

Article

Factors Affecting Cash Demand in South Africa [†]

Randheer Ramsoomer ^{*}, Hermann Azemtsa Donfack  and Adri Drotskie

Kingsway Campus, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa;
hadonfack@uj.ac.za (H.A.D.); adrid@uj.ac.za (A.D.)

^{*} Correspondence: randheerr79@gmail.com

[†] This paper originates from a PhD study by Randheer Ramsoomer (220130181) at the University of Johannesburg.

Abstract

Physical cash remains a critical component of payment systems worldwide due to its accessibility, liquidity, anonymity, and role in promoting financial inclusion and resilience during systemic shocks. Despite rapid digitalisation, cash retains its relevance in economies such as South Africa, where it supports both formal and informal market activity. Understanding the determinants of cash demand is therefore essential for managing operational and policy risks faced by central banks. This study examines the factors influencing cash demand in South Africa and their implications for the South African Reserve Bank's (SARB) currency management and risk mitigation strategies. Using a Vector Error Correction Model (VECM), Impulse Response Functions (IRFs), and advanced forecasting techniques, the analysis integrates key macroeconomic and technological variables, including GDP, interest rates, mobile penetration, ATMs, EFTs, and tax ratios. The study also benchmarks its results against international empirical evidence to contextualise South Africa's evolving cash dynamics. The results highlight the significant impact of payment technology, especially mobile banking, on reducing cash usage. While ATMs and bank branches still support cash demand to some extent, the growing preference for digital transactions, notably through EFTs and mobile platforms, is reshaping financial behaviour. Macroeconomic variables like GDP and interest rates continue to influence demand, but their role is increasingly mediated by digital adoption. The forecasting analysis reveals that neural network models, particularly NNETAR, outperform traditional linear models (like VECM and Exponential Smoothing), especially over longer horizons. These models better capture non-linearities and evolve structural dynamics in cash usage. These insights hold material implications for SARB's operational and financial risk frameworks. As cash demand becomes more unpredictable and technology-driven, adaptive forecasting and policy strategies are required to ensure efficient currency management and financial system stability.

Keywords: cash demand; forecasting; monetary risk policy; South Africa; payments

JEL Classification: E41; C32; E52



Academic Editor: Thanasis Stengos

Received: 6 November 2025

Revised: 25 January 2026

Accepted: 3 February 2026

Published: 4 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Background

Cash remains a crucial component of South Africa's financial ecosystem. Despite the proliferation of digital and mobile payment channels, the resilience of cash is evident in the steady rise in currency in circulation (CIC) year-on-year. According to the South African Reserve Bank (SARB) statistics, the value of CIC has grown on average by 7–9% annually over the past two decades. For many unbanked and underbanked communities, cash is not only a means of

transaction but also a store of value and a symbol of trust and accessibility. This dual economy, combining sophisticated digital infrastructure with persistent informal sector reliance on cash, poses unique challenges for monetary policy and currency supply management.

From a risk and financial management perspective, understanding cash demand is critical for central banks and financial institutions. Inaccurate cash demand forecasts expose the financial system to operational risk, liquidity mismatches, and increased currency handling costs. Excess supply raises storage and security risks, while undersupply threatens payment continuity and public confidence. Periods of economic stress, such as power outages, labour disruptions, or systemic shocks, further amplify the importance of cash as a risk-mitigating instrument. In this context, cash functions not only as a medium of exchange but also as a resilience mechanism within the broader financial system.

South Africa's informal sector, which accounts for over 25% of employment, often operates outside formal banking channels (Statistics South Africa, 2025). Limited digital literacy, concerns over transaction fees, and trust issues further sustain cash usage. Moreover, cash plays an essential role as a contingency tool during crises such as natural disasters, political instability, or widespread power outages that may disrupt digital payment systems.

1.2. Problem Statement

While developed economies report a decline in CIC relative to GDP, South Africa defies this trend. This paradox indicates gaps in existing models, which often fail to account for structural, technological, and behavioural drivers unique to emerging markets. The SARB, responsible for ensuring an adequate and cost-effective supply of currency, needs robust and context-specific forecasting frameworks to mitigate the risks and costs associated with cash supply shortfalls or excesses.

Thus, the central research question is the following: What are the main determinants of cash demand in South Africa, and how can their dynamic interactions be modelled to enhance forecasting accuracy?

1.3. Research Objectives

- To identify the macroeconomic, technological, and institutional factors influencing cash demand.
- To quantify both the short-run and long-run relationships between cash demand and its key determinants in South Africa using advanced econometric models.
- To compare traditional econometric models with modern machine learning approaches for forecasting CIC.
- To provide actionable insights and recommendations for the SARB to optimise cash supply planning and risk policy development.

1.4. Paper Structure

This paper is organised as follows: Section 1 begins with a background to cash demand in South Africa. Section 2 follows with a review of both the theoretical and empirical literature relevant to the topic. The research methodology and modelling strategy are presented in Section 3, while Section 4 reports the empirical findings. In Section 5, the managerial and risk policy implications are discussed. Finally, Section 6 concludes by summarising the study's key contributions, noting its limitations, and offering suggestions for future research.

2. Literature Review

2.1. Theoretical Foundations

The Classical Quantity Theory of Money, associated with economists such as Irving Fisher and later refined by Milton Friedman, posits a direct and proportional relationship

between the quantity of money in an economy and the general price level. Fisher's equation of exchange formalised this relationship, while Friedman's contemporary interpretation treated money as an asset similar to other goods, focusing on the quantity individuals wish to hold rather than the reasons for holding it. In this view, money provides a flow of services as a readily available purchasing power but is subject to diminishing marginal returns (Endale, 2023).

Building on these foundations, Keynes advanced the theory of money demand by identifying three distinct motives for holding money: transactions, precautionary, and speculative. This approach highlighted the role of income and interest rates in shaping demand, with the speculative motive particularly sensitive to interest rate changes. While some scholars have questioned the relative importance of the transactional and precautionary motives, Keynes's framework broadened the understanding of money demand beyond price levels, paving the way for later models such as the Baumol–Tobin approach, which formalised the trade-off between liquidity and interest earnings (Keynes, 1936).

The seminal Baumol–Tobin model (Baumol, 1952; Tobin, 1956) lays the foundation for understanding transactional demand for money. The model conceptualises the trade-off between the opportunity cost of holding cash (interest foregone) and the transaction costs of converting non-cash assets into cash. According to the model, higher income levels and transaction volumes increased cash demand, while higher interest rates reduce it. However, the model assumes a relatively stable, formalised economy with minimal technological disruption. In practice, various factors, including financial innovation, socio-economic conditions, and institutional trust, significantly influence cash holding behaviour.

2.2. Empirical Insights

Empirical studies have extended the classical framework by integrating country-specific factors. For example, Rauf and Khan (2012) analysed money demand in Pakistan, finding GDP and inflation to be positive drivers, while interest rates exert a negative effect. Masciandaro (2018b) explored the role of shadow economies in sustaining cash demand despite digital trends. In Sub-Saharan Africa, studies by Mlambo and Msosa (2020) reveal mixed effects of ATMs and mobile money. In some contexts, ATMs facilitate cash withdrawals, increasing CIC, while mobile payments reduce the need for physical currency. Studies on Kenya's M-Pesa revolution highlight how mobile payments can transform cash demand (Jack & Suri, 2011). However, differences in financial inclusion levels and infrastructure maturity mean these findings are not directly transferable to South Africa.

International research on cash and money demand has evolved significantly in response to financial innovation and digitalisation. Studies in advanced economies generally document a declining transactional role for cash, driven by widespread adoption of electronic payments, low inflation, and high financial inclusion. However, these studies also emphasise the continued importance of cash as a store of value and as a precautionary asset during periods of economic uncertainty (Bordo, 2022). In contrast, evidence from emerging and developing economies highlights the persistence of cash usage despite rapid technological change. Factors such as limited access to banking services, informality, trust considerations, and infrastructural constraints often weaken the substitutability between cash and digital payments. As a result, financial innovation may coexist with, rather than fully displace, cash demand (Shirai & Sugandi, 2019).

Empirical research on money and cash demand in South Africa has historically been situated within the broader monetary economics and inflation-targeting literature. Early South African studies primarily examined the stability of money demand functions and their relevance for monetary policy transmission, often focusing on aggregate monetary

aggregates, output, prices, and interest rates within conventional econometric frameworks aligned with the South African Reserve Bank's policy mandate (SARB, 2022a, 2022b).

Subsequent work increasingly acknowledged the structural characteristics of the South African economy, including high income inequality, persistent informality, and uneven access to financial services. Studies examining financial innovation and payment-system development highlighted the growing role of ATMs, bank branches, and electronic payment instruments, while also noting that these developments coexist with sustained demand for physical cash, particularly among lower-income households and informal sector participants (Fish & Whymark, 2015; Shirai & Sugandi, 2019).

More recent South African and regional evidence suggests that financial innovation does not uniformly displace cash usage. Mlambo and Msosa (2020), examining selected African economies including South Africa, find that mobile subscriptions and financial innovation proxies are negatively related to cash demand, while traditional cash access points such as ATMs retain a complementary role. These findings align with broader African evidence showing that digital payment adoption may reduce transactional cash demand without eliminating the precautionary and store-of-value functions of cash (Masciandaro, 2018a; Bordo, 2022).

Despite these advances, the domestic literature remains fragmented. Much of the South African research either concentrates on macroeconomic determinants of money demand or examines payment innovation in isolation, without jointly modelling institutional, technological, fiscal, and behavioural drivers of cash demand. Moreover, forecasting-oriented analyses of cash demand, particularly those incorporating non-linear dynamics or machine learning techniques, remain limited in the South African context (Endale, 2023).

Various countries have recently reported that traditional money demand connections and determinants may have changed due to the impact of financial innovations (Dunne & Kasekende, 2016). Cash demand functions that do not incorporate financial technology are incorrectly specified and are commonly referred to as missing money (Aliha et al., 2018). Cash demand levels remain high for a variety of reasons. One key reason is that cash provides individuals with control over their liquidity. Other reasons include the anonymity and security of cash as compared to other payment systems which may lead to too much information being shared online, or cybercrime. It is also widely accepted, especially in developing countries, where payment infrastructures may not be readily available (Tagat & Trivedi, 2020). Cash remains a resilient payment mechanism. The understanding of cash demand as a result remains critical for economies to function. Determinants which impact cash demand have evolved over time, especially given the rise in non-traditional payment mechanisms, e.g., digital payments.

South African studies highlight the dual role of technology; while EFTs and mobile banking reduce reliance on cash, ATM density and bank branch access continue to play a critical role, particularly in peri-urban and rural areas. The informal sector, estimated at 30% of GDP, remains highly cash-intensive due to limited trust in digital alternatives and practical barriers such as unreliable electricity supply. Moreover, the SARB's Cash Management Strategy (SARB, 2021) underlines the importance of forecasting CIC accurately to avoid over- or under-supplying cash, which has cost and security implications.

Little is known about the empirical validity of determinants of cash demand in South Africa. There have been previous studies on the cash demand of other countries, using country panel data or time series data. While international studies increasingly adopt extended money demand frameworks that integrate digitalisation, payment infrastructure, and institutional factors, South African applications of such approaches remain scarce. Existing studies provide valuable insights into specific aspects of money demand or payment behaviour but do not offer a unified framework capable of explaining both the persistence of cash and its evolving role amid rapid digital transformation.

In particular, there is limited South African evidence that simultaneously accounts for macroeconomic variables, payment technology adoption, physical cash infrastructure, and fiscal influences within a long-run equilibrium and short-run adjustment framework. This gap constrains the ability of policymakers and currency managers to assess structural shifts in cash demand and to develop robust forecasting strategies tailored to local conditions (SARB, 2021; Tan et al., 2020).

This study addresses these gaps by extending established money demand frameworks to the South African context through the integration of institutional, technological, and fiscal determinants of cash demand. Building on prior South African research on monetary aggregates, financial innovation, and payment systems, the analysis explicitly incorporates payment infrastructure, digital transaction channels, and taxation effects, while accounting for the country's dual economic structure and persistent reliance on cash.

By combining long-run econometric modelling with forecasting analysis, the study contributes locally grounded evidence relevant to the South African Reserve Bank's monetary policy formulation, currency management, and operational planning.

However, model specification and methodologies differ from study to study based on countries, economies, economic cycles, data availability and research questions analysed (Rauf & Khan, 2012). For instance, Mlambo and Msosa (2020) found, for selected African states, that the variables of GDP, mobile subscriptions and ATMs, categorised as financial innovation, had a negative relationship with cash demand. Inflation and interest had a positive relationship with cash demand. Conversely, Rauf and Khan (2012) found, for Pakistan, that bank branches and ATMs have a positive impact on cash demand, whilst interest/deposit rates and low financial inclusion have a negative impact on cash demand.

Despite a growing international literature, three gaps remain. First, much of the empirical evidence is concentrated in advanced economies, limiting its applicability to countries with dual economic structures and high informality. Second, existing studies often examine macroeconomic determinants or payment innovation in isolation, rather than within an integrated framework. Third, relatively few studies combine long-run econometric analysis with forecasting-oriented evaluation tailored to operational and risk management needs.

This study advances the literature by addressing these gaps through a South Africa-specific empirical framework that integrates macroeconomic, institutional, and technological determinants of cash demand. By combining Vector Error Correction Modelling with forecasting analysis, the study provides both theoretical insight and practical relevance, extending existing knowledge beyond descriptive or replication-based approaches.

3. Forecasting Approaches

Econometric models such as Autoregressive Integrated Moving Average (ARIMA) and Vector Error Correction Model (VECM) have long been used to forecast money demand, especially when variables are cointegrated. However, structural breaks and non-linear relationships challenge the assumptions underlying these models. Recent studies have explored machine learning techniques, including Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), which can better accommodate complex, non-linear data patterns (Petropoulos et al., 2019).

Combining traditional econometric and ML approaches enhances forecasting robustness. For example, Christoffersen and Diebold (2006) argue that hybrid frameworks yield more reliable results in volatile environments.

4. Research Methodology

4.1. Research Approach

This study adopts a positivist, quantitative approach, appropriate for modelling relationships among macroeconomic and technological variables using time series data. By combining classical econometric techniques with modern machine learning, the research addresses both the linear and non-linear dynamics of cash demand.

The empirical analysis follows a structured, stepwise approach consistent with standard econometric practice in cash and money demand studies. First, the time series properties of the data are examined through unit root testing. Second, long-run relationships are assessed using Johansen cointegration techniques. Third, a Vector Error Correction Model (VECM) is estimated to capture both long-run equilibrium relationships and short-run dynamics. Fourth, Granger causality is evaluated within the VECM framework. Finally, alternative forecasting models, including machine learning approaches, are assessed and compared based on out-of-sample performance. This sequential structure ensures transparency and facilitates a clear mapping between the methodological approach and the empirical results.

This study also adopts the Baumol–Tobin cash demand model as its primary theoretical framework, as it explicitly captures the optimisation trade-off between the transaction costs of obtaining cash and the opportunity cost of holding it, making it more suitable than the earlier Neo-Classical and Keynesian theories for analysing South Africa’s mixed formal informal economy. The Baumol–Tobin model provides a foundational framework for understanding transaction-based money demand by emphasising the trade-off between transaction costs and the opportunity cost of holding liquid balances. While the original formulation addresses money demand broadly, its core intuition has informed subsequent extensions that examine specific components of money holdings, including cash, particularly in contexts where alternative payment instruments are imperfect substitutes. In this study, the Baumol–Tobin framework is therefore not applied directly to model cash demand. Instead, it serves as a conceptual reference point that motivates the inclusion of, and opportunity costs and payment-system variables within, an extended empirical cash demand framework adapted to South African institutional conditions.

To operationalise this framework, a Vector Error Correction Model (VECM) is employed, which is appropriate for cointegrated time series and enables the analysis of both short-run adjustments and long-run equilibrium relationships. The model incorporates traditional determinants of cash demand (GDP, inflation, interest rates) alongside novel variables, cellular subscriptions, electronic fund transfers, and the tax-to-GDP ratio, identified using LASSO regression to ensure empirical relevance and minimise multicollinearity. This combined theory–econometric approach provides a rigorous basis for understanding and forecasting cash demand in South Africa.

4.2. Data Collection and Description

Monthly data spanning January 2000 to December 2019 was sourced from the South African Reserve Bank (SARB), Statistics South Africa, and the World Bank. The dependent variable, currency in circulation, is available at a monthly frequency, while several macroeconomic variables (including GDP and selected fiscal indicators) are only reported quarterly. To ensure consistency within a unified monthly modelling framework, quarterly series were converted to monthly frequency using temporal interpolation techniques. Specifically, a proportional interpolation approach was applied, ensuring that the aggregated monthly values align with observed quarterly totals.

This harmonisation strategy is commonly adopted in applied monetary and cash demand studies where higher-frequency modelling is required to capture operational and seasonal dynamics, particularly in currency issuance and cash logistics. While interpola-

tion may smooth short-term fluctuations, it enables the integration of key macroeconomic variables without altering their long-run informational content. The Baumol–Tobin model represented seminal work in our understanding of cash demand. Moreover, while the Keynesian and Neo-Classical theories provide valuable macroeconomic foundations, the Baumol–Tobin approach is better suited for integrating variables such as financial innovation. Rauf and Khan (2012) adopted an extension and modification of the model. The analysis of economic theory and past evidence uses a model for testing determinants affecting cash demand. The basis of the current model uses the Rauf model variables, which include Automated Teller Machines (ATMs), bank branches, number of bank accounts (as a proxy of financial innovation), and interest rates. (Financial innovation was used as an alternative to bank accounts, given the rapid rise in electronic payments).

The dependent variable is currency in circulation (CIC), used as a proxy for cash demand. The main independent variables include:

GDP (Gross Domestic Product): proxy for transactional motives.

Inflation Rate: represents store-of-value motives.

Interest Rate (Repo Rate): reflects the opportunity cost of holding cash.

ATM Ratio: number of ATMs per 100,000 adults.

Bank Branch Ratio: branches per 100,000 adults.

Mobile Subscriptions: cellular subscriptions per 100 people.

Electronic Fund Transfers (EFTs): value of EFTs relative to GDP.

4.3. Econometric Framework

As the data sets are time series-based, there are potential econometric pitfalls. The understanding of the stationarity of variables with unit root tests is critical. However, in addition, times series data often shows indicators of autocorrelation and heteroskedasticity. Therefore, an additional preceding step is included before the unit root: tests for autocorrelation and heteroskedasticity are required. For the paper, we use the Durbin–Watson test for autocorrelation hypothesis tests and the Breusch–Pagan test for heteroskedasticity hypothesis tests. Based on the outcome of the results, simple OLS regressions may be inadequate as this model does not capture shocks or cyclic events, autocorrelation or heteroskedasticity, and thus results between two or more variables could be spurious. Cointegration tests identify scenarios where two or more “non-stationary” time series are integrated in a way that they cannot deviate from equilibrium in the long term (they are not a spurious regression). The cointegration approach is used to accommodate such deviations. These tests identify scenarios where two or more non-stationary time series variables are integrated in a way that does not deviate substantially from the equilibrium in the long term.

4.3.1. Unit Root and Cointegration Tests

Stationarity of the variables was tested using the Augmented Dickey–Fuller (ADF) test. Variables found to be $I(1)$ were tested for cointegration using the Johansen cointegration test, confirming the presence of a stable long-run relationship among variables, justifying the VECM approach.

4.3.2. Vector Error Correction Model (VECM)

The VECM captures both the short-run dynamics and long-run equilibrium. The error correction term (ECT) measures the speed at which deviations from long-run equilibrium are corrected. The lag structure of the Vector Error Correction Model (VECM) was determined using standard information criteria, including the Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC). Given the persistence, seasonality, and institutional inertia inherent in cash demand dynamics, a relatively rich lag structure was retained to adequately capture delayed adjustment effects across economic, fiscal, and

payment-system variables. It is acknowledged that higher-dimensional lag specifications may increase the risk of parameter instability in short-run dynamics. However, the selected specification prioritises capturing the full adjustment process typical of monetary aggregates rather than enforcing strict parsimony at the expense of dynamic completeness.

4.3.3. Impulse Response Functions (IRFs)

IRF analysis tracks the effect of a one-standard-deviation shock in each independent variable on cash demand over a 12-month horizon. This offers intuitive insights into how shocks propagate.

4.3.4. Forecast Error Variance Decomposition (FEVD)

FEVD quantifies the proportion of the forecast error variance for CIC attributable to each independent variable, helping understand the relative importance of shocks over time.

4.4. Machine Learning Forecasting

Forecasting is a crucial tool for reducing decision risk by providing insights into uncertain outcomes. After collecting data for a time series, analysts select an appropriate model. Differentiating between trends and random errors in a time series is the basis of time series forecasting. The objective is to isolate the underlying pattern from the noise by recognising patterns, long-term changes, and seasonality arising from variables such as fluctuations in consumption and demand. There are several time series forecasting methods at one's disposal (Atoyebi et al., 2023).

In order to address the study objectives, the VECM forecast technique will also be compared against machine learning techniques. Machine learning techniques are a modern approach to forecasting, utilising large datasets and complex patterns. These techniques include neural networks, particularly Long Short-Term Memory (LSTM) networks, and ARIMA predictions. Neural networks are also non-linear. For many years, linear statistics have been used for forecasting. The Box–Jenkins or ARIMA methods, two conventional techniques for time series prediction, make the assumption that the time series being studied are produced by linear processes. One advantage of linear models is that they are simple to describe and use, and they can be thoroughly examined and understood. However, in the event that the underlying process is non-linear, they could be completely unsuitable. It is irrational to presume that a linear process is responsible for a specific realisation of a given time series (Zhang et al., 1998).

Model training and validation are crucial to tune hyperparameters and avoid overfitting. Forecast comparison is used to evaluate performance and accuracy by comparing machine learning models with traditional econometric models. These techniques help capture non-linear relationships in data and improve predictions. Ensuring projections are correct through out-of-sample testing, which offers a thorough evaluation of the model's performance. An important way to determine if a model can be applied to fresh data is to measure its prediction accuracy on data that was not utilised in the estimate process, or out-of-sample forecast accuracy. To do this, divide the data into two sets: a test set for assessment and a training set for model estimate. Using the data from the training set, the VECM is first calculated. Then, using R's predict function to define the prediction horizon, forecasts are produced for the test set period. In order to evaluate forecast precision, accuracy metrics like Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Theil are used on the out-of-sample data. These metrics provide both absolute and relative measurements of accuracy.

The MAPE is another measure of prediction accuracy of forecasting, and it is computed using the following equation:

$$\text{MAPE} = (1/\text{sample size}) \times \sum [(|\text{actual} - \text{forecast}|) / |\text{actual}|] \times 100 \quad (1)$$

Mean Absolute Percentage Error (MAPE) is a metric that defines the accuracy of a forecasting method. It represents the average of the absolute percentage errors of each entry in a dataset to calculate how accurate the forecasted quantities were in comparison with the actual quantities. The dataset used in the R code completely excluded all data in the forecast period. The MAE is another measure of prediction accuracy of forecasting, and it is computed using the following equation:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - x| \quad (2)$$

where n is the number of errors, Σ = summation symbol and $|x_i - x|$ = the absolute errors.

To complement the VECM, an NNETAR (Neural Network Autoregressive) model was developed using a 70/30 train–test split. This approach captures non-linearities and structural breaks often missed by linear models. Forecast accuracy was assessed using:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)
- MAPE (Mean Absolute Percentage Error)

4.5. Hypotheses

The following hypotheses were formulated for the study:

H0. *No significant relationship exists between the selected variables and cash demand.*

H1. *There is a significant relationship between the selected variables and cash demand.*

NULL HYPOTHESIS: H_0

$$H_0 : \beta_0, \beta_{ATM}, \beta_{BRANCH}, \beta_{EFUND}, \beta_{INTREAL} = 0 \text{ (hypothesis 1.1)}$$

ALTERNATIVE HYPOTHESIS: H_1

$$H_a : \beta_0, \beta_{ATM}, \beta_{BRANCH}, \beta_{EFUND}, \beta_{INTREAL} \neq 0 \text{ (hypothesis 1.2)}$$

4.6. Ethics

All data used is secondary and publicly available. No primary data collection involving human subjects was required; thus, ethical approval was not applicable.

5. Results and Findings

This section presents the empirical results and findings of the study. The presentation of results follows the same sequential order as outlined in the Methodology section. Unit root and cointegration results are reported first, followed by long-run and short-run VECM estimates, causality analysis, and finally forecasting performance. This alignment enhances clarity and allows the reader to trace each empirical outcome directly to its corresponding methodological step.

For each independent variable, the expected impact on cash demand was specified based on the theoretical framework and literature review. These expectations are subsequently evaluated through formal statistical hypothesis testing. The analysis draws on data for both the RAUF model and the South African model, sourced from reputable institutions including the World Bank, Statistics South Africa, and the South African Reserve Bank.

5.1. Descriptive Statistics

Initial descriptive analysis indicates that CIC has grown steadily in absolute terms, mirroring GDP growth but with variations linked to inflationary cycles and shifts in payment technologies.

5.2. Estimation of Results

5.2.1. Relevance of Variables

As mentioned, simple OLS-type regressions that do not capture shocks and cyclic events are not considered. An extension of the OLS regression is LASSO, which performs both variable selection and regularisation through a shrinkage factor. LASSO is an established method to identify and highlight potential independent variables that may have an influence on the dependent variable (cash in circulation). It informs us of independent variables that seem to warrant inclusion in multivariate time series models for further investigation. Based on the results, variables cellY, cpiM, elecFundValM and gdpQ demonstrate LASSO significance. This provides initial validation of the inclusion of the new variables into the extended model (refer to Appendix A).

5.2.2. Time Series Hypothesis Tests

A preliminary investigation of the selected variables justified performing hypothesis tests to determine the identification of appropriate econometric multivariate time series models to explain potential interaction between the selected variables.

5.2.3. The Unit Root Tests

As mentioned above and given the data set, the Augmented Dickey–Fuller Test (ADF) was performed to determine the stationarity of the variables. The series may be integrated of order zero, one that is I (0), one that is I (1) or a mixture. The results are shown in Table 1, which indicate that all the variables are non-stationary at level, except for interest rates and tax-to-income ratio. Given the results, the next step would require a test for cointegration.

Table 1. Data description and acronyms (source: authors).

Variable Type	Description	R Code	Hypothesis Code
Dependent (Y)	South African Reserve Bank liabilities: Notes and coin in circulation	cashCirc	CIC_t
Independent (X)	Automated teller machines (ATMs) (per 100,000 adults)	atmRatioY	ATM_t
Independent (X)	Commercial bank branches (per 100,000 adults)	branchRatioY	$BRANCH_t$
Independent (X)	Mobile cellular subscriptions	cellY	$CELL_t$
Independent (X)	Consumer Price Index (December 2016 = 100)	cpiM	CPI_t
Independent (X)	Value of electronic fund transfers processed during the period	elecFundValM	$EFUND_t$
Independent (X)	Gross domestic product at market prices	gdpQ	GDP_t
Independent (X)	Real interest rate (%)	intRealY	$INTREAL_t$
Independent (X)	Tax payable by persons and individuals as percentage of total revenue	taxRatioQ	TAX_t

5.2.4. Testing for Autocorrelation

Durbin–Watson hypothesis testing to determine the presence of autocorrelation was performed. The presence of significant autocorrelation (at the 5% level of significance) was determined across all variable data ranges (refer to Appendix B).

5.2.5. Testing for Heteroskedasticity

Breusch–Pagan hypothesis testing to determine the presence of heteroskedasticity was performed. The presence of significant heteroskedasticity (at the 5% level of significance) was determined across variables excluding bank branches (branchRatioY), GDP (gdpQ) and taxRatio (taxRatioQ). Given the results, the next step would require a test for cointegration (refer to Appendix C).

5.3. Model Specification

Given the above statistical results and characteristics of multivariate time series data, historically, VECMs are used. Pure linear regression looks to the p -values for key decisions, which may lead to spurious results (Kreinovich et al., 2019). To select the most suitable model, Johansen hypothesis tests are conducted to determine if there is cointegration amongst variables. The VECM would be the most suitable approach if there is at least one cointegrated relationship. The next section follows a logical sequence for model selection.

5.3.1. Lag Selection

Determination of optimal lags is required to perform the Johansen hypothesis test (also required for VECM modelling.) The VARselect function (in R) and Akaike’s Information Criterion (AIC) were used to determine optimal lags. AIC is an estimator of prediction error, and therefore the relative quality of statistical models for the lag order P for VECM is equal to optimal lag minus one. (VECM uses the first difference between periods.) The optimal lag for the 20-year data range was determined as 14 and the K -value is 13.

5.3.2. Testing for Cointegration

For the Johansen test, the optimal lag to be utilised is lag minus one. In the outcomes of the Johansen test for the 20 years indicated, as can be seen below, the “ $r \leq 5$ ” row shows that there may be at least five cointegration relationships. The null hypothesis is therefore rejected as there are cointegrating relationships. In the row “ $r \leq 5$ ”, the test column’s value, 52.91, is larger than 48.28 at a significance of 5%. Based on the results in Table 2 below, we would continue with the VECM’s analysis.

Table 2. Augmented Dickey–Fuller (ADF) level unit root test (source: authors).

Level Variables	ADF Statistics Level (Probability)	First Difference
cashCirc	−1.9643 (0.5909)	−10.141 (<0.01)
atmRatioY	−1.5988 (0.7446)	−4.6943 (<0.01)
branchRatioY	−2.2883 (0.4546)	−5.3553 (<0.01)
cellY	−2.1943 (0.4941)	−5.6043 (<0.01)
cpiM	−2.2619 (0.4657)	−5.6456 (<0.01)
elecFundValM	−3.0849 (0.1196)	−6.3892 (<0.01)
gdpQ	−1.0154 (0.9344)	−8.8034 (<0.01)
intRealY	−4.402 (<0.01)	−6.055 (<0.01)
taxRatioQ	−4.448 (<0.01)	−8.1557 (<0.01)

5.4. Vector Error Correction Model (VECM) Results

A cointegration vector is a mathematical process of combining independent non-stationary time series variables (with cointegrating relationships among the variables) in such a way that their linear combination is stationary. VECM utilises the first difference between consecutive periods in time series.

5.4.1. Results: VECM Long-Run Relationship

Rauf and Khan (2012) normalised cointegrated coefficients and found the long-run relationships between the currency in circulation to deposit ratio, bank branches, deposit rate, and the number of account holders to be negative and significant. However, ATMs had a positive relationship. The Engle–Granger 2OLS (ordinary least squares) estimator VECM results can be used to construct the appropriate VECM equations. The cointegration output for the full 20-year period is represented as follows:

Model 1:

$$ECT_{t-1} = CIC_{t-1} + 0.03050662ATM_{t-1} - 0.2357148BRANCH_{t-1} - 0.8377822EFUND_{t-1} + 0.03985564INTREAL_{t-1} \quad (3)$$

The model demonstrates that ATMs and interest rates have a positive relationship, whilst bank branches and electronic fund transfers have a negative relationship to cash demand in South Africa. The extended equation represented by Model 2 includes four additional variables results, which are represented as follows:

Model 2:

$$ECT_{t-1} = CIC_{t-1} + 0.08317883ATM_{t-1} + 0.05211919BRANCH_{t-1} - 0.1085082CELL_{t-1} - 0.714757CPI_{t-1} - 0.1084856EFUND_{t-1} - 0.4059556GDP_{t-1} + 0.006323035INTREAL_{t-1} + 0.03667117TAX_{t-1} \quad (4)$$

The results indicate that mobile subscriptions, inflation, electronic fund transfers and GDP negatively impact cash demand. The determinants of ATM, bank branches, interest and taxation positively impact cash demand. Based on South African data and variables, the Table 3 represents the consolidated output from the Johansen cointegration test and Table 4 represent the ECT test results.

Table 3. Johansen cointegration procedure (source: authors).

Nine Variables—Full Twenty Years (2000–2019)				
$r \leq$	Test	10% Significance	5% Significance	1% Significance
$r \leq 8$	0.50	6.50	8.18	11.65
$r \leq 7$	11.10	15.66	17.95	23.52
$r \leq 6$	30.32	28.71	31.52	37.22
$r \leq 5$	52.91	45.23	48.28	55.43
$r \leq 4$	89.30	66.49	70.60	78.87
$r \leq 3$	139.41	85.18	90.39	104.20
$r \leq 2$	192.15	118.99	124.25	136.06
$r \leq 1$	255.64	151.38	157.11	168.92
$r = 0$	352.52	186.54	192.84	204.79

Table 4. ECT comparison (source: authors).

Variables	(Model 2) Estimated ECT Coefficient 9 Variables	(Model 1) Estimated ECT Coefficient 5 Variables	RAUF Estimated ECT Coefficient 5 Variables
cashCirc	1	1	1
atmRatioY	0.08317883	0.03050662	1.29889
branchRatioY	0.05211919	−0.2357148	−1.869161
cellY	−0.1085082	-	-
cpm	−0.714757	-	-
elecFundValM	−0.1084856	−0.8377822	−1.255711
gdpQ	−0.4059556	-	-
intRealY	0.006323035	0.03985564	−3.237118
taxRatioQ	0.03667117	-	-

Interestingly, the following results can be observed:

The results reveal several key drivers of cash demand in South Africa. ATMs show a positive relationship with cash demand in both models, consistent with [Rauf and Khan \(2012\)](#) but contrary to [Mlambo and Msosa \(2020\)](#). Bank branches display a negative effect in Model 1, aligning with prior findings, but a positive effect in Model 2. Cellular subscriptions, as a proxy for digitisation, negatively impact cash demand, suggesting reduced reliance on physical currency. Inflation and electronic fund transfers both show negative relationships, indicating that rising prices and digital payment adoption reduce cash usage. Surprisingly, GDP also has a negative effect, possibly reflecting increased financial inclusion and digital transitions. Real interest rates positively affect cash demand, consistent with previous studies, likely due to greater disposable income. Finally, the tax-to-income ratio shows a positive relationship, possibly reflecting a shift toward cash transactions to avoid formal taxation.

5.4.2. Results: VECM Short-Run Relationship

The short-run relationships provide insights into the long-run relationships. The following observations could be made:

Equation:

$$\begin{aligned}
 CIC_{t-1} = & 0.0157 - 0.0749 ECT - 0.5625 \text{ cashCirc}_{t-1} \\
 & - 0.2004 \text{ cashCirc}_{t-2} - 0.0502 \text{ elecFundValM}_{t-2} \\
 & + 0.1047 \text{ atmRatioY}_{t-3} + 0.0500 \text{ elecFundValM}_{t-4} \\
 & - 0.0496 \text{ elecFundValM}_{t-6} - 0.0689 \text{ elecFundValM}_{t-8} \\
 & + 0.0086 \text{ intRealY}_{t-8} - 0.2204 \text{ cashCirc}_{t-10} \\
 & - 0.0682 \text{ branchRatioY}_{t-11} + 0.5319 \text{ cashCirc}_{t-12} \\
 & + 0.4301 \text{ cashCirc}_{t-13}
 \end{aligned} \quad (5)$$

The short-run dynamics of cash demand in South Africa, as captured by the VECM over the 2000–2019 period, reveal several statistically significant lagged effects. A positive short-run coefficient at lag three for ATMs suggests that increased ATM availability leads to higher cash usage within a few months. For bank branches, a significant negative coefficient at lag eleven indicates that the presence of branches may reduce cash demand, albeit with a delayed effect. Electronic fund transfers exhibit mixed short-run impacts: significant negative coefficients at lags two, six, and eight point to a consistent decline in cash demand following increased EFT activity, while a positive coefficient at lag four suggests a temporary increase in cash use. Lastly, real interest rates have a significant positive effect at lag eight, implying that higher real returns may lead to greater cash holding after a few months. These lagged effects underscore the dynamic and delayed nature of short-term cash demand responses to financial

infrastructure and macroeconomic variables. Short-run coefficient estimates in the VECM should be interpreted with caution. In high-dimensional monetary systems, short-run parameters may exhibit instability or counterintuitive signs due to multicollinearity, overlapping adjustment paths, and contemporaneous interactions among regressors. Such coefficients primarily reflect transitional dynamics rather than structural economic relationships.

Accordingly, the analysis places greater emphasis on the estimated long-run cointegrating relationships, which are theoretically grounded and exhibit greater stability. These long-run results provide the primary basis for economic interpretation and policy inference.

5.5. Granger Results

Similar to Rauf and Khan's (2012) analysis, a test of Granger causality was conducted. The data range was divided equally to demonstrate any changes over the period. Table 5 results indicate that Granger causality was identified for the variables of ATMs, interest, electronic fund transfers and Gross Domestic Product. It is important to note the increased impact and significance of electronic fund payments over the period. This possibly indicates the growing significance of alternative payment mechanisms in South Africa, as electronic fund transfers have seen increased prominence over the latter part of the full data range.

Table 5. Granger causality test (source: authors).

Compare the First Ten and Last Ten Years		
Variables	Variables First 10 Years	Variables Last 10 Years
atmRatioY	0.5062	0.3479
branchRatioY	0.4238	0.7485
cellY	0.03534 *	0.01792 *
cpiM	0.007718 **	1.463 ***
elecFundValM	0.517	4.002 ***
gdpQ	0.0001344 ***	3.846 ***
intRealY	0.898	0.3259
taxRatioQ	0.3182	0.01513 *

Significant codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

5.6. Impulse Response Function (IRF), Forecast Error Variance Decomposition (FEVD), Orthogonalised Impulse Response Function (OIRF) Results

The link between ATM usage and cash demand is comprehensively explained by the combined findings from the IRF, OIRF, and FEVD investigations. Both the IRF and the OIRF show a distinct pattern of first unfavourable effects, followed by a partial recovery, and finally a stabilisation of the response. The FEVD data supports this story by demonstrating that, whereas ATM shocks initially contribute little to the variation in cash demand, over time they become increasingly substantial. This increasing impact implies that the trends seen in the IRF and OIRF are not just transient variations but rather a more extensive, long-lasting pattern. The more variance that ATM shocks can account for, the more significant an influence they have on the dynamics of cash demand, especially when such shocks compound over time. This transitional behaviour is reinforced by the FEVD results, which show that, although ATM shocks initially explain a small portion of the variation in cash demand, their influence increases steadily. This trend implies that ATMs, over time, become a more integral part of consumers' interaction with cash, underscoring their evolving role in a changing financial ecosystem (refer to Appendix D for graphical representation).

In contrast, the effect of bank branch density (branchRatioY) on cash demand is more subdued. Both IRF and OIRF analyses reflect limited immediate impact, with only minor negative responses observed in early periods. These results suggest that changes in physical

banking infrastructure do not disrupt cash usage significantly in the short term. Over time, FEVD results confirm this interpretation by attributing only a modest portion of variance in cash demand to branch density. This likely reflects a gradual shift away from reliance on traditional banking channels in favour of digital alternatives.

The relationship between mobile subscriptions (cellY) and cash demand is nuanced. The IRF, OIRF, and FEVD results collectively suggest that shocks to mobile subscriptions (cellY) have a complex but ultimately diminishing impact on cash circulation (cash demand). Initially, changes in mobile subscriptions lead to minor fluctuations in cash demand, as seen in the negative responses in the early periods. However, as the economy adjusts, the impact turns positive, indicating that higher mobile usage may stimulate certain economic activities that impact cash demand.

Despite this short-term increase, the FEVD shows that the overall contribution of cellY to variations in cash demand remains relatively small. Moreover, both the IRF and OIRF indicate that this impact diminishes over time, with the response approaching zero in the later periods. This convergence suggests that the effects of changes in mobile subscriptions on cash demand are transient and that the system tends to return to equilibrium after the initial shock.

The IRF, OIRF, and FEVD results collectively illustrate a dynamic and oscillatory relationship between inflation (cpiM) and cash circulation (cash demand). Initially, inflation shocks have a modest impact on cash demand, as indicated by the minimal response in the early periods. However, as time progresses, the impact of inflation becomes more pronounced, with significant fluctuations in cash demand. The most substantial impact is observed around Periods 5 and 6, where inflation shocks lead to both the largest decrease and subsequent increase in cash demand. This oscillatory behaviour reflects the economy's attempts to adjust to inflationary pressures, with periods of contraction followed by rebounds in cash demand.

The IRF and OIRF indicate that the immediate impact of electronic fund transfer shocks on cash demand is minimal, with no significant response in the first period. However, by the second period, cash demand decreases, reflecting a shift towards electronic payments. The largest negative impact is observed in Period 2, followed by a series of oscillations, with the largest positive response occurring in Period 8. These oscillations suggest that the financial system is adjusting to the increased use of electronic fund transfers, with temporary reductions in cash demand followed by recoveries.

The FEVD results reveal that the influence of electronic fund transfer shocks on cash demand is relatively significant compared to other variables, with a notable contribution to the variance in cash demand from the start. The impact peaks in Periods 3 and 5, after which it begins to stabilise and gradually declines. This pattern suggests that, while electronic fund transfers significantly influence cash demand in the short- to medium-term, their impact diminishes over time as the system adapts.

According to the FEVD results, GDP shocks have a first-rate impact on cash demand, with GDP playing a major role in the variation in cash demand. Period 5 is when this influence peaks, following which it starts to stabilise and progressively decrease. The trend shows that, while short- to medium-term GDP shocks have a major impact on cash demand over time, as the system stabilises and adjusts to the changes in GDP, this impact decreases.

According to the study, changes in real interest rates have a negligible and erratic impact on the demand for cash. These shocks initially cause only slight changes in cash demand; a brief positive effect is followed by a longer-lasting negative reaction. Cash demand, however, starts to stabilise as the system corrects itself, showing a few hints of improvement in later decades.

The overall study indicates that the shocks to the tax ratio have a major and rapid effect on cash demand, as seen by the comparatively strong FEVD contribution in the first period. The first negative reaction in the IRF and OIRF data indicates a drop in cash demand after the shock, most likely as a result of higher taxes lowering disposable income or changing consumer spending patterns. Positive reactions in subsequent periods, however, show that this influence eventually fades and the system starts to heal.

6. Machine Learning Forecasting Performance

A cash in circulation forecast was generated by identifying appropriate variables and applying modelling for the period January 2000 to December 2019. The forecast generated was “out of sample”. The forecast horizon was over a 12-, 24-, and 36-month period. Table 6 represents the results for the 12 month period.

The Vector Error Correction Model (VECM) emerged as the most effective forecasting method amongst the evaluated techniques for the first 12 months of the forecast. The results showed the minimum error deviations across all performance metrics, including a MAPE of 3.65%, an MAE of R6.034 ZAR, an RMSE of R6.843 ZAR, and a Thiel Inequality Coefficient of 0.0206. These results suggest that VECM is highly accurate in predicting cash demand, with minimal deviation from actual figures over the 12-month forecast time horizon.

Table 6. Comparison of forecast methods for 12 months (source: authors).

Comparison of Various Forecasting Methods (over 12 Months)					
Source Data: SARB Cash in Circulation in Rand Millions. (Monthly History from Jan. 2000 to Dec. 2019)					
Out-of-Sample Forecast Period: Monthly from Jan. 2020 to Dec. 2020 (Rand Millions)					
RANK	Forecasting Method	MAPE	MAE	RMSE	THIEL
		Mean Absolute Percentage Error	Mean Absolute Error	Root Mean Square Error	Thiel Inequality Coefficient
1	(VECM) Vector Error Correction Model	3.65%	6034	6843	0.0206
2	(MLE SPLINE) Maximum Likelihood Spline	4.08%	6443	8548	0.0252
3	(GRNN) General Regressive Neural Network	4.26%	7208	8209	0.0249
4	(NNETAR) Feedforward Neural Network	4.33%	7396	8917	0.0272
5	(ARIMA) Autoregressive Moving Average	4.43%	7610	9712	0.0298
6	(BATS) Box–Cox Arma Trend Seasonal	4.64%	7936	10,134	0.0310
7	(RWF) Random Walk Forecast	5.32%	8777	9562	0.0288
8	(SES) Simple Exponential Smoothing	5.44%	9061	9868	0.0298
9	(ETS) State Space Exponential Smoothing	6.24%	10,720	13,250	0.0410
10	(THETA) Exponential Smoothing with Drift	6.47%	11,123	13,778	0.0428

Following VECM, the Maximum Likelihood Spline (MLE SPLINE) and General Regressive Neural Network (GRNN) also performed well, ranking second and third, respectively. MLE SPLINE had a MAPE of 4.08% and slightly higher MAE and RMSE values than VECM, indicating a marginal decrease in accuracy. The GRNN, with a MAPE of 4.26% and slightly higher error values than MLE SPLINE, still demonstrated solid forecasting capability, particularly in more complex patterns in the currency data.

Other neural network-based methods, such as the Feedforward Neural Network (NNETAR), also performed relatively well, ranking fourth with a MAPE of 4.33%. These methods generally outperformed traditional statistical techniques, such as the Autoregressive Moving Average (ARIMA) and the Box–Cox Arma Trend Seasonal (BATS) models, which had higher error metrics, reflecting their limitations in capturing the dynamics of cash demand. Table 7 below represents the results for the 24 month period.

Table 7. Comparison of forecast methods for 24 months (source: authors).

Comparison of Various Forecasting Methods (over 24 Months)					
Source Data: SARB Cash in Circulation in Rand Millions. (Monthly History from Jan. 2000 to Dec. 2019)					
Out-of-Sample Forecast Period: Monthly from Jan. 2020 to Dec. 2021 (Rand Millions)					
	Forecasting Method	MAPE	MAE	RMSE	THIEL
		Mean Absolute Percentage Error	Mean Absolute Error	Root Mean Square Error	Thiel Inequality Coefficient
RANK					
1	(NNETAR) Feedforward Neural Network	2.90%	4924	6624	0.0200
2	(GRNN) General Regressive Neural Network	3.04%	5124	6393	0.0192
3	(ARIMA) Autoregressive Moving Average	3.16%	5384	7443	0.0226
4	(BATS) Box–Cox Arma Trend Seasonal	3.18%	5419	7646	0.0232
5	(RWF) Random Walk Forecast	3.55%	5903	7429	0.0223
6	(SES) Simple Exponential Smoothing	3.88%	6520	7927	0.0239
7	(VECM) Vector Error Correction Model	4.22%	7024	7740	0.0229
8	(THETA) Exponential Smoothing with Drift	5.83%	9889	11,673	0.0360
9	(ETS) State Space Exponential Smoothing	6.10%	10,336	11,838	0.0365
10	(MLE SPLINE) Maximum Likelihood Spline	7.20%	11,840	13,824	0.0400

The forecasting performance for cash demand in South Africa was also evaluated over a longer 12-month time period (as shown in the first set of results), and a 24-month period (as shown in the second set of results). The following results highlight the changes in performance for the top forecasting method as the time horizon is extended.

1. NNETAR (Feedforward Neural Network)

12-month Results: MAPE: 4.33%, MAE: 7396, RMSE: 8917, Thiel: 0.0272

24-month Results: MAPE: 2.90%, MAE: 4924, RMSE: 6624, Thiel: 0.0200

The NNETAR method showed improved performance over the 24 months with lower errors in all metrics, suggesting better adaptability and predictive accuracy over a longer forecast horizon.

Given that neural network models can learn and adjust to intricate patterns in the data, models such as NNETAR and GRNN often perform better across longer-predicting horizons. Non-linear correlations and interactions between variables, which may become increasingly evident and significant over long periods of time, can be captured using these models. Given additional data points, they perform better, which reduces mistakes and improves ranks in the 24-month projection.

The rankings of forecasting methods for the 36 months exhibit notable shifts compared to shorter forecast horizons, reflecting the evolving dynamics of the methods over extended periods. Several factors can explain these differences.

Neural network models like NNETAR and GRNN continue to perform strongly, with NNETAR ranking first in the 36-month forecast. This steady top performance highlights the neural networks' resilience in identifying long-term patterns and non-linear correlations within the data. Neural networks are particularly good at adjusting to complicated patterns, which probably grow more important with time and result in lower errors in measures like RMSE and MAPE.

Methods such as VECM and THETA Exponential Smoothing with Drift show a decline in performance as the forecast period extends. VECM, which performed relatively well in shorter forecasts, ranks lower in the 36 months, likely due to its reliance on linear relationships and its sensitivity to structural changes in the data over time. The increased errors suggest that VECM might struggle with capturing the evolving nature of cash demand over a longer horizon. Table 8 below represents the results for the 36 month period.

Table 8. Comparison of forecast methods for 36 months (source: authors).

Comparison of Various Forecasting Methods (over 36 Months)					
Source Data: SARB Cash in Circulation in Rand Millions. (Monthly History from Jan. 2000 to Dec. 2019)					
Out-of-Sample Forecast Period: Monthly from Jan. 2020 to Dec. 2022 (Rand Millions)					
Forecasting Method	RANK	MAPE	MAE	RMSE	THIEL
		Mean Absolute Percentage Error	Mean Absolute Error	Root Mean Square Error	Thiel Inequality Coefficient
(NNETAR) Feedforward Neural Network	1	2.35%	3998	5584	0.0167
(ARIMA) Autoregressive Moving Average	2	2.63%	4490	6410	0.0192
(BATS) Box-Cox Arma Trend Seasonal	3	2.75%	4702	6723	0.0201
(GRNN) General Regressive Neural Network	4	3.24%	5460	6448	0.0191
(RWF) Random Walk Forecast	5	3.36%	5653	7111	0.0213
(SES) Simple Exponential Smoothing	6	3.86%	6547	7887	0.0237
(THETA) Exponential Smoothing with Drift	7	5.50%	9333	10,784	0.0330
(VECM) Vector Error Correction Model	8	5.99%	10,070	11,410	0.0331
(ETS) State Space Exponential Smoothing	9	6.40%	10,862	11,959	0.0367
(MLE SPLINE) Maximum Likelihood Spline	10	10.22%	17,106	19,642	0.0555

When compared to shorter projections, the ARIMA and BATS models continue to perform well; however, their error rates have somewhat increased. Even though these models are good at simulating trends and seasonality, they may become less accurate over longer periods as new or unexpected patterns start to appear. Although these models are still dependable, the minor rise in error metrics suggests that they may not be able to fully accommodate the additional complexity of a 36-month projection.

Simpler methods like SES (Simple Exponential Smoothing) and Random Walk Forecast (RWF) also experience a drop in ranking as the forecast horizon increases. These methods are generally less adaptive to changes in the data, which becomes more apparent over extended periods. The increased errors suggest that these methods might be more suited to short-term forecasting where data patterns are more stable.

The MLE SPLINE ranks the lowest in the 36-month forecast. The spline method, which is often used for its smoothness in fitting data, may not handle the unpredictability and variability of long-term forecasts well. The high MAPE and RMSE values indicate significant deviations from actual values, reflecting the method's limitations in extended forecasting scenarios.

It is important to note that during the forecast period there were a number of significant structural changes. The forecast period includes possible impacts of COVID-19, civil unrest, electricity outages, war and global economic slow-down. Despite these significant structural changes, the forecast for cash in circulation was remarkably accurate.

7. Managerial Implications

The findings of this study provide critical insights for the South African Reserve Bank (SARB) and other risk policymakers responsible for managing the country's currency supply and payment systems.

The results of this study reinforce findings from both the international and South African literature that cash demand remains shaped by a combination of macroeconomic, institutional, and technological factors. The long-run cointegrating relationships highlight the persistent role of physical cash infrastructure alongside the growing influence of digital payment technologies.

Consistent with prior studies, digital transaction channels are associated with declining cash demand over the long term, while access points such as ATMs and *bank* branches

continue to support cash usage, reflecting complementarity rather than full substitution. From a risk and financial management perspective, these findings underscore the importance of distinguishing between short-run volatility and long-run structural drivers when forecasting cash demand and planning currency operations.

Further, the persistent role of GDP and inflation in driving cash demand highlights the importance of closely monitoring macroeconomic conditions when planning cash supply and printing schedules. During periods of economic expansion, proactive adjustments to cash supply can prevent shortages that disrupt commerce, especially in the informal sector.

In addition, the results confirm that mobile penetration and EFTs are effective substitutes for cash, suggesting that policies encouraging financial inclusion through mobile banking can gradually reduce the cost burden associated with printing, storing, and distributing physical cash. However, the moderate positive impact of ATMs and bank branches on cash demand indicates that traditional infrastructure continues to serve communities with limited digital access. Therefore, SARB's Cash Management Strategy must balance digital advancements with investments in physical infrastructure to ensure equitable access.

Moreover, the superior forecasting accuracy of the NNETAR model demonstrates the value of integrating machine learning into SARB's operational forecasting toolkit. Hybrid forecasting frameworks can improve scenario planning and stress testing, particularly during periods of heightened uncertainty, such as pandemics, political unrest, or technological disruptions.

Finally, the insights can inform contingency planning. Cash remains vital in crises when digital channels may fail due to power outages, cyber-attacks, or network disruptions. Maintaining an appropriate balance between cash reserves and digital transaction capabilities is essential to protect economic resilience and public confidence in the financial system.

8. Conclusions, Limitations and Future Research

This study extends the classical Baumol–Tobin framework by integrating modern technological variables and advanced forecasting techniques to examine cash demand in an emerging market context. Using a VECM framework alongside IRF, FEVD, and neural networks, the research shows that GDP and inflation continue to sustain cash demand while digital payment technologies reduce it. The main contributions of this research are threefold:

Empirical: It provides updated estimates of cash demand determinants specific to South Africa, addressing a notable gap in local monetary policy research.

Methodological: It demonstrates the comparative advantage of machine learning models in forecasting non-linear patterns in CIC.

Practical: It offers actionable insights for SARB to manage risks across costs, plan reserves, and align digital transformation goals with cash supply strategies.

However, the study is not without limitations. The data range (2000–2019) does not account for structural shifts brought by COVID-19, which saw a paradoxical increase in cash hoarding despite digital surges. A limitation of the forecasting analysis is that the out-of-sample evaluation is centred on a period of exceptional economic disruption. While this enhances policy relevance, future research could extend the analysis by comparing forecast performance across multiple evaluation windows, including pre-pandemic periods, to further assess the stability of model rankings under different economic conditions.

A limitation of the study is the absence of a direct proxy for illicit or underground economic activity. While cash usage is often associated with such activity, data constraints prevent the inclusion of reliable high-frequency indicators for South Africa. Future research may explore alternative modelling approaches or newly available datasets to quantify

these effects explicitly. Additionally, the informal economy remains difficult to measure accurately, and proxy variables may understate its effect on CIC.

While the modelling framework captures both long-run equilibrium relationships and short-run dynamics, certain limitations should be acknowledged. The use of interpolated quarterly variables may smooth high-frequency variation, and the relatively rich lag structure introduces a trade-off between dynamic completeness and parameter parsimony. Future research may explore alternative mixed-frequency frameworks or parsimonious specifications to further assess robustness. Despite these limitations, the core long-run relationships and qualitative findings remain consistent with monetary theory and comparable empirical studies.

Alternative causality testing approaches, such as the Toda–Yamamoto procedure, offer advantages by avoiding cointegration pre-testing and may yield complementary insights. While such methods are not employed in this study, future research could apply them alongside VECM-based causality tests to further validate the direction and robustness of causal relationships.

Future research should:

- Incorporate post-pandemic payment behaviour to validate or adjust findings.
- Explore the implications of Central Bank Digital Currencies (CBDCs) on cash demand.
- Utilise real-time transaction data and big data analytics for more dynamic forecasting.

By understanding the evolving dynamics of cash demand, South Africa can ensure monetary stability while navigating the transition towards a more digital economy.

Author Contributions: R.R. for all components; Supervision, H.A.D. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Randheer Ramsoomer.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available from the 771 Refinitiv database access in December 2024. No new data was created during this research.

Acknowledgments: The authors gratefully acknowledge the University of Johannesburg, A. Drotskie, and H.A Donfack for their guidance and constructive feedback throughout this study.

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

Appendix A. (LASSO Regression) (Source: Calculated by Author)

(2000–2019)					
Variable	Estimated	Standard	Df	z Value	Pobability
	Coefficient	Error			Pr(> z)
(Intercept)	0.361945	0.414267	1.000	0.874	0.3823
atmRatioY	−0.029086	0.018863	0.790	−1.542	0.1231
branchRatioY	−0.031640	0.019117	0.827	−1.655	0.0979
cellY	0.130637	0.024445	1.000	5.344	9.08×10^{-8} ***
cpim	0.698261	0.093487	1.000	7.469	8.08×10^{-14} ***
elecFundValM	0.098266	0.017454	1.000	5.630	1.80×10^{-8} ***
gdpQ	0.342929	0.082866	0.999	4.138	3.50×10^{-5} ***
intRealY	−0.001911	0.004905	0.860	−0.390	0.6968
taxRatioQ	−0.026180	0.018145	0.774	−1.443	0.1491

Significant codes: *** $p < 0.001$.

Appendix B. Durbin–Watson (DW) Autocorrelation Test (Source: Calculated by Author)

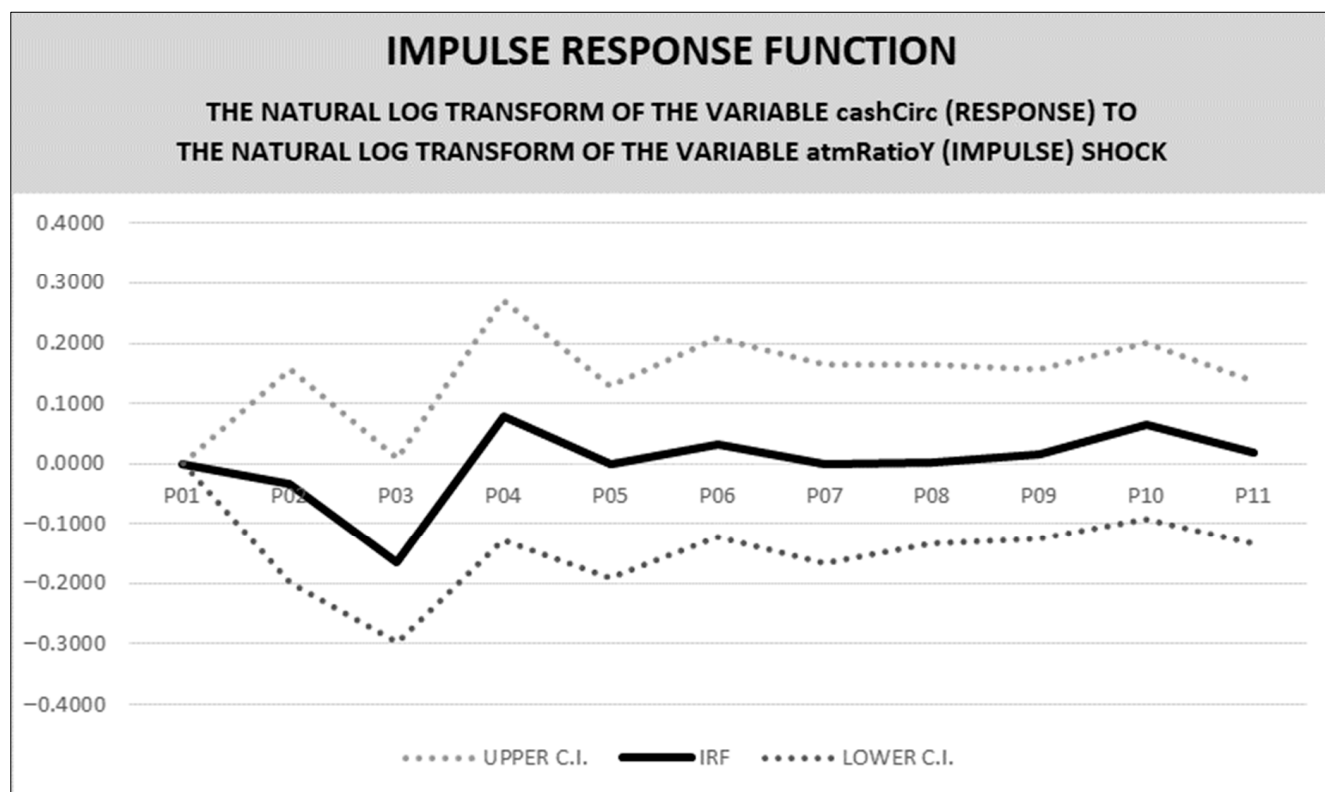
Variables	DW Statistics	(<i>p</i> -Value) (Probability)	Hypothesis Test
cashCirc	0.46828	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
atmRatioY	0.028901	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
branchRatioY	0.081277	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
cellY	0.021982	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
cpiM	0.042498	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
elecFundValM	0.25564	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
gdpQ	0.10579	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
intRealY	0.29782	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance
taxRatioQ	0.79782	($<2.2 \times 10^{-16}$)	Autocorrelation at 1% level of significance

Appendix C. Breusch–Pagan (BP) Heteroskedasticity Test (Source: Calculated by Author)

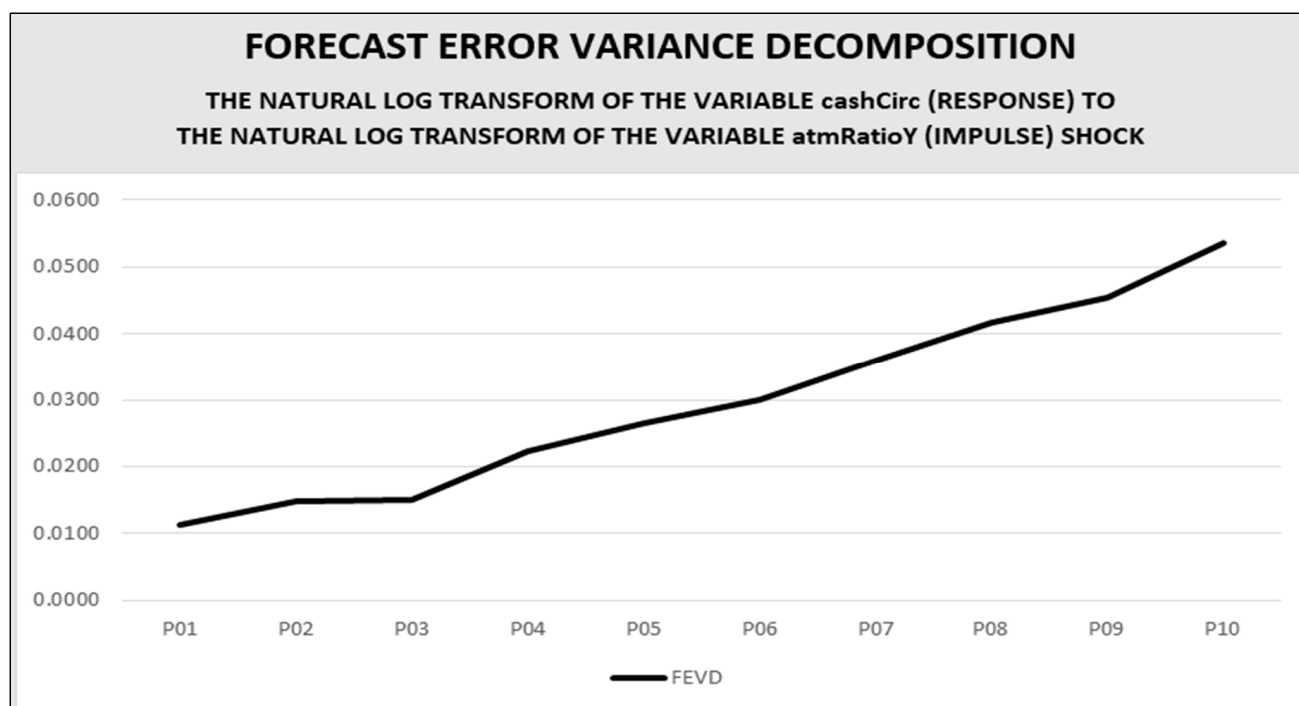
Variables	DW Statistics	df	(<i>p</i> -Value) (Probability)	Hypothesis Test
cashCirc	3.5586	1	(0.05924)	Heteroskedasticity at 10% level of significance
atmRatioY	4.6826	1	(0.03047)	Heteroskedasticity at 5% level of significance
branchRatioY	0.12054	1	(0.7284)	Not heteroskedastic
cellY	31.41	1	(2.089×10^{-8})	Heteroskedasticity at 1% level of significance
cpiM	26.471	1	(2.676×10^{-7})	Heteroskedasticity at 1% level of significance
elecFundValM	109.62	1	($<2.2 \times 10^{-16}$)	Heteroskedasticity at 1% level of significance
gdpQ	1.4099	1	(0.2351)	Not heteroskedastic
intRealY	8.04	1	(0.004576)	Heteroskedasticity at 1% level of significance
taxRatioQ	17.58	1	(2.755×10^{-5})	Not heteroskedastic

Appendix D. IRF/FEVF/OIRF

Cash Demand Response to ATM (atmRatioY) IRF

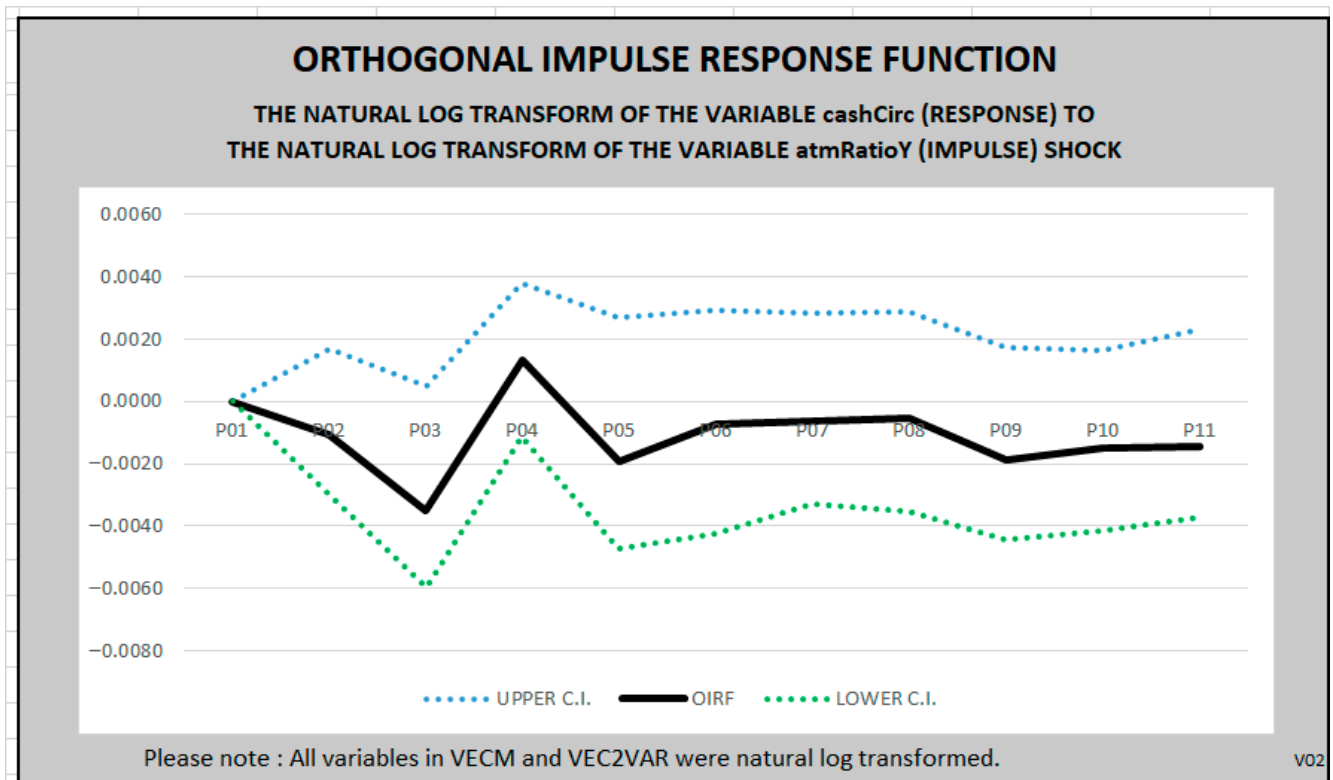


Cash Demand Response to ATM (atmRatioY) FEVD



Similarity Summary Page

Cash Demand Response to ATM (atmRatioY) OIRF



References

- Aliha, M. P., Sarmidi, T., & Said, F. F. (2018). *Investigating the effect of financial innovations on the demand for money in Australia using dols and fmols and comparing their predictive powers*. Available online: <https://ideas.repec.org/a/hrs/journal/vxy2018i1p47-61.html> (accessed on 20 March 2021).
- Atoyebe, S. B., Olayiwola, M. F., Oladapo, J. O., & Oladapo, D. I. (2023). Forecasting currency in circulation in Nigeria using Holt-Winters exponential smoothing method. *South Asian Journal of Social Studies and Economics*, 20(1), 25–41. [CrossRef]
- Baumol, W. J. (1952). The transactions demand for cash: An inventory theoretic approach. *Quarterly Journal of Economics*, 66(4), 545–556. [CrossRef]
- Bordo, M. D. (2022). Central bank digital currency in historical perspective: Another crossroad in monetary history. *capitalism: A journal of history and economics* (Vol. 3, pp. 421–442). University of Pennsylvania Press. [CrossRef]
- Christoffersen, P. F., & Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52(8), 1273–1287. [CrossRef]
- Dunne, J. P., & Kasekende, E. (2016). *Financial innovation and money demand in sub-Saharan Africa*. Available online: <https://econrsa.org/publications/research-briefs/financial-innovation-and-money-demand-sub-Saharan-Africa> (accessed on 2 April 2023).
- Endale, M. (2023). *Essays on international economics, forecasting money demand and stock markets*. Ph.D. thesis, Clark University, Worcester, Massachusetts. ProQuest Dissertations Publishing, 30485268.
- Fish, T., & Whymark, R. (2015). How has cash usage evolved in recent decades? What might drive demand in the future? *Bank of England Quarterly Bulletin*, Q3, 216–227.
- Jack, W., & Suri, T. (2011). Mobile money: The economics of M-Pesa. In *NBER Working Paper No. 16721*. National Bureau of Economic Research.
- Keynes, J. M. (1936). *The general theory of employment, interest and money*. Palgrave MacMillan.
- Kreinovich, T. V., Trung, D. N., & Thanh, V. D. (2019). *Beyond traditional probabilistic methods in economics*. Springer. [CrossRef]
- Masciandaro, D. (2018a). Central bank digital cash and cryptocurrencies: Insights from a new baumol-friedman demand for money. *Australian Economic Review*, 51(4), 540–550. Available online: <https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=ecn&AN=1790180&site=ehost-live&scope=site> (accessed on 29 August 2023)..
- Masciandaro, D. (2018b). Missing money: Cash demand and shadow economy. *Economic Modelling*, 69, 244–253.

- Mlambo, C., & Msosa, S. K. (2020). The effect of financial technology on money demand: Evidence from selected African states. *European Research Studies*, 23(1), 143–150.
- Petropoulos, F., Makridakis, S., & Assimakopoulos, V. (2019). Statistical and machine learning forecasting methods: Concerns and ways forward. *European Journal of Operational Research*, 272(1), 1–21.
- Rauf, A., & Khan, M. A. (2012). Determinants of money demand in Pakistan. *International Journal of Economics and Finance*, 4(10), 146–152.
- Shirai, S., & Sugandi, E. A. (2019). Growing global demand for cash. *International Business Research*, 12, 74.
- South African Reserve Bank (SARB). (2021). *National payment system regulatory and oversight Report 2020/21*. South African Reserve Bank. Available online: <https://www.resbank.co.za> (accessed on 29 August 2023).
- South African Reserve Bank (SARB). (2022a). *Annual report 2021/22*. South African Reserve Bank. Available online: <https://www.southafricanreservebank.co.za/en/home/publications/publication-detail-pages/reports/annual-reports/2022/SARB-Annual-Report-2021-22> (accessed on 29 August 2023).
- South African Reserve Bank (SARB). (2022b). *Quarterly bulletin No. 303, March 2022*. South African Reserve Bank. Available online: <https://www.southafricanreservebank.co.za/en/home/publications/publication-detail-pages/quarterly-bulletins/quarterly-bulletin-publications/2022/FullQuarterlyBulletinNo303March2022> (accessed on 29 August 2023).
- Statistics South Africa. (2025). *South Africa's informal economy: A lifeline for millions*. News release, 26 March 2025. Statistics South Africa. Available online: <https://www.statssa.gov.za/?p=18255> (accessed on 29 August 2023).
- Tagat, A., & Trivedi, P. L. (2020). Demand for cash: An econometric model of currency demand in India. *Macroeconomics and Finance in Emerging Market Economies*, 13(2), 205–222. [CrossRef]
- Tan, W., Puah, C., Wong, S., & Chong, M. (2020). Economic uncertainty and the demand for broad money in South Africa. *Business and Economic Research*, 10, 123–133.
- Tobin, J. M. (1956). The interest-elasticity of transactions demand for cash. *The Review of Economics and Statistics*, 38(3), 241–247. [CrossRef]
- Zhang, G., Eddy Patuwo, B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.