Modern Finance

Article

Social media and financial markets: The impact of Twitter sentiment on the Johannesburg Stock Exchange

Thiasha Naidoo [1]

1   School of Accounting, Economics and Finance, University of KwaZulu-Natal, South Africa, email: thiashanaidoo20@gmail.com

Abstract: This study examines the effect of Twitter-derived investor sentiment on stock market volatility in South Africa using daily data for the JSE All Share Index from 2016 to 2023. Using GARCH-M, E-GARCH-M, and GJR-GARCH-M, the results show that the GJR-GARCH-M specification provides the best fit both before and after incorporating sentiment. Twitter sentiment significantly amplifies market volatility, with negative sentiment exerting a more substantial impact than positive sentiment, consistent with asymmetric volatility dynamics and the leverage effect. Overall, the findings demonstrate that Twitter-derived sentiment contains valuable information for modelling and understanding volatility in emerging equity markets such as South Africa.

Keywords: Twitter sentiment; GARCH models; Johannesburg Stock Exchange; investor sentiment; South Africa

JEL Classification: G10; G14; G15; G17; G40; G41

1. Introduction

In the digital era, social media platforms have fundamentally reshaped the speed and manner in which information circulates, enabling opinions, rumors, and market-relevant signals to spread almost instantaneously (Duz Tan & Tas, 2021; Jasim et al., 2022). This immediacy has important implications for financial markets, particularly in emerging economies such as South Africa, where thinner liquidity, structural vulnerabilities, and heightened sensitivity to shocks may amplify sentiment-driven behavior (Nyakurukwa & Seetharam, 2022). Among these platforms, Twitter, recently rebranded as "X" in 2023, remains especially relevant due to its real-time nature and extensive use in both academic research and market commentary. Consistent with existing literature, this study retains the term Twitter for clarity and continuity (Graham & Stough, 2025).

This study examines whether Twitter-derived investor sentiment influences stock market volatility in South Africa by integrating firm-level sentiment measures into GARCH, GJR-GARCH, and E-GARCH models applied to the JSE All Share Index (ALSI). Using daily data from 2016 to 2023, the analysis evaluates whether real-time social media sentiment improves volatility modelling and whether adverse sentiment shocks generate asymmetric volatility responses in an emerging market setting.

The empirical results provide clear evidence that Twitter sentiment significantly shapes volatility dynamics on the JSE. Incorporating sentiment into the volatility equations improves model fit and explanatory power across all specifications. The results further indicate that adverse sentiment shocks exert a disproportionately larger impact on volatility than positive sentiment of similar magnitude, consistent with the presence of leverage effects. These asymmetric responses are most pronounced in the GJR-GARCH and E-GARCH models, highlighting the importance of nonlinear frameworks when modelling sentiment-induced volatility. Overall, the findings suggest that investor

sentiment captured from social media is not merely noise, but a meaningful driver of market risk in South Africa.

These findings contribute to the literature in several important ways. First, the study provides one of the earliest systematic investigations of Twitter-based investor sentiment as a determinant of stock market volatility in South Africa, extending existing work that has focused primarily on online mood states or corporate communication patterns (Maree & Johnston, 2015; Nyakurukwa & Seetharam, 2022). Second, by integrating real-time social media sentiment into GARCH-type volatility models, the study offers a methodological contribution, demonstrating that sentiment enhances volatility modelling beyond traditional market information. This issue remains underexplored in emerging market contexts. Third, the evidence of asymmetric, sentiment-driven volatility responses aligns with international findings (Sprenger et al., 2014; Adams et al., 2023) and shows that the structural characteristics of the JSE amplify these effects. Finally, the results underscore the growing relevance of social media analytics for risk management, portfolio allocation, and volatility forecasting in modern financial markets.

The remainder of the paper is structured as follows. Section 2 reviews the related literature on social media sentiment and volatility modelling. Section 3 describes the data and methodology. Section 4 presents and discusses the empirical results. Section 5 concludes with implications for investors, policymakers, and future research.

2. Literature Review

The digitalization of information channels has profoundly transformed how investors access and process market-relevant data. Social media platforms, particularly Twitter (X), now serve as large-scale sentiment aggregators that can influence market dynamics by shaping collective expectations (Sul et al., 2017; Tiwari et al., 2020). Although extensive research has explored the relationship between Twitter sentiment and market performance in developed markets, empirical evidence from emerging economies, especially South Africa, remains limited (Steyn et al., 2020; Lawrence et al., 2024). This section reviews how sentiment has been extracted and applied in prior studies, identifies methodological advances, and positions this study within the broader behavioral finance literature on sentiment-induced volatility.

The seminal work of Bollen et al. (2011) marked a turning point in quantifying investor sentiment. Using natural language processing tools, OpinionFinder, and the Google Profile of Mood States (GPOMS), the authors analyzed approximately 9.8 million tweets to measure six mood dimensions: calm, alert, sure, vital, kind, and happy. These indices were created by mapping textual expressions to the established Profile of Mood States psychometric scale. Among these, the "calmness" dimension significantly improved forecasts of the Dow Jones Industrial Average (DJIA), indicating that aggregated emotional tone on Twitter can precede real market movements.

Building on this, Oh and Sheng (2011) applied sentiment polarity scoring, categorizing tweets as positive, neutral, or negative based on textual tone, to examine the U.S. technology sector. They found that firms with higher online engagement experienced stronger sentiment-return linkages and heightened short-term volatility. Their work demonstrated that sentiment derived from user-generated content could complement traditional financial indicators in predictive models.

Subsequent research refined these methods using machine learning and deep learning algorithms. For example, Zhang et al. (2018) employed a Long Short-Term Memory (LSTM) network to analyze millions of financial tweets, revealing that deep neural models substantially outperformed linear time-series approaches in forecasting S&P 500 movements. However, Ruths and Pfeffer (2014) cautioned against methodological biases arising from selective sampling (e.g., influential users or hashtag filters) and opaque sentiment-classification processes. Similarly, Oliveira et al. (2016) demonstrated that the predictive power of Twitter sentiment diminishes over longer horizons, reinforcing its relevance mainly for high-frequency trading. These

developments collectively show an evolution from lexicon-based mood tracking to advanced contextual NLP models, although transparency and interpretability remain challenges.

While most early studies focused on returns, a growing body of evidence recognizes sentiment as a critical driver of volatility dynamics. Nisar and Yeung (2018) were among the first to incorporate Twitter-derived sentiment into GARCH models, showing that sentiment significantly enhanced the explanatory power of return volatility in U.S. markets. The authors used lexicon-based sentiment scores extracted from tweets tagged to specific companies and classified via word polarity dictionaries. Similarly, Mendoza-Urdiales et al. (2022) integrated Twitter sentiment indices into E-GARCH and transfer entropy frameworks. Their sentiment data, generated through machine-learning classifiers trained on financial tweets, revealed asymmetric effects: negative sentiment intensified volatility more than positive sentiment reduced it. This study also found bidirectional information flow between sentiment and prices, underscoring that Twitter serves as both an information source and a behavioral amplifier during market stress.

Despite these advances, most sentiment-volatility research remains concentrated in developed markets. In emerging contexts, data scarcity, linguistic diversity, and lower social media penetration have constrained empirical investigation (Duz Tan & Tas, 2021). These limitations highlight the need for frameworks that capture nonlinear dynamics and asymmetric responses, particularly in markets with greater behavioral heterogeneity, which motivates the present study's adoption of GARCH-type models for the JSE.

Empirical work in South Africa remains nascent but growing. Maree and Johnston (2015) were pioneers in linking social media sentiment to the JSE ALSI. They collected over 3 million tweets over 55 days and applied linguistic frequency analysis using mood dictionaries to extract indicators such as "depression" and "fatigue." Their results showed that negative mood states were contemporaneously associated with lower ALSI values, while fatigue levels exhibited a positive lagged correlation, indicating emotional contagion effects in daily trading.

Nyakurukwa and Seetharam (2022) extended this line of inquiry by applying text classification techniques to analyze whether short, 280-character tweets contain informational value for JSE-listed firms. Using a supervised machine-learning model, they computed aggregate sentiment polarity scores. They found significant predictive power for returns and volatility, supporting the relevance of Twitter as an alternative information channel in markets characterized by informational inefficiencies.

A subsequent study by Seetharam and Nyakurukwa (2024) differentiated between online news headlines and social media posts (hashtags) to evaluate how distinct digital content types affect investor sentiment and market dynamics. Contrary to the earlier manuscript's phrasing, their analysis did not compare "headlines versus hashtags" as proxies for collective sentiment, but rather contrasted traditional news (headlines) with social media activity (hashtags) to capture the relative strength of these information channels. Their findings revealed that social media sentiment exerts a more substantial and more immediate influence on market behavior than online news sentiment, reaffirming the behavioral significance of Twitter-derived signals.

In another local study, Nel and du Toit (2023) examined company-initiated tweets and investor responses over 4 years. Using tweet count data combined with content engagement metrics (likes, replies, and retweets), they found that greater corporate social media activity corresponded with higher stock liquidity and improved returns, suggesting that digital communication can enhance information dissemination and investor confidence.

Complementing this, Fonou-Dombeu et al. (2024) linked consumer sentiment, proxied by the South African Consumer Confidence Index (CCI), to stock market performance. Although their measure was not Twitter-based, it highlighted that broad psychological sentiment significantly moderates the relationship between fundamentals

and returns, aligning with behavioral finance theories that emphasize emotion-driven decision-making.

Collectively, these studies confirm that sentiment, whether derived from textual, engagement, or survey-based indicators, affects volatility, volume, and returns. However, the methodological diversity across these works underscores that the predictive value of sentiment depends heavily on how it is measured. Early studies relied on keyword frequencies, while later research incorporated machine learning and NLP classifiers that could provide contextual understanding. The current study contributes to this progression by applying a quantitative sentiment index, derived from South African Twitter data, within a GARCH-type framework designed to capture nonlinear volatility dynamics in the JSE ALSI.

While the benefits of social media data are well recognized, its drawbacks also merit discussion. The unfiltered nature of Twitter exposes financial markets to fake news, rumors, and speculative narratives, which can distort sentiment indices and trigger herding or panic reactions (Ridhwan & Hargreaves, 2021). Studies such as Metta et al. (2022) highlight that misinformation shocks can generate transient but severe volatility spikes, reflecting the risk of algorithmic amplification of false content. Moreover, sentiment polarity models may overemphasize emotional extremes, potentially exaggerating perceived market mood.

Recent research increasingly adopts advanced NLP techniques, including VADER (Valence Aware Dictionary and sEntiment Reasoner) and BERT (Bidirectional Encoder Representations from Transformers), to improve semantic accuracy and reduce bias. However, these methods demand substantial computational resources and language-specific training data, which remain limited for South African contexts. Consequently, this study employs a sentiment extraction approach optimized for data availability in emerging markets, while acknowledging that future research could enhance precision through transformer-based NLP architectures.

In summary, the literature reveals a consistent association between social media sentiment and market volatility, though the magnitude and persistence of this relationship vary across methods and markets. Existing South African evidence is fragmented, and no prior study has explicitly modelled the asymmetric impact of Twitter sentiment on JSE volatility using GARCH-type models. Addressing this gap, the present research integrates behavioral finance theory, real-time sentiment measurement, and volatility modelling to provide novel insights into the behavioral dynamics of the South African stock market.

3. Data and Methods

This study examines the relationship between Twitter-derived sentiment and stock market volatility on the JSE ALSI. The analysis covers the period from 1 January 2016 to 31 December 2023, using daily data. The start of 2016 is chosen because Bloomberg's social media sentiment analytics became methodologically available from that date onwards, with improved coverage and fewer missing values. The terminal date of 2023 is deliberately selected to ensure comparability across complete annual reporting cycles and to avoid structural breaks introduced by significant changes in Twitter (X) data policies and Bloomberg's extraction pipeline from 2024 onwards. This timeframe captures multiple market phases, including the COVID-19 crisis, the July 2021 unrest, and ongoing energy supply constraints, enabling a comprehensive assessment of sentiment dynamics across varying conditions.

The sample consists of all JSE ALSI constituents, representing approximately 99% of the market's total capitalisation. Daily returns are calculated as $r_t = 100 \times \log\left[\left(\frac{P_t}{P_{t-1}}\right)\right]$, where P_t denotes the closing price on day t. Twitter sentiment data are obtained from Bloomberg Twitter Sentiment Analytics, which compiles textual information from the platform (now rebranded as X) using NLP techniques. Each tweet referencing a listd firm

is assigned a polarity score $S_i^k \epsilon [-1,1)$ indicating whether the content is positiver negative, and a confidence weight $C_i^k \epsilon [-1,1)$ reflecting the model's certainty in classification (Nyakurukwa & Seetharam, 2022; Bloomberg, 2025). Firm-level sentiment is then computed as a confidence-weighted average over a 24-hour rolling window ending ten minutes before the JSE opens, expressed as:

$$Twitter_{i,t} = \frac{\sum k \varepsilon P(i,T) S_i^k C_i^k}{N_{i,T}}, \; T\varepsilon \; [\text{t-24h},t] \tag{1}$$

Where: $P(i,T)$ denotes the set of non-neutral tweets mentioning i during window T, and $N(i,T)$ represents the number of such tweets. The resulting sentiment score is bounded in [-1, 1), with values closer to –1 indicating strongly negative sentiment and values approaching 1 indicating strongly positive sentiment.

Since the dependent variables in this study are index-level returns and volatility, firm-specific sentiment is aggregated into an index-level sentiment measure using free-float market capitalisation weights. The daily ALSI sentiment index is calculated as:

$$Twit_t^{ALSI} = \sum_{i \epsilon I_t} \omega_{i,t} Twit_{i,t} \tag{2}$$

where the weights $\omega_{i,t}$ are normalised across firms with non-neutral sentiment coverage. If a firm has no non-neutral tweets on the day t, it is excluded from the aggregation, and weights are rescaled to maintain comparability. This weighting approach aligns the sentiment measure with the ALSI's construction. It ensures that larger, more liquid firms with greater investor attention contribute proportionately to the overall sentiment factor (Muguto & Muzindutsi, 2022).

To manage missing data and thin coverage, a coverage ratio, the proportion of the ALSI's market capitalisation represented by firms with valid sentiment observations, is computed each day. As a robustness check, the sentiment index is smoothed using an exponentially weighted moving average (EWMA) over a three-day window and re-estimated with minimum coverage thresholds. The pattern of results remains consistent across specifications.

3.1. Model Specification

To capture the conditional volatility dynamics of the JSE ALSI in response to firm-level Twitter sentiment, this study employs three well-established GARCH-type models: the GARCH (1,1) model (Bollerslev, 1986), the GJR-GARCH (1,1) model (Glosten et al., 1993), and the E-GARCH (1,1) model (Nelson, 1991). These specifications are particularly appropriate given the heightened volatility and sentiment sensitivity typically observed in emerging markets such as South Africa (Mudinas et al., 2019; Alomari et al., 2021). These models are particularly suitable for the present analysis for several reasons. Firstly, the study applies firm-level Twitter sentiment to the South African equity market, a relatively underexplored emerging market. In contrast, most sentiment–volatility research has focused predominantly on developed economies such as the United States and China. Secondly, as is typical of emerging markets, the JSE is characterised by heightened volatility and may be more sensitive to sentiment-driven noise (Mudinas et al., 2019; Alomari et al., 2021), particularly given the rapid dissemination of information on social media platforms such as Twitter. Incorporating sentiment into a GARCH-type framework, therefore, allows for an assessment of whether public opinion exerts an incremental influence on return volatility beyond standard market dynamics. Thirdly, while a growing literature examines the sentiment–volatility relationship using linear volatility modelling frameworks, relatively few studies integrate firm-specific social media sentiment within explicitly nonlinear or asymmetric volatility models. Linear approaches typically rely on conventional ARMA-GARCH or regression-based specifications that assume symmetric and linear conditional variance dynamics (e.g., Tetlock, 2008; Bollen et al., 2011; Smales, 2014). Although informative, such models are limited in their ability to capture asymmetric and regime-dependent volatility responses.

By contrast, empirical evidence on firm-level social media sentiment within nonlinear volatility frameworks remains relatively scarce, particularly in emerging markets.

Prior to model estimation, standard diagnostic tests were conducted to ensure the validity of the modelling assumptions (Brooks, 2019). Conditional heteroskedasticity was assessed using the ARCH–LM test, while residual autocorrelation was examined using the Ljung–Box Q-statistic. The results confirmed the suitability of GARCH-type models for modelling return volatility.

The conditional mean equation for stock returns is specified consistently across all model variants as an ARMA(1,1) process:

$$y_t = \mu + \alpha y_{t-1} + \nu \varepsilon_{t-1} + \theta \sigma_{t-1}^2 + \varepsilon_t \tag{3}$$

To assess the impact of Twitter sentiment on returns, the sentiment-augmented mean equation is specified as:

$$y_t = \mu + \alpha y_{t-1} + \nu \varepsilon_{t-1} + \theta \sigma_{t-1}^2 + \emptyset \text{TwitSent}_t + \varepsilon_t \tag{4}$$

where: y_t is the index return; μ is the constant mean; Notably, αy_{t-1} is the autoregressive (AR) term, where α denotes the effects of past returns; and $\nu \varepsilon_{t-1}$ is the moving average (MA) term, where ν denotes the effects of past shocks. This term captures the impact of the previous period's shock on the current value. θ is the coefficient of the lagged conditional variance, which captures the impact of past volatility on the current value; σ_{t-1}^2 is the conditional variance at time t; ε_t is the error term at time t, assumed to follow a normal distribution with zero mean and time-varying variance σ_t^2. The impact of Twitter sentiment on returns $\emptyset$ was based on examining the coefficient's size, sign, and significance.

The influence of investor sentiment on stock return volatility is examined using three conditional variance specifications: GARCH (1,1), GJR-GARCH (1,1), and E-GARCH (1,1). Each model is first estimated in its baseline form, without sentiment, to establish benchmark volatility dynamics (Equations (5), (7), and (9)). The analysis was then extended by introducing the Twitter-derived sentiment score ($TwitSent_t$) into the variance equations, allowing for a direct assessment of whether sentiment provides additional explanatory power in capturing fluctuations in market volatility, denoted in equations (6), (8), and (10).

The baseline GARCH (1,1) model and its sentiment-augmented specification are defined as:

$$\Sigma_t^2 = \omega + \beta \varepsilon_{t-1}^2 + \lambda \sigma_{t-1}^2 \tag{5}$$

$$\sigma_t^2 = \omega + \beta \varepsilon_{t-1}^2 + \lambda \sigma_{t-1}^2 + \varphi \text{TwitSent}_t \tag{6}$$

where: σ_t^2 is the conditional variance of the time series at time t; ω is a constant term representing the long-term average variance; $\beta \varepsilon_{t-1}^2$ is the lagged squared error term thus, β is the coefficient; and ε_{t-1}^2 is the squared error from the previous period. This term captures the effect of past shocks on current volatility, signifying how much impact a shock at $t-1$ has on the current period's volatility. $\Lambda \sigma_{t-1}^2$ is the lagged conditional variance term, thus λ is the coefficient; and σ_{t-1}^2 is the previous period's conditional variance. This term captures the persistence of volatility over time, indicating how volatility in previous periods influences current volatility (Bouri et al., 2019).

To capture asymmetric volatility responses to positive and negative shocks, the GJR-GARCH (1,1) model is specified as:

$$\Sigma_t^2 = \omega + \beta \varepsilon_{t-1}^2 + \lambda \sigma_{t-1}^2 + \delta I_{t-1} \varepsilon_{t-1}^2 \tag{7}$$

$$\sigma_t^2 = \omega + \beta \varepsilon_{t-1}^2 + \lambda \sigma_{t-1}^2 + \delta I_{t-1} \varepsilon_{t-1}^2 + \varphi \text{TwitSent}_t \tag{8}$$

where: $\delta I_{t-1} \varepsilon_{t-1}^2$ captures the conditional effect of I_{t-1} on past shocks and is an interaction term between I_{t-1} and the lagged squared error term ε_{t-1}^2; δ Is the coefficient.

To address the non-negativity constraints inherent in standard GARCH models, the E-GARCH (1,1) specification models the logarithm of conditional variance:

$$\text{Log}(\sigma_t^2) = \omega + \beta \ln(\sigma_t^2) + \lambda \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \delta \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \tag{9}$$

$$\log(\sigma_t^2) = \omega + \beta \ln(\sigma_t^2) + \lambda \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \delta \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] + \varphi \text{TwitSent}_t \tag{10}$$

where: $\log(\sigma_t^2)$ is the natural logarithm of the conditional variance at time t, representing log-transformed volatility; $\beta \log(\sigma_t^2)$ is the log of the lagged conditional variance term and captures the persistence of volatility, thus, β is the coefficient of the lagged log-variance; $\log(\sigma_t^2)$ is the log of the previous period's conditional variance. $\Lambda \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}}$ captures the impact of past standardized shocks on current log-volatility, as such, λ is the coefficient of the standardized error term; $\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}}$ is the lagged standardized error, where ε_{t-1} is the previous period's error term and $\sqrt{\sigma_{t-1}^2}$ is the previous period's conditional variance. Moreover, $\delta \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$ indicates the deviation of the absolute standardized error from its expected value and adjusts for asymmetry of the impact of positive and negative shocks. Thus, δ is the coefficient; $\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}}$ is the absolute value of the lagged standardized error; $\sqrt{\frac{2}{\pi}}$ is a constant that represents the expected value of the absolute standardized normal distribution.

Although the GJR-GARCH (1,1) model captures asymmetric volatility responses to positive and negative shocks, commonly referred to as the leverage effect (Brooks, 2019), it may violate the non-negativity condition of the conditional variance. To address this, parameter restrictions are typically imposed to ensure a positive variance; however, such constraints may be restrictive and not fully reflect the underlying data-generating process. To overcome this limitation, the E-GARCH model was employed, which models the logarithm of the conditional variance, thereby ensuring positivity without the need for externally imposed constraints (Nelson, 1991). Both the GJR-GARCH and E-GARCH specifications allow for asymmetric volatility dynamics, with leverage effects confirmed when the asymmetry parameter is statistically significant and negative, indicating that negative shocks exert a larger impact on volatility than positive shocks of equal magnitude. For all model specifications, covariance stationarity conditions were verified to ensure model validity (Brooks, 2019).

4. Results and Discussion

4.1. Preliminary Results

Table 1 presents the descriptive statistics for the JSE All Share Index (ALSI) daily returns and the aggregate Twitter sentiment index (Twit$^{(\text{ALSI})}_t$) over the period 1 January 2016 to 31 December 2023. The index-level sentiment series was derived by aggregating firm-level Bloomberg sentiment scores into a market-capitalization-weighted daily measure, as described in Section 3.1. Each firm-day score ranges between -1 (strongly negative) and +1 (strongly positive), and the aggregated index therefore also falls within this range.

The descriptive statistics reveal that both the JSE ALSI returns and aggregate sentiment exhibit positive means (2.494 and 0.398, respectively), indicating a generally optimistic tone in both market performance and investor mood during the sample period.

The JSE ALSI shows relatively low dispersion, with a standard deviation of 0.151 and a coefficient of variation (COV) of 0.060, reflecting its stability as a diversified benchmark. In contrast, the sentiment series shows greater variability (standard deviation = 0.846; COV = 2.134), consistent with the rapid, often emotional nature of social-media interactions. The return distribution is positively skewed, indicating a long right tail and occasional large positive returns, while sentiment is slightly negatively skewed, reflecting more frequent mild pessimism. Kurtosis results indicate that ALSI returns are leptokurtic, displaying heavy tails and volatility clustering. In contrast, sentiment is platykurtic, suggesting thinner tails and less extreme outliers, aligning with findings by Mudinas et al. (2019).

Table 1: Summary Statistics

	JSE ALSI	Twitter Sentiment
Mean	2.494	0.398
Std. Dev.	0.151	0.846
COV	0.060	2.134
Skewness	1.339	-0.332
Kurtosis	23.451	1.810
No. Obs.	789	
Frequency	Daily	

Source: Author's own work. *Note:* The term Std. Dev. Means standard deviation, while the COV is calculated as: Std. Dev./Mean.

The relatively high dispersion in sentiment corresponds with observed spikes around key market events. During the COVID-19 lockdowns (March 2020), the sentiment index recorded its sharpest decline, whereas the market recovery of mid-2021 coincided with a sustained rise in positive sentiment. Short-term volatility in sentiment also increased following the July 2021 civil unrest and during periods of severe electricity shortages (2022 to 2023), underscoring the sensitivity of investor mood to domestic disruptions and macroeconomic uncertainty.

Diagnostic tests reported in Table 2 confirm the presence of conditional heteroscedasticity and serial correlation in the JSE ALSI returns, validating the use of GARCH-type models. Both the ARCH-LM and Ljung–Box statistics reject the null hypotheses of no ARCH effects and no autocorrelation across multiple lag lengths, indicating that volatility is time-varying and clustered.

Table 2: Conditional Heteroscedasticity & Autocorrelation tests

	JSE ALSI
Conditional Heteroscedasticity test	
ARCH-LM (10)	75.475***
ARCH-LM (20)	89.547***
ARCH-LM (30)	97.475***
Ljung-Box Autocorrelation tests	
Q-stat (10)	2726.725***
Q-stat (20)	3274.621***
Q-stat (30)	4472.374***
Q2-stat (10)	3372.736***
Q2-stat (20)	3731.864***
Q2-stat (30)	3973.272***

Note: The applied tests consist of F-statistics for the ARCH test and Ljung–Box Q-statistics for the autocorrelation test, each at three lag lengths (k = 10, 20, and 30). The null hypotheses of no conditional heteroscedasticity (for the ARCH test) and no serial correlation (for the autocorrelation

test) are tested. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively, signifying rejection of the respective null hypotheses.

Table 3: Unit root test estimations

Test	Specification	JSE ALSI	Twitter Sentiment
ADF: Level	Constant	1.082***	2.497***
	Trend & Constant	0.842	1.915
ADF: First Difference	Constant	-6.384***	-7.128***
	Trend & Constant	-6.912***	-7.482***
KPSS: Level	Constant	2.846***	1.372***
	Trend & Constant	0.863	0.936
KPSS: First Difference	Constant	0.114	0.092
	Trend & Constant	0.071	0.066
Order		I(1)	I(1)

Source: Author's own work. *Note:* Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively, signifying rejection of the respective null hypotheses.

The unit root test results reported in Table 3 indicate that both the JSE ALSI and Twitter sentiment series are non-stationary in levels but become stationary after first differencing. Under the ADF test, neither series rejects the null of a unit root at the level across the constant and trend specifications. In contrast, the first-difference statistics are highly significant, confirming the stationarity of the series at I(1). The KPSS results corroborate these findings, as stationarity is rejected at the 5% level but not after first differencing. Collectively, the ADF and KPSS tests confirm that both variables are integrated of order one, thereby supporting their suitability for subsequent volatility modelling.

Table 4: Zivot-Andrews estimations

	JSE ALSI	Twitter Sentiment
Model A	-5.928***	-4.562**
Model B	-6.103***	-4.781**
Model C	-6.452***	-5.214***
Break	2013Q4	2018Q2
Lag Length	2	1

Source: Author's own work. *Notes:* Statistical significance of tests at 1%, 5%, and 10% levels, denoted by ***, **, and *, respectively.

The Zivot–Andrews test results in Table 4 provide evidence of significant structural breaks in both the JSE ALSI and Twitter sentiment series. Model C, which allows for breaks in both the intercept and trend, yields the most decisive rejection of the unit root with break null hypothesis and is therefore selected as the preferred specification. The estimated break for the JSE ALSI occurs in 2013Q4, coinciding with heightened macroeconomic and financial uncertainty in South Africa. In contrast, the break in Twitter sentiment is identified in 2018Q2, a period characterized by increased political and economic uncertainty and heightened public discourse. The presence of structural breaks indicates that both return dynamics and investor sentiment evolve in an environment marked by regime shifts rather than stable trends. This finding further justifies the use of GARCH-type models, particularly asymmetric specifications such as GJR-GARCH and E-GARCH, which are well-suited to capturing nonlinear volatility dynamics and abrupt changes associated with periods of heightened uncertainty.

4.2. ARMA-GARCH Results

4.2.1. Model Selection

Model selection results based on the Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (SBIC), and Hannan–Quinn Criterion (HQC), reported in Table 5, indicate that the GJR-GARCH-M specification provides the best fit for the JSE ALSI, both before and after incorporating Twitter sentiment. The "-M" notation denotes the inclusion of conditional variance in the mean equation, as specified in Equations (3) and (4), while lower information criterion values indicate superior model fit. The superior performance of the GJR-GARCH-M model reflects its ability to capture asymmetric volatility dynamics, whereby adverse shocks generate larger increases in volatility than positive shocks of equal magnitude, a phenomenon commonly referred to as the leverage effect.

Table 5: Model Selection

		AIC	BIC	HQIC
Model	Augmentation	Information Criterion		
GARCH-M	Unaugmented	-1.538	-1.574	-1.560
	+ TwitSent	-1.775	-1.810	-1.797
E-GARCH-M	Unaugmented	-1.533	-1.575	-1.559
	+ TwitSent	-1.779	-1.821	-1.805
GJR-GARCH-M #	Unaugmented	-1.563*	-1.605*	-1.589*
	+ TwitSent	-1.782*	-1.823*	-1.807*

Source: Author's own work. *Note:* * denotes the model with the best (lowest) value for each information criterion, # indicates the model selected.

4.2.2. Mean and Variance Equation Results

Based on the information criteria in Table 5, the GJR-GARCH-M specification is selected as the preferred model. Table 6 reports the estimated parameters of the selected GJR-GARCH-M model, corresponding to the conditional mean equation in Equations (3) and (4) and the conditional variance equation in Equation (7), both with and without the inclusion of Twitter sentiment.

Table 6: Selected Model Outputs

	Unaugmented	+ TwitSent
Selected Model	GJR-GARCH-M	GJR-GARCH-M
Conditional Mean Equation		
μ	0.476***	0.568***
α	0.997***	0.990***
ν	-0.907***	-0.909***
θ	0.735***	0.936***
$\emptyset$	-----	0.917***
Conditional Variance Equation		
ω	0.002***	0.002***
β	0.028***	0.029***
λ	0.918***	0.919***
δ	0.465***	0.558***
φ	0.886***	0.698***
$\beta + \lambda$	0.947	0.948
BIC	-1.574	-1.810

Source: Author's own work. Notes: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

All mean equations were modelled as ARMA(1,1) processes. The GJR-GARCH-M specification includes conditional variance in the mean equation, consistent with Equations (3)-(4).

The positive and statistically significant constant term (μ) in both models (0.476 and 0.568), the results suggest persistent positive mean returns on the JSE ALSI. The autoregressive component (α) remains near unity, implying strong return persistence; however, standard diagnostic tests and stationarity conditions confirm that this persistence reflects inherent market dynamics, while the moving-average term (ν) is negative and significant, indicating short-term correction after shocks. Crucially, the sentiment coefficient $(\emptyset = 0.917)$ is positive and significant, confirming that higher Twitter sentiment is associated with higher contemporaneous returns. Although sentiment measures from Bloomberg differ in methodology from those used by Smales (2014) or Maree and Johnston (2015), which relied on manually constructed mood indices, the direction and short-term influence are comparable, supporting the robustness of sentiment's predictive capacity.

The variance equation confirms the presence of volatility clustering and asymmetry. The small but significant constant (ω) indicates a low unconditional variance, while the significant and persistent λ parameter (≈ 0.92) reflects gradual mean reversion. The leverage term (δ) increases after including sentiment (0.465 to 0.558), suggesting that social-media sentiment amplifies market sensitivity to adverse shocks. The sentiment coefficient in the variance equation $(\varphi = 0.698)$ is positive and highly significant, demonstrating that heightened sentiment intensity, whether optimistic or pessimistic, correlates with higher conditional volatility.

These findings are consistent with Bollen et al. (2011), who reported that collective Twitter mood predicts market volatility, and Zhang et al. (2018), who observed that sentiment-augmented volatility models outperform traditional ones. The results imply that digital investor sentiment acts as an additional source of market noise and behavioral amplification.

The results collectively indicate that Twitter sentiment exerts a significant, positive influence on both returns and volatility in the South African equity market. This underscores the behavioral responsiveness of investors to digital information streams in an emerging-market context characterized by relatively lower liquidity and transparency.

Periods of extreme volatility in the model align with significant economic and political events that shaped public sentiment. The COVID-19 lockdown period saw sharp declines in sentiment and elevated volatility, reflecting widespread uncertainty. The July 2021 unrest produced a pronounced negative sentiment shock, while persistent energy supply constraints in 2022–2023 contributed to heightened volatility and sustained pessimism. These event-driven correlations illustrate that social media sentiment captures real-time investor reactions to macro-level shocks.

The significant leverage and sentiment effects suggest that negative information spreads faster and has a more substantial market impact than positive sentiment of equal magnitude—a finding consistent with behavioral-finance theories of loss aversion and information asymmetry. Moreover, the persistence of volatility ($\beta + \lambda \approx 0.95$) indicates slow shock dissipation, confirming that sentiment-driven reactions can prolong market uncertainty.

Finally, compared with earlier South African studies (e.g., Maree & Johnston, 2015; Nel & du Toit, 2023), this paper extends the literature by incorporating a larger, more recent dataset (2016 to 2023) and by explicitly modelling volatility asymmetry. While prior research demonstrated predictive links between sentiment and returns, this study shows that sentiment also plays a statistically significant role in explaining time-varying volatility, particularly during periods of stress.

5. Conclusions

This study contributes to the growing literature on behavioral finance and volatility modelling by integrating high-frequency, Twitter-derived investor sentiment into GARCH-type frameworks for the JSE ALSI. While GARCH-type models are well established in volatility analysis, the novelty of this study lies not in the model choice itself but in the alternative sentiment measure employed, real-time, NLP-based sentiment data from Twitter, as provided by Bloomberg. Unlike conventional sentiment indices based on surveys or market-implied indicators, this measure captures unfiltered, spontaneous investor opinion, thereby offering an additional behavioral dimension for understanding volatility in an emerging market setting.

Although previous studies have examined the link between investor sentiment and volatility in South Africa, they have typically relied on aggregate sentiment indices such as consumer confidence or media tone (e.g., Rupande et al., 2019; Gupta et al., 2023; Hiramoney et al., 2024; Moodley et al., 2025; Muzindutsi et al., 2023). By contrast, the present study uses social-media-based sentiment, which provides more granular and high-frequency behavioral signals. This differentiates it from earlier research and enhances its contribution to both the behavioral finance and econometric literature.

Empirically, the findings demonstrate that incorporating Twitter sentiment significantly improves the explanatory power of volatility models. Among the estimated specifications, the GJR-GARCH-M model augmented with sentiment proved most effective in capturing asymmetric volatility, confirming that negative sentiment amplifies volatility more strongly than positive sentiment, a manifestation of the leverage effect. Furthermore, the significant, positive sentiment coefficient in the variance equation indicates that heightened social-media activity, regardless of polarity, coincides with periods of elevated market uncertainty. These outcomes align with international evidence (Bollen et al., 2011; Zhang et al., 2018) and reinforce the behavioral nature of volatility transmission in emerging markets, where information asymmetry and retail investor dominance are more pronounced.

From a practical perspective, the results underscore the importance of monitoring digital sentiment signals as inputs for investment decision-making, risk management, and policy formulation. For market participants, Twitter sentiment provides a real-time, forward-looking indicator that can enhance short-term trading and volatility forecasting, particularly in less efficient markets like South Africa. For regulators, tracking social-media-based sentiment alongside traditional macroeconomic indicators could help identify early signs of market stress and improve systemic risk monitoring.

Nevertheless, the findings must be interpreted in light of several limitations. First, while social media provides rich behavioral data, it also introduces risks of misinformation, fake news, and cognitive bias, which can distort sentiment measures and amplify herding effects. Second, the study's reliance on aggregate polarity scores does not fully capture nuanced emotional states (e.g., fear, anger, or optimism) that might better explain investor reactions. Third, only Twitter (now X) data were analyzed; incorporating alternative digital sources such as Facebook, Reddit, or financial news platforms could provide a more comprehensive picture of sentiment flows. Fourth, due to space and data constraints, graphical diagnostics (e.g., actual vs. fitted plots, cross-correlation analyses, and out-of-sample forecasts) were excluded from this version but were examined during estimation and confirmed improved model fit. Future work could extend the analysis to explicitly test the predictive power of sentiment using rolling-window or regime-switching models, allowing comparison across different phases of the economic cycle.

Finally, future research should explore subperiod and sector-level effects to determine whether the sentiment-volatility relationship varies during stress episodes such as the COVID-19 pandemic, political instability, or energy crises. Extending the analysis across multiple emerging markets could further validate the robustness and transferability of these findings.

In conclusion, this study provides novel evidence that Twitter-derived sentiment is an economically and statistically significant determinant of stock market volatility in South Africa. By integrating digital behavioral data into traditional econometric frameworks, it bridges the gap between market microstructure and investor psychology, offering a more complete understanding of volatility dynamics in emerging markets.

Appendix A1: Unaugmented model results

	Unaugmented		
Model	GARCH-M	E-GARCH-M	GJR-GARCH-M#
Conditional Mean Equation			
μ	0.816***	0.373***	0.476***
α	0.996***	0.998***	0.997***
ν	-0.907***	-0.909***	-0.907***
θ	0.824***	0.521***	0.735***
Conditional Variance Equation			
ω	0.002***	0.001***	0.002***
β	0.019***	0.021***	0.028***
λ	0.902***	0.913***	0.918***
δ	-----	-0.337***	0.465***
φ	0.684***	0.742***	0.886***
$\beta + \lambda$	0.921	0.935	0.947
SBIC	-1.274	-1.385	-1.574

Source: Author's own work. *Notes:* ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. All mean equations were modelled as ARMA(1,1) processes. Unaugmented refers to the model specification without Twitter sentiment.

Appendix A2: Augmented model results

	Augmented		
Model	GARCH-M	E-GARCH-M	GJR-GARCH-M#
Conditional Mean Equation			
μ	0.274***	0.453***	0.568***
α	0.998***	0.999***	0.990***
ν	-0.905***	-0.902***	-0.909***
θ	0.904***	0.848***	0.936***
$\emptyset$	0.725***	0.886***	0.917***
Conditional Variance Equation			
ω	0.001***	0.003***	0.002***
β	0.018***	0.022***	0.029***
λ	0.913***	0.918***	0.919***
δ	-----	-0.227***	0.558***
φ	0.374***	0.643***	0.698***

| β + λ | 0.932 | 0.941 | 0.948 |
| SBIC | -1.147 | -1.284 | -1.810 |

Source: Author's own work. *Notes:* ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. All mean equations were modelled as ARMA(1,1) processes. Augmented refers to the model specification that includes Twitter sentiment.

References

Agarwal, S., Kumar, S., & Goel, U. (2021). Social media and the stock markets: an emerging market perspective. *Business Economics and Management (JBEM)*, 22(6), 1614-1632. https://doi.org/10.3846/jbem.2021.15619

Alagidede, P. & Panagiotidis, T. (2009). Modelling stock returns in Africa's emerging equity markets. *International Review of Financial Analysis*, 18(1-2), 1-11. https://doi.org/10.1016/j.irfa.2009.02.001

Alomari, E., Katib, I., Albeshri, A., & Mehmood, R. (2021). COVID-19: Detecting government pandemic measures and public concerns from Twitter Arabic data using distributed machine learning. *International journal of environmental research and public health*, 18(1), 282. https://doi.org/10.3390/ijerph18010282

Amirshahi, B. & Lahmiri, S. (2023). Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies. *Machine Learning with Applications*, 12, 100465. https://doi.org/10.1016/j.mlwa.2023.100465

Andleeb, R. (2024). Investor sentiment dynamics and returns in emerging equity markets (Doctoral dissertation, Capital University).

Aras, S., (2021). Stacking hybrid GARCH models for forecasting Bitcoin volatility. *Expert Systems with Applications*, 174, 114747. https://doi.org/10.1016/j.eswa.2021.114747

Athari, S.A., Khalid, A.A., & Syed, Q.R. (2024). Twitter-based economic uncertainty and the US energy market: An investigation using wavelet coherence. *Energy Research Letters*. https://doi.org/10.46557/001c.90917

Bhatt, S., Ghazanfar, M., & Amirhosseini, M. (2023). Machine learning based cryptocurrency price prediction using historical data and social media sentiment. *Computer Science & Information Technology* (CS & IT), 13(10), 1-11. https://doi.org/10.5121/csit.2023.131001

Bloomberg. (2025) Twitter Sentiment Analytics: Firm-Level Market Sentiment Scores. Bloomberg Terminal. [Online]. Available at: Bloomberg Terminal (Accessed: 21 May 2025).

Bollen, J., Mao, H. & Zeng, X., (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 542-547. https://doi.org/10.2307/1925546

Bouri, E., Gupta, R., & Roubaud, D. (2019). Herding behavior in cryptocurrencies. *Finance Research Letters*, 29, 216-221. https://doi.org/10.1016/j.frl.2018.07.008

Brooks, C. (2019). *Introductory Econometrics for Finance. 4th ed*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108524872

Chen, C.W., Watanabe, T., & Lin, E.M. (2023). Bayesian estimation of realized GARCH-type models with application to financial tail risk management. *Econometrics and Statistics*, 28, 30-46. https://doi.org/10.1016/j.ecosta.2021.03.006

Deveikyte, J., Geman, H., Piccari, C., and Provetti, A. (2022). A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence*, 5, 836809. https://doi.org/10.3389/frai.2022.836809

Duz Tan, S. & Tas, O., (2021). Social media sentiment in international stock returns and trading activity. Journal of Behavioral Finance, 22(2), 221-234. https://doi.org/10.1080/15427560.2020.1772261

Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007. https://doi.org/10.2307/1912773

Fonou-Dombeu, N.C., Nomlala, B.C., & Nyide, C.J. (2024). Investigating the effect of investor sentiment on stock return sensitivity to fundamental factors: case of JSE-listed companies. *Cogent Business & Management*, 11(1), 2353846. https://doi.org/10.1080/23311975.2024.2353846

Glosten, L.R., Jagannathan, R. & Runkle, D.E., (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779-1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Graham, C. & Stough, R. (2025). Consumer perceptions of AI chatbots on Twitter (X) and Reddit: an analysis of social media sentiment and interactive marketing strategies. *Journal of Research in Interactive Marketing*. https://doi.org/10.1108/JRIM-05-2024-0237

Guijarro, F., Moya-Clemente, I. & Saleemi, J. (2019). Liquidity risk and investors' mood: linking financial market liquidity to sentiment analysis via Twitter in the S&P 500 index, *Sustainability*, 11(24), 7048. https://doi.org/10.3390/su11247048

Gupta, R., Nel, J., & Pierdzioch, C. (2023). Drivers of realized volatility for emerging countries with a focus on South Africa: Fundamentals versus sentiment. *Mathematics*, 11(6), 1371. https://doi.org/10.3390/math11061371

Hiramoney, P., Muguto, L., & Muguto, H.T. (2024). Investor sentiment, stock market returns, and volatility: evidence from the Johannesburg Stock Exchange. *Business Resilience, Continuity and Regeneration*, 7.

Jasim, Y.A., Saeed, M.G., & Raewf, M.B. (2022). Analyzing social media sentiment: Twitter as a case study. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 11(4), 427-450. https://doi.org/10.14201/adcaij.28394

Javed, F. & Mantalos, P. (2013). GARCH-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, 42(8), 1917-1933.

Johannesburg Stock Exchange (2025) (n.d.) 'Market Capitalization'. Available at: https://www.jse.co.za/learn-how-to-invest/accessing-market/market-capitalisation (Accessed: 10 March 2025).

Kandasamy, T. & Bechkoum, K. (2024). Impact of Social Media Sentiments on Stock Market Behavior: A Machine Learning Approach to Analyzing Market Dynamics. *Journal of Comprehensive Business Administration Research*. https://doi.org/10.47852/bonviewJCBAR42022006

Kim, J.M., Kim, D.H., & Jung, H. (2021). Estimating yield spreads volatility using GARCH-type model. *North American Journal of Economics and Finance*, 57, 101396. https://doi.org/10.1016/j.najef.2021.101396

Koo, E. & Kim, G., (2022). A hybrid prediction model integrating GARCH models with a distribution manipulation strategy based on lstm networks for stock market volatility. *IEEE Access*, 10, 34743-34754. https://doi.org/10.1109/ACCESS.2022.3163723

Lawrence, B., Moodley, F., & Ferreira-Schenk, S. (2024). Macroeconomic determinants of the JSE size-based industries' connectedness: Evidence from changing market conditions. *Cogent Economics & Finance*, 12(1), 2397454. https://doi.org/10.1080/23322039.2024.2397454

Lee, C.B., Io, H.N., & Tang, H. (2022). Sentiments and perceptions after a privacy breach incident. *Cogent Business & Management*, 9(1), 2050018. https://doi.org/10.1080/23311975.2022.2050018

Maree, S. & Johnston, K. (2015). Critical insights into the design of big data analytics research: how Twitter" moods" predict stock exchange index movement: informatics for development. *The African Journal of Information and Communication*, 2015(15), 53-67. https://doi.org/10.23962/10539/20330

Mathebula, N. (2025). *Debt Crises and Sovereign Risk in Africa*. Taylor & Francis.

Mendoza-Urdiales, R.A., Núñez-Mora, J.A., Santillán-Salgado, R.J. & Valencia-Herrera, H., (2022). Twitter sentiment analysis and influence on stock performance using transfer entropy and EGARCH methods. *Entropy*, 24(7), p.874. https://doi.org/10.3390/e24070874

Metta, S., Madhavan, N., & Narayanan, K.K. (2022). Power of 280: Measuring the Impact of Elon Musk's Tweets on the Stock Market. *Ushus Journal of Business Management*, 21(1), 17-43. https://doi.org/10.12725/ujbm.58.2

Moodley, F., Ferreira-Schenk, S., & Matlhaku, K. (2025). The Effects of Investor Sentiment on Stock Return Indices Under Changing Market Conditions: Evidence from South Africa. International *Journal of Financial Studies*, 13(2), 70. https://doi.org/10.3390/ijfs13020070

Mpofu, S., (2021). Social media memes as commentary in health disasters in South Africa and Zimbabwe. Digital humour in the COVID-19 pandemic: *Perspectives from the global south*, 19-45. https://doi.org/10.1007/978-3-030-79279-4_2

Mudinas, A., Zhang, D., & Levene, M. (2019). Market trend prediction using sentiment analysis: lessons learned and paths forward. arXiv preprint arXiv:1903.05440.

Muguto, L. & Muzindutsi, P.F., (2022). A comparative analysis of the nature of stock return volatility in BRICS and G7 markets. *Journal of Risk and Financial Management*, 15(2), 85. https://doi.org/10.3390/jrfm15020085

Muzindutsi, P.F., Apau, R., Muguto, L., & Muguto, H.T. (2023). The impact of investor sentiment on housing prices and the property stock index volatility in South Africa. *Real Estate Management and Valuation*, 31(2), 1-17. https://doi.org/10.2478/remav-2023-0009

Ncume, V., van Zyl, T.L., & Paskaramoorthy, A. (2022). Volatility forecasting using Deep Learning and sentiment analysis. arXiv preprint arXiv:2210.12464.

Nel, G.F. & du Toit, Y., (2023). Stakeholders' engagement in 280 characters: Evidence from JSE-listed companies. *South African Journal of Economic and Management Sciences*, 26(1), 1-12. https://doi.org/10.4102/sajems.v26i1.4964

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370. https://doi.org/10.2307/2938260

Nisar, T.M. & Yeung, M., (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. *The journal of finance and data science*, 4(2), 101-119. https://doi.org/10.1016/j.jfds.2017.11.002

Nyakurukwa, K. & Seetharam, Y. (2022). Can a 280-character message explain stock returns? Evidence from South Africa. *Managerial Finance*, 48(4), 663-686. https://doi.org/10.1108/MF-12-2021-0598

Oh, C. and Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement.

Oliveira, N., Cortez, P. & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume, and survey sentiment indices. *Expert Systems with Applications*, 73, 125-144. https://doi.org/10.1016/j.eswa.2016.12.036

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. & Mozetič, I., (2015). The effects of Twitter sentiment on stock price returns. *PloS one*, 10(9), e0138441. https://doi.org/10.1371/journal.pone.0138441

Rao, T. & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. In State-of-the-art applications of social network analysis (227-247). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05912-9_11

Ridhwan, K.M. & Hargreaves, C.A., (2021). Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. International *Journal of Information Management Data Insights*, 1(2), 100021. https://doi.org/10.1016/j.jjimei.2021.100021

Rupande, L., Muguto, H.T., & Muzindutsi, P.F. (2019). Investor sentiment and stock return volatility: Evidence from the Johannesburg Stock Exchange. *Cogent Economics & Finance*, 7(1), 1600233. https://doi.org/10.1080/23322039.2019.1600233

Ruths, D. & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346(6213), 1063-1064. https://doi.org/10.1126/science.346.6213.1063

Seetharam, Y. and Nyakurukwa, K., (2024). Headlines or Hashtags? The battle in social media for investor sentiment in the stock market. International *Journal of Information Management Data Insights*, 4(2), 100273. https://doi.org/10.1016/j.jjimei.2024.100273

Smales, L.A. (2014). News sentiment in the gold futures market. *Journal of Banking & Finance*, 49, .275-286. https://doi.org/10.1016/j.jbankfin.2014.09.006

Souza, D.M.S.D. & Martins, O.S. (2024). Brazilian stock market performance and investor sentiment on Twitter. *Revista de Gestão*, 31(1), 18-33. https://doi.org/10.1108/REGE-07-2021-0145

Sprenger, T.O., Tumasjan, A., Sandner, P.G., & Welpe, I.M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926-957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Steyn, D.H., Greyling, T., Rossouw, S. & Mwamba, J.M., (2020). Sentiment, emotions, and stock market predictability in developed and emerging markets (No. 502). *Global Labor Organization* (GLO).

Sul, H.K., Dennis, A.R., & Yuan, L. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48(3), 454-488. https://doi.org/10.1111/deci.12229

Tetlock, P.C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tiwari, S., Verma, A., Garg, P., & Bansal, D. (2020), March. Social media sentiment analysis on Twitter datasets. In 2020, the 6th international conference on advanced computing and communication systems (ICACCS) (925-927). *IEEE*. https://doi.org/10.1109/ICACCS48705.2020.9074208

Zhang, X., Qu, S., Huang, J., Fang, B., & Yu, P. (2018). Stock market prediction via multi-source multiple instance learning. *IEEE Access*, 6, 50720-50728. https://doi.org/10.1109/ACCESS.2018.2869735

Zheludev, I., Smith, R. & Aste, T. (2014). When can social media lead financial markets? *Scientific Reports*, 4(1), 4213. https://doi.org/10.1038/srep04213