Data Science in Finance and Economics

http://www.aimspress.com/journal/dsfe

DSFE, 2(3): 294–320.
DOI:10.3934/DSFE. 2022015
Received: 10 July 2022
Revised: 03 August 2022
Accepted: 09 August 2022
Published: 07 September 2022

Research article

Examining stylized facts and trends of FTSE/JSE TOP40: a parametric and Non-Parametric approach

Katleho Makatjane[1,*] **and Ntebogang Moroke**[2]

[1] Department of Statistics, University of Botswana, Private Bag UB 0022, Gaborone, Botswana

[2] Department of Statistics and Operations Research, North-West University, Private Bag X2046, Mmabatho, Republic of South Africa

* **Correspondence:** Email: makatjanek@ub.ac.bw; Tel: +267 355 2695.

Abstract: It is critical in risk and portfolio management to identify groups or classes of financial returns. Portfolio diversification is one of the first decisions made during the portfolio construction phase, and it entails allocating assets among various asset class groups to maximize the risk/reward trade-off. Therefore, this research provides a detailed examination of empirical analysis concerning the characterization of financial markets. In this study, we use parametric and non-parametric approaches to look at stylized facts and patterns of the FTSE/JSE Top40, which comprises the top 40 holdings companies in the South African financial market. To the best of our knowledge, this is the first time a model of this type has been used to create a map that characterizes this index. Our findings indicated that the majority of the properties of the data were valid including among others, clustering volatility, monthly seasonal effects and significant autocorrelation (or serial correlation) on logarithmic returns. Moreover, we found that intra-week trend effects exist, whereas the weekend effect has practically vanished in the FTSE/JSE Top40. With regard to the transition probabilities of the MS(2)-GJR-GARCH (1,1) model, the FTSE/JSE Top40 index had a 98.8% chance of exhibiting long memory, while the volatility had a 99.6% chance of exhibiting long memory.

Keywords: anomalies; Fat-Tail; markov-switching; seasonality; South Africa; volatility clustering

JEL Codes: C24, D53, E32, E44

1. Introduction

Financial markets experienced a paradigm shift due to high-frequency trading. High-frequency data is intra-day data for financial metrics collected on a relatively short time frame and delivered intermittently across time. This data can be used to characterize the micro-structure of the financial markets and to make informed real-time decisions. Low-frequency data, such as monthly, quarterly, or annual data, comes in a variety of patterns, while ultra-high-frequency data, such as intra-day, intra-hour, intra-minute, or intra-seconds data, has its own set of attributes and spectral characteristics (Aquilina et al. 2021). Emerging trends in high-frequency literature show that this data has significant promise for market analysts, brokers, scholars, regulatory agencies and shareholders in the fields of financial re-engineering, management of risks, volatility modeling and forecasting (Dufrenot and Matsuki 2021). Although statistical properties of stock prices, commodity prices and market indices have been studied for more than a half-century using data from various markets and instruments, the availability of large data sets of high-frequency price series, as well as the application of computer-intensive methods for analyzing their properties, has opened up new horizons for empirical finance research (Arratia and Lopez-Barrantes 2021). The investigation of these multiple datasets has resolved several long-standing debates about the quality of the data, and it has raised interesting questions. One of the most important issues is the ability to synthesize and meaningfully depict the information and attributes inherent in this massive volume of data. In the literature, empirical studies showed and classified a set of stylized facts that appear in different techniques, markets and time frames (Arora 2017).

Since all market transactions are sporadic, the most essential feature of high-frequency data is that it is time-spaced unevenly. In time series analysis, this is regarded as a homogeneous time series. In financial markets, adjustments in transactions are explicit and limited to a set of principles because different transactions have policies that limit price swings to ensure financial stability and functionality (Zumbach and Muller 2000). Sensible traders, on the other hand, rarely experience extreme price swings; hence, variations in transaction value only influence a limited set of data, and kurtosis is widespread (Restocchi et al. 2019). Against this backdrop, the purpose of this study is to model and identify the characteristics of financial time series. The paper also introduces the reader to novel statistical insights that are later used in empirical finance. In the main, we use non-parametric and parametric procedures that allow the data to speak for itself as much as possible. The proposed approaches, as advised by Porras (2017) and Nystrup et al. (2015), are based on qualitative and quantitative assumptions about the stochastic process that generates the data.

Theoretically, the non-parametric approach has the advantage of not requiring a model. This means that the method requires fewer assumptions about the data, and it consequently proves better in situations where the true distribution is unknown or cannot be easily approximated using a probability distribution. According to Nickl and Ray (2020), assumptions are made about a distribution: and this continuity and /or symmetry. The major advantages of non-parametric procedures compared to parametric methods are that (1) they can be applied to a large number of situations; (2) they can be more easily understood intuitively; (3) they can be used with smaller sample sizes; (4) they can be used with different data types; (5) they need fewer or less stringent assumptions about the nature of the population distribution; (6) they are generally more robust and not often seriously affected by extreme values; and (7) they have, in many cases, a high level of asymptotic relative efficiency compared to the classical parametric methods (Jiang 2022). Nevertheless, many critics of non-parametric procedures have pointed out some major drawbacks: (1) They are usually neither as powerful nor as efficient as the parametric procedures; (2) they are not as precise or as accurate in many cases (for instance, ranking tests with a large number of ties); (3) because of lack of precision, non-parametric procedures may lead to making type I or type II errors; (4) since they transform observed data into ranks and groups, utilization of data by these methods is inadequate; and (5) the sampling distribution and distribution tables for non-parametric analysis are too numerous and often cumbersome, which lead to limitation to small sample sizes. With a parametric approach, it is possible to add constraints in assumptions made, meaning that they cannot be altered by mistake later in the design process. In other words, constraints are a way of ensuring that any modifications made are done so with design intent in mind (Verma and Abdel-Salam 2019). The way that these constraints work is that they often have references that can work against each other. While this can be in a design and in a model, it can also prove to be restrictive. For instance, a change to a feature in the model can have a dramatic effect on the subsequent features. Sometimes, the unintended effects can even cause a model to fail.

We outline the stylized statistical truths that are similar to a wide range of financial time series. Despite the prevalence of recent contributions related to methodological issues in time series data literature, the primary contribution of this study is to depict the attributes of financial markets and to provide new evidence on their relationship with business cycles in the South African financial market. Additionally, the study investigates a few possible economic ruses that could explain similar empirical findings. We also answer the following three questions on stylized financial time series facts: (1) How do the basic characteristics of financial and business cycles differ? (2) How closely do the financial and business cycles resemble each other? (3) Which economic considerations could explain financial cycles' characteristics, as well as their differences from economic cycles? Another purpose of the study is to contrast Markov-switching generalized autoregressive conditional heteroscedasticity (MS-GARCH) models for identifying and classifying stylistic elements in stock returns with Hiebert et al. (2018), who employed the traditional version of the approach to comprehensively compare financial and business cycles. The proposed MS-GARCH family models can encapsulate the volatility of the mean and variance caused by structural changes, and it pacifies extreme GARCH projections throughout tumultuous time frames. This infused model also considers a real-time prediction system, which improves the accuracy of non-linear business and financial cycle prediction and classification. This makes it simply possible to spot changes at the moment, which is particularly useful when dealing with continuously changing financial conditions. When it comes to forecasting stock market indices, most analysts overlook this aspect.

The paper proceeds as follows. Section 1.1 presents a review of literature that is related to this study. Section 2 presents the study's data, materials and methods. The presentation comprises non-parametric and parametric approaches for detecting and classifying the stylized facts of South Africa's stock market. The scientific results are discussed in Section 3 , while Section 4 discuses the results and findings of the study. The conclusion and recommendations are presented in Section 4.

1.1. Literature review

High-quality high-frequency financial time series data is currently of interest by most researchers. In recent years, low-frequency data, such as annual, bi-annual, quarterly and monthly data, is no longer used by market analysts and researchers. Scholars and analysts have started focusing on high-frequency data, which incorporates weekly, intra-day, intra-hour and intra-minute data. The interest in high-frequency data is expanding as a result of new financial and economic development and progression of electronic innovation in financial markets. Market analysts need to utilize high-frequency (tick-by-tick) data to settle on more powerful trade methodologies and choices. However, Shakeel and Scrivastava (2021) analyzed high-frequency data to concentrate on different market micro-structures and modeling of real-time market dynamics. Da Cunha and Da Silva (2020) tested Bitcoin for a set of stylized facts in their study of relevant stylized facts about Bitcoin. These authors have established that Bitcoin performs statistically as most other assets. For illustration, it exhibits aggregation, Gaussianity and oscillation scaling. This further proves that Bitcoin obeys the laws of Omori and Gutenberg-Richter. Global stability, originally defined for spiral systems, produces power-law deportment with a profile similar to that seen in stock markets. On the other hand, Bariviera et al. (2017) tested for long memory in a return time series from 2011 to 2017 using transaction data from the Bitcoin platform and calculating the Hurst exponent with the detrended volatility analysis method. The sliding window was for measuring distance. The findings of these scholars showed that Hurst exponents fluctuated significantly in the early years of Bitcoin and tended to stabilize more recently. Furthermore, their multivariate analysis revealed a similar behavior of the Hurst exponent, indicating a self-similar process.

Kufenko and Geiger (2017), on the other hand, empirically investigated and theoretically reflected on the generality of stylized facts that are discussed in business cycle analysis. Business cycles together with three models for capturing the core macroeconomic relations were estimated using OECD data spanning the period 1960 to 2010 and validated by relevant statistical tests. The results showed that the basic coefficients of the relationship, such as the slope of the Phillips curve, are significantly associated with some of these variables, but the results are not uniform. In detailed theoretical discussions and interpretations, the observed differences between countries cast doubt on the universality of stylized facts, but these differences include the institutional agent variables used here. It is argued that this cannot be explained. Nevertheless, stylized facts on the asset pricing properties of cryptocurrencies, summary statistics on cryptocurrency return properties and measures of common variation for secondary market returns on 222 digital coins have been presented by Hu et al. (2019). The returns on the secondary market of all other currencies have indicated high correlations with returns on Bitcoin. Both the returns and volatilities that are documented by these authors were high. In addition, they demonstrated that cryptocurrencies, in aggregate, carry a common source of systematic risk correlated with Bitcoin returns. This has important implications for portfolio diversification and risk assessment. The development of an agent-based speculation game for higher reproducibility of financial stylized facts was done by Katahira et al. (2019). The authors proposed a speculation game agent-based model for better reproducibility of the stylized facts. With the discovered features, the developed model produced 10 out of the currently well-studied 11 stylized facts under a single parameter setting. Moreover, by combing the density of Balanced Flows with Bottlenecks Patterson et al. (2020) discovered the main stylized facts that are observed in financial systems together with their counterparts in a mechanical system. The experimental model of these scholars accurately reproduces financial properties such as the scaling of the price fluctuation volatility clustering, and multi-scaling.

Examining stability, disparity and cyclical, quasi-cyclical and predictable chaotic dynamics, Cavalli et al. (2017) used analytical and quantitative tools to pinpoint the extent of the interaction of accelerator parameters and financial interventions with the intact market. Power can influence the flow of economic and financial markets. Simulation studies by these authors have shown that the proposed model can identify many statistical properties and stylized facts observed in real financial markets. These include stable and high volatility fat-tailed return distribution, volatility clustering, and positive autocorrelation of absolute returns. Kim and Shin (2022) further explained how financial data is heavily tailed, arguing that log returns have finite moments to account for stylized facts of high-frequency financial data, such as volatility clustering, intra-day, U-shapes and leverage. The integrated daily volatility of the proposed volatility process implemented a generalized autoregressive conditional heteroscedasticity (GARCH) framework with an asymmetric effect on log returns and developed the Huber regression estimator to optimize asymptotic properties and adjusted losses.

2. Materials and methods

The data used in this study are five business day financial time series exchange/Johannesburg stock exchange (FTSE/JSE) Top40 indexes from 4 January 2010 to 30 June 2021. To avoid fluctuations in exchange rates, the index is kept in its original currency. The data were obtained from the South African Stock Exchange and was accessed on July 3, 2021. This index comprises the top 40 holdings companies in the South African financial market.

Assuming that the FTSE/JSE Top40 returns are represented as in Equation 1 (Bee and Trapin 2018), r_t is the depreciation percentage of closing stock price between $t-1$ and t. Thus, returns can be described as in Equation 2.

$$r_t = 100 \times \left(\frac{\ln(X_t) - \ln(X_{t-1})}{\ln(X_{t-1})} \right) \tag{1}$$

where X_t denotes the current value of the stock price at time t, and X_{t-1} presents the past value of a stock price at time $t-1$. Then, the returns in Equation 1 can be modeled by

$$r_t = \mu_t + \varepsilon_t. \tag{2}$$

In Equation 2, μ_t is a time-varying mean and ε_t is the measurement error (Boer et al. 2019), which can be written as in Equation 3.

$$\varepsilon_t = \nu_t \sigma_t, \tag{3}$$

where σ_t is the time-varying dynamics, and ν_t is a process of independently and identically distributed (i.i.d) residuals with mean zero and roman familyVar $(\nu_t) = \sigma_t^2$. This study focuses on the distribution of ν_t, particularly its tails. As mentioned in the introduction, the study uses both non-parametric and parametric methods to examine the behavior and features of these stylized facts.

2.1. Non-parametric approach to stylized facts

Descriptive statistics are used to display a picture of the overall nature of data. We use the mean and median to determine where the distribution's center is. Additionally, normal histograms with kernel densities and normal quantile-quantile (Q-Q) plots are used to detect data quality and observe whether the FTSE/JSE Top40 index follows a more asymmetric distribution than a normal distribution. According to Restocchi et al. (2019), if the distribution does not display a normal distribution, the resulting distribution is an extreme value which signify rare events that occur at time t, and this is a way for expressing skewness and kurtosis in terms of prescribed asymmetry and peakedness metrics. These non-parametric representations are used to demonstrate how the FTSE/JSE Top40 deviates from normality. Since extreme returns are caused by unexpected events or shocks, these methods can also be used to determine the return series' heteroscedasticity and volatility clustering. Extreme highs and lows are more prevalent than normal distribution norms. A tail index is used to describe the tail of a distribution. A normal distribution's tail index is one; however, tail indices for normally high-frequency data are two to four. If the marginal cumulative density function (MCDF) F has a probability density function (PDF), a simple histogram is a well-known estimator of this PDF. The form of a histogram is dependent on the number of locations in its cells, rendering it an inadequate density estimator (Washington et al. 2020). Given the sample X_i and the observation x, the kernel density estimator (KDE), on the other hand, is expected to offer a much better estimate than a histogram, and according to Duong (2022), KDE is represented in Equation 4.

$$\hat{F}(x) = \frac{1}{bn} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right). \tag{4}$$

Here, b is a bandwidth that determines the resolution of an estimator, and $K(\bullet)$ is the kernel function with the following restrictions: $\int K(u)\,du = 1, K(u) \geq 0$. Small values of b are used because they permit a high degree of variation as to detect satisfactory features of the true density (Ledl 2016). Basically, the KDE smooth each data point X_i into a small density bumps and then sum all these small bumps together to obtain the final density estimate. Ruppert and Matteson (2015) opined that suitable values of b rest on both the sample size n and the true density. Several authors, notably Li et al. (2020) and Silverman (2018), criticize histograms because they fail to convey information regarding the distribution center. Some core bars, according to these authors, contain far too much information. Furthermore, the heights of the bars that make up the outliers on the left and right of a histogram are quite low relative to the peaks of center bars, culminating in the extreme values on the left and right of a histogram not always appearing (Li et al. 2020). As a result of this limitation, researchers choose to undertake their studies using a kernel technique rather than a histogram. Despite its more refined appearance, the estimate suffers from the same shortcoming. The contributions to the graph from both ends of the distribution (known as tails) become overpowered by the core region of the distribution. In addition, when the sample size is large enough, such as $n \geq 30$, many estimators become uniformly distributed. The theorem of Bruce et al. (2020) is used in this study to check the presence of exceptional quantiles. Extreme quantiles are points that segregate the tail of a PDF into seamless intervals with comparable probabilities (Katz 2013). Washington et al. (2020) are of the opinion that irrespective of whether one is working with a statistical population or a sample, the computation of these distributions varies. Using the theorem of Bruce et al. (2020), we let our time series be a series of i.i.d with the cumulative density function F that is continuous and positive. For a large sample that is greater than thirty, any quantile q is approximately equal to the population quantile denoted by $F^{-1}(q)$, while σ^2 is the variance that is estimated as in Equation 5

$$\sigma^2 = \frac{q\,(q-1)}{n\left[f\left(F^{-1}\left(q\right)\right)\right]^2}. \tag{5}$$

In Equation 5, n is the sample size with an unknown density f at any given point x.

2.2. Parametric approach to stylized facts

A variation from an asymmetrical bell curve is referred to as skewness. This measure is used as the third moment of dispersion in this study to emphasize asymmetries in the FTSE/JSE Top40 empirical distribution. Symmetric distributions might depart from normality if the center is abnormally peaked or flat or if the center is heavy-tailed or light-tailed. These types of deviations can be detected using the kurtosis coefficient. As a result, kurtosis is defined as a distribution's fourth moment, and it is used in this study to classify the returns distribution as leptokurtic, mesokurtic, or platykurtic using the results of this statistic (Ruppert and Matteson 2015).

2.3. Test for highly correlated returns

Correlation analysis is used to assess the strength of the relationship between two or more correlated indexes and price changes in this study. Risk analysts strive to discover if the relative and squared returns are related in any way. Montgomery et al. (2015) proposed partial autocorrelation functions (PACF) and autocorrelation functions (ACF) to achieve this objective. The FTSE/JSE Top40 index is evaluated based on how close it resembles the preceding FTSE/JSE Top40 index. Table 1 summarizes the conceptual features of ACF and PACF for a stationary series. ACF and PACF patterns must concur with these hypothetical trends to provide accurate findings.

Table 1. Theoretical behavior of ACF and PACF.

	$Lag(P,0)_s$	$Lag(0,Q)_s$	$Lag(P,Q)_s$
ACF	Tails off at lags ks	Cuts off after lag Qs	Tails off at lags k
PACF	Cuts after lag Ps	Tails off at lag ks	Tails off at lag k

Source: Montgomery et al. (2015)

Wulff (2017) describe the ACF as in Equation 6

$$\rho_k = \frac{\text{roman familyCov } (X_t, X_{t-k})}{\text{roman familyVar } (X_t)}, \qquad k = 0, 1, 2, 3, \cdots \tag{6}$$

where roman familyCov (X_t, X_{t-k}) is the covariance between the present index and the lag index. This measure is used to test whether the returns series' residuals are zero, i.e., $\rho_k = 0$, and k is the number of lags used. X_t

denotes a time series at time t, and X_{t-k} represents a time series at $t-k$. As a result, Hart (2013) developed a Portmanteau statistic, which is shown in Equation 7

$$Q*(m) = n\sum_{k=1}^{m}\hat{\rho}_k^2. \tag{7}$$

Equation 7 is used to test the hypothesis under the assumption that the returns are i.i.d. sequences and that certain moment conditions are met. Silverman (2018), Wulff (2017) and Tsay (2015) demonstrated that Equation 7 is an asymptotically chi-squared random variable with m degrees of freedom, i.e, $\chi^2_{1-\alpha,m.}$ that boosts the power of a test statistic. We, therefore, reject the null hypothesis if the calculated probability value is less than the conventional probability value and conclude that the return series is highly correlated.

2.4. Markov-switching generalized autoregressive conditional heteroscedasticity

The generalized Autoregressive Conditional Heteroscedasticity models exhibit rapid shifts in data, resulting in non-stationary time-varying parameters and poor hazard estimations. Cavicchioli (2021) and Maaziz and Kharfouchi (2018) proposed the MS-GARCH as a mechanism for resolving this disagreement. The model parameters oscillate later due to the unconstrained discrete Markov process. According to Ardia et al. (2019), the MS-GARCH models can swiftly fine-tune to variations in the conditional volatility level, improving risk forecasting. Variable unconditional variance periods are permitted as ascertained because of a haphazard adjustment in the variability approach. These models, according to Alemohammad et al. (2020), can switch between a variety of regimes with different volatilities and simulate time series with shock variants. The structure of the model generates dynamic behavior in each regime, which reacts to various kinds of shocks in different ways. For example, in the high volatility regime, the model can respond in a variety of ways to very high shocks, whereas in the high volatility domain, the model may respond in a variety of ways to moderate low shocks. Cavicchioli (2021), on the other hand, presumed that these barriers, frequently called regime transitions, are characterized by short-term fluctuations whenever returns increase or decline rapidly and that they may account for asymmetries and deviations in stock return series for computations and classification. They can compute and detect asymmetries and deviations in market returns series. This allows all characteristics to fluctuate from varying volatility regimes, resulting in distinct parameters for each regime.

3. Results

Following the methods and procedures outlined in the previous section, this section provides and discusses the data analysis. We employ a pictorial view of data for non-parametric and parametric data. The FTSE/JSE Top40 index's time-series dynamics are explored. Tables and figures are used to report the results.

3.1. Non-parametric approach to stylized facts

The five-day returns on the FTSE/JSE Top40 are depicted in Figure 1. The series clearly shows periods of high and low volatility, as well as values that are exceptionally large and small. This indicates that the FTSE/JSE Top40 returns have seasonal variations and volatility clustering. These times of volatility clustering are attributed to the market panic caused by global financial crises, and this financial market has the most condensed stock return losses, as shown in Figure 1. Additionally, Figure 1 depicts a potential benefit when conditional heteroscedasticity is taken into account. Two main issues are emphasized here: the source of weight loss and the unpredictability of weight loss. According to the latter, irregular shocks in the real business economy have a greater impact on subsequent volatility, while the former asserts that these shocks are followed by downturn volatility rather than extreme volatility. Nevertheless, Wang et al. (2020) declared that these large losses are caused by a contractionary monetary policy that was implemented by the South African Reserve bank (SARB). In that context, Moema and Bonga-Bonga (2020) also pointed out four events that are linked to these losses which are:

1. The surge in a number of sub-prime mortgages as a response to high level of housing speculations and building up of the bubble.
2. The creation of new financial instruments which was risky, hard to assess, and shifted the accountability between agents.
3. The fall of real interest rate, combined with the Federal Reserves' expansionary monetary policy.
4. Global financial imbalances caused by COVID-19 pandemic.
   Firstly, it was assumed that the COVID-19 pandemic would be localized in China only. However, it later spread across the world through the movement of people. The economic pain became severe as people were asked to stay at home. The effect of the pandemic was further felt in various sectors of the economy with travel bans affecting the aviation industry, sporting event cancellations that affected the sports industry and the prohibition of mass gatherings that affected events and entertainment industries (Ozili and Arun 2020). In 2020, the macroeconomic slowdown led to an increase in non-performing loans within the banking sector by 250 basis points. Private sector banks had the very best exposure to credit risk during this outbreak. According to Financial Times (2022), non-performing loans arose from the ones issued to small and medium-scale enterprises (SMEs), airlines, hotels, tour operators, restaurants, retail, construction and assets businesses. During the COVID-19 pandemic, there was a general decline in the volume of bank transactions, a decline in card payments and a fall in the use of ATM cash machines worldwide. This led to fewer fees collected by banks, which negatively affected banks' profits, and other South African businesses were also affected. Some businesses witnessed very low patronage by consumers, resulting in a loss of revenue and profits, which negatively affected the equity investment of venture capitalists that funded existing and new South African firms. This made many venture capitalists begin to hoard new equity, which led to the dehydration of financing for a few South African businesses. On the opposite hand, the lockdowns thanks to the coronavirus outbreak resulted in higher demand for a few forms of online services like online shopping, hence the condensed volatility clustering from the year 2020, as shown in Figure 1. Although the oil price war, in which Russia and Saudi Arabia were driving down oil prices by increasing oil production, played a role in the fall in stock market indices, the subsequent fall in stock market indices in March was mainly due to investors' flight to safety during the coronavirus pandemic (Ozili and Arun 2020).

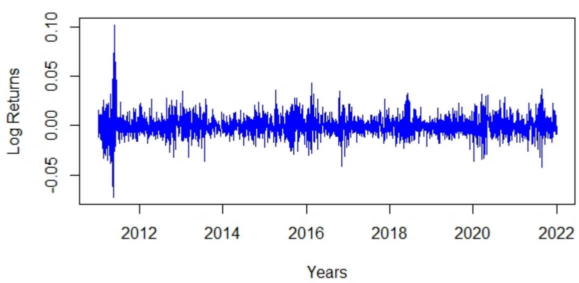

Figure 1. Returns on FTSE/JSE Top40 index.

3.1.1. Assessing dependence in FTSE/JSE Top40 returns

Although the autocorrelation (also known as serial correlation, serial reliance, mean aversion or mean return) underlying price changes (hence logarithmic returns) is insignificant, as shown in Figure 2, there are a few subtle but noteworthy anomalies. Given the temporal accuracy of the weak form of the market efficiency, we discovered significant but not zero outcomes in FTSE/JSE Top40 logarithmic returns; these are not unusual in all other financial market returns. According to Dias et al. (2015), the autocorrelation function of log-returns is typically positive, yet we found the same positive ACF results as displayed in Figure 2. Although the theoretical ACF is useful for characterizing the features of certainly expected dynamics of a time series, most parametric research requires high sample data. This is a concern from a traditional statistical standpoint since we may not always have i.i.d series with which to calculate correlation functions, hence the need for some sort of transformation to make the returns i.i.d. However, autocorrelations grow negative for 2-year returns, reach a minimum for 3-5-year returns and then return to zero for longer periods (Fama and French 1988). This is what we have found in this study making the ACF function to be significant. According to Liu et al. (2020), intra-seconds, hourly, intra-day and daily stock returns have a weakly negative autocorrelation, which is the case in this study, where the returns on the FTSE/JSE Top40 index are weakly negative.

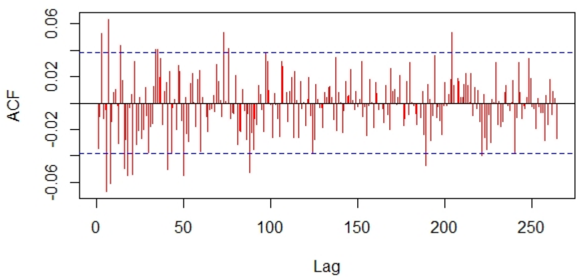

Figure 2. ACF plots for FTSE/JSE Top40 returns.

This appears to be a contradictory finding, but Liu et al. (2020) claims that it can only signify that there are high positive cross-autocorrelations throughout particular returns over time. Returns on the FTSE/JSE Top40 index are unclear and have insignificant autocorrelations as per the results of ACF, and this implies that the Top40 index analyzed in this study is chronically volatile during the sampling time and has a negative relevance. If the previous return value was below average, the next result is more likely to be higher, which is the case found in this study. Emerging economies, such as the South African economy, are known for their high volatility since the underlying economy undergoes a dramatic shift and regulatory experimentation. Their interest rates serve as a conduit for fiscal and monetary policies to reach the financial markets. During a period of low interest rates, a surge in stock market investments will occur, and vice versa. Significant autocorrelations also indicate erratic and volatile results.

3.1.2. Autocorrelation in absolute and squared returns

The autocorrelation of absolute and squared returns, in contrast to the lack of dependency in returns, always seems to be positive yet unimportant. The function decays gradually as a result of variations in time lag following an exponentiated power of $\beta \in [0 : 2; 0 : 4]$. The lack of dependence can be misinterpreted as a symptom of long-term reliance. Similarly, Figure 3 illustrates that relative returns have a larger autocorrelation than equivalent squared returns. The absolute squared returns have autocorrelation that decays slowly, and this is evidenced in this figure. The implication here is that the FTSE/JSE Top40 index is non-stationary, indicating high variation or volatility clustering.

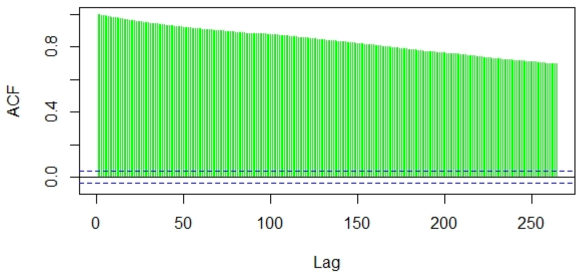

Figure 3. ACF of absolute squared returns.

3.1.3. Intra-Month U-shaped pattern

The other property that we discover in the FTSE/JSE Top40 index is the U-shape pattern. Phylaktis and Manalis (2013) opined that during a trading year, bi-annual, month, quarter, week, day, hour, minute and second, the U-shaped pattern is seen in a myriad of traded assets, from stocks to commodities. Figure 4 presents a one-month U shape pattern of FTSE/JSE Top40 returns. Each dot on this visualization denotes the day of a month from 25 March to 23 April. It is worth revealing that from the 30^{th} of March to the 3^{rd} of April, the U shape pattern is visible. The same features are also seen between the 16^{th} of April and 19^{th} of April. Moreover, from the 4^{th} of April to the 6^{th}, there are sharper U shape characteristics, which are followed by the ones between the 9^{th} to the 11^{th}. These features will be seen each year due to the significance of cyclical components in the studied time series, where in our case it is FTSE/JSE Top40 returns.

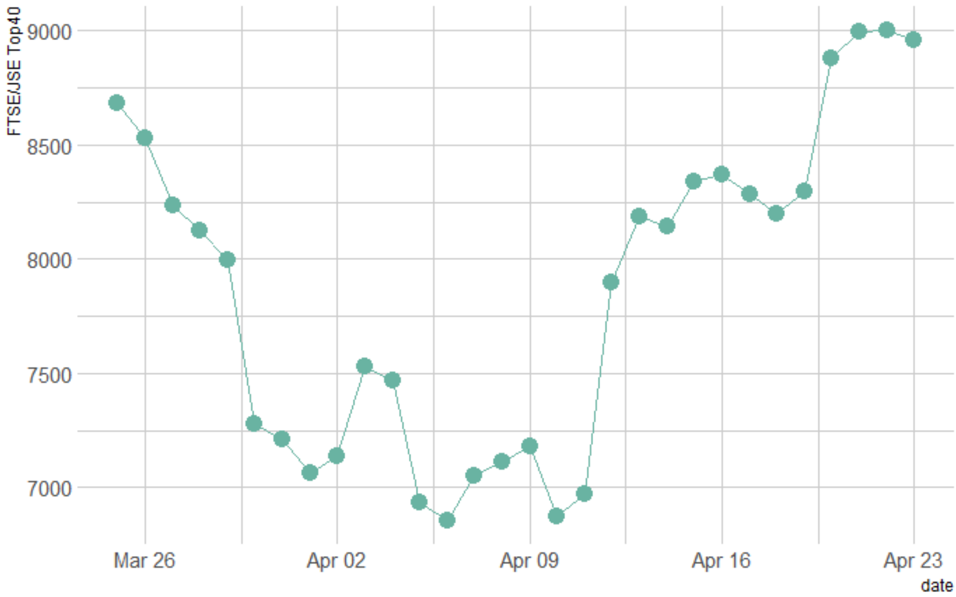

Figure 4. U-shape one month pattern.

The economic implication of these characteristics is that both business and financial cycles are the same, because, at the top of this U shape, is the beginning of a new day, a new business day which represents a new financial cycle of a trading day or stock price in that particular day and period. Here, business and financial cycles have identical co-movement. Without the opening of the business day, there is no opening of a financial day. The same can be interpreted from the bottom of the U shape characteristic, where it represents the closing of the trading business, and the stock prices begin to increase as there is no activity in the stock market from that period till morning. This rise is due to the demand and supply theory, where at the closing period, more people are relaxed from their long day at work, and they want to buy a stock than sell it hence the rise of the stock prices and vice versa. The stock market at the beginning of the day has high opening stock prices, and during the day it gradually declines, hence the presence of a downward trend in Figure 4. As spending occurs, stocks of physical and financial assets are altered, with each sector adjusting existing positions toward long-run equilibrium targets through the interrelated real business economy and financial flows positions toward long-run equilibrium targets through interrelated real and financial flows. The flows are inherently intertwined, producing side-by-side fluctuations in real and financial variables. Before the market closes, all traders and investors make appropriate changes based on information garnered during the trading session, causing the market to become extremely volatile. Finally, towards the closing of a business day, the stock price gradually increases; therefore, an upward trend is seen after closing stock trade or businesses. However, Olbrys and Oleszczak (2020) declared there are other patterns in high-frequency data, such as trading volume, transaction costs, order flows, depths, spreads, price returns and stock market resiliency. These patterns are also correlated with business cycles and time.

3.1.4. Distribution of FTSE/JSE Top40

The returns on financial markets have a leptokurtic distribution. Catastrophic events are more likely to take place when compared to the average distribution. The power-law or Pareto-like tail is presented with a definite tailed index of greater than two or less than five for the majority of analyzed time series (Liu et al. 2020). This rules out steady conditions with unbounded variance and the normal probability plot in particular. On the other hand, the precise shape of the tails is difficult to determine. The normal or Gaussian distribution is extensively utilized in financial analysis for risk simulation and evaluation of performance. The central limit theorem is used for estimating the normality of the return distribution. Events that deviate from the mean by more than five standard deviations are relatively rare in a normal distribution.

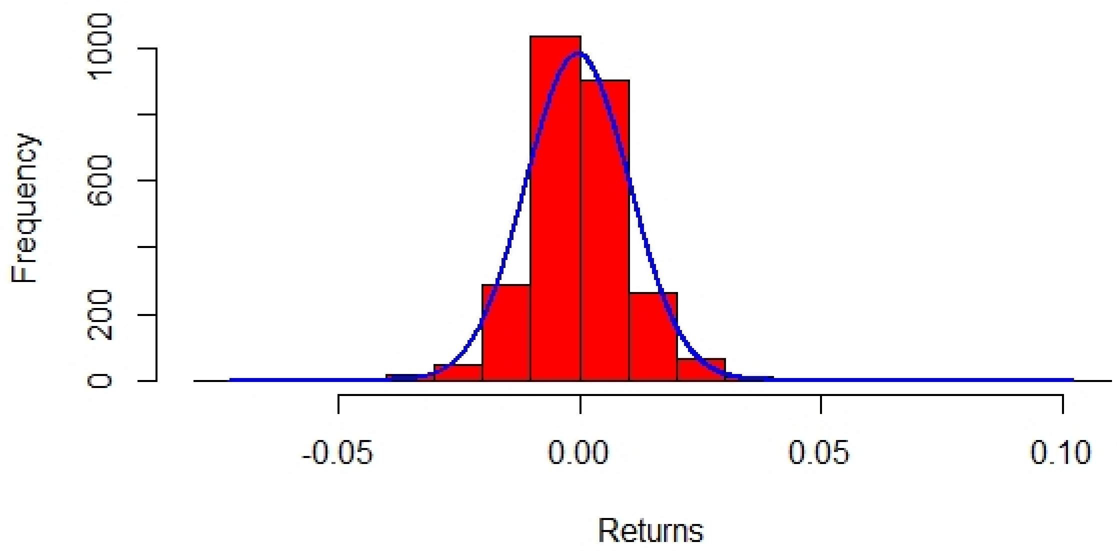

Figure 5. Normal densities.

Financial returns are non-stationary by their very nature. This means that the asymptotic distributions (PDF) of returns have symmetry and have substantial outliers as opposed to the normal distribution. With increased data frequency, the distributions become fatter-tailed (smaller interval sizes). Investment returns are around average (Liu et al. 2020). We see heavy tails in the FTSE/JSE Top40 index, which is followed by extremely large and small tails of a normal histogram with kernel densities in Figure 5 and a Q-Q plot in Figure 6. Although there is a major deviation from linearity, Table 3 shows that kurtosis is more than three, with skewness below zero. As a result, the distribution of the FTSE/JSE Top40 index lacks stationarity, as evidenced in Figure 1, and is asymmetric, with growing progressive kurtosis and a Pareto-like tail as the time interval shrinks. Fat-tails are a good approximation based on our findings, and they can be calculated using Hill estimators. The Pareto tails of the FTSE/JSE-Top 40 are plotted in Figure 7.

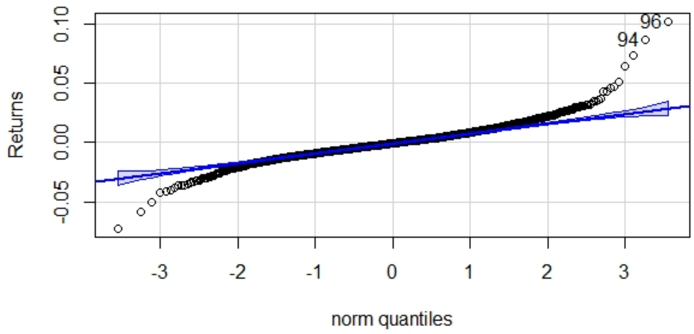

Figure 6. Quantile-quantile plots for FTSE/JSE Top40 returns.

According to Korkpoe and Junior (2018), price-influencing events that occur daily are the source of the deviation from normality. However, the chances of these events being covered in media are rising. The distribution can be explained by taking into account the media's tendency to enhance and embellish. When the normal distribution is augmented by a distribution dependent on the probability, longevity or influence of media reports, the result is a considerably fatter-tailed distribution than a Gaussian. Therefore, the financial sector had to act as an amplifier and be the source of these shocks that triggered business cycle fluctuations. At this moment, the household balance sheets, firms and banks give rise to various pro-cyclical mechanisms (such as the financial accelerator). In this manner, demand shocks are amplified through corresponding changes in the value of the collateral (such as residential or commercial property) and the real value of nominally fixed debt. Hence, Ha et al. (2020) disclosed that this leads to credit and asset price-driven cyclical fluctuations which are expected to yield higher peaks and lower troughs than normal business cycles, possibly with more prolonged periods of boom and bust.

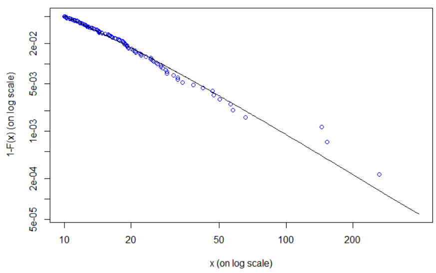

Figure 7. Pareto tails of FTSE/JSE-Top 40 returns.

3.1.5. Monthly seasonal effects

Seasonality or robust recurring trends are common among high-frequency data. Volatility, trading frequency and turnover reflect seasonality in various markets. A U-shaped pattern can be seen for the majority of exchanged commodities. In intra-day and intra-week data, OTC (Over-the-Counter) commodities such as foreign exchange show significant seasonality (Schmid 2009). Seasonal effects, also referred to as calendar effects, are recessionary anomalies in returns that follow a calendar pattern. The most significant calendar irregularities are (or were) the January impact and weekend effects (Strohsal 2019). This section includes information about annual anomalies. It should be noted that when data mining the calendar effects, repeating the same data set to construct and test hypotheses induces data mining prejudices that, if not taken into consideration, imperil the assumptions that support conventional statistical analysis (Sullivan et al. 2001). The importance of calendar market rules, according to the authors, is substantially weaker when seen from the perspective of a universe of rules, which might have been reviewed. They are correct in emphasizing the hazards of data mining, but they fail to acknowledge that traditional statistical inference has already been intimidated. A better perspective is to consider that an astounding finding necessitates further research, as Bayes' theorem clearly shows and opined by (Strohsal 2019). Understandably, a planetary effect would require more proof than a fiscal sales effect. Since their discovery, numerous calendar influences have either declined, vanished entirely or even reverted. Figure 8 highlights the influence of the 12 months' seasonal effects on FTSE/JSE Top40 stock returns. Monthly seasonal effects are detected. It can be seen from Figure 10 that there are some up and down seasonal trends for each year with the year 2016 having the highest seasonal trend, above 100 units from January to April and starting to decline thereafter. The same trend is seen in 2014 where the highest seasonal effect is seen above 100 units in July.

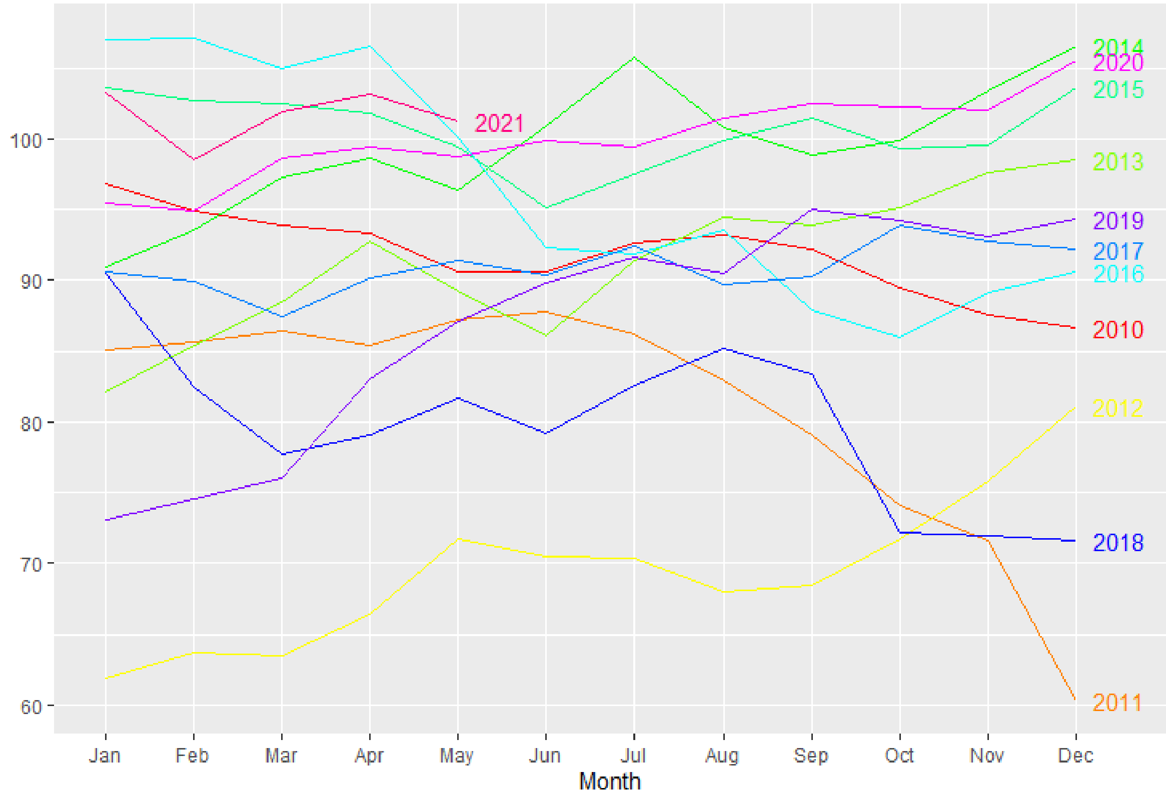

Figure 8. Monthly seasonal effect.

In general, the monthly seasonal effects are low and according to Statistics South Africa (2022) this began in 2014 when South Africa's economy grew by 1.5%, which marks down from 2.2% in 2013, according to preliminary estimates of real gross domestic product (GDP) released by Stats SA. Eight of the ten industry groups experienced some growth during the year, while two industries shrank in size. This can also be seen in Figure 10 that, after the year 2014, the economy was detrending. Economic activity within the mining and electricity industries decreased by 1.6% and 0.9% respectively, while manufacturing showed very little change. The mining industry was interrupted by widespread strikes during the first half of 2014, resulting in a decline in mining activity in the first quarter (with a decline of -22.8%) and second quarter (with a -3.0% decline). This was followed by positive growth during the second half of the year, with mining expanding by 3.9% in the third quarter and 15.2% in the fourth quarter (Statistics South Africa 2022). Moreover, the results of the estimated monthly seasonal effects are reported in Table 2 where all the seasonal moths are significant confirming the results presented in Figure 10.

Table 2. Monthly seasonal estimates.

Months	Estimate	Std.Error	t-value	p-value	Months	Estimate	Std.Error	t-value	p-value
Jan	90.032	3.256	27.65	0.002	Jul	91.054	3.400	26.78	0.002
Feb	89.432	3.256	27.47	0.002	Aug	90.857	3.400	26.72	0.002
Mar	89.889	3.256	27.61	0.002	Sep	90.269	3.400	26.55	0.002
Apr	91.642	3.256	28.15	0.002	Oct	88.924	3.400	26.15	0.002
May	91.220	3.256	28.02	0.002	Nov	89.499	3.400	26.32	0.002
Jun	89.319	3.400	26.27	0.002	Dec	90.083	3.400	26.49	0.002

3.1.6. Intra-Weekly trend pattern analysis

Trends are also evident in high-frequency data during the week, month, quarter etc. On weekends and holidays, almost all exchanges are closed. Therefore, no trading takes place. The degree of activity on weekdays is significantly high. Across FTSE/JSE-Top40 returns, a day-of-the-week impact is noticed in Figure 10. Generally, market activity is lowest on Monday and highest on the last two working days of the week. To be precise, trading activity starts growing progressively from Monday to Friday. On weekends and holidays, there are no trading activities, and Figure 9 shows all the holidays and weekends. The highlighted days are the holidays that are available in South Africa during the year 2021, and on these days, no trading activities happened, leading to low opening stock prices on Monday. In practical analysis, the impact of each day in a week should always be considered. However, the weekly seasonality in different markets may differ. In a case where a week has one or two holidays, the situation becomes much more convoluted, especially for those sectors that are highly linked and have distinct holidays in various countries. This is the contrast with this study and it is evidenced in Figure 9. The weekend effect (also called the Monday effect, the day-of-the-week impact, or the Monday seasonal) describes how stocks tend to perform better on Fridays than on Mondays (Strohsal 2019).

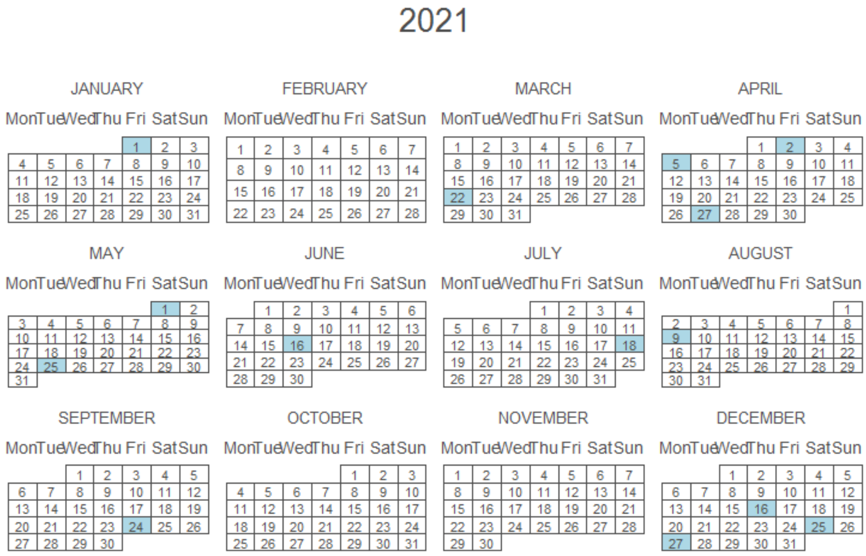

Figure 9. South African calendar holidays in 2021.

It is a surprising phenomenon since returns from Monday cover three days, and one would anticipate returns for Monday to be better than other days of the week due to the longer length and heavier risk (Zhai et al. 2020). Figure 10 depicts the significance of the weekly trend, from the first week of January 2020 to the last week of January 2020. Because the first of January 2020 was on Wednesday, and there was no trading activity, we, therefore, begin the full week on the 06^{th} as depicted in Figure 10. Note that the weekends and holidays are not included, as there is no stock market trading on these days. The FTSE/JSE Top 40 stock is de-trending as of the first week of January 2020 and shows a significant downward trend pattern on week 2, leading to the closing stock prices value of 27500 and slightly increasing to 27800 at the beginning of the third week of January 2020. It is worth noting that Monday returns should be three times the expected returns of the other days of the week, but, under the trading time hypothesis, the expected returns for each day of the week are the same; hence the high records at each beginning of the week, and this is also evidenced in Figure 4 and Figure 8, respectively. According to Atsin and Ocran (2015) and Fama and French (1988), the average returns for the other four days of the week are positive, while the average returns for Monday are noticeably negative, hence the significant negative autocorrelation of the log returns as depicted in Figure 2.

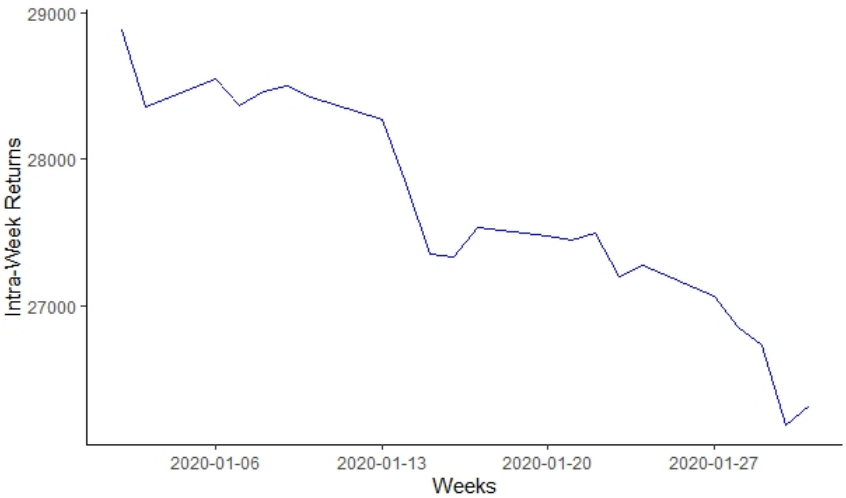

Figure 10. Intra–Weekly trend analysis.

3.2. Parametric approach to stylized facts

As proposed by Cox (2017), we use central tendency and location metrics and normality tests to elucidate the observed features of the returns. The assumption of a normal distribution underpins many statistical procedures. According to Hu and Plonsky (2021) and Garson (2012), non-normally distributed data is corrected using a variety of transformations.

3.2.1. Descriptive statistics

The results in Table 3 reveal that the unconditional standard deviation of the FTSE/JSE Top40 index is 20%. This implies that the index has a fairly insignificant variation in volatility. The implication is that the size of the average of the returns is not fairly large. As evidenced in Table 3, the reported Kurtosis is greater than three, indicating that the FTSE/JSE Top40 index is leptokurtic, with instances that do not even fit into a typical distribution. The presence of negative skewness is another feature of the returns series (Xaba et al. 2021). According to Cheng et al. (2018), all of the Jarque-Bera (J-B) and Shapiro-Wilk (S-W) tests do not confirm the normality null hypothesis, confirming that the index is asymmetric. Significant negative swings in stock prices and market indexes are common, whereas large positive swings are uncommon. This distinguishing feature does not apply to exchange rates because up and down transitions have more symmetry. This means that the empirical returns distribution is slanted to the left, with negative returns occurring more frequently than positive returns. It is also clear that the five banking stock indices have a positive mean. The deduction here is that the overall closing stock price of FTSE/JSE-TOP40 is increasing during the period in question. The magnitude of the average return was very small compared to the standard deviation. Beytell (2016) in his study reported the same results. All in all, one can infer that the mean of returns is somehow smaller than the standard deviation of the returns.

Table 3. Descriptive statistics for returns series.

	Mean	Median	Skewness	Std.Dev	Kurtosis	J-B test	S-W test
FTSE/JSE Top40	0.005	0.061	-0.536701	0.2015	11.26593	7735.4(0.0002)	0.936(0.0002)

Note: The S-W and J-B probabilities are represented by the statistics in parenthesis.

3.2.2. High correlated returns

The ACF is used as a tool for planning. Table 4 depicts the connection between observations of stock returns at various time frames. As a follow-up, the McLeod-Li test is used to validate the condition of autocorrelation (Xaba et al. 2021). The maximum propagation delay chosen with various degrees of freedom is 5%. Table 3 clearly shows that the returns series are volatile at a 5% significance level. As a result, the FSE/JSE Top40 returns are highly correlated from 4 January 2010 to 30 June 2021.

Table 4. McLeod-Li test correlated stock returns.

FTSE/JSE Top40					
Lags	1	2	3	4	5
Statistic	3.1855	3.4819	10.847	11.21	11.28
p-value	0.02431	0.07429	0.1754	0.01258	0.04609

3.2.3. Discreteness asymmetry and high kurtosis

High-frequency data is, by definition, discrete. Price changes have implications on only a few values. The high degree of discretion in transaction prices hampered this study on volatility and reliance. For example, it is partly to blame for the negative first-order autocorrelation of returns. The discreteness also causes a significant amount of kurtosis. The current study follows in the footsteps of Atsin and Ocran (2015), who used the MS-AR model for the parametric test of discreteness and kurtosis. In this study, however, the extended GARCH models with two regimes are used (Jakata and Chikobvu 2022); and four different types of MS-GARCH models namely, the MS-GARCH, MS-TGARCH MS-EGARCH, and MS-GJR-GARCH are estimated using distorted t-distribution error to capture the anticipated fat-tail behavior of the data. The models' parameters are calculated using the Markov-chain Monte-Carlo (MCMC) method containing 1500L MCMC replicates. To come up with the best model that describes the discreteness properties and best mimics FTSE/JSE-Top40 returns to produce fewer forecasts, a prediction comparative analysis exercise is done. We, therefore, use Akaike information criteria (AIC), Schwarz Bayesian criteria (SBC), mean absolute error (MAE), log-likelihood (LL) and mean absolute percentage error (MAPE), which are defined as statistical loss functions. The well-known information criteria that are usually used for model selection remain as the criterion used for the prediction ability of the estimated models, that is, they are used here to select the model that best predicts the discreteness, asymmetry and high kurtosis in FTSE/JSE-Top40 returns. Meanwhile, error metrics are used for forecasting performance. For instance, the log-likelihood selected the best model as MS-TGARCH. Looking at AIC, the best-performing model is selected as MS-GJR-GARCH, while MAE selected the MS-EGARCH model as the best model. Therefore, Raihan (2017) in his study, ranked the models according to their statistical loss functions to overcome contradicting results. The author used 10 statistical loss functions. Following this author, we only used five statistical loss functions, and the models are ranked from 1 to 4. On that note, rank 1 denotes the best model while 4 denoted the poorest model. Table 5 gives much evidence that the MS-GARCH (1,1) model has the poorest performance, as the frequency of rank 4 is higher than the other three models by looking at their statistical loss functions. Nonetheless, the MS-GJR-GARCH (1,1) model outperformed all the models, as the frequency of rank 1 is higher than the other three models as it recorded 3 out of 5 statistical loss functions.

Table 5. Comparative analysis.

Test	MS(2)-GARCH(1,1)	MS(2)-TGARCH(1,1)	MS(2)-EGARCH(1,1)	MS(2)-GJR-GARCH(1,1)
LL	108	110	102	95
Rank	2	1	3	4
AIC	-1710	-1712	-1719	-1760
Rank	4	3	2	1
SBC	-1794	-1799	-1802	-1807
Rank	4	3	2	1
MAE	0.9023	0.988	0.8774	0.7866
Rank	2	3	1	4
MPAE	1.247	1.004	1.002	0.986
Rank	4	3	2	1

Finally, we, therefore, proceed with the estimated MS(2)-GJR-GARCH(1,1). All of the model's parameters are significant, indicating that the latent Markov chains (MC) used to assess the data's unknown discreteness worked properly. Xaba et al. (2021) used a Bayesian approach to estimate discreteness, and the authors discovered that as a result, their Markov chains are more robust. Furthermore, multiple measures of volatility exhibit better serial correlation across several days, implying that high-volatility incidences are concentrated throughout the business period, indicating that there is a direct positive correlation between financial and business cycles. The results of the estimated MS(2)-GJR-GARCH(1,1) are reported in Table 6, where the estimate of a time-invariant mean parameter is denoted by μ and is statistically significant. The ARCH α and GARCH estimates β are significant, indicating the presence of conditional heteroscedasticity characteristics in each regime and volatility of the previous period is strongly correlated with the fluctuations of the later period, i.e. the volatility clustering. Because both α and β are not close to the unit, this shows that there is no strong persistence of volatility shocks in the FTSE/JSE-Top4 returns (Sigaukea et al. 2014). Finally, the gamma estimate γ is statistically significant, meaning that the effect of negative returns' shocks on the conditional variance is higher in FTSE/JSE Top40 returns. Moreover, the value of $1 - \alpha - \beta - 0.5\xi$ is greater than zero, in both regimes, it is concluded that the estimated MS(2)-GJR-GARCH (1,1) model is stationary for FTSE/JSE-Top40 returns. The leverage ξ of FTSE/JSE-Top40 returns is positive for both regimes and regime one has high leverage than regime two. As FTSE/JSE-Top40 rose by more than 15 times in less than a month at the end of 2020 we speculate that the difference in leverage effects in both regimes may be related to fluctuations in the rise and fall of themselves and also related to the high record COVID-19 cases in South Africa during this period of the year 2020.

Nonetheless, the transition probability matrix suggests that the probability of the returns on the Top40 index in the low regime is higher than that of the high regime, with about 2.14%. This implicitly articulates that when the returns on Top40 index are in regime 0 (Low regime), the probability that this series switches to the upper regime denoted as regime 1 is $\Pr(S_t = 1 \mid S_{t-1} = 0) = 0.0064$, which is lower than that of regime 1. The average duration of each regime also supports these findings. Based on the expected duration, the non-turbulent period, which is regime 0, has 20.40 days (approximately 20 days) while the turbulent period has 93.55 (which is approximately 94 days). This denotes that the South African financial market will be in the low returns state for an average of 20 years, giving high losses, as indicated by the upper regime. We, therefore, conclude that there is a significant regime shift in FTSE/JSE-Top40, and it can be shown using the filtered and smoothed probabilities in Figure 11. It is worth noting that Table 6 along with Figure 11 suggests that the statistical characterization of the South African financial cycle afforded by the MS(2)-GJR-GARCH(1,1) model is inadequate. Our findings are in line Bilgili et al. (2020) but in contrast to those of **?**, suggesting that the business-cycle interpretation of the model becomes rather financial because the estimation period includes the end of financial crises.

Table 6. Estimated MS(2)-GJR-GARCH(1,1).

	Regime 1				Regime 2			
Coefficient	Estimate	Std Error	t-value	p-value	Estimate	Std Error	t-value	p-value
μ	0.1575	0.0069	22.826	0.000	0.1518	0.0070	21.686	0.000
γ	0.0430	0.0014	30.714	0.000	0.8052	0.0040	20.130	0.000
α	0.1356	0.0020	67.8	0.000	0.044	0.0030	14.666	0.000
β	0.8050	0.5701	1.412	0.000	0.9536	0.5654	1.687	0.000
ν	1.7229	0.8808	1.956	0.0001	5.6399	0.6491	8.688	0.001
ξ	0.8574	0.0822	1.043	0.002	0.0908	0.0401	2.264	0.002
Transition Matrix								
	P_{11}		0.9936			P_{12}		0.0064
	P_{21}		0.0278			P_{22}		0.9722

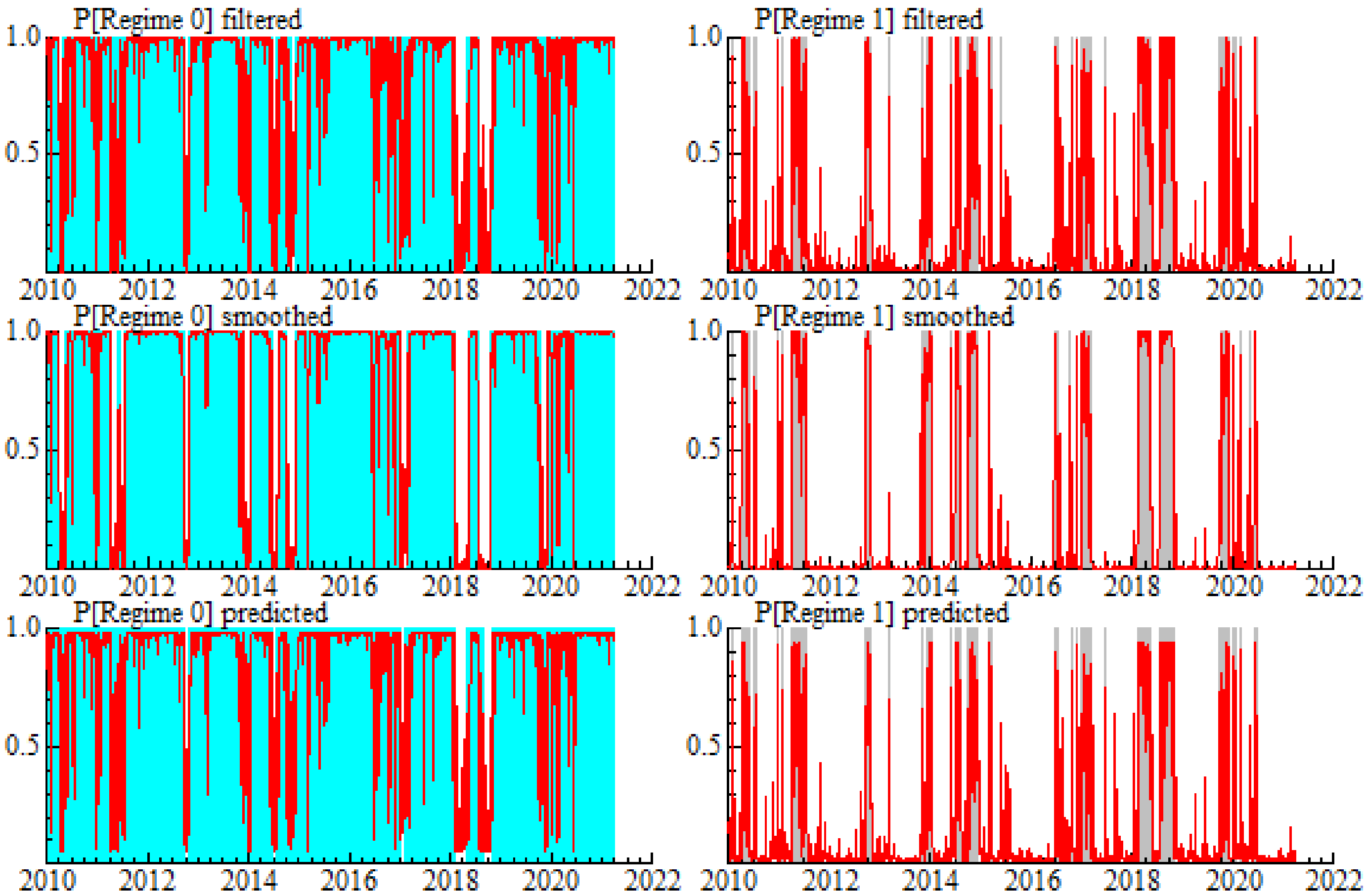

Figure 11. Filtered, smoothed and predicted regime probabilities.

Lastly, a weekly business and financial cycle response rates and behavior are reported in Figure 12. It is seen that there is a difference between business and financial cycles. To begin with, business cycles outnumber finance cycles. Moreover, they are longer, and finally, they are more symmetrical, with much longer recessions making both financial and business cycles not precisely synced, owing to symmetry discrepancies. Their movement is not in the same direction, indicating some sort of discrete inverse relationship that can be seen with the jumps in Figure 12. It is interesting to note that business cycles are slightly continuous, while financial ones are discrete, as denoted in Figure 12. The discreteness of the financial cycles is caused by the closing of business days, while the business markets are continuing to operate even after the financial market has closed. Furthermore, on weekends and holidays, stock markets are closed while businesses are sometimes operational. The macro-economic sector, specifically the monetary sector, is also operational, hence the continuous discovery. The implementation of various macro-economic interventions by the monetary policy committee (MPC) distorts the financial markets. The high-interest rates, repo rates and exchange rates impose reluctance of financial risk management expects and risk analysts to critically analyze the stock market in real time making financial markets developed be distorted (i.e. discrete) cycles.

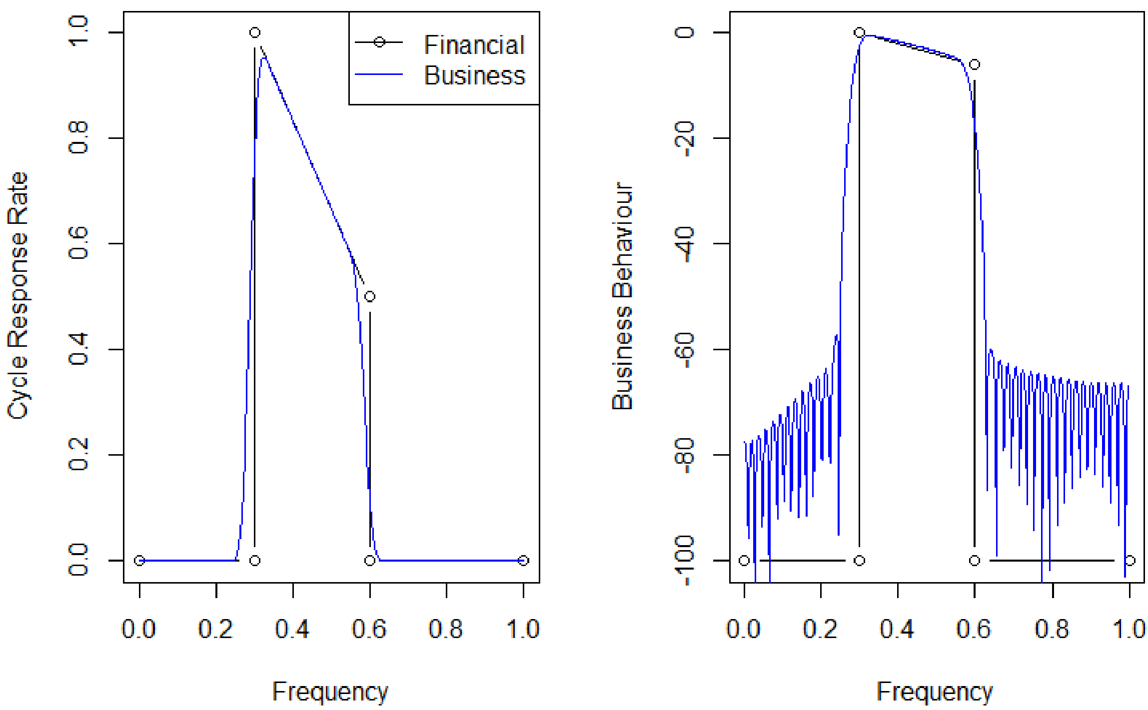

Figure 12. Business and financial cycle.

4. Discussion and conclusions

It is critical in risk and portfolio management to identify groups or classes of financial returns. Portfolio diversification is one of the first decisions made during the portfolio construction phase, and it entails allocating assets among various asset class groups to maximize the risk/reward trade-off. The statistical methodology used in this study is a useful tool for identifying groups of comparable financial assets. The study of stylized facts and patterns of financial time exchange/Johannesburg stock exchange closing Top40 indexes is a statistical extension of this work. The results demonstrated both non-parametric and parametric techniques for reliably achieving our goal. Graphical representations are used in non-parametric analysis to display all aspects of the series for the studied data from the FTSE/JSE Top40 index. Only a few studies have attempted to analyze stylized facts and patterns in financial time series in South Africa. To the best of our knowledge, this is the first time a model of this type has been used to create a map that characterizes the FTSE/JSE Top40 index. Volatility clustering, loss asymmetry, heavy tails, slow autocorrelation decay in absolute returns, monthly seasonality, a U-shape one-month pattern, discreteness and kurtosis, and an intra-weekly trend pattern are discovered as anomalies. This study finds that business and financial cycles vary in three different ways. To begin with, business cycles outnumber finance cycles. Moreover, they are longer, and finally, they are more symmetrical, with much longer recessions. This study reveals that financial and business cycles are not precisely synced, owing to symmetry discrepancies. Finally, as the study by (Hiebert et al. 2018) found, the length of financial cycle downturns is an important feature of divergence. Based on monthly data, Xaba et al. (2019) discovered that the financial cycles of the BRICS stock markets are more symmetrical as opposed to their business cycles. In contrast, we discovered April, is the most profitable month as it had the most price appreciation as indicated by the seasonal estimate in Table 5. The findings contradict the January effect theory's hypothesis, implying that there was no January effect on the JSE, invalidating Jooste's argument (Jooste 2006). The current study, however, observed the same results as Atsin and Ocran (2015). According to the analysis, the distribution of the Top40 index is significantly skewed. The one-month U-shape pattern is also observed and proved to be extremely important, because it provides the trend and significant seasonality patterns. According to our findings, autocorrelation in returns is primarily significant and unfavorable for this high-frequency data. This is because intra-day effects persist, the weekend effect is nearly gone, the month effects exist, the January impact does not exist, and holiday effects are also present, as shown in Figure 9.

Financial risk encompasses a wide range of threats related to finance, including financial transactions that involve corporate credit on the verge of default. Risk analysts can use the approaches presented to quantify their financial losses and assess the risk associated with those losses. When forecasting loss risk using loss distributions, associated risk factors should be considered. Insurance companies can also use loss distributions to guide their risk management strategy concerning their loss on investment and loss return levels. The estimated MS(2)-GJR-GARCH(1,1) model in this study can be used by stock markets to forecast the overall investment risk. The discovered characteristics can be used to forecast daily operational risk factors for the South African financial sector Jakata and Chikobvu (2022) and Xaba et al. (2021). Other financial institutions may be able to use the findings to assess credit risk. To investigate interdependence and extreme interactions, future research should integrate a multivariate time series methodology with a multivariate copular method. It would be intriguing to investigate what kinds of results could be obtained by filtering the series using a machine learning approach and investigating the stylized facts in a multivariate time series context, performing a comprehensive evaluation with time-varying parameters and extreme value distributions. Another area where prospective research is needed is a probabilistic description and modeling of the effects of news and foreign exchange loads on stock market participants utilizing the Poison point process. The strategy can help to figure out how often peak dangers occur. A daily risk sensitivity analysis is also recommended, and the creation of two-stage stochastic integer recourse models to maximize return distribution is an exciting future research area. The limitation of this study is the absence of data values for business days that were holidays. Therefore, interpolation was used to fill the gaps. This highlights the risks of data mining. As a result, the findings of the study must be interpreted with caution.

Acknowledgments

The Johannesburg Stock Exchange (JSE) deserves special recognition for granting us permission to use the high-frequency five-day data in this study. The authors are also thankful for the constructive and meaningful guiding comments from anonymous reviewers.

Conflict of interest

The authors declare no conflicts of interest in this paper.

References

Alemohammad N, Rezakhah S, Alizadeh SH (2020) Markov-switching Asymmetric GARCH model: Stability and Forecasting. *Stat Pap* 61: 1309–1333. https://doi.org/10.1007/s00362-018-0992-2

Aquilina M, Budish E, O'Neill P (2021) Quantifying the High-frequency Trading "Arms Race". *Q J Econ* 137: 493–564. https://doi.org/10.1093/qje/qjab032

Ardia D, Bluteau K, Boudt K, et al. (2019) Markov-switching GARCH Models in R: The MSGARCH Package. *J Stat Software* 91. https://dx.doi.org/10.2139/ssrn.2845809

Arora H (2017) Stylized Facts and Trends of High-frequency Data in Financial Markets. *Asian J Res Bus Econ Manage* 7: 303–317. https://doi.rog/10.5958/2249-7307.2017.00115.3

Arratia A, Lopez-Barrantes AX (2021) Do Google Trends Forecast Bitcoins? Stylized Facts and Statistical Evidence. *J Banking Financ Technol* 5: 45–57. https://doi.org/10.1007/s42786-021-00027-4

Atsin JA, Ocran MK (2015) Calendar Effects and Market Anomalies on the Johannesburg Stock Exchange. *MPRA Paper No. 87448. Munich: University Library of Munich*. Available from: https://mpra.ub.uni-muenchen.de/87448/

Bariviera AF, Basgall MJ, Hasperué W (2017) Some Stylized Facts of the Bitcoin Market. *Physica A* 484: 82–90. https://doi.org/10.1016/j.physa.2017.04.159

Bee M, Trapin L (2018) Estimating and Forecasting Conditional Risk Measures with Extreme Value Theory: A Review. *Risks* 6: 1–16. https://doi.org/10.3390/risks6020045

Beytell D, The Effect of Extreme Value Distributions on Market Risk Estimation. MCom. (Financial Economics), Faculty of Finance and Economic Science: University of Johannesburg, 2016. Available from: https://ujcontent.uj.ac.za/vital/access/manager/Index?sitename=Research%20Output

Bilgili F, Ulucak R, Koçak E, et al. (2020) Does Globalization Matter for Environmental Sustainability? Empirical Investigation for Turkey by Markov Regime Switching Models. *Environ Sci Pollut Res* 27: 1087–1100. https://doi.org/10.1007/s11356-019-06996-w

Boer PH, Munapo E, Chanza M, et al. (2019) Exchange market pressure in South Africa and Kenya: An analysis using parametric and non-parametric extreme value theory. *J Econo Financ Sci* 12: 1–15. https://doi.org/10.4102/jef.v12i1.202

Bruce P, Bruce A, Gedeck P (2020) *Practical Statistics for Data Scientists: 50+ Essential Concepts using R and Python.* roman familySecond ed. O'Reilly Media. http://oreilly.com/

Cavalli F, Naimzada A, Pecora N (2017) Real and Financial Market Interactions in a Multiplier-accelerator Model: Non-linear Dynamics, Multi-stability and Stylized Facts. *Chaos: An Interdiscip J Non-linear Sci* 27: 103120. https://doi.org/10.1063/1.4994617

Cavicchioli M (2021) Markov Switching Garch Models: Higher Order Moments, Kurtosis Measures, and Volatility Evaluation in Recessions and Pandemic. *J Bus Econ Stat* 1–12. https://doi.org/10.1080/07350015.2021.1974459

Chen YF, Ling XM, Liu MM, et al. (2018) Statistical Distribution of Hydraulic Conductivity of Rocks in Deep-incised Valleys, Southwest China. *J Hydrol* 566: 216–226. https://doi.org/10.1016/j.jhydrol.2018.09.016

Cox V (2017) Translating Statistics to Make Decisions: A Guide for the Non-statistician. New York: Apress. https://doi.org/10.1007/978-1-4842-2256-0

Da Cunha CR, Da Silva R (2020) Relevant Stylized Facts About Bitcoin: Fluctuations, First Return Probability, and Natural Phenomena. *Physica A* 550: 124155. https://doi.org/10.1016/j.physa.2020.124155

Dias JG, Vermunt JK, Ramos S (2015) Clustering Financial Time Series: New Insights from an Extended Hidden Markov Model. *Eur J Oper Res* 243: 852–864. https://doi.org/10.1016/j.ejor.2014.12.041

Dufreno G, Matsuki T (2021) *Recent Econometric Techniques for Macroeconomic and Financial Data*. Berlin Springer. https://doi.org/10.1007/978-3-030-54252-8

Duong T (2022) Ks: Kernel Density Estimation for Bivariate Data. *J Stat Software*. Available from: http://freebsd.yz.yamagata-u.ac.jp/pub/cran/web/packages/ks/vignettes/kde.pdf

Fama E, French K (1988) Permanent and Temporary Components of Stock Prices. *J Polit Econ* 96: 246–273. https://doi.org/10.1086/261535

Financial Times (2022) Coronavirus threatens India's Banking Recovery Before it Even Starts. Available from: https://www.ft.com/content/153f2922-6e15-11ea-89df-41bea055720b

Garson GD (2012) Testing Statistical Assumptions. *Asheboro, NC: Stat Assoc Publishing*. Available from: http://www.statisticalassociates.com/assumptions.pdf

Ha J, Kose MA, Otrok C, et al.(2020) Global Macro-financial Cycles and Spillovers. *Natl Bur Econ Res*. htpps://doi.org/10.3386/w26798

Hart JD (2013) *Nonparametric Smoothing and Lack-of-fit Tests,* roman familyBerlin Springer. https://doi.org/10.1007/978-1-4757-2722-7

Hiebert P, Jaccard I, Schuler Y (2018) Contrasting Financial and Business Cycles: Stylized Facts and Candidate Explanations. *J Financ Stab* 38: 72–80. https://doi.org/10.1016/j.jfs.2018.06.002

Hu Y, Plonsky L (2021) *Second Language Research* 37: 171-184. https://doi.org/10.1177/0267658319877433

Hu AS, Parlour CA, Rajan U (2019)/ Cryptocurrencies: Stylized Facts on a New Investible Instrument. *Financ Manage* 48: 1049–1068. https://doi.org/10.1111/fima.12300

Jakata O, Chikobvu D (2022) Extreme Value Modelling of the Monthly South African Industrial Index (J520) Returns. *Stat Optim Inf Comput* 10: 383–400. https://doi.org/10.19139/soic-2310-5070-906

Jiang J (2022) Non-parametric Statistics. *In Large Sample Techniques for Statistics* 379–415, Springer Cham. https://doi.org/10.1007/978-3-030-91695-4_11

Jooste L (2006) Cash Flow Ratios as a Yardstick for Evaluating Financial Performance in African Businesses. *Managerial Financ* 32: 569-576. https://doi.org/10.1108/03074350610671566

Katahira K, Chen Y, Hashimoto G, et al. (2019) Development of an Agent-based Speculation Game for Higher Reproducibility of Financial Stylized Facts. *Physica A* 524: 503–518. https://doi.org/10.1016/j.physa.2019.04.157

Katz RW (2013) Statistical methods for Non–stationary Extremes. *Extremes in a changing climate. Water Science and Technology Library,* 65. Dordrecht Springer. https://doi.org/10.1007/978-94-007-4479-0_2

Kim D, Shin M (2022) Volatility Models for Stylized Facts of High-Frequency Financial Data. *arXiv pre-print arXiv:2205.15808*. Available from: https://doi.org/10.48550/arXiv.2205.15808

Korkpoe CH, Junior PO (2018) The Behavior of Johannesburg Stock Exchange all-share index Returns–An Asymmetric GARCH and News Impact Effects Approach. *SPOUDAI J Econ Bus* 68: 26–42. Available from: https://spoudai.unipi.gr/index.php/spoudai/article/view/2634.

Kufenko V, Geiger N (2017) Stylized Facts of the Business Cycle: Universal Phenomenon, or Institutionally Determined? *J Bus Cycle Res* 13: 165-187. https://doi.org/10.1007/s41549-017-0018-5

Ledl T (2016) Kernel Density Estimation: Theory and Application in Discriminant Analysis. *Austrian J stat* 33: 267–279.

Li Y, Abdel-Aty M, Yuan J, et al. (2020) Analyzing Traffic Violation Behavior at Urban Intersections: A Spatio-temporal Kernel Density Estimation Approach using Automated Enforcement System Data. *Acci Anal Prev* 141: 105509. https://doi.org/10.1016/j.aap.2020.105509

Liu Z, Shang P, Wang Y (2020) Characterization of Time Series Through Information Quantifiers. *Chaos Solitons Fractals* 132: 109565. https://doi.org/10.1016/j.chaos.2019.109565

Maaziz M, Kharfouchi S (2018) Parameter Estimation of Markov-switching Bilinear Model using the (EM) Algorithm. *J Stat Plan Inference* 192: 35–44. https://doi.org/10.1016/j.jspi.2017.07.002

Morema K, Bonga-Bonga L (2020) The Impact of Oil and Gold Price Fluctuations on the South African Equity Market: Volatility Spillovers and Financial Policy Implications. *Resour Policy* 68: 101740. https://doi.org/10.1016/j.resourpol.2020.101740

Montgomery DC, Jennings CL, Kulahci M (2015) *Introduction to Time Series Analysis and Forecasting*, New Jersey: John Wiley and Sons. https://doi.org/10.2307/2938260

Nickl R, Ray K (2020) Non-parametric Statistical Inference for Drift Vector Fields of Multi-dimensional Diffusions. *Ann Stat* 48: 1383–1408. https://doi.org/10.1214/19-AOS1851

Nystrup P, Madsen H, Lindstrom E (2015) Stylized Facts of Financial Time Series and Hidden Markov Models in Continuous Time. *Quant Financ* 15: 1531-1541. https://doi.org/10.1080/14697688.2015.1004801

Olbryś J, Oleszczak A (2020) Intra–day Patterns in Trading Volume. Evidence from High Frequency Data on the Polish Stock Market. In Saeed K, Dvorský J.(eds) Computer Information Systems and Industrial Management. CISIM 2020. *Lecture Notes in Computer Science*, 390–401. Springer Cham. https://doi.org/10.1007/978-3-030-47679-3_33

Orlando G, Zimatore G (2020) Business Cycle Modeling between Financial Crises and Black Swans: Ornstein–Uhlenbeck Stochastic Process versus Kaldor Deterministic Chaotic Model. *Chaos Interdiscip J Nonlinear Sci* 30: 083129. https://doi.org/10.1063/5.0015916

Ozili P, Arun T (2020) Spillover of COVID-19: Impact on the Global Economy. University Library of Munich, Germany. Available from: https://mpra.ub.uni-muenchen.de/99850/

Patterson GA, Sornette D, Parisi DR (2020) Properties of Balanced Flows with Bottlenecks: Common Stylized Facts in Finance and Vibration-driven vehicles. *Phy Rev E* 101: 042302. https://doi.org/10.1103/PhysRevE.101.042302

Phylaktis K, Manalis G (2013) Futures Trading and Market Micro–structure of the Underlying Security: A High-frequency Experiment at the Single Stock Future Level. *Borsa Istanbul Rev* 13: 79–92. https://doi.org/10.1016/j.bir.2013.10.012.

Porras ER (2017) *Stylized Facts of Financial Markets and Bubbles*. Bubbles and Contagion in Financial Markets, 2: 53–70.Palgrave Macmillan, London. https://doi.org/10.1057/978-1-137-52442-3_2

Raihan T (2017) Performance of Markov-Switching GARCH Model Forecasting Inflation Uncertainty. In Res-Report. Munich: University Library of Munich.

Restocchi V, Mcgroarty F, Gerding E (2019) The Stylized Facts of Prediction Markets: Analysis of Price Changes. *Physica A* 515: 159–170. https://doi.org/10.1016/j.physa.2018.09.183

Ruppert D, Matteson D (2015) *Statistics and Data Analysis for Financial Engineering: with R examples*. Berlin Springer. https://doi.org/10.1007/978-1-4939-2614-5

Schmid F (2009) High-frequency Financial Markets Data Cleaning and Stylized Facts in Financial Markets Data. Paper Presentation. Cologne: Seminar of Economic and Social Sciences at the University of Cologne. Available from: https://hughchristensen.com/papers/academic_papers/High%20frequency%20data%20analysis.pdf

Shakeel M, Srivastava B (2021) Stylized Facts of High-frequency Financial Time Series data. *Global Bus Rev* 22: 550–564. htpps://doi.org/10.1177/0972150918811701

Strohsal T, Proaño CR, Wolters J (2019) Characterizing the Financial Cycle: Evidence from a Frequency Domain Analysis. *J Bank Financ* 106: 568–591. https://doi.org/10.1016/j.jbankfin.2019.06.010

Silverman BW (2018) *Density Estimation for Statistics and Data Analysis*, Routledge, New York. https://doi.org/10.1201/9781315140919

Sigaukea C, Rhoda MM, Maseka L (2014) Modeling Conditional Heteroscedasticity in JSE Stock Returns using the Generalized Pareto Distribution. *Afr Rev Econ Financ* 6: 41–55. https://www.ajol.info/index.php/aref/issue/view/11344

Statistics South Africa, Economic growth slows in 2014, 2022, Available from: statssa.gov.za/?p=4184. [Accessed on 31 May 2022].

Sullivan R, Timmermann A, White H (2001) Dangers of Data Mining: the Case of Calendar Effects in Stock Returns. *J Econometrics* 105: 249–286.https://doi.org/10.1016/S0304-4076(01)00077-X

Tsay RS (2015) Financial time series. New Jersey: John Wiley and Sons. https://doi.org/10.1002/9781118445112.stat03545.pub2

Verma JP, Abdel-Salam ASG (2019) Testing Statistical Assumptions in Research. 1 edn, John Wiley and Sonshttps://dl.uswr.ac.ir

Wang L, Ma F, Niu T, et al.(2020) Crude Oil and BRICS Stock Markets under Extreme Shocks: New Evidence. *Econ Modell* 86: 54–68. https://doi.org/10.1016/j.econmod.2019.06.002

Washington S, Karlaftis M, Mannering F, et al. (2020) Statistical and econometric methods for transportation data analysis, Chapman and Hall/CRC. https://doi.org/10.1201/9780429244018

Wulff SS (2017) Time Series Analysis: Forecasting and Control. *J Qual Technol* 49: 418. https://doi.org/10.1080/00224065.2017.11918006

Xaba LD, Moroke ND, Metsileng LD (2021) Performance of MS-GARCH models: Bayesian MCMC-based Estimation. *Handbook of Research on Emerging Theories, Models, and Applications of Financial Econometrics* 323–356. Berlin: Springer. https://doi.org/10.1007/978-3-030-54108-8_14

Xaba D, Moroke ND, Rapoo I (2019) Modeling Stock Market Returns of BRICS with a Markov-switching Dynamic Regression Model. *J Econ Behav Stud* 11: 10–22. https://doi.org/10.22610/jebs.v11i3(J).2865

Zhai L, Wu Y, Yang J, et al. (2020) Characterizing Initiation of Gas–liquid Churn flows using Coupling Analysis of Multivariate Time Series. *Physica A* 540: 123099. https://doi.org/10.1016/j.physa.2019.123099

Zumbach GO, Müller U (2000) Operators on in Homogeneous Time Series. *Olsen Associates Working Paper No. 324*. http://dx.doi.org/10.2139/ssrn.208278