L.E.; formal analysis: E.M.L.E., G.E.H., N.C.K.; investigation: E.M.L.E., N.C.K. G.E.H.; writing: E.M.L.E., N.C.K. G.E.H.; original draft preparation: E.M.L.E., N.C.K., G.E.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Acknowledgments: The authors are grateful to the Editor and the referees for their comments and suggestions that led to the further improvement of the overall argumentation of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

Primary Sources

Aristophanes. 1994. *Frogs, Engl. Trnsl. M. Dillon*. New York: Routledge.

Aristotle. 1944. *Politics, Engl. Trnsl. H. Rackham*. Cambridge: Harvard University Press; London: William Heinemann Ltd., Perseus Digital Library.

Aristotle. 1952. *Athenian Constitution, Engl. Trnsl. H. Rackham*. Cambridge: Harvard University Press; London: William Heinemann Ltd., Perseus Digital Library.

Demosthenes. 1939. *Speeches: XXIV, Against Timocrates. Engl. Trnsl. A. T. Murray*. Cambridge: Harvard University Press; London: William Heinemann Ltd., Perseus Digital Library.
Isocrates. 1980. *Speeches: Panegyricus, 4, Engl. Trnsl. G. Norlin*. Cambridge: Harvard University Press; London: William Heinemann, Ltd., Perseus Digital Library.
Lysias. 1930. *Speeches, XII: Against Eratosthenes, Engl. Trnsl. W. R. M. Lamb*. Cambridge: Harvard University Press; London: William Heinemann Ltd., Perseus Digital Library.
Plato. 1969. *Plato in Twelve Volumes, Vols. 5 & 6. Republic, Engl. Trnsl. P. Shorey*. Cambridge: Harvard University Press; London: William Heinemann Ltd., Perseus Digital Library.
Pseudo-Xenophon. 1984. *Constitution of the Athenians, Engl. Trnsl. E. C. Marchant*. Cambridge: Harvard University Press; London: William Heinemann, Ltd.
Xenophon. 1914. *Cyropaedia, Engl. Trnsl. W. Miller*. Cambridge: Harvard University Press; London: William Heinemann, Ltd., Perseus Digital Library.
Xenophon. 1923. *Memorabilia*. Cambridge: Harvard University Press; London: William Heinemann, Ltd., Perseus (Xenophon 1914) Digital Library.
Xenophon. 1925. *Minor Works: Ways and Means*. Cambridge: Harvard University Press; London: William Heinemann, Ltd., Perseus Digital Library.

Secondary Sources

Acemoglu, Daron, and James Robinson. 2013. *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. New York: Crown Business.
Acton, Peter H. 2014. *Poiesis: Manufacturing in Classical Athens*. Oxford: Oxford University Press.
Adler, Paul S. 2001. Market, Hierarchy and trust: The Knowledge Economy and the Future of Capitalism. *Organisation Science* 12: 215–34. [CrossRef]
Akerlof, George A. 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84: 488–500. [CrossRef]
Amemiya, Takeshi. 2004. The Economic Ideas of Classical Athens. *The Kyoto Economic Review* 73: 57–74.
Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: W. W. Norton & Company.
Baloglou, Christos P. 2012. The Tradition of Economic Thought in the Mediterranean World from the Ancient Classical Times through the Hellenistic Times Until the Byzantine Times and the Arab Islamic World. In *Handbook of History of Economic Thought. Insights of the Founders of Modern Economics*. Edited by J. Georg Backhaus. New York, Heidelberg and London: Springer, pp. 7–92.
Bingham, Tom. 2011. *The Rule of Law*. London: Penguin UK.
Bitros, George C., and Anastassios Karayiannis. 2008. Values and Institutions as Determinants of Entrepreneurship in Ancient Athens. *Journal of Institutional Economics* 4: 205–30. [CrossRef]
Bitros, George C., Emmanouil M. L. Economou, and Nicholas C. Kyriazis. 2020. *Democracy and Money: Lessons for Today from Athens in Classical Times*. London and New York: Routledge.
Bresson, Alain. 2016a. *The Making of the Ancient Greek Economy: Institutions, Markets, and Growth in the City-States*. Princeton: Princeton University Press.
Bresson, Alain. 2016b. Aristotle and Foreign Trade. In *The Ancient Greek Economy: Markets, Households and City States*. Edited by Edward M. Harris, David M. Lewis and Mark Woolmer. New York: Cambridge University Press, pp. 41–65.
Coase, Ronald H. 1960. The Problem of Social Cost. *Journal of Law and Economics* 3: 1–44. [CrossRef]
Cohen, Edward E. 1992. *Athenian Economy and Society: A Banking Perspective*. Princeton: Princeton University Press.
Economou, Emmanouil M. L. 2020. *The Achaean Federation in Ancient Greece. History, Political and Economic Organization, Warfare and Strategy*. Cham: Springer.
Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2017. The emergence and the evolution of property rights in ancient Greece. *Journal of Institutional Economics* 13: 53–77. [CrossRef]
Economou, Emmanouil M. L., and Nicholas C. Kyriazis. 2019. *Democracy and Economy: An Inseparable Relationship Since Ancient Times to Today*. Newcastle Upon Tyne: Cambridge Scholars Publishing.
Economou, Emmanouil M. L., Nicholas C. Kyriazis, and Nikolaos A. Kyriazis. 2021. Money Decentralization under Direct Democracy Procedures. The Case of Classical Athens. *Journal of Risk and Financial Management* 14: 30. [CrossRef]
Elliot, Colin P. 2018. The Role of Money in the Economies of Ancient Greece and Rome. In *Handbook of the History of Money and Currency*. Edited by Stephano Battilossi, Youssef Cassis and Kazuhiko Yago. Singapore: Springer Nature Singapore Private Limited, pp. 1–20.
Engen, Darel T. 2005. Ancient Greenbacks, Athenian Owls, the Law of Nicophon, and the Greek Economy. *Historia* 54: 359–81.
Ferguson, Niall. 2008. *The Ascent of Money: A Financial History of the World*. New York: The Penguin Press.
Figueira, T. J. 1998. *The Power of Money: Coinage and Politics in the Athenian Empire*. Philadelphia: University of Pennsylvania Press.
Finley, Moses I. 1973. *The Ancient Economy*. Berkeley: University of California Press.
Finley, Moses I. 1983. *Economy and Society in Ancient Greece*. New York: Penguin Non Classics.
Fudenberg, Drew, and Jean Tirole. 1993. *Game Theory*. London: The MIT Press.
Fukuyama, Francis. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. New York: Free Press.
Gibbons, Robert. 1992. *A Primer in Game Theory*. New York: Harvester Wheatsheaf.

Gowder, Paul. 2016. *The Rule of Law in the Real World*. Cambridge: Cambridge University Press.
Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge and New York: Cambridge University Press.
Halkos, George E., and Nicholas C. Kyriazis. 2010. The Athenian Economy in the Age of Demosthenes. *European Journal of Law and Economics* 29: 255–77. [CrossRef]
Halkos, George E., and Argyro Zisiadou. 2020. Is Investors' Psychology Affected Due to a Potential Unexpected Environmental Disaster? *Journal of Risk and Financial Management* 13: 151. [CrossRef]
Harris, Edward M., and David M. Lewis. 2016. Introduction: Markets in Classical and Hellenistic Greece. In *The Ancient Greek Economy: Markets, Households and City States*. New York: Cambridge University Press, pp. 1–40.
Hodgson, Geoffrey M. 2015. *Conceptualizing Capitalism: Institutions, Evolution, Future*. Chicago: University of Chicago Press.
Kramer, Roderick M. 1999. Trust and Distrust in Organisations: Emerging Perspectives, Enduring Questions. *Annual Review of Psychology* 50: 569–98. [CrossRef]
O'Halloran, Barry. 2018. *The Political Economy of Classical Athens. A Naval Perspective*. Leiden and Boston: Brill.
Ober, Josiah. 2008. *Democracy and Knowledge. Innovation and Learning in Classical Athens*. Princeton University Press: Princeton.
Puga, Diego, and Daniel Trefler. 2014. International Trade and Institutional Change: Medieval Venice's Response to Globalization. *The Quarterly Journal of Economics* 129: 753–821. [CrossRef]
Ryan, John, and John Loughlin. 2018. Lessons from Historical Monetary Unions—Is the European Monetary Union Making the Same Mistakes? *International Economics and Economic Policy* 15: 709–25. [CrossRef]
Stewart, Edmund, Harris Edward, and Lewis David. 2020. Specialization and the Division of Labour in the Ancient City. In *Skilled Labour and Professionalism in Ancient Greece and Rome*. Edited by Edmund Stewart, Edward Harris and David Lewis. Cambridge: Cambridge University Press, pp. 127–202.
Tridimas, George. 2019. The Debate about the Nature of the Ancient Greek Economy: 19th Century German Scholarship and Modern Economic Theory. In *German-Greek Yearbook of Political Economy*. Edited by Manfred J. Holler and George Tridimas. München: Verlag Holler, vol. 2, pp. 4–25.
van Alfen, Peter G. 2011. Hatching Owls: Athenian Public Finance and the Regulation of Coin Production. In *Quantifying Monetary Supplies in Greco-Roman Times*. Edited by François de Callatay. Bari: Edipuglia, pp. 127–49.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Journal of Risk and Financial Management Editorial Office
E-mail: jrfm@mdpi.com
www.mdpi.com/journal/jrfm

Academic Open
Access Publishing

www.mdpi.com

ISBN 978-3-0365-8368-6