Mean Reversion Lessens Mean Blur: Evidence from the S&P Composite Index

Luigi Buzzacchi 1,* and Luca Ghezzi 2

1 Interuniversity Department of Regional and Urban Studies and Planning & FULL, Politecnico di Torino, Viale Mattioli 39, 10125 Torino, Italy
2 Department of Integrated Business Management, Università Carlo Cattaneo, Corso Matteotti 22, 21053 Castellanza, Italy
* Correspondence: luigi.buzzacchi@polito.it

Abstract: This study makes use of a very long time series of the S&P Composite Index, checking once more that the rates of return benefit from aggregational normality. It performs unit root tests as well as elementary statistical tests that take advantage of normality. It finds that mean blur is not consistent with the hypothesis of random walk with constant parameters, because the means of the annual real rates of linear return can be estimated as usual. It gives further evidence that the rates of return on the S&P Composite Index are mean-reverting.

Keywords: stock index; mean reversion; mean blur

1. Introduction

Mean reversion is a popular topic dealt with in informative books such as Siegel (2014) and Shiller (2015). According to Siegel (2014, p. 6), mean reversion in stock market returns means that “periods of above-average returns tend to be followed by periods of below-average returns and vice versa.” Although short-term stock market returns are very volatile, long-term stock market returns are seemingly not volatile enough to meet the alternative hypothesis of random walk.

As summarised by the Royal Swedish Academy of Sciences (2013), the empirical analyses of Eugene F. Fama, Lars Peter Hansen, and Robert J. Shiller have shown that stock prices are unpredictable over a period of a few days or weeks, and somewhat predictable over, say, the following 3–5 years. The 2013 Nobel laureates have also provided both rational and behavioural interpretations of such predictability. Nevertheless, mean reversion is hard to test, since very long time series of stock prices and returns are needed (Spierdijk and Bikker 2017).

This study reconsiders mean blur, which entails that the mean rates of return cannot be determined accurately enough by using statistical methods (Luenberger 2014). On stressing that mean blur occurs in discrete-time random walks with constant parameters owing to high risk/reward ratios, it finds that mean reversion seemingly lessens mean blur by using a very long time series of the S&P Composite Index.

Keep in mind that mean reversion cannot be disregarded when forming long-run capital market expectations on various asset classes, the last task but one in the planning step of a rational portfolio management process (Maginn et al. 2007). As shown by Sharpe et al. (2007), long-run capital market expectations play an important role when determining a strategic asset allocation, the last task in the above-mentioned planning step.

The plan of our study is as follows: the scientific literature on mean reversion is reviewed in Section 2, the data set is described in Section 3, the statistical properties of the S&P Composite Index are reviewed in Section 4, our statistical findings are presented in Section 5, and conclusions are drawn in Section 6.
2. Literature Review

Evidence either in favour or against the random walk hypothesis was found by Lo and MacKinlay (1988) by using weekly stock returns from 1962 to 1985. More precisely, the one-week serial correlation was positive and statistically significant for five size-sorted portfolios and two US stock indices, being the largest for the smallest quintile of size-sorted portfolios; on average, it was slightly negative and statistically insignificant for individual stocks.

Meanwhile, mean reversion in stock returns was examined by Poterba and Summers (1988) and Fama and French (1988a). Their empirical analyses rest on the probabilistic model of Summers (1986), where stock prices include two components, a permanent and a transitory one. Since the latter component is stationary and slowly decaying, stock returns should display negative serial correlation at all non-zero lags. Such a mean reversion in stock returns can have two different, though not alternative, origins. On the one hand, stock prices may take long-lasting swings away from fundamental values also owing to noise trading. On the other, risk factors, and hence the returns required by rational investors, may be time-varying.

Fama and French (1988a) provided stronger empirical evidence than Poterba and Summers (1988); they made use of monthly stock returns from 1926 to 1985, let $T$ range from 1 to 10 years, and obtained $T$-year overlapping returns for several size and industry portfolios. According to their regression analysis, serial correlation becomes negative for $T = 2$ and takes its lowest values for $T = 3, 4, 5$. Therefore, 3–5-year returns are somewhat predictable, especially for portfolios of small stocks; negative serial correlation may be due to common rather than firm-specific factors.

In turn, Fama and French (1988b) provided stronger empirical evidence than Fama and French (1988a) by using monthly stock returns from 1927 to 1986; they considered two US portfolios, either value- or equal-weighted, and run simple linear regressions of portfolio returns on dividend/price ratios that explain a considerable proportion of the variances of 2–4-year returns.

Subsequently, Campbell and Shiller (1998) performed a statistical analysis of the S&P 500 stock index; their data set went back to 1872. A cyclically adjusted price-earnings ratio was derived by dividing the real S&P 500 by an appropriate 10-year moving average of its real earnings. The CAPE ratio was also shown to fluctuate within its historical range, to have little power in forecasting the ten-year growth in cyclically adjusted earnings, and to have much more power in forecasting the ten-year growth in stock prices. Campbell and Shiller (1998, p. 19) claimed accordingly that US stock prices, “rather than dividends and earnings appear to adjust to bring abnormal valuation ratios back to historical average level.” When the CAPE ratio is above average, the 10-year real stock return tends to be below average, and vice versa. Upsides and downsides of the CAPE ratio are outlined in Siegel (2014, chap. 11).

Cornell (2010) explored the link between earnings growth and GDP growth under the assumption that the ratio of aggregate earnings to GDP is stationary but not constant, the numerator being more volatile than the denominator. His preliminary validation was based on two data sources for US earnings: the US national income and product accounts, and the S&P 500 companies. Two ratios were obtained from 1986 to 2008, the latter being smaller than the former. On recalling that technical advance and population growth drive the growth of real per-capita GDP, Cornell (2010, p. 59) claimed, “Although the data largely support the hypothesis that E/GDP is stationary, it is far from constant. [My] Figure 2 shows that corporate profits vary between 3% and 11% of GDP. The variability of the ratio for S&P 500 earnings is even greater. This variability suggests that when earnings are low relative to GDP, they grow more quickly; the reverse is true when earnings are relatively high. This mean reversion in the growth rate of earnings maintains the stationarity of E/GDP.” As reported by Loomis and Buffett (1999) and Buffett and Loomis (2001), Warren Buffett, chairman of Berkshire Hathaway, had told his audiences earlier that those after-tax corporate profits had varied between 4% and 8% of GDP over the previous 60 years. Remarkably, both Campbell and Shiller (1998) and Loomis and Buffett (1999) were bearish.
Spierdijk and Bikker (2017) provided a comprehensive review of the literature on mean reversion in stock prices and stock returns. They observed that mean reversion is hard to test, since very long time series are needed. They concluded that the empirical evidence in favour of mean reversion is thin. Indeed, Golez and Koudijs (2018) achieved statistical significance by combining the annual returns of the Amsterdam, London, and New York stock markets over the very long historical period 1629–2015. They found that dividend/price ratios are stationary and tend to increase (decrease) before a period of above-average (below-average) returns. They also found that dividend/price ratios tend to grow in a recession.

The above-mentioned findings of Lo and MacKinlay (1988) were corroborated by Nguyen et al. (2022), who made use of the monthly returns of three stock market indices, which run from January 1926 to December 2016 for the US and from December 1989 to December 2015 for both the UK and Japan. According to a battery of unit root tests, all three stock indices are trend-stationary rather than difference-stationary like random walks. Moreover, one or two structural breaks were detected in each stock market.

3. Data Set

Data were downloaded in winter 2022 from the webpage of Professor Robert Shiller, Yale University. He is a 2013 Nobel Laureate in Economic Sciences. Such monthly data span the historical period 1872–2020 and deal with the S&P Composite Index, S&P dividends, S&P earnings, and the US Consumer Price Index.

As explained by Professor Shiller, monthly dividends and earnings were linearly interpolated either from annual data until 1925 or from quarterly data since 1926. As recalled by Siegel (2014), the S&P 500 Index, launched in 1957, expanded on the S&P Composite Index, a 90-stock index since 1926.

In this study, a total return version of the S&P Composite Index is used, either in nominal or real terms, under the assumption that gross dividends were reinvested at the end of each month. Both 1788 monthly rates of return and 149 annual rates of return are computed from 1872 to 2020; both linear and logarithmic rates of return are considered.

The sample moments of the monthly real rates of return are reported in Table 1, whereas the sample moments of the annual real rates of return are reported in Table 2; both linear and logarithmic rates of return are considered. The relationship between mean, median, and skewness is thoroughly explained by von Hippel (2005).

Table 1. Monthly real rates of return, 1872–2020.

<table>
<thead>
<tr>
<th></th>
<th>Linear Rate</th>
<th>Logarithmic Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst rate</td>
<td>−26.19%</td>
<td>−30.36%</td>
</tr>
<tr>
<td>Median</td>
<td>0.94%</td>
<td>0.94%</td>
</tr>
<tr>
<td>Best rate</td>
<td>52.43%</td>
<td>42.15%</td>
</tr>
<tr>
<td>Negative rates</td>
<td>39.21%</td>
<td>39.21%</td>
</tr>
<tr>
<td>Mean</td>
<td>0.65%</td>
<td>0.56%</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>4.10%</td>
<td>4.08%</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.58</td>
<td>−0.36</td>
</tr>
<tr>
<td>Kurtosis excess</td>
<td>18.24</td>
<td>11.38</td>
</tr>
</tbody>
</table>


Notice that linear and logarithmic rates display, by definition, the same percentage of negative rates, with monthly real rates being more often negative than annual real rates; monthly linear rates display a considerable kurtosis excess, or leptokurtosis, with their empirical distribution having a high peak and fat tails, while annual linear rates display a negligible kurtosis excess.
Table 2. Annual real rates of return, 1872–2020.

<table>
<thead>
<tr>
<th></th>
<th>Linear Rate</th>
<th>Logarithmic Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst rate</td>
<td>−39.29%</td>
<td>−49.90%</td>
</tr>
<tr>
<td>Median</td>
<td>10.01%</td>
<td>9.54%</td>
</tr>
<tr>
<td>Best rate</td>
<td>53.25%</td>
<td>42.69%</td>
</tr>
<tr>
<td>Negative rates</td>
<td>30.87%</td>
<td>30.87%</td>
</tr>
<tr>
<td>Mean</td>
<td>8.55%</td>
<td>6.73%</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>17.98%</td>
<td>17.63%</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.29</td>
<td>−0.83</td>
</tr>
<tr>
<td>Kurtosis excess</td>
<td>0.03</td>
<td>0.92</td>
</tr>
</tbody>
</table>


4. Preliminary Analysis

4.1. Aggregational Normality

As is well known, the S&P Composite Index benefits from aggregational normality. Accordingly, the null hypothesis of normally distributed rates of return has been rejected for monthly linear rates of return, both nominal and real, whereas it has not been rejected for annual linear rates of return, both nominal and real. The null hypothesis has also been rejected for monthly and annual logarithmic rates of return, both nominal and real. Our outcomes are in accordance with Nguyen et al. (2022), where the null hypothesis is rejected for the monthly returns on the stock market indices of the US, the UK, and Japan.

Jarque-Bera and Shapiro-Wilk tests have been performed using RStudio software: when a p-value is greater than 0.05, the null hypothesis is not rejected. Whatever the monthly rate of return, the p-values of both tests are lower than \(2.2 \times 10^{-16}\). The p-values for all annual rates are reported in Table 3.

Table 3. Annual rates of p-values.

<table>
<thead>
<tr>
<th>Rate</th>
<th>Jarque-Bera</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>0.3494</td>
<td>0.5082</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>3.741 \times 10^{-6}</td>
<td>0.00021</td>
</tr>
<tr>
<td>Real Linear</td>
<td>0.3595</td>
<td>0.3214</td>
</tr>
<tr>
<td>Real Logarithmic</td>
<td>2.501 \times 10^{-5}</td>
<td>0.0001341</td>
</tr>
</tbody>
</table>

Quantile-quantile plots have been constructed using RStudio software. For instance, the empirical distribution of the annual real rates of linear return is compared with a normal distribution in Figure 1. The confidence level of the shaded area is 0.95.

Non-overlapping multi-year rates of logarithmic return have also been examined. The p-values of Jarque-Bera and Shapiro-Wilk tests and the attendant quantile-quantile plots are not reported. The null hypothesis of normally distributed rates of return has been rejected for 2- and 3-year rates of logarithmic return, both nominal and real, whereas it has not been rejected for 4-year rates of logarithmic return, both nominal and real.

Therefore, the linear and logarithmic rates of return benefit from aggregational normality on two different time scales.
4.2. Serial Correlation

Sample autocorrelation functions have been estimated using RStudio software. The autocorrelation function for the monthly real rates of linear return is displayed in Figure 2. The unreported autocorrelation function for the monthly real rates of logarithmic return has a similar pattern. The one-month autocorrelation is spurious; it is brought about by the S&P Composite data, which are monthly averages of daily closing values (Working 1960). Nonetheless, a few sample correlations lie outside the 0.95 confidence interval, the 5-month autocorrelation being positive and the 13-, 14-, 15-, 20-, and 21-month autocorrelations being negative.

The autocorrelation function for the annual real rates of linear return is displayed in Figure 3. The unreported autocorrelation function for the annual real rates of logarithmic return has a similar pattern. Only the 2-year autocorrelation lies outside the 0.95 confidence interval, being negative. Since Figures 2 and 3 taken together are hard to interpret, a basic rescaled range analysis is performed in Section 4.3, whereas stationarity tests are performed in Section 5.2.

Figure 1. Annual real rates of linear return. Quantile-quantile plot.

Figure 2. Monthly real rates of linear return. Autocorrelation function.

Figure 3. Annual real rates of linear return. Autocorrelation function.
The one-month autocorrelation is spurious; it is brought about by the S&P Composite data, which are monthly averages of daily closing values (Working 1960). Nonetheless, a few sample correlations lie outside the 0.95 confidence interval, the 5-month autocorrelation being positive and the 13-, 14-, 15-, 20-, and 21-month autocorrelations being negative.

The autocorrelation function for the annual real rates of linear return is displayed in Figure 3. The unreported autocorrelation function for the annual real rates of logarithmic return has a similar pattern. Only the 2-year autocorrelation lies outside the 0.95 confidence interval, being negative. Since Figures 2 and 3 taken together are hard to interpret, a basic rescaled range analysis is performed in Section 4.3, whereas stationarity tests are performed in Section 5.2.

4.3. Rescaled Range Analysis

Use has been made of \( n = 149 \times 12 = 1788 \) monthly real rates of logarithmic return from January 1872 to December 2020. The whole historical period has been divided into \( \frac{1788}{m} \) subperiods, where \( m \) is the number of monthly observations that fall into each subperiod. Following Peters (1994), we have set \( m \geq 10 \), since a small \( m \) is not likely to be matched by a reliable estimate of the rescaled range \( R/S_m \). Iterating the procedure outlined in the Appendix A, the rescaled range \( R/S_n \) has been estimated for \( n \) equal to \( m = 10, 20, 30, \ldots, 90, 100 \) months.

According to Hurst (1951), a British hydrologist, the exponent \( H \) obeys the equation

\[
R/S_m = c \times m^H
\]

so that it can be estimated by running the simple regression

\[
\log_{10}(R/S_m) = \log_{10}(c) + H \times \log_{10}(m) + \text{error}_m
\]

If both randomness and normality are assumed, the standard deviation of the resulting estimate \( \hat{H} \) is approximately \( \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{1788}} = 0.02365 \). Recall that the null hypothesis of normality has been previously rejected for all monthly rates of return.
The original theory allows one to distinguish between randomness ($H = 0.5$), a persistent trend ($0.5 < H \leq 1$), and an antipersistent trend ($0 \leq H < 0.5$). Randomness is represented by a sequence of independent and identically distributed random variables, in a persistent trend the next deviation from the mean is likely to have the present sign, and in an antipersistent trend the next deviation from the mean is likely to reverse the present sign. Persistent time series also display dependence between the remote past and the distant future owing to long-term memory; antipersistent time series reverse direction more often than random time series. For instance, Peters (1994) finds that the Dow Jones Industrial Average is a persistent time series by using the daily closing prices from 1888 to 1991. In contrast, he finds that the realized volatility of the S&P Composite Index is an antipersistent time series by using the daily prices from 1928 to 1989.

Unfortunately, a biased estimate $\hat{H}$ may be obtained, since $\hat{H} > 0.5$ may be due to short-term memory rather than long-term memory. However, a heuristic remedy, reported by Peters (1994) and references therein, can reduce the risk of a biased estimate $\hat{H}$. Accordingly, the monthly real rates of logarithmic return have been replaced with the residuals of their autoregression of order 1; next, ten rescaled ranges have been estimated by iterating the procedure outlined in the Appendix A. Finally, the following Hurst exponent has been obtained from the simple regression (2):

$$\hat{H} = 0.64803$$

along with the coefficient of determination $R^2 = 0.9972$. This outcome is hard to interpret owing to the heuristic procedure.

Suppose that the monthly real rates of logarithmic return form a sequence of identically distributed random variables. Adapting from Corazza and Malliaris (2002), we can now realize that for $0.5 < H \leq 1$, such random variables may be either independent and Pareto-Lévy stable so that their variance is undefined, or not independent though fat-tailed. As stated by Campbell et al. (1997), if population variance is undefined and sample size is sequentially increased, sample variance does not approach a limiting value; Pareto-Lévy stability is preserved under addition. Therefore, the assumption of independent and Pareto-Lévy stable random variables is not tenable in light of the statistical tests of Section 4.1, whereby the 4-year real rates of logarithmic return are normally distributed.

Nonetheless, further analysis is required to ascertain whether the above-mentioned random variables are not independent though fat-tailed.

Borrowing from Hurst (1956), we have reconsidered the monthly real rates of logarithmic return and readily computed the running sum of the monthly departures from the mean, which is equal to 0.56%. Figure 4 displays such cumulative departures from the mean, which slowly fluctuate around the horizontal axis. In our opinion, Figure 4 complements Siegel (2014, p. 6), who observes: “Note that stocks fluctuate both below and above the trendline but eventually return to the trend. Economists call this behavior mean reversion, a property that indicates that periods of above-average returns tend to be followed by periods of below-average returns and vice versa.” Recall that such a trendline cannot be rigorously estimated by using elementary statistical methods.
5. Statistical Findings

5.1. Mean Blur

Let time be measured in years and $SP_t$ be the market value of a total real return stock index at the end of year $t$. Suppose that the total real return stock index follows a discrete-time random walk with constant parameters

$$\ln(SP_t) = \ln(SP_{t-1}) + r_t$$

where the error $r_t$ is an annual real rate of logarithmic return. Assume away the occurrence of structural breaks and suppose at first that the annual real rates of logarithmic return form a sequence of independent and normally distributed random variables. Let $\mu_{\log}$ and $\sigma_{\log}$ be their constant and known population mean and standard deviation. Let $n$ be the sample size and 0.95 be the confidence level. As is well known, the sample mean $\bar{r}_{\log}$ is normally distributed so that its confidence interval is

$$\mu_{\log} - 1.96 \frac{\sigma_{\log}}{\sqrt{n}} \leq \bar{r}_{\log} \leq \mu_{\log} + 1.96 \frac{\sigma_{\log}}{\sqrt{n}}$$

Therefore, the larger the sample size $n$, the narrower is the confidence interval. Unfortunately, increasing the sampling frequency is pointless. As hinted by Black (1993) and explained by Luenberger (2014), using a smaller unit of time does not help, since only the number of available years of data matters. For instance, if the monthly real rates of logarithmic return are considered, the sample size is larger and equal to $12 \times n$. In contrast, the constant and known population mean and standard deviation are smaller and respectively equal to $\frac{\mu_{\log}}{12}$ and $\frac{\sigma_{\log}}{\sqrt{12}}$. The confidence interval of the sample mean $\bar{r}_{\log}$ is

$$\frac{1}{12} \left( \mu_{\log} - 1.96 \frac{\sigma_{\log}}{\sqrt{n}} \right) \leq \bar{r}_{\log} \leq \frac{1}{12} \left( \mu_{\log} + 1.96 \frac{\sigma_{\log}}{\sqrt{n}} \right)$$

It is readily realised that the width of the confidence interval (6) is one-twelfth of the width of the confidence interval (5).

Suppose now that the annual real rates of logarithmic return form a sequence of independent and identically distributed random variables. Let now $\mu_{\log}$ and $\sigma_{\log}$ be their
constant and unknown population mean and standard deviation. Let \( s_{\log} \) be the sample standard deviation. As is well known, we have approximately

\[
\tau_{\log} - 1.96 \frac{s_{\log}}{\sqrt{n}} \leq \mu_{\log} \leq \tau_{\log} + 1.96 \frac{s_{\log}}{\sqrt{n}}
\]  

(7)

provided that the sample size \( n \) is large enough.

According to Table 2, we have \( n = 149 \) as well as \( \tau_{\log} = 6.73\% \) and \( s_{\log} = 17.63\% \). Substituting such values into the inequalities (7) gives

\[
3.87\% \leq \mu_{\log} \leq 9.59\%
\]  

(8)

which seems to be a case of mean blur, a definition given by Luenberger (2014). On considering only individual stocks, he concludes that reliable estimates of population means cannot be derived from historical rates of return, since too many years of data would be needed. However, he does not stress that mean blur is tacitly implied by the hypothesis of random walk (4) owing to high risk/reward ratios.

Both individual stocks and asset classes are considered by Haugen (1997, p. 171), who claims that, “For classes of securities, sample means of past rates of return are a good starting point for estimating future expected rates of return. Analysts usually start there, and then they make subjective adjustments to the estimates based on contemporary economic conditions which are now different from the past.” He adds that, “While sample mean returns serve as a good starting point in estimating the expected returns of classes of securities, they are poor indicators of differentials in expected return between securities within a class, such as common stocks. The past returns to individual stocks have been affected by myriad idiosyncratic events which are unlikely to repeat themselves in the future.”

5.2. Stationarity Tests

The hypothesis of random walk has been tested on 1788 monthly rates of logarithmic return. Unit root tests have been carried out by using RStudio software. According to both augmented Dickey-Fuller and Phillips-Perron tests, the null hypothesis of a unit root is rejected with a confidence level of 0.99 for both nominal and real rates of logarithmic return. Therefore, the total return version of the S&P Composite Index is trend-stationary rather than difference-stationary, both in nominal and real terms. Our outcomes are consistent with the broader evidence provided by Nguyen et al. (2022) for the monthly returns on the stock market indices of the US, the UK, and Japan.

Additional stationarity tests have been performed to gain additional insight into mean blur. Use has been made of annual real rates of linear return, which meet the assumption of normality. Let \( \mu_{\text{lin}} \) and \( \sigma_{\text{lin}} \) be their unknown population mean and standard deviation. At first, the historical period 1872–2020 has been divided into the two subperiods 1872–1935 and 1936–2020, respectively consisting of 64 and 85 years. The attendant sample moments are reported in Table 4.
Table 4. Annual real rates of linear return, 1872–2020, two subperiods.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst rate</td>
<td>−39.29%</td>
<td>−36.66%</td>
<td>−39.29%</td>
</tr>
<tr>
<td>Median</td>
<td>10.01%</td>
<td>4.90%</td>
<td>11.98%</td>
</tr>
<tr>
<td>Best rate</td>
<td>53.25%</td>
<td>53.25%</td>
<td>49.16%</td>
</tr>
<tr>
<td>Negative rates</td>
<td>30.87%</td>
<td>32.81%</td>
<td>29.41%</td>
</tr>
<tr>
<td>Mean</td>
<td>8.55%</td>
<td>8.61%</td>
<td>8.51%</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>17.98%</td>
<td>18.70%</td>
<td>17.53%</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.29</td>
<td>−0.04</td>
<td>−0.52</td>
</tr>
<tr>
<td>Kurtosis excess</td>
<td>0.03</td>
<td>0.01</td>
<td>0.10</td>
</tr>
</tbody>
</table>


Our first null hypothesis is that the sample means for both subperiods come from populations that have the same mean. The first null hypothesis is not rejected with a confidence level of 95 percent, since the two-sample \( t \) statistic takes the value

\[
t = 0.032 < 1.655	ag{9}
\]

Our second null hypothesis is that the sample variances for both subperiods come from populations that have the same variance. The second null hypothesis is not rejected with a confidence level of 95 percent, since the \( F \) statistic takes the value

\[
F = 1.139 < 1.468
\]

Next, a single factor analysis of variance (ANOVA) has been performed. Therefore, the historical period 1872–2020 has been divided into the three subperiods 1872–1911, 1912–1970, and 1971–2020, respectively consisting of 40, 59 and 50 years. The attendant sample moments are reported in Table 5.

Table 5. Annual real rates of linear return, 1872–2020, three subperiods.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst rate</td>
<td>−39.29%</td>
<td>−27.94%</td>
<td>−36.66%</td>
<td>−39.29%</td>
</tr>
<tr>
<td>Median</td>
<td>10.01%</td>
<td>7.96%</td>
<td>7.45%</td>
<td>12.29%</td>
</tr>
<tr>
<td>Best rate</td>
<td>53.25%</td>
<td>40.53%</td>
<td>53.25%</td>
<td>34.99%</td>
</tr>
<tr>
<td>Negative rates</td>
<td>30.87%</td>
<td>27.50%</td>
<td>35.59%</td>
<td>28.00%</td>
</tr>
<tr>
<td>Mean</td>
<td>8.55%</td>
<td>8.67%</td>
<td>8.82%</td>
<td>8.14%</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>17.98%</td>
<td>14.27%</td>
<td>21.39%</td>
<td>16.51%</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.29</td>
<td>−0.04</td>
<td>−0.09</td>
<td>−0.94</td>
</tr>
<tr>
<td>Kurtosis excess</td>
<td>0.03</td>
<td>−0.05</td>
<td>−0.55</td>
<td>0.69</td>
</tr>
</tbody>
</table>


Our null hypothesis is now that the sample means for all three subperiods (or groups) come from populations that have the same mean. The null hypothesis is provisionally not rejected with a confidence level of 95 percent, since the \( F \) statistic takes the value

\[
F = \frac{\text{variance between groups}}{\text{variance within groups}} = 0.002 < 3.058
\]

\[
(11)
\]
Unfortunately, the null hypothesis of homoskedasticity is rejected with a confidence level of 95 percent, since Bartlett’s statistic takes the value

\[ \chi^2 = 8.1039 > 0.103 \]  

(12)

The attendant \( p \)-value is equal to 0.01736, implying that heteroskedasticity is not strong. Indeed, the range of the sample means in Table 5 is narrow, whereas the range of the sample standard deviations in Table 5 is broad. However, small samples are involved so that the long-run stationarity of variance is an issue that would require further examination. In doing so, the role played by structural breaks should not be disregarded.

Nonetheless, Equation (8) seems to clash with Equations (9) and (10) as well as with the previous unit root tests. In our opinion, this is further evidence that the annual real rates of return are not independent owing to mean reversion so that the unknown standard deviation of the sample mean is lower than

\[ \frac{\sigma_{lin}}{\sqrt{n}} \]  

(13)
as tacitly assumed.

Remarkably, other evidence was provided by Poterba and Summers (1988), who also analyzed the performance of the S&P Composite Index over the years 1871–1985 by making a comparison between different multi-year real rates of logarithmic return and the one-year real rates of logarithmic return. According to the variance-ratios of their Table 3, the annualized variances of the former are lower than the variance of the latter. Unfortunately, statistical significance is hard to attain. Recall that the above-mentioned variances are proportional to the sampling interval under the assumption of random walk, whereas they are less than proportional to the sampling interval under the assumption of mean reversion.

6. Conclusions

This study has used a very long time series of the S&P Composite Index, checking once more that the annual rates of linear return, both nominal and real, benefit from aggregational normality. It has stressed that the hypothesis of random walk with constant parameters implies that means are very blurred owing to high risk/reward ratios so that they cannot be estimated by using statistical methods. It has performed unit root tests on the S&P Composite Index as well as elementary statistical tests that take advantage of normality. It has ascertained that the means of the annual real rates of linear return are not very blurred. This is further evidence that the annual real rates of return are not independent, owing to mean reversion.

Therefore, when it comes to stock indices (and well-diversified stock portfolios), mean reversion may lessen mean blur. Although our statistical evidence is preliminary, it can be corroborated by performing the same elementary statistical tests on very long time series of other stock indices. Keep in mind that our study has the limitation of not dwelling on past structural breaks.

Owing to their clarity and simplicity, the above-mentioned elementary statistical tests may be useful both in teaching and business practice. Indeed, the late Fischer Black was keen on telling the truth and explaining quantitative models to traders in simple terms (Dunbar and Dunbar 2001).

Author Contributions: Conceptualization, L.B. and L.G.; Methodology, L.B. and L.G.; Writing—original draft, L.B. and L.G. 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: We are grateful to Elena Conti for her editorial support and to Laura Ballestra for her bibliographical assistance. We also thank two anonymous reviewers for their useful comments and suggestions. As usual, we are solely responsible for any remaining errors.

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

Appendix A

Let \( m = 12 \) so that \( \frac{1788}{12} = 149 \) subperiods are available. Compute each year a monthly mean, a monthly standard deviation, \( m = 12 \) monthly deviations from the mean, a partial (or running) sum of such deviations, a range, namely the difference between the positive maximum and the negative minimum of the running sum, and the ratio between range and standard deviation.

Next, obtain the rescaled range \( R/S_{12} \) by averaging 149 ratios between range and standard deviation.

Note

As remarked by an anonymous reviewer, the above-mentioned unit root tests may be misleading whenever data are affected by structural breaks. According to the Zivot-Andrews test, the null hypothesis of a unit root is rejected once more with a confidence level of 0.99 for real rates of logarithmic return. A structural break occurring in July 2000 has been detected, which is in line with Table 5 in Nguyen et al. (2022). The same authors summarize the relevant econometric literature.

References

Nguyen, James, Wei-Xuan Li, and Clara Chia-Sheng Chen. 2022. Mean Reversions in Major Developed Stock Markets: Recent Evidence from Unit Root, Spectral and Abnormal Return Studies. *Journal of Risk and Financial Management* 15: 162. [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.