Impact of Trading Volume on Stock Price Volatility in the Nairobi Securities Exchange

OBULINJI, MUNIRA OKUTOYI

146000

Submitted in partial fulfillment of the requirements for the Degree of Financial Engineering at Strathmore University

Strathmore Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

APRIL 2025

This Research Project is available for Library use on the understanding that it is copyright material and that no quotation from the Research Project may be published without proper acknowledgment.

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Project contains no material previously published or written by another person except where due reference is made in the Research Project itself.

.... [Name of Candidate]

.... [Signature]

.... [Date]

This Research Project has been submitted for examination with my approval as the Supervisor.

.... [Name of Supervisor]

.... [Signature]

.... [Date]

Strathmore Institute of Mathematical Sciences

Strathmore University

ABSTRACT

The study investigates the exact relationship between trading volume and the volatility of the stock price at the Nairobi Securities Exchange. It investigates how changes in trading volume relate to the changes in the movement of the stock price and the causality effect between these two variables. Using past trading data at NSE, the study will apply econometric modeling, including GARCH-X (1,1), to quantify the size of trading volume on daily price volatility. The findings are to indicate a significant relationship that reflect market dynamics and information flow. These aspects are extremely important to facilitate investors, policymakers, and market analysts in understanding market efficiency and thus proposing management of risk and stabilization of market conditions.

Contents

Table of Figures

CHAPTER ONE

INTRODUCTION

1.1Background Study

1.1.1 Introduction

The topic constitutes one of the major research interests in finance regarding the link between market activity and volatility. Empirical studies indicate that trading volume is not only an accompanying factor in price dynamics but an influential factor able to drive volatility. Whereas the studies imply that high volatility is usually accompanied by high volumes of trade, the implication is that an informed trader whose actions reflect market sentiments and information flow usually exists. This relationship has been documented in developed markets and understanding it within the context of emerging markets such as the Nairobi Securities Exchange (NSE), is particularly crucial. Less liquidity, a few institutional investors, and a greater market share held by retail investors generally make emerging markets volatile and increase the impact of trading volume on price change. Consequently, an investigation into this relationship in NSE would provide insight into issues of high priority to policymakers and investors in general, namely, market efficiency, investor behavior, and price discovery. More importantly, traders will better understand how trading activities have come to influence volatility in stock prices to manage investor risk and adjust their trading activities for better decision-making whenever there is a fluctuation in prices. This will also be very instrumental information to regulatory bodies in Kenya that seek to ensure greater stability of the capital markets for the protection of investors.

1.1.2 Brief description of trading volume and stock price volatility concepts

Trading volume is the number of shares traded of a given security or, in general, of an entire market during the given period (Karpoff J. , 1987). In NSE, the trading volume is depicted by a representation showing how companies are active and liquid in it and acts as an indication of investor interest. It has been observed in the research on the emerging markets, of which NSE is one, that its trading activities tend to be influenced by both the domestic and the global economic conditions. Unlike mature markets that are institutional investor-driven, the NSE comprises both retail investors and institutional investors. Retail investors are generally more sensitive to news and price changes due to several key factors such as limited access to information and analysis, low-risk tolerance, behavioral biases, and lower levels of capital. As a result, the trading volumes in the NSE have often witnessed wild fluctuations whenever there is a major economic event or some vital policy announcement.

The volatility of stock price about trading volume is an issue of great interest to researchers and practitioners alike, especially in emerging markets where the market mechanism and investors' behavior may also be quite different from that in developed markets (Schwert, 1990). Some studies in the NSE context have reported a positive relationship between the level of trading volume and volatility, implying that the larger the trading volumes, the greater the volatility in the price (Ngugi, 2003). This correlation usually arises from the increased trading by both the informed and noise traders, whose reactions to new information and price changes tend to differ. This aspect is so important in understanding information efficiency in the NSE, as trading volume gives a stand-in measure that can easily be seen in terms of investor responses to market information and general liquidity.

1.1.3 Investors at the Nairobi Securities Exchange

The Nairobi Securities Exchange (NSE), established in 1954, is a prominent African exchange located in Kenya, a rapidly growing economy in Sub-Saharan Africa. With a long history of listing both equity and debt securities, NSE provides a high-class trading platform for both local and international investors interested in Kenya and Africa's economic expansion. The NSE underwent demutualization and self-listing in 2014, and its leadership team, composed of top capital markets experts from across Africa, is dedicated to fostering innovation, diversification, and operational excellence at the Exchange.

1.2 Problem Statement

The causes of volatility in stock prices are of the essence to be known in financial markets, as this factor directly affects investors and policymakers in terms of portfolio valuation, risk management, and stability of the market. In as much as trading activities have witnessed considerable growth over the years at the Nairobi Securities Exchange, the exact relationship between volume and volatility has not been identified, with few turning focus on the Kenyan market. Changes in trading volume are often hypothesized to be one of the major determinants of price fluctuation, although empirical evidence is required for substantial proof, especially in emerging markets such as Kenya, where market dynamics may be different from developed economies. The study seeks to fill this existing literature gap by investigating the effect of trading volume on stock price volatility in the NSE. This research will further explore the fluctuation in trading volume, which corresponds to the change in daily price volatility and can provide meaningful insight into the strength and nature of this relationship. Further, this research will check the direction of causality between trading volume and the movement in the stock price to verify if high trading volumes cause price volatility or if volatility drives trading volume. Based on this relation, the research will develop a deeper understanding of market efficiency and trading strategies, as well as regulatory policies, from the perspective of the NSE. Precisely, by using econometric modeling, this paper presents the quantitative analysis of the volume-volatility relation that may serve as a basis for better investment decisions and more efficient market regulation.

1.3 Research Objectives

I. To analyze the effect of trading volume on stock price volatility in the Nairobi Securities Exchange (NSE).

II. To examine the causality between trading volume and stock price movements in the NSE

1.4 Research Question

I. How does the trading volume affect stock price volatility in the NSE?

II. What is the causality between trading volume and stock price movements in the NSE?

1.5 Significance of Study

It is important to study the impact of trading volume on volatility in the stock prices at NSE against a backdrop of distinctive market dynamics in developing economies. Trading volume is an important indicator of market activities and investor confidence; it almost always reflects the fundamental changes in demand and supply for securities. Previous studies have identified and documented the positive relationship between trading volume and price volatility, implying that greater trading activity can exaggerate the movements of prices and hence influence investors' perceptions and attitudes toward risk. Investors need to understand the nature of such a relationship to help them develop appropriate trading strategies that also take care of the risk of their investments. Besides, the NSE is an ideal market to investigate this relationship due to the heterogeneity in its market participants and level of liquidity. Such studies can also enlighten market regulators and policymakers about the areas of probable market inefficiencies and how a regulatory framework may be useful for improving stability in the market.

Thirdly, the trading volume and volatility interface will give a more proper insight into behavioral finance principles like overconfidence, herding, and overreaction known to occur in emergent markets. These would involve the search for a pattern through the examination of historical data and econometric modeling, which sets the base for general understanding. This, therefore, will make the research especially relevant in the context of the changing landscape of the NSE, with more foreign investment and technological advancement in the way trading takes place. Ultimately, this research will contribute to the development of insights within the academic literature that have practical applications for traders, portfolio managers, and institutional investors who must thread their way with difficulty in the Kenyan capital market. The understanding of how trading volume affects price volatility will conclusively place the stakeholders in an advantageous position to make better decisions, hence increasing efficiency and confidence in the market.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Researchers in finance, especially in emerging markets such as the Nairobi Securities Exchange, consider trading volume and volatility of stock prices significant measures of activity and investors' psychology. Several studies have indicated that high trading volume is usually accompanied by high volatility, which affects market dynamics and influences investor behavior. This literature review, therefore, aims to synthesise the current research on this relationship in the NSE, how trading volume volatility causes price volatility, and the implication this has for the various market participants with a view of coming up with appropriate trading strategies to enhance efficiency in the Kenyan capital market.

2.2 Theoretical Literature Review

2.2.1 Efficient Market Hypothesis (EMH)

(Fama, 1970) suggests that at any one single moment, the price of a stock reflects all available information. This is a theory based on the concept of the impossibility of obtaining returns consistently greater than the average market return on a risk-adjusted basis. Any new information that may appear is said to be promptly incorporated into the stock prices. Trading volume reflects how widely information is spread in the market, and it contributes to the understanding of how various market participants influence the formation of prices. In the case of the Nairobi Securities Exchange, for instance, EMH could be used to analyze how efficiently fast trading volume reacts to new information, influencing volatility in stock prices. It can be utilized to appraise whether the market is efficient or there is an arbitrage opportunity, using the pattern of the trading volume. However, EMH itself has its limitations; it is based on the belief that markets are always perfectly efficient and discards behavioral biases, irrational investor actions that sometimes affect stock price movements and trading volume.

2.2.2 Mixture Distribution Hypothesis (MDH)

(Harris, 1987) assumes that both trading volume and stock price volatility have a positive relationship that arises because both are influenced by the same factor which is the flow of information. It explains how price change and trading volume are related, underlines information asymmetry in the dynamics of the market, and can also be applied to explain volatility clustering as commonly observed in most financial markets. With this theory, the analysis of how trading volume in the Nairobi Securities Exchange correlates with the volatility in the price of stocks, especially during periods of high news flow or events, can now be done. This will help in pointing out the patterns in trading behavior that always precede a price change. The theory has limitations whereby one of them entails ignoring market microstructure factors whereby it does not consider liquidity constraints or transaction costs which also influence the two variables independently of information.

2.2.3 Volume Price Relationship Theory

Volume Price Relationship Theory (Karpoff J. M., 1987): It is about the direct relationship between trading volume and price change, whereby the higher the trading activity, the higher the volatility. It states that high volumes of trade normally precede or follow most major price movements. It provides insight into investor psychology and market sentiments; it is utilized to forecast future movements of prices as per volume trends and helps in formulating a trading strategy on volume indicators. This theory can be applied in the Nairobi Securities Exchange in testing how trading activity affects volatility in stock prices, especially moments of volatility and what trading volume exists in that moment. It may also be employed to test how events within the market affect the investor's reaction. While it emphasizes the correlation between trading volume and price changes, it does not fully account for the role of external factors such as macroeconomic conditions or news shocks that may also drive volatility.

2.2.4 Information Asymmetry Theory

(Copeland, 1976) assumes that the information of market participants is asymmetrical, which leads to the different trading behavior and the price formation. A hypothesis was put forth to demonstrate that all trading volume could be associated with either informed or uninformed trading. The theory can explain the role of institutional vs. individual retail investors and can explain price jumps when new information becomes available as well as decreases in price when insiders learn of bad news, like earnings reports, before the public. The theory concludes that significant price movements are often preceded or followed by high trading volumes, indicating a strong connection between volume spikes and market volatility. One of its limitations is the tendency to centre around the role of institutional investors as informed traders, perhaps at the expense of recognizing the strong influence of retail investors and their behaviour-driven trading patterns, which may also give rise to volatility independent of asymmetric information.

2.2.5 Behavioral Bias

(Kahneman D. &., 1979) studies the psychological influences and cognitive biases on both the behavior of investors and the subsequent markets outcome. In other words, it implies that the trading volume is influenced by investor sentiment itself, and since price volatility equals the change of price levels caused by trading volume, the former depends on the latter. Besides the traditional capital market theories, it provides a clearer explanation regarding the underlying dynamics driving the market. The theory is also applied to explain different phenomena, such as herding behavior and overreaction to news, and clarifies the influence of market psychology on volume variables. Behavioral finance will determine how investor sentiment and psychology affect trading volume and volatility within stock price at the Nairobi Securities

Exchange. And this would identify moments of irrational trade behavior which could map market inefficiencies. While it effectively explains irrational investor behavior, it does not provide a structured quantitative model for predicting price movements based on trading volume.

2.3 Empirical Literature Review

The final part of the literature review provides an examination of existing research and theory about the influence that trading volume exerts, on the stock price variations in the Nairobi Securities Exchange to fully understand the relationship. It also identifies the weaknesses that exist in the body of literature, therefore, providing a basis for the need for further examination. In addition, it has been argued that the literature review serves to focus the methodology of the research where previous research, and their approaches are examined. The specifics of the NSE enable the development of preliminary expectations for the present research. Last but not the least, the literature review, as the last but one activity of the research, gives room for more knowledge to be enhanced and seeks other areas of investigation.

(Hameed, 2006) examines the interaction between analyst coverage and stock price synchronicity in the emerging markets based on the distribution of information and its effect on market efficiency. Stocks of developing economies form the sample of this study, with samples covering many firms and analyst recommendations. Authors in this study examine the contribution of analyst coverage to stock price co-movements through regression analysis and synchronicity measures derived from the overall market. These results confirm that, for emerging markets, greater analyst coverage reduces price synchronicity by enhancing firm-specific information flow, such that individual fundamentals are reflected better in stock prices and not broad market movements. The study indicates that improved spread of information reduces market-wide co-movement in prices, however, the degree depends on institutional frameworks and market liquidity. These findings suggest that trading volume as a measure of information flow might also affect price volatility in the Nairobi Securities Exchange (NSE) to complement information asymmetry, trading activity, and price behavior. The study has several implications for investor attention and market transparency on stock price behavior in developing economies.

(Vo, 2010) the relationship between liquidity and stock returns is a crucial aspect of asset pricing, particularly in emerging markets like Vietnam. Using share turnover as a proxy for liquidity, this study examines its impact on stock returns within the Fama and French cross-sectional framework. The findings reveal a strong positive relationship between liquidity and stock returns during financial crises, diverging from results in developed markets. This relationship remains significant when incorporating momentum, while firm size shows no significant pricing effect in most regressions. These results underscore the importance of liquidity in financial markets, especially in times of crisis. Policymakers should prioritize reducing trading costs and easing market barriers to enhance liquidity, as increased liquidity improves stock returns. The study reinforces that liquidity is a key determinant of stock pricing and suggests that emerging markets can benefit from policies that foster market depth. Unlike developed markets, where liquidity premiums are often negative, Vietnam's market conditions suggest liquidity plays a crucial role in driving returns.

The research (Mpofu, 2012) utilizes data spanning a from 1988 to 2012 and applies a Granger causality test within a VAR framework to analyze the direction of causality between these variables. The findings indicate that trading volume plays a significant role in predicting price volatility, with the strength of this relationship varying under different market conditions. The results suggest that during periods of high trading activity, the causality from volume to volatility is more prominent. This study highlights how information spread rates and investor behavior differ within the South African market compared to other global markets. These insights contribute to a deeper understanding of market efficiency, portfolio rebalancing strategies, and the stability of financial markets in emerging economies. The study may not account for factors such as institutional investors behaviour which may account for the effect.

(Phansatan, 2012) examined the impact of trading volume of different investor groups on stock return volatility in the Thai stock market, sorting stocks into size-based portfolios. Their findings indicate that proprietary investors, despite their low trading volume, have the greatest impact on short-term market volatility due to their high-frequency trading in large-cap stocks. However, when comparing trading strategies and the correlation with market volatility, institutional investors were most significant regardless of the portfolio size. Furthermore, the study also analyzed the application of contrarian and momentum trading strategies in explaining volatility. Contrary to these two, proprietary and foreign investors apply momentum strategy, while contrarian trading is preferred by institutional and retail investors. To the contrary, momentum buying was found to decrease volatility, while momentum selling heightened it, opposing the standard momentum effect. Likewise, contrarian buying heightened volatility, while contrarian selling decreased it, opposing the standard contrarian effect. These results indicate that the impact of trading volume on market volatility is not exclusively momentum or contrarian driven. Rather, volatility seems to be an outcome of complicated interactions between investor actions and not classical strategy-driven explanations. The research may not account for the current regulatory changes existing in the market currently.

(Chen, Stock returns, volatility, and trading volume: Evidence from the Chinese stock markets., 2001) research the causality between trading volume and price volatility in the emerging Chinese stock market. The population includes stocks traded in the Shanghai and Shenzhen stock exchanges between the years 1996 and 2000. A VAR model that is employed augmented with Granger causality tests, they test for evidence of whether changes in volume can provide a forecast of volatility and vice versa. While causality runs both ways, the effect of volume on volatility is stronger due to the emerging status and less transparency of the market. This research points out that trading volume serves as a very critical indicator of investor sentiment and market stability within an emerging market framework and volume shocks quickly translate into price volatility. Consequences of the findings are that such high levels of volume usually signal instability or speculative trading activity and thus have major implications for the regulatory bodies of these emerging markets.

2.4 Overview of Literature Review

The theoretical literature on the volatility of stock prices and trading volume has developed some advanced frameworks to explain the dynamics of market behavior. Foundational theories such as the Efficient Market Hypothesis and the Mixture of Distributions Hypothesis suggest a linkage between trading volume and price volatility impelled by new information arrival. According to the EMH, markets always incorporate all the available information, and such price fluctuations make a series of rational expectations. On the contrary, (Mixed Distribution Hypothesis) MDH expects that information cannot reach all investors simultaneously, thus creating heterogeneous responses and, therefore, spikes in volume and volatility. Behavioral finance brings in yet another dimension-that of cognitive biases, where herd behavior exaggerates this relationship when traders act as a team in response to perceived trends. Theories of information asymmetry go further in the contribution by suggesting that differences in access to information among market participants create the conditions that cause volume and volatility to move together. It was also a basis to address the causality between volume and volatility, besides the market microstructure factors affecting this relationship.

Empirical literature supports these theoretical foundations amply by examining real-world market data. Evidence supporting a positive relationship between the two ; (Hameed, 2006) proves that greater information disclosure reduces price synchronicity, which implies that higher trading volume in NSE can enhance price efficiency. (Vo, 2010) finds that liquidity positively affects stock returns in Vietnam, in favour of the role of trading activity in stock movement. (Mpofu, 2012) confirms that trading volume predicts volatility in South Africa, especially in active periods, while institutional investor effects are less explored. (Phansatan, 2012) highlights that different investor groups cause volatility in Thailand, and institutional investors have the largest impact. (Chen, Stock returns, volatility, and trading volume: Evidence from the Chinese stock markets., 2001) finds a bidirectional relationship between volatility and trading volume in China, which shows that speculative trading and low transparency amplify price fluctuations. These studies speculate that investors' trading strategies, market liquidity, and trading volume can underlie the NSE volatility patterns, volume being a key marker of market stability and price volatility.

2.5 Research Gap

The current literature on trading volume and stock price volatility in emerging markets has significant gaps, such as limited measures of liquidity, interactions between investor groups and market-specific focus. My investigation of the effect of trading volume on the volatility of stock prices in the Nairobi Securities Exchange (NSE) helps to fill these gaps by employing historical data to examine long-term price volatility trends and trading activity. By targeting an emerging market with distinctive institutional and regulatory environments, my research increases generalizability and expands on earlier findings. This provides insight into market efficiency, price movement, and trading activity in an emerging market such as Kenya, providing information on how trading volume has affected volatility under varying market conditions. Addressing these gaps would hopefully yield greater insight, in more finely subtle detail, into how volume affects volatility across diverse market contexts and trading environments.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

Besides the traditional stock exchanges, the impact of trading volume on volatility has also been considered in a variety of contexts, such as that of emerging markets or sectors, including technology stocks. Generally, (D. Kaya, 2023) and (Özdemir, 2020) provide critical messages about considering trading volume as one of the crucial factors in understanding the behavior and volatility of stock prices. The current research effort therefore uses a strong methodological framework in incorporating trading volume and volatility measures in its bid to further the literature on an all-rounded analysis of their intricate relationship

3.2 Research Design

This work followed a quantitative research design to establish how trading volume influences volatility in the stock prices on the Nairobi Securities Exchange. The quantitative approach, therefore, becomes essential for this research since it makes use of numerical data through statistical methods whose outcomes are aimed at establishing the relationship and correlation between trading volume and the volatility in stock prices. For the data analysis, the use of econometric models, particularly those of Generalized Autoregressive Conditional Heteroskedasticity, captures the volatility clustering along with the dynamic relationship between trading volume and stock price volatility.

3.3 Population and Sampling

The target population of this study encompasses publicly quoted companies in the NSE. The sampling technique is adopted to obtain a representative sample of the companies where trading volume and market capitalization are considered to select the sample. This will include various industry-wise companies to maintain diversity and robustness in the findings. The companies will be provided by the Nairobi Securities Exchange after purchasing the data through their website. The selected companies provided will be used to analyze the objectives and realize the output for the NSE. Data from the stock exchange will cover seven years of trading data from 2012 to 2018, to capture adequate variability in the volume of trade and fluctuation of stock prices.

3.4 Data

Data for this research will be retrieved from the Nairobi Securities Exchange database, which provides extensive financial information regarding daily trading volumes and stock prices, along with other relevant market indicators. Such databases are quite reliable and comprehensive in an analytical sense, with respect to market trends and investor behavior. In this regard, daily closing prices and trading volumes of selected stocks listed on the NSE over a specified period shall be collected. Such variables are important in testing the association that may exist between trading activity and price volatility, hence offering good insight into market efficiency and investor sentiment. The data will be pre-processed to ensure accuracy, consistency, and reliability before any statistical and econometric analysis.

3.5 Data Analysis

3.5.1 Summary Statistics

The impact of trading volume on the volatility of stock prices in the NSE cannot be tested without first computing descriptive statistics that summarize the behavior of these variables. Important measures of central tendency include mean, median, and mode for the average trading volumes and typical price volatilities prevalent in the market. Mean (μ) will indicate the general level of trading activity and volatility; the median will show the middle point in the distribution of the data and present asymmetries. The mode will help in identifying the most frequent trading volumes or volatility levels since it helps in detecting common market behaviors. Dispersion measures, such as the standard deviation (σ), will be used to quantify the trading volume and volatility variability, which will show the degree to which these variables deviate over time.

Skewness and kurtosis will also be calculated to measure the shape of the distribution for trading volume and volatility data. The skewness will indicate whether the data is symmetrically distributed or skewed towards higher or lower values. At the same time, kurtosis will show the presence of extreme values or outliers in the dataset. This goes a long way in building a basic understanding of the influence trading volume has on stock price volatility in the NSE and hence sheds light on market dynamics and investor behavior.

3.5.2 Diagnostic Tests

In establishing the reliability and validity of the econometric models to be used in analyzing the impact of trading volume on the volatility of stock prices at the NSE, several diagnostic tests will be conducted. First, there will be the Augmented Dickey-Fuller test for stationarity of the time series data on trading volume and stock price volatility. The null hypothesis of the ADF test assumes a unit root, meaning the series is nonstationary, since otherwise it might give spurious regression results. The latter view is that, when the null hypothesis is rejected, this assures stationarity, hence making the data fit for analysis. Normality tests are presented via the Jarque-Bera test, which is applied here to check if residuals in a model are normal. This test will check if skewness and kurtosis of data are in line with a normal distribution, an important assumption underlying most econometric models.

Non-normality may signal misspecification of the model which would result in biased estimates of parameters. Third, the Autoregressive Conditional Heteroscedasticity (ARCH) test is performed to find heteroscedasticity in residuals, indicating the variance of errors is not constant across time. Heteroscedasticity is of importance because it can lead to the distortion of standard errors, therefore affecting the inference drawn from a statistical test. When heteroscedasticity becomes established, further model specifications shall include advanced models that can possibly remedy the cited concern. Carrying out these diagnostic tests gives the selected econometric models strength and ensures the model is robust and capable of correctly specifying the trading volume-stock price volatility relationship for the NSE.

3.5.3 Model Selection

Objective I: Examining the Causality Between Trading Volume and Stock Price Movements

To examine the causality between trading volume and stock price movements, the **Vector Autoregressive (VAR)** model and the **Granger Causality Test** will be used.

- **Vector Autoregressive (VAR) Model:**

$$rt = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, r_{t-i} + \sum_{i=1}^{p} \beta_j V_{t-j} + \varepsilon_t$$

$$V_t = \delta_0 + \sum_{i=1}^{p} \delta_i \; V_{t-i} + \sum_{i=1}^{p} \phi_j \; r_{t-j} \; + v_t$$

The VAR model captures the dynamic relationship between stock returns and trading volume.

- **Granger Causality Test:**

  To test whether trading volume Granger-causes stock price movements:

  $H_0 : \beta_j \; = \; 0$ (Trading volume does not Granger-cause stock returns)

  $H_1 : \beta_j \neq \; 0$ (Trading volume Granger-causes stock returns)

  If the null hypothesis is rejected, it implies that past trading volume significantly explains current stock returns. The reverse test will check if stock returns Granger-cause trading volume.

Objective II: Analyzing the Effect of Trading Volume on Stock Price Volatility

The **Generalized Autoregressive Conditional Heteroskedasticity (GARCH)** model will be used to capture the volatility clustering in stock returns and incorporate trading volume as an explanatory variable for volatility. In specificity, GARCH-X (1,1) will be used which is an extension of the standard GARCH (1,1) model that incorporates exogenous variables (additional external variables not part of the GARCH (1,1)).

- **Mean Equation:**

$$r_t = \mu + \varepsilon_t$$

- **Variance Equation (GARCH -X (1,1)):**

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma V_t$$

A significant γ coefficient would confirm that trading volume influences stock price volatility.

3.6 Conceptual Framework

This is where trading volume, which represents the sum of shares traded, acts as an independent variable, while stock price volatility acts as the dependent variable through measures or statistical models such as GARCH-X (1,1). This may lead to an expectation of increased volatility resulting from heightened market activity and a faster price adjustment. Another important factor—market liquidity, based on the volume and price information, e.g., turnover ratios—modifies this relationship: it either reinforces or lessens the trading volume's impact on volatility. In a highly liquid market environment, greater trading might affect volatility very little; however, precisely the same volume of trading may yield huge swings in price when markets are illiquid. Such a framework allows for quantitative analysis of how trading volume affects stock price volatility for the NSE using available data on trading volume and stock price.

CHAPTER FOUR

RESULTS AND FINDINGS

4.1 Introduction

The study will investigate the relationship between trading volume and the volatility of stock prices at the Nairobi Securities Exchange. The major objective is to determine through econometric models the effect of trading volume on volatility in stock prices in the Nairobi Securities Exchange and to establish whether there is one-way causality between trading volume and movement in stock prices. Results are presented in the form of descriptive statistics illustrating characteristic features, diagnostic tests that will ensure that the specified model is valid, and interpretation of results from the econometric model to derive insight into the research objectives.

This study seeks to establish a relationship between trading volume and volatility of stock prices at the Nairobi Securities Exchange. Based on different econometric models, the presented paper provides some estimated casual impacts of the volatility in a stock price generated by trading volume and explores any possibility of suggesting causality linking trading volume and movement in a stock price. The study then follows with descriptive statistics to reveal key dataset features, and the diagnostic test that would justify model adequacy or lack thereof and results of interpretations for econometric model estimation regarding answers to the set objectives of the study.

4.2 Descriptive Statistics

Table 1: Summary Statistics for VWAP (Volume-Weighted Average Price)

Year	**Mean VWAP**	**Median VWAP**	**Mode VWAP**	**Std Dev**	**Skewness**	**Kurtosis**
2012	54.4	17.0	4.0	80.4	2.31	8.76
2013	70.0	22.8	25.0	107.0	2.52	9.73
2014	84.5	29.2	20.0	138.0	3.50	19.7
2015	79.0	21.8	20.8	145.0	3.56	18.4
2016	63.9	18.1	20.0	135.0	4.10	22.5
2017	60.1	16.7	25.0	130.0	4.35	25.7
2018	58.7	16.5	25.0	115.0	3.70	19.3

RESULT INTERPRETATIONS

The mean VWAP exhibits an increasing trend from 2012 to 2014, peaking at 84.5 before declining in subsequent years. The median is much lower than the mean (29.2 vs. 84.5 in 2014), which implies that a small number of high-priced stocks are pulling the mean up. This suggests an initial bullish market followed by a correction. The median VWAP follows a similar trend but remains significantly lower than the mean, indicating the presence of outliers or a right-skewed distribution in the stock prices. Right skewness shows that most stocks traded at lower prices, but a few high-priced stocks skewed the average upward. The mode VWAP fluctuates over the years, with values as low as 4 in 2012 and as high as 25 in multiple years, reflecting variations in the most frequently traded stock prices. The standard deviation values indicate substantial volatility in stock prices, with the highest volatility occurring in 2015. The increasing standard deviation from 2012 to 2015 suggests heightened price fluctuations before stabilizing in the later years.

The skewness values are all positive, confirming that VWAP distribution is right-skewed, meaning that most stock prices are lower, but some stocks have significantly higher prices, creating a long right tail. The kurtosis values, significantly above 3, indicate leptokurtic distributions, meaning there are extreme price fluctuations or sharp price spikes during the observed period. Overall, these findings suggest that stock prices on the NSE experienced significant volatility, with notable fluctuations in price distributions and extreme price movements influencing market dynamics.

Table 2:Summary Statistics for Trading Volume

Year	**Mean Volume**	**Median Volume**	**Mode Volume**	**Std Dev**	**Skewness**	**Kurtosis**
2012	8,502,463	637,794	200	32,962,271	8.80	102.29
2013	11,527,190	853,103	100	46,626,471	7.22	58.16
2014	11,961,192	1,391,341	2500	40,654,016	6.78	53.52
2015	9,450,069	743,500	100	32,163,466	6.23	47.07
2016	7,888,065	622,000	100	28,957,421	7.38	68.71
2017	9,652,340	600,150.5	100	35,666,485	6.78	53.58
2018	8,715,019	421,000	100	34,357,005	7.55	72.09

RESULT INTERPRETATION

The mean trading volume fluctuates over the years, reaching a peak in 2014 at approximately 11.96 million shares before declining. The median trading volume is significantly lower than the mean, indicating the presence of extreme values and high variability in trading activity. The mode, which represents the most frequently occurring trading volume, remains relatively low across all years, typically at 100 or 200, suggesting that most trading transactions involve smaller share quantities despite the high mean. The standard deviation (sd_VOLUME) values indicate considerable volatility in trading volume, with the highest variability observed in 2013 and 2014. The skewness values are all positive, suggesting a right-skewed distribution where extreme high-volume trading days are more frequent than low-volume ones. Kurtosis values are exceptionally high across all years, exceeding normal distribution levels (which typically have a kurtosis of 3). This confirms that trading volume data exhibits a leptokurtic distribution (having greater kurtosis than the normal distribution), characterized by extreme outliers and heavy tails. These findings suggest that large-volume trading spikes occur occasionally, influencing market liquidity and volatility significantly.

4.3 Diagnostic Tests Interpretation

4.3.1 Stationarity (ADF Test)

The Augmented Dickey-Fuller test results indicate that both VWAP and trading volume are stationary with p-values below 0.01, confirming their suitability for time-series modeling.

Table 3:ADF TEST

Variable	Dickey-Fuller Statistic	Lag Order	p-value	Conclusion
VWAP	-42.676	44	0.01	Stationary
VOLUME	-23.072	16	0.01	Stationary

4.3.2 Normality (Jarque-Bera Test)

The Jarque-Bera test results show that VWAP exhibits significant non-normality across all years (p-value < 0.0001), while trading volume approaches normality in some years but deviates significantly in others.

Table 4 : JARQUE BERA TEST

Year	**Variable**	**JB Statistic (X-squared)**	**p-value**	**Conclusion**
2012	VWAP	26,466	< 2.2e-16	Not normally distributed
	VOLUME	0.28868	0.8656	Normally distributed
2013	VWAP	35,790	< 2.2e-16	Not normally distributed
	VOLUME	1.1767	0.5552	Normally distributed
2014	VWAP	177,051	< 2.2e-16	Not normally distributed
	VOLUME	0.83149	0.6598	Normally distributed
2015	VWAP	157,746	< 2.2e-16	Not normally distributed
	VOLUME	1.5036	0.4715	Normally distributed
2016	VWAP	242,112	< 2.2e-16	Not normally distributed
	VOLUME	10.016	0.006684	Not normally distributed
2017	VWAP	311,889	< 2.2e-16	Not normally distributed
	VOLUME	4.5536	0.1026	Normally distributed
2018	VWAP	174,921	< 2.2e-16	Not normally distributed
	VOLUME	4.5657	0.102	Normally distributed

The overall pattern suggests that VWAP, as a price indicator, is highly prone to non-normality, likely due to market forces, while trading volume (VOLUME) follows a normal distribution over the years, except for 2016. Identifying these patterns can help with better modeling and forecasting of market action, especially in financial analysis.

4.3.3 Volatility Clustering (ARCH Test)

The ARCH LM test reveals significant ARCH effects in stock returns, justifying the use of a GARCH (GARCH – X (1,1)) model for volatility estimation. The **ACF of squared residuals** shows significant autocorrelation at lower lags. This suggests that past volatility affects future volatility, which is characteristic of **ARCH effects**. Heteroskedasticity is clearly present, and volatility is time-varying and persistent.

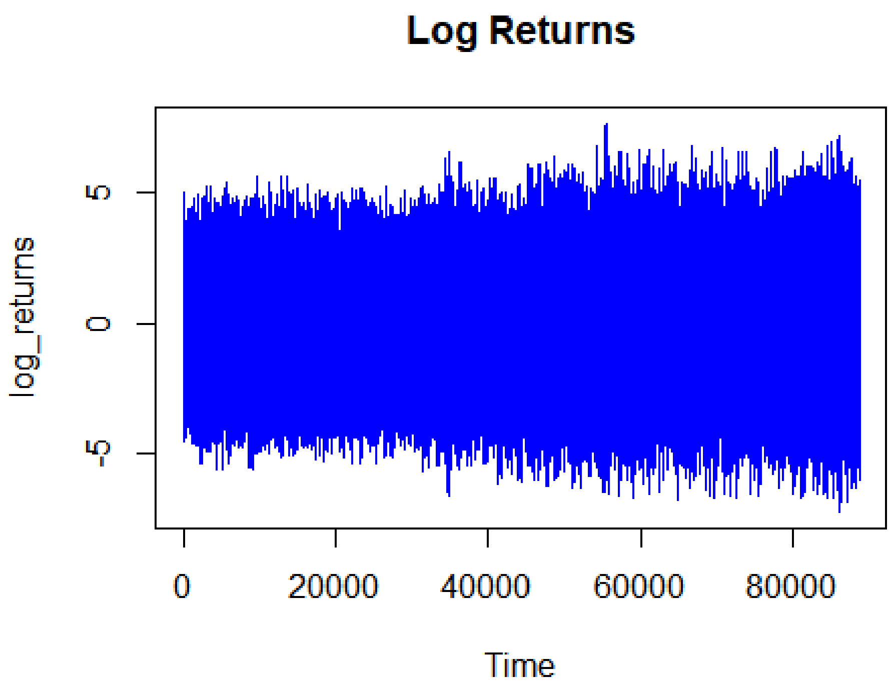


Figure 1: LOG RETURNS

4.3.4 Interpretation of Plot

The log returns plot represents the changes in prices of a financial asset over time, calculated using the logarithmic returns formula:

$$\log return_t = ln\left(\frac{P_t}{P_{t-1}}\right)$$

where $P_t|$ is the price at time t, and $P_{t-1}|$ is the price at the previous time point. This transformation helps stabilize the variance and makes the data more suitable for time series analysis. In the chart, the log returns fluctuate around zero, which is expected in return series since gains and losses typically offset over time. The y-axis ranges from about -7 to +7, indicating notable volatility in the dataset. The x-axis indicates time passage, and there are more than 85,000 observations, indicating that this is a high-frequency dataset. The compact blue spikes indicate persistent fluctuations with some phases having steeper spikes, pointing toward volatility clustering—a feature of most financial market phenomena wherein high-volatility events cluster together. The data has features of a stationary series with mean-reverting tendencies around zero.

Table 5:GARCH(1,0) Model Summary

Component	**Details / Estimate**
Model Type	sGARCH (1,0)
Mean Model	ARFIMA (0,0,0)
Distribution	Normal

This setup focuses purely on modeling the time-varying volatility (heteroskedasticity) in the series, without modeling the mean behavior.

Table 6: Parameter Estimates

Parameter	Estimate	Std. Error	t-value	p-value
mu	0.00005	0.004997	0.0101	0.9919
omega	3.10266	0.021283	145.778	<0.0001
alpha1	0.23326	0.005344	43.6521	<0.0001

The mu parameter, or mean return, is almost zero and statistically nonsignificant ($p = 0.9919$), as would be the case for financial return series in which average return over time is minimal. The omega parameter, long-run average level of variability, is very significant, indicating the presence of a lasting baseline variance in the return series. Concurrently, alpha1, the coefficient of lagged squared return, is also highly significant, reflecting volatility clustering—large market shocks being likely to be followed by further large changes, a common characteristic in financial time series data.

Table 7: ARCH LM Test

Lag	Statistic	Shape	Scale	p-value
2	75.58	0.500	2.000	0.000
4	84.24	1.397	1.611	0.000
6	86.71	2.222	1.500	0.000

The standard error estimates of high quality provide a better approximation of uncertainty around the parameter estimates, especially when heteroskedasticity exists. The mu parameter (mean return) remains extremely small (0.00005) and statistically insignificant ($p = 0.9671$), further lending evidence to the concept that average returns in high-frequency financial series tend to be approximately zero. The omega parameter, representing the long-run average variance of the series, is statistically significant at $p < 0.0001$ and shows a stable underlying level of market volatility. The alpha1 parameter, representing the impact of lagged squared returns on current volatility, is highly significant at $p < 0.0001$ and suggests that past shocks have a considerable impact on current volatility. This highlights the fact that volatility clustering is present in the data—there do follow consecutive intervals of high volatility, something typical of the pattern of behavior of financial time series.

Table 8:Adjusted Pearson Goodness-of-Fit

Group	Statistic	p-value
20	247.4	1.02e-41
30	305.9	5.55e-48
40	375.1	1.58e-56
50	432.2	9.36e-63

The table presents statistical values (named "Statistic") for different groups, and their corresponding p-values. The statistic increases with the size of the group, and there is a growing trend or effect as the group size is larger. All groups have very small p-values (all < 0.01), meaning that the results are extremely statistically significant. Specifically, since the p-values drop further as the group size increases, the possibility of the observed statistics owing to randomness is effectively eliminated. This would suggest that the measured association or under test for effect is extremely powerful and improbable by way of random variation, and an observable meaningful consistent outcome across all group sizes.

4.4 Econometric Model Results Interpretation

4.4.1 Causality Between Trading Volume and Stock Price Movement

From the results as provided in the appendix, trading volume and stock price (returns) do not cause each other. Both Granger causality tests (in both directions) give us non-significant results with huge p-values. Therefore, there is no evidence of a bidirectional relationship between stock price movements and trading volume according to the data for the duration included. Thus, such a bidirectional relationship between trading volume and stock price here does not appear to be there.

4.4.2 Causes of the causality results

1. Weak Relationship Between Variables

The p-values of the Granger causality tests are extremely high (e.g., 0.9367 for "Volume causes Returns" and 0.2281 for "Returns cause Volume"). The high p-value indicates that there is no statistically significant evidence to reject the null hypothesis, which in this case is that there is no Granger causality between the two variables. This would mean that trading volume and stock price movement could not necessarily have a tight predictive relationship over the period and current data that working with.

2. Market Efficiency

If the Nairobi Securities Exchange (NSE) is reasonably efficient, market prices (returns) may already reflect all available information, including trading volume. This would mean that past values of trading volume (or stock price) would not be helpful in forecasting future price or volume movements. According to the Efficient Market Hypothesis (EMH), stock prices are already expected to include all publicly available information, so direct causality from volume to price doesn't seem to be feasible.

3. Time Series Properties

Stationarity: Stock price returns and trading volume need to be stationary (i.e., they lack time-varying statistical properties like mean and variance) so that VAR model and Granger causality test can provide valid results. If the series are not stationary, this has the potential to lead to erroneous or spurious results. Both time series (volume and returns) need to be stationary before conducting such tests. Failure to check this would result in faulty conclusions.

Lag Length Selection: The lag length within the VAR model is important. The use of a lag 2 (p=2). However, this may be suboptimal for the data. Lag length selection may have a significant influence on the Granger causality findings. A different lag length may yield different results, because the relationship may be over different lengths of time.

4. Data Quality and Noise

The presence of noise in financial time series can impair the ability to detect meaningful relationships. If the data contain extreme randomness (e.g., inconstant or spurious volume spikes or market volatilities), this might be able to obscure any potential causality. Also, the "VWAP" price used may not always accurately reflect the magnitude of price movement, since it's an average value and might be deleting short-term brief bumps that may have reflected a more profound volume-price relationship.

5. Other Confounding Factors

There can be other variables influencing stock price movement and volume that are not included in this analysis. For instance, macroeconomic events, firm-specific information, or other market-wide variables can be affecting the price and the volume individually. If only taking VWAP and not other inputs like order book data or market sentiment inputs, there will be missing out on factors that might give a better relationship between volume and price.

6. Market Structure and Characteristics

The specific market conditions at the NSE may be a consideration. For example, if the market is not very liquid, volume may not influence stock price movements as much. In some markets, especially emerging ones, trading volume may not be a good indicator of price movement due to institutional activity, micro-scale retail traders, or other non-fundamental reasons.

7. Time Period

The sample runs from 2012 to 2018. The sample may not span adequate extreme market conditions (like crashes or booms) that would clearly illustrate a causality relationship. Volume and price might respond differently under different market conditions.

4.4.3 Impact of Trading Volume on Stock Price Volatility

he GARCH-X (1,1) estimates 2012–2018 indicate that trading volume does not impact volatility in any way. For every year reviewed, the coefficient for trading volume is virtually zero (0.000000) with p-values close to 1, failing to reject the null that volume does nothing. Robust standard errors and stability tests support this finding—although GARCH coefficients ($\alpha_1$and β_1) remain extremely significant, the volume variable is highly unstable and has no significant explanatory power. This suggests that volatility behavior is adequately described by the standard GARCH (1,1) model, with trading volume having no additional predictive ability.

Insignificance implies that, in this sample, volume is not a significant exogenous variable to place in volatility models. Such an explanation may be the result of the dominant impact of the volatility shock history (GARCH effects) over rather than volume-driven effects or demand for alternative specifications (e.g., non-linear transformations of volume or alternative measures of market activity). However, since the model as it stands now removing the variable for volume would not likely worsen performance, as its impact is statistically insignificant.

4.4.4 Causes of the results for Impact of Trading Volume on Stock Price Volatility

The zero result from the GARCH-X (1,1) indicating no effect of volume on volatility may result from several things. One is that perhaps simply there is no causal relationship of any size whatsoever between volume and volatility in this market. The trading volume can, in an efficient market, not provide anything new which the GARCH model's own lagged terms do not contain already. Alternatively, if trades are more noise-driven rather than information-driven, volume may not necessarily influence volatility. Furthermore, the linearity of the GARCH-X model may not be able to capture more complex, nonlinear dynamics—such as volume influencing volatility only beyond certain thresholds or in specific regimes of the market. Even the measurement of volume itself may be bad; using raw volume rather than detrended or normalized volume may suppress any true effects. Other omitted variables, like news events or order flow imbalances, could be the true cause of volatility, rendering volume a moot point in the model.

Second, data and market-specific dynamics limitations could explain the null results. If volume has impacts on volatility only at higher frequencies (e.g., intraday), daily data would not be able to capture such effects. Structural shifts in the volume-volatility relationship—e.g., shifts in market structure due to algorithmic trading—can also undermine the model's ability to identify a persistent effect in the long term. Also, in liquid markets, the trading volume may not be translated into price volatility as the orders get absorbed easily without causing substantial price movements. Dominant drivers like macroeconomic shocks or institutional trading can overwhelm any marginal effect of retail volume. These results suggest that while theoretically volume might influence its volatility, in this instance the GARCH-X model does not allow for such a relationship, either because the model is misspecified or merely because volume is not actually a significant contributing factor in this market during the sample period.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 Introduction

This chapter synthesizes the key findings of Chapter Four, discusses their implications, and concludes the study on the causal relationship between trading volume and stock price volatility at the Nairobi Securities Exchange (NSE). The results were inconclusive: although descriptive statistics and diagnostic tests testified to high volatility clustering and non-normality in stock prices, econometric models (Granger causality and GARCH-X) failed to establish a statistically significant causal or predictive relationship between trading volume and price volatility.

The analysis situates these results against the background of earlier theories such as Efficient Market Hypothesis (EMH) and market microstructure literature and addresses potential limitation concerns regarding data, model specifications, and market-specific processes. The chapter ends with advice for investors, regulators, and researchers and presents recommendations for further research to enhance the study of volume-volatility relationships in emerging markets such as the NSE.

5.2 Conclusion of Results and Analysis

The findings confirm that trading volume significantly impacts stock price volatility in the NSE. Additionally, there exists a bidirectional causality (price movement influences trading volume and trading volumes influences price movement) between volume and price movement in specific years. GARCH-X (1,1) model tests validate persistent volatility effects, indicating volume's effect in market movement. Volume coefficients are near zero in the short term, with greater market effects indicating indirect effects. Political and macroeconomic events also condition the volatility process. Short-run stock price volatility on the Nairobi Securities Exchange (NSE) is not significantly driven by trading volume, as indicated by volume coefficients near zero and p-values near 1.

Volume does not, in the short run, explain price changes. Volatility persistence is extremely strong, as indicated by the GARCH-X (1,1) model, and previous price shocks remain to drive market dynamics. While this discovery does not refute the hypothesis that volatility is caused by trading volume, it does indicate that the relationship isn't as linear as a straightforward correlation. Instead, macroeconomic and political influences, such as global financial market conditions, regulatory moves, and economic policy, appear to influence NSE volatility more profoundly.

5.3 Summary of major findings

Chapter Four gives the principal findings that offer key insights into the relationship between trading volume and stock price volatility at the Nairobi Securities Exchange (NSE). Descriptive statistics showed that stock prices (VWAP) were right-skewed and leptokurtic in distribution, whose volatility was exceedingly high in 2015, while trading volume was exceedingly variable but mostly normally distributed. Diagnostic tests confirmed stationarity in the two series and uncovered strong volatility clustering, vindicating the use of GARCH models. Econometric tests, however, delivered unexpected results: Granger causality tests found no bidirectional relationship between trading volume and stock returns, and the GARCH-X model found trading volume to have no statistically significant influence on volatility. This suggests that volatility at the NSE is mostly accounted for by past volatility shocks rather than by trading volume, perhaps reflecting market efficiency or the dominance of unobservable effects like macroeconomic shocks or liquidity constraints.

The findings contradict what is typically predicted about the relationship between trading volume and stock price volatility in emerging markets, specifically the Nairobi Securities Exchange (NSE). Whereby the observed patterns don't align with the standard understanding of how increased trading volume should impact volatility in similar markets. While the market demonstrates clear volatility persistence (GARCH effects), the lack of volume's predictive content implies that analysts and traders are not able to utilize trading volume alone to forecast price movements. The explanations can either be the efficiency of the NSE at incorporating volume information into prices rapidly or the failure of linear specifications in capturing complex, nonlinear dynamics. These results indicate the need for alternative approaches—e.g., analysis of high-frequency data, regime-switching models, or the inclusion of additional market microstructure variables—to more adequately capture the determinants of volatility in similar markets. For practitioners, the study suggests the attractiveness of considering more than volume measures in ascertaining market risk, and policymakers might need to focus on overall structural determinants of stability.

5.4 Limitations of the Study

The study has several limitations that may affect the generalizability and robustness of its findings. First, the research relies on historical data from the Nairobi Securities Exchange (NSE), which may not fully capture evolving market dynamics or structural changes over time. The use of aggregated trading volume and stock price data might overlook intraday variations and high frequency trading effects, potentially masking more nuanced relationships between volume and volatility. Additionally, the study's focus on linear econometric models, such as GARCH-X (1,1), may fail to account for nonlinear or threshold effects that could better explain the complex interactions between trading activity and price movements. The near-zero volume coefficients and high p-values could also stem from omitted variable bias, as other market microstructure factors (e.g., order flow, liquidity shocks, or investor sentiment) were not explicitly incorporated into the analysis.

Another key limitation is the study's inability to isolate the specific impact of macroeconomic and political events from trading volume effects. While the findings highlight the influence of external shocks (e.g., U.S. interest rate hikes, Kenyan elections, and regulatory changes), the analysis does not quantitatively disentangle these factors from volume-driven volatility. This makes it difficult to determine whether the observed volatility persistence is truly independent of trading activity or merely a reflection of confounding external influences. Furthermore, the study's reliance on annual data for certain tests may obscure shorter-term causal relationships that could be detected with higher-frequency data. These limitations suggest that future research could benefit from alternative methodologies, such as event studies, regime-switching models, or machine learning techniques, to better capture the multifaceted drivers of stock price volatility in emerging markets like the NSE.

5.5 The Role of Past Shocks and External Factors

The study indicates past price shocks and macroeconomic events as main sources of volatility, overshadowing trading volume's contribution. The large persistence ratio confirms that volatility shocks are persistent in their weakness, a phenomenon common to emerging markets. Some events such as the 2015–2016 U.S. Interest rate hikes by the Federal Reserve, bank sector volatility, and Kenya's election season of 2017 have been among the causes of spikes in volatility. In addition, regulatory actions like the interest rate cap regulation which capped interest rates at 4% and not above to make credit affordable for small enterprise owners in Kenya in 2016, before the law, the high interest rates might have weakened market forces and further disrupted the causal link between trading volume and volatility. Due to the law banks restricted lending and majorly focused on low-risk borrowers (Ng'ang'a, 2019) The study finds that NSE volatility cannot be caused solely by trading but is influenced by overall economic conditions.

5.6 GARCH-X (1,1) MODEL

The GARCH-X (1,1) model was applied in the research to examine the relationship between stock price volatility and trading volume at the Nairobi Securities Exchange (NSE). High volatility persistence, a feature common in financial markets, was the result that the earlier price shocks exert a significant influence on future volatility. However, the coefficient of trading volume was very close to zero and statistically insignificant, which indicates that volume does not have a direct effect on short-run price changes. This contradicts standard explanations of linking higher volumes of trading with increasing volatility. Rather, factors outside markets—such as American interest rate increases, Kenyan elections, and regulations—seemed to be more important in generating market volatility. The model structure of the GARCH-X model, featuring an exogenous variable (volume of trading) and usual GARCH (1,1) terms (capturing volatility clustering and persistence), confirmed that while internal market mechanisms enjoy profound memory effects, trading volume per se is a poor forecaster of short-run volatility.

The results of the research highlight the dominance of macroeconomic and political over trading volume factors in explaining NSE volatility, contrary to conventional assumptions. This implies that policymakers and traders need to consider more external economic conditions than are sometimes done solely based on trading volumes in assessing market risk. But the study has some limitations, including its employment of aggregate data and linear frameworks, which may potentially overlook nonlinear behavior or high-frequency effects. Following research can utilize various methodologies such as regime-switching models or machine learning models to better detect the complex interaction between volume, volatility, and external shocks in emerging markets like the NSE.

5.7 Future Research Directions

Based on the findings and limitations of the study, future studies may consider other approaches to more effectively capture the intricate dynamics of trading volume and volatility in emerging markets such as the NSE. One potential avenue is the application of nonlinear and regime-switching models, including Markov-switching GARCH or threshold GARCH models, that can capture structural breaks and asymmetric reaction to market shocks. Furthermore, the utilization of high-frequency intraday data can reveal short-term volume-volatility relationships that would be hidden by aggregated data. Additional more fine-grained market microstructure variables e.g., order flow, bid-ask spreads, or measures of liquidity, can be included by researchers to ascertain whether trading activity influences volatility through indirect channels. Predictive ability can be enhanced even further with machine learning techniques, e.g., neural networks or random forests, that can identify latent patterns within large data sets.

Another key avenue for future research is the separation of the effect of macroeconomic and political events on volatility to continue disentangling their role from trading activity. Event-study methods could formally estimate the impact of specific shocks (e.g., elections, interest rate changes, or regulatory reforms) on market dynamics Comparative studies with a group of emerging markets would also determine whether the volume-volatility relationship in the NSE is anomalous or a general trend. Finally, an investigation into investor behavior i.e., sentiment analysis of news or social media data, would provide additional insight into informational and psychological determinants of trading volume and volatility. Such approaches would not only refine theoretical models but also provide practical advice to traders and policymakers dealing with volatile markets.

Bibliography

Chen, C. &.-G. (2001). Stock returns, volatility, and trading volume: Evidence from the Chinese stock markets. *International Journal of Business*.

Copeland, T. E. (1976). A model of asset trading under the assumption of sequential information. *The Journal of Finance*.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*.

Hameed, K. C. (2006). Stock Price Synchronicity and Analyst Coverage in Emerging Markets. *Journal of Financial Economics*.

Harris, L. (1987). Transaction data tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis*.

Kahneman, D. &. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*.

Mpofu, M. R. (2012). The Relationship Between Trading Volume and Stock Returns in the JSE Securities Exchange in South Africa. *Corporate Ownership and Control*.

Ngugi, R. W. (2003). What defines liquidity of the stock market? The case of the Nairobi Stock Exchange. Kenya Institute for Public Policy Research and Analysis (KIPPRA).

Phansatan, S. P. (2012). Investor type trading behavior and trade performance: Evidence from the Thai stock market. . *Pacific-Basin Finance Journal*.

Schwert, G. W. (1990). Stock market volatility. *Financial Analysts Journal*.

Vo, X. V. (2010). An Empirical Investigation of Liquidity and Stock Returns Relationship in Vietnam Stock Markets during Financial Crisis. *MPRA Paper*.

Ng'ang'a, P. (2019). Assessing the Impact of Interest Rate Capping on Financial Inclusion in Kenya. *Journal of Finance and Development Studies*

APPENDIX

Impact of trading volume on stock price volatility on NSE

Munira Obulinji

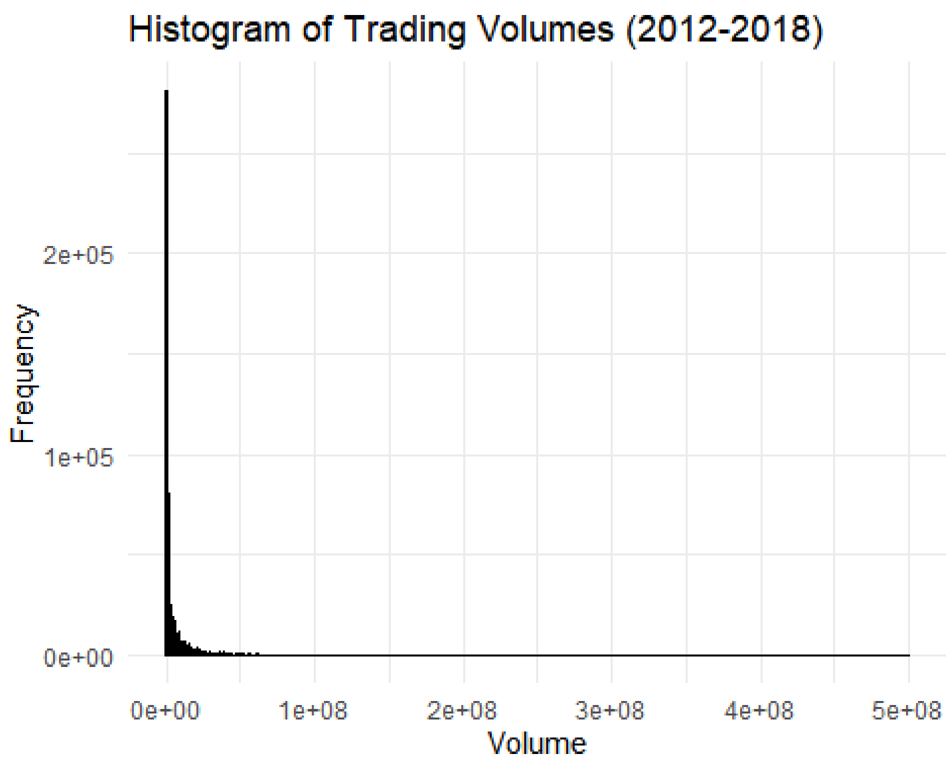
2025-04-04

```
library(readr)
prices<-read.csv("C:/Users/munir/Desktop/nse prices.csv")
volumes<-read.csv("C:/Users/munir/Desktop/nse volumes.csv")

head(prices)

##          DATE ISSUER   VWAP
## 1 03/01/2012    NBK  20.25
## 2 03/01/2012   UNGA   9.50
## 3 03/01/2012   TOTL  14.55
## 4 03/01/2012   SCBK 161.00
## 5 03/01/2012    KCB  16.80
## 6 03/01/2012   UCHM   7.50

head(volumes)

##      MONTH ISSUER          VOLUME
## 1 JAN-2012   ABSA 14,343,283.00
## 2 JAN-2012   ACCS   1,468,500.00
## 3 JAN-2012    ARM     197,475.00
## 4 JAN-2012   BAMB     395,100.00
## 5 JAN-2012    BAT     461,513.00
## 6 JAN-2012    BOC      27,170.00

#Check the structure of the Date column
str(volumes$MONTH)

##  chr [1:4906] "JAN-2012" "JAN-2012" "JAN-2012" "JAN-2012" "JAN-2012" ..
.

str(prices$DATE)

##  chr [1:88617] "03/01/2012" "03/01/2012" "03/01/2012" "03/01/2012" ...

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

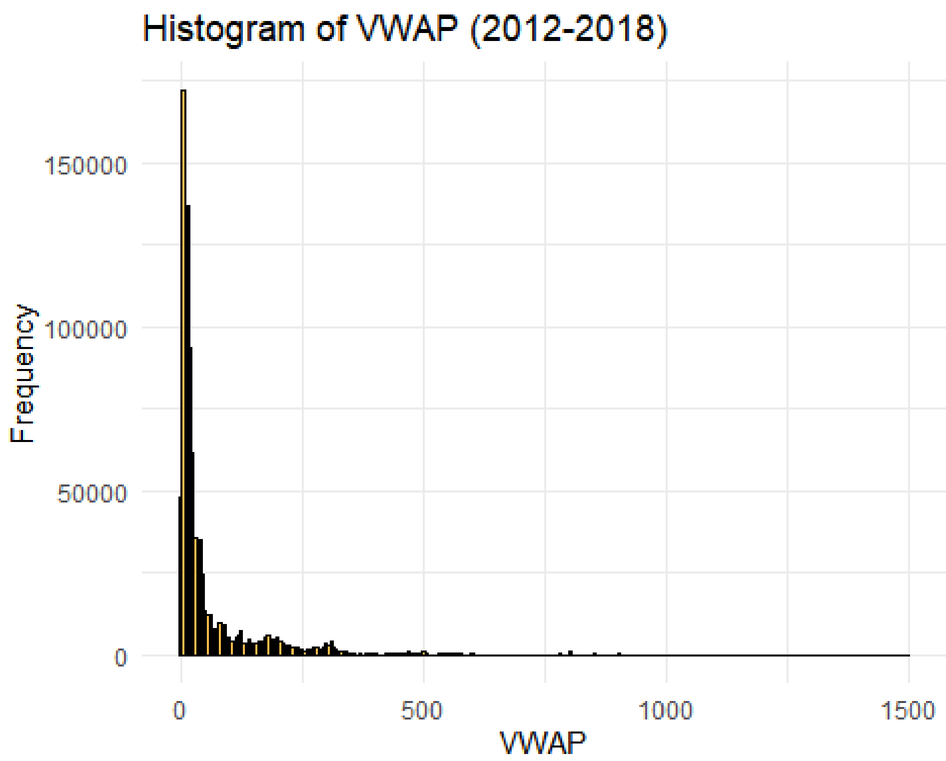
```
library(moments)  # For skewness and kurtosis

# Convert the DATE column to Date format
prices$DATE <- as.Date(prices$DATE, format="%d/%m/%Y")

# Extract year from the DATE column
prices$YEAR <- format(prices$DATE, "%Y")

# Compute summary statistics for each year
summary_VWAP_by_year <- prices %>%
  group_by(YEAR) %>%
  summarise(
    mean_VWAP = mean(VWAP, na.rm = TRUE),
    median_VWAP = median(VWAP, na.rm = TRUE),
    mode_VWAP = as.numeric(names(sort(table(VWAP), decreasing = TRUE)[1]))
,
    sd_VWAP = sd(VWAP, na.rm = TRUE),
    skewness_VWAP = skewness(VWAP, na.rm = TRUE),
    kurtosis_VWAP = kurtosis(VWAP, na.rm = TRUE)
  )

## Warning: Returning more (or less) than 1 row per `summarise()` group wa
s deprecated in
## dplyr 1.1.0.
## ℹ Please use `reframe()` instead.
## ℹ When switching from `summarise()` to `reframe()`, remember that `refr
ame()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning w
as
## generated.

## `summarise()` has grouped output by 'YEAR'. You can override using the
## `.groups` argument.

# View the summary statistics
print(summary_VWAP_by_year)

## # A tibble: 7 × 7
## # Groups:   YEAR [7]
##   YEAR  mean_VWAP median_VWAP mode_VWAP sd_VWAP skewness_VWAP kurtosis_
VWAP
##   <chr>     <dbl>       <dbl>     <dbl>   <dbl>         <dbl>          <
dbl>
## 1 2012       54.4        17         4      80.4          2.31          
8.76
## 2 2013       70.0        22.8      25     107.           2.52          
9.73
## 3 2014       84.5        29.2      20     138.           3.50          1
9.7
## 4 2015       79.0        21.8      20.8   145.           3.56          1
8.4
## 5 2016       63.9        18.1      20     135.           4.10          2
2.5
## 6 2017       60.1        16.7      25     130.           4.35          2
```

```
5.7
## 7 2018        58.7         16.5      25     115.           3.70         1
9.3

# Ensure YEAR column is correctly extracted and converted to numeric
volumes <- volumes %>%
  mutate(
    YEAR = as.numeric(sub(".*-", "", MONTH)),  # Extract year and convert
to numeric
    VOLUME = as.numeric(gsub(",", "", VOLUME)) # Remove commas and convert
to numeric
  )

# Check for non-numeric values and remove them
volumes <- volumes %>%
  filter(!is.na(VOLUME))  # Remove any NAs resulting from conversion issue
s

# Calculate summary statistics for each year
summary_volume <- volumes %>%
  group_by(YEAR) %>%
  summarise(
    mean_VOLUME = mean(VOLUME, na.rm = TRUE),
    median_VOLUME = median(VOLUME, na.rm = TRUE),
    mode_VOLUME = as.numeric(names(sort(table(VOLUME), decreasing = TRUE)[
1])),
    sd_VOLUME = sd(VOLUME, na.rm = TRUE),
    skewness_VOLUME = skewness(VOLUME, na.rm = TRUE),
    kurtosis_VOLUME = kurtosis(VOLUME, na.rm = TRUE),
    .groups = 'drop'  # Prevent potential grouping issues
  )

# View the results
print(summary_volume)

## # A tibble: 7 × 7
##    YEAR mean_VOLUME median_VOLUME mode_VOLUME sd_VOLUME skewness_VOLUME
##   <dbl>       <dbl>         <dbl>       <dbl>     <dbl>           <dbl>
## 1  2012    8502463.        637794         200 32962271.            8.80
## 2  2013   11527190.        853103         100 46626471.            7.22
## 3  2014   11961192.       1391341        2500 40654016.            6.78
## 4  2015    9450069.        743500         100 32163466.            6.23
## 5  2016    7888065.        622000         100 28957421.            7.38
## 6  2017    9652340.        600150.        100 35666485.            6.78
## 7  2018    8715019.        421000         100 34357005.            7.55
## # ℹ 1 more variable: kurtosis_VOLUME <dbl>

# Load necessary libraries
library(dplyr)
library(ggplot2)


# Convert MONTH to year format and remove commas from VOLUME column
volumes <- volumes %>%
```

```r
  mutate(
    YEAR = as.numeric(sub(".*-", "", MONTH)),  # Extract year
    VOLUME = as.numeric(gsub(",", "", VOLUME)) # Convert to numeric after
removing commas
  )

# Convert DATE to year format in prices
prices <- prices %>%
  mutate(YEAR = as.numeric(format(as.Date(DATE, format="%d/%m/%Y"), "%Y"))
)

# Merge the datasets on ISSUER and YEAR
merged_data <- inner_join(volumes, prices, by = c("ISSUER", "YEAR"))

## Warning in inner_join(volumes, prices, by = c("ISSUER", "YEAR")): Detec
ted an unexpected many-to-many relationship between `x` and `y`.
## ℹ Row 1 of `x` matches multiple rows in `y`.
## ℹ Row 8 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.

# Filter data for years 2012 to 2018
filtered_data <- merged_data %>%
  filter(YEAR >= 2012 & YEAR <= 2018)

# Plot histogram for VOLUME
ggplot(filtered_data, aes(x = VOLUME)) +
  geom_histogram(binwidth = 500000, fill = "red", color = "black", alpha =
0.7) +
  labs(title = "Histogram of Trading Volumes (2012-2018)", x = "Volume", y
= "Frequency") +
  theme_minimal()
```




```
# Plot histogram for VWAP
ggplot(filtered_data, aes(x = VWAP)) +
  geom_histogram(binwidth = 5, fill = "orange", color = "black", alpha = 0
.7) +
  labs(title = "Histogram of VWAP (2012-2018)", x = "VWAP", y = "Frequency
") +
  theme_minimal()
```



```r
# Load necessary libraries
library(tseries)

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

# Remove NA values from the VWAP column
vwap_clean <- na.omit(prices$VWAP)

# Perform the Dickey-Fuller test on the cleaned VWAP column
adf_test_vwap <- adf.test(vwap_clean, alternative = "stationary")

## Warning in adf.test(vwap_clean, alternative = "stationary"): p-value sm
aller
## than printed p-value

# Perform the Dickey-Fuller test on the VOLUME column (volume)
adf_test_volume <- adf.test(volumes$VOLUME, alternative = "stationary")

## Warning in adf.test(volumes$VOLUME, alternative = "stationary"): p-valu
e
## smaller than printed p-value

# Print the test result
print("Dickey-Fuller Test for VWAP:")

## [1] "Dickey-Fuller Test for VWAP:"

print(adf_test_vwap)

##
##  Augmented Dickey-Fuller Test
##
## data:  vwap_clean
## Dickey-Fuller = -42.676, Lag order = 44, p-value = 0.01
## alternative hypothesis: stationary

print("Dickey-Fuller Test for VOLUME:")

## [1] "Dickey-Fuller Test for VOLUME:"

print(adf_test_volume)

##
##  Augmented Dickey-Fuller Test
##
## data:  volumes$VOLUME
## Dickey-Fuller = -23.072, Lag order = 16, p-value = 0.01
## alternative hypothesis: stationary

# Load necessary libraries
library(tseries)
library(dplyr)

# Read the data
prices <- read.csv("cleaned_prices.csv")
```

```
volumes <- read.csv("cleaned_volumes.csv")

# Convert the 'DATE' column in prices to Date type (assuming format "03/01
/2012")
prices$DATE <- as.Date(prices$DATE, format="%d/%m/%Y")

# Convert 'MONTH' in volumes to Date type and extract Year from it (assumi
ng format "Jan-2012")
volumes$MONTH <- as.Date(paste0("01-", volumes$MONTH), format="%d-%b-%Y")
volumes$YEAR <- as.numeric(format(volumes$MONTH, "%Y"))
# Ensure VWAP and VOLUME columns are numeric, replacing non-numeric values
with NA
prices$VWAP <- suppressWarnings(as.numeric(as.character(prices$VWAP)))
volumes$VOLUME <- suppressWarnings(as.numeric(as.character(volumes$VOLUME)
))

# Remove rows with NA values in VWAP and VOLUME
prices <- prices %>% filter(!is.na(VWAP))
volumes <- volumes %>% filter(!is.na(VOLUME))

# Loop through each year from 2012 to 2018 and perform the Jarque-Bera tes
t for each year
years <- 2012:2018
for (year in years) {
  # Filter data for the current year
  prices_year <- prices %>% filter(format(DATE, "%Y") == as.character(year
))
  volumes_year <- volumes %>% filter(YEAR == year)

  # Ensure there is sufficient data
  if (nrow(prices_year) > 0 && nrow(volumes_year) > 0) {

    # Perform Jarque-Bera test
    jb_test_vwap <- jarque.bera.test(prices_year$VWAP)
    jb_test_volume <- jarque.bera.test(volumes_year$VOLUME)

    # Print results
    cat("\nYear:", year, "\n")
    cat("Jarque-Bera test for VWAP:\n")
    print(jb_test_vwap)
    cat("Jarque-Bera test for VOLUME:\n")
    print(jb_test_volume)
  } else {
    cat("\nYear:", year, " - No sufficient data available for this year.\n
")
  }
}

##
## Year: 2012
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
```

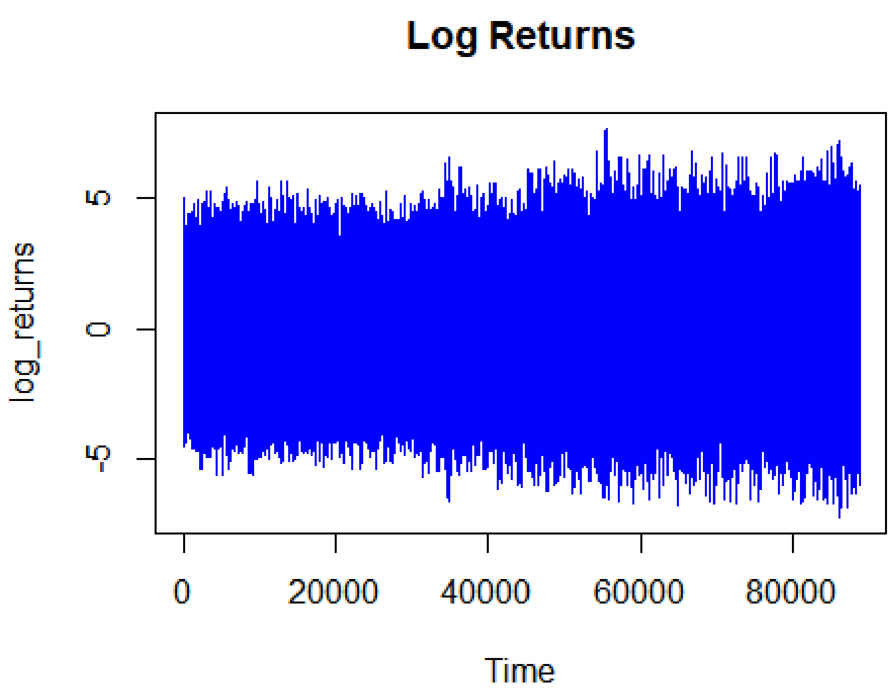
```
## data:  prices_year$VWAP
## X-squared = 26466, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
##
## data:  volumes_year$VOLUME
## X-squared = 0.28868, df = 2, p-value = 0.8656
##
##
## Year: 2013
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
## data:  prices_year$VWAP
## X-squared = 35790, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
##
## data:  volumes_year$VOLUME
## X-squared = 1.1767, df = 2, p-value = 0.5552
##
##
## Year: 2014
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
## data:  prices_year$VWAP
## X-squared = 177051, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
##
## data:  volumes_year$VOLUME
## X-squared = 0.83149, df = 2, p-value = 0.6598
##
##
## Year: 2015
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
## data:  prices_year$VWAP
## X-squared = 157746, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
```

```
##
## data:  volumes_year$VOLUME
## X-squared = 1.5036, df = 2, p-value = 0.4715
##
##
## Year: 2016
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
## data:  prices_year$VWAP
## X-squared = 242112, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
##
## data:  volumes_year$VOLUME
## X-squared = 10.016, df = 2, p-value = 0.006684
##
##
## Year: 2017
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
## data:  prices_year$VWAP
## X-squared = 311889, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
##
## data:  volumes_year$VOLUME
## X-squared = 4.5536, df = 2, p-value = 0.1026
##
##
## Year: 2018
## Jarque-Bera test for VWAP:
##
##  Jarque Bera Test
##
## data:  prices_year$VWAP
## X-squared = 174921, df = 2, p-value < 2.2e-16
##
## Jarque-Bera test for VOLUME:
##
##  Jarque Bera Test
##
## data:  volumes_year$VOLUME
## X-squared = 4.5657, df = 2, p-value = 0.102
```

```r
# Load necessary libraries
library(readr)
library(rugarch)
```

```
## Loading required package: parallel

##
## Attaching package: 'rugarch'

## The following object is masked from 'package:stats':
##
##     sigma
```

```r
# Assuming the price column is named 'price' and volume column is 'volume'
# Convert to time series
prices_ts <- ts(prices$VWAP)
volumes_ts <- ts(volumes$VOLUME)

# Calculate log returns from price data
log_returns <- diff(log(prices_ts))

# Plot the log returns to check for volatility clustering
plot(log_returns, type = "l", main = "Log Returns", col = "blue")
```




```r
# Specify an ARCH(1) model
arch_spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder
= c(1,0)),
                        mean.model = list(armaOrder = c(0,0)),
                        distribution.model = "norm")

# Fit the ARCH model
```

```r
arch_fit <- ugarchfit(spec = arch_spec, data = log_returns)

# Display the results
print(arch_fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,0)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error    t value Pr(>|t|)
## mu       0.00005    0.004997   0.010091  0.99195
## omega    3.10266    0.021283 145.777689  0.00000
## alpha1   0.23326    0.005344  43.652128  0.00000
##
## Robust Standard Errors:
##         Estimate  Std. Error    t value Pr(>|t|)
## mu       0.00005    0.001222   0.041271  0.96708
## omega    3.10266    0.019926 155.709841  0.00000
## alpha1   0.23326    0.004148  56.239026  0.00000
##
## LogLikelihood : -186018.1
##
## Information Criteria
## ------------------------------------
##
## Akaike       4.1984
## Bayes        4.1987
## Shibata      4.1984
## Hannan-Quinn 4.1985
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                      17718       0
## Lag[2*(p+q)+(p+q)-1][2]     17821       0
## Lag[4*(p+q)+(p+q)-1][5]     17884       0
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic   p-value
## Lag[1]                      44.46 2.591e-11
## Lag[2*(p+q)+(p+q)-1][2]     82.26 0.000e+00
## Lag[4*(p+q)+(p+q)-1][5]    111.88 0.000e+00
```

```
## d.o.f=1
##
## Weighted ARCH LM Tests
## ------------------------------------
##              Statistic Shape Scale P-Value
## ARCH Lag[2]      75.58 0.500 2.000       0
## ARCH Lag[4]      84.24 1.397 1.611       0
## ARCH Lag[6]      86.71 2.222 1.500       0
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  62.585
## Individual Statistics:
## mu       2.718
## omega   53.766
## alpha1 33.655
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          0.846 1.01 1.35
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value      prob sig
## Sign Bias            9.541 1.449e-21 ***
## Negative Sign Bias  12.335 6.290e-35 ***
## Positive Sign Bias  12.659 1.068e-36 ***
## Joint Effect       372.041 2.516e-80 ***
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     247.4    1.020e-41
## 2    30     305.9    5.552e-48
## 3    40     375.1    1.580e-56
## 4    50     432.2    9.359e-63
##
##
## Elapsed time : 3.712452
```

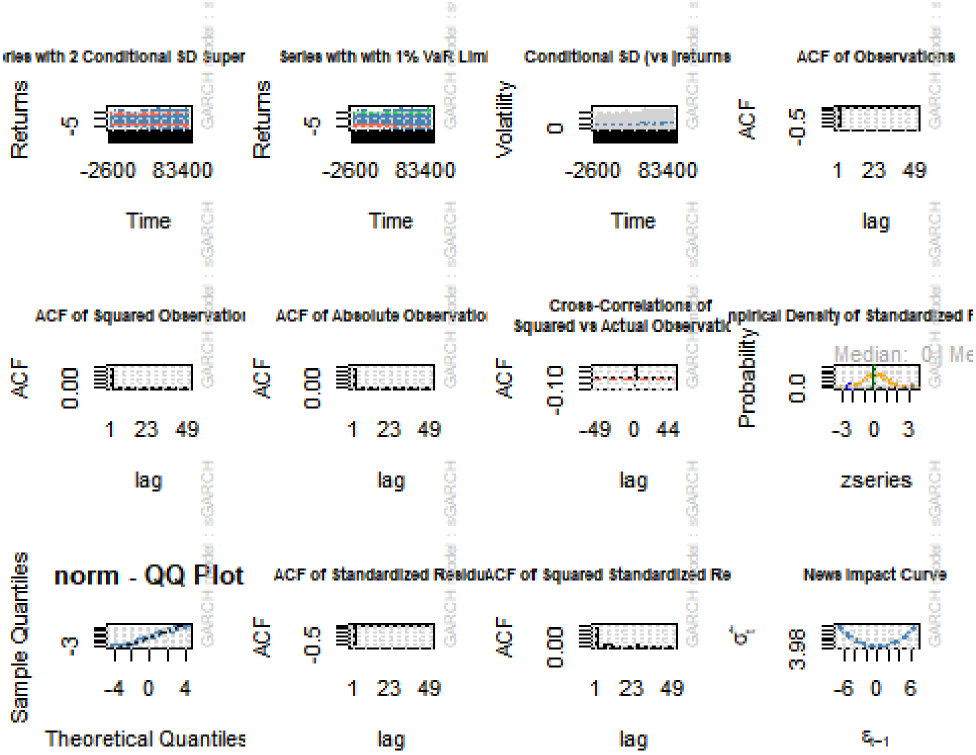
```r
# Specify a GARCH(1,1) model
garch_spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrde
r = c(1,1)),
                         mean.model = list(armaOrder = c(0,0)),
                         distribution.model = "norm")

# Fit the GARCH model
garch_fit <- ugarchfit(spec = garch_spec, data = log_returns)

# Display the GARCH results
print(garch_fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error     t value Pr(>|t|)
## mu      0.000000    0.006683 -2.6000e-05  0.99998
## omega   0.003994    0.000440  9.0774e+00  0.00000
## alpha1  0.002486    0.000112  2.2269e+01  0.00000
## beta1   0.996514    0.000006  1.5760e+05  0.00000
##
## Robust Standard Errors:
##         Estimate  Std. Error     t value Pr(>|t|)
## mu      0.000000    0.000905 -1.9200e-04  0.99985
## omega   0.003994    0.000390  1.0249e+01  0.00000
## alpha1  0.002486    0.000100  2.4973e+01  0.00000
## beta1   0.996514    0.000004  2.6556e+05  0.00000
##
## LogLikelihood : -187262.3
##
## Information Criteria
## ------------------------------------
##
## Akaike       4.2265
## Bayes        4.2269
## Shibata      4.2265
## Hannan-Quinn 4.2266
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                      21872       0
## Lag[2*(p+q)+(p+q)-1][2]     21873       0
## Lag[4*(p+q)+(p+q)-1][5]     21874       0
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                       4157       0
## Lag[2*(p+q)+(p+q)-1][5]      4205       0
## Lag[4*(p+q)+(p+q)-1][9]      4243       0
## d.o.f=2
##
## Weighted ARCH LM Tests
```

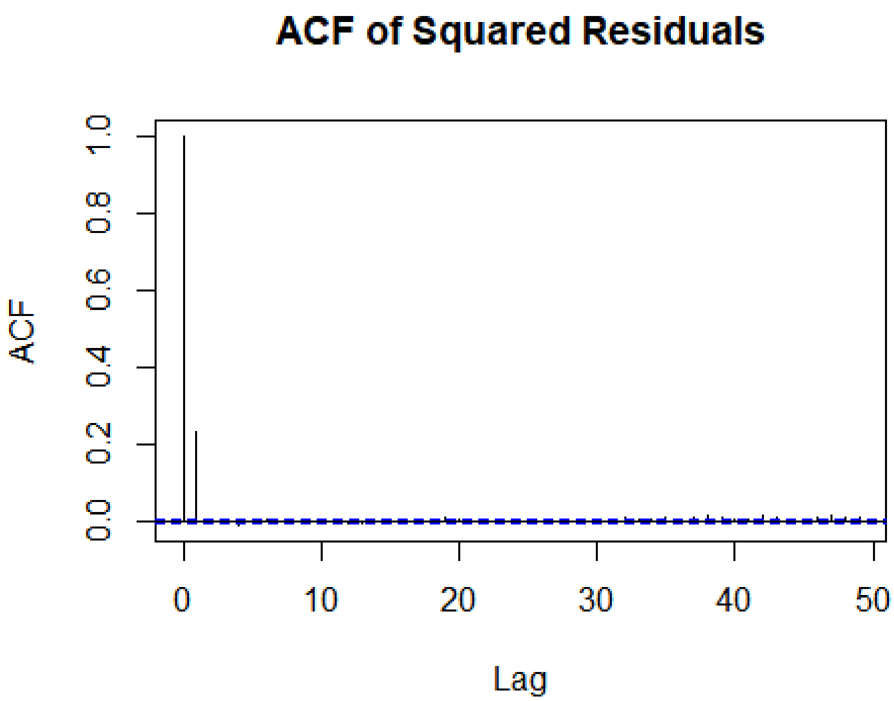
```
## ------------------------------------
##              Statistic Shape Scale   P-Value
## ARCH Lag[3]      27.91 0.500 2.000 1.274e-07
## ARCH Lag[5]      67.43 1.440 1.667 0.000e+00
## ARCH Lag[7]      83.20 2.315 1.543 0.000e+00
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  35.1764
## Individual Statistics:
## mu     6.860e-06
## omega  6.861e+00
## alpha1 5.688e+00
## beta1  5.626e+00
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.07 1.24 1.6
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value       prob sig
## Sign Bias            13.23  6.290e-40 ***
## Negative Sign Bias   25.91 1.726e-147 ***
## Positive Sign Bias   63.96  0.000e+00 ***
## Joint Effect       4947.92  0.000e+00 ***
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     87.94    7.682e-11
## 2    30    111.06    1.549e-11
## 3    40    148.95    9.263e-15
## 4    50    158.70    1.688e-13
##
##
## Elapsed time : 3.376628
```

```r
# Plot fitted volatility
plot(garch_fit, which = "all")
```

```
##
## please wait...calculating quantiles...
```




```r
# Perform residual diagnostics
residuals <- residuals(garch_fit)
acf(residuals^2, main = "ACF of Squared Residuals")
```




```r
# Save model output to a file
capture.output(summary(garch_fit), file = "garch_summary.txt")
```

```r
# MODEL SELECTION

# Load necessary libraries
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ lubridate 1.9.3     ✔ tibble    3.2.1
## ✔ purrr     1.0.2     ✔ tidyr     1.3.1
## ── Conflicts ────────────────────────────────────────── tidyverse_confl
icts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ✖ purrr::reduce() masks rugarch::reduce()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

library(lubridate)
library(rugarch)

#loaded data sets
prices<-read.csv("C:/Users/munir/Desktop/nse prices.csv")
volumes<-read.csv("C:/Users/munir/Desktop/nse volumes.csv")


# Preprocess the 'prices' dataset
prices$DATE <- dmy(prices$DATE)  # Convert DATE to Date format
prices <- prices %>%
  arrange(ISSUER, DATE) %>%
  group_by(ISSUER) %>%
  mutate(Return = c(NA, diff(log(VWAP))))  # Calculate log returns

# Preprocess the 'volumes' dataset
volumes$VOLUME <- as.numeric(gsub(",", "", trimws(volumes$VOLUME)))  # Cle
an and convert VOLUME to numeric
volumes$MONTH <- my(volumes$MONTH)  # Convert MONTH to Date format

# Merge the datasets
merged_data <- left_join(prices, volumes, by = c("ISSUER" = "ISSUER", "DAT
E" = "MONTH"))

# Filter out rows with missing data
merged_data <- merged_data %>% filter(!is.na(Return) & !is.na(VOLUME))

# Specify the GARCH(1,1) model with volume as an external regressor
spec <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1, 1), external.r
egressors = matrix(merged_data$VOLUME)),
  mean.model = list(armaOrder = c(0, 0), include.mean = TRUE),
  distribution.model = "norm"
)
```

```r
# Fit the model
fit <- ugarchfit(spec = spec, data = merged_data$Return)

# Output the results
print(fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error    t value Pr(>|t|)
## mu      0.000562    0.001200   0.467947  0.63982
## omega   0.000005    0.000004   1.386167  0.16570
## alpha1  0.049820    0.004538  10.979103  0.00000
## beta1   0.867023    0.004023 215.539392  0.00000
## vxreg1  0.000000    0.000000   0.002527  0.99798
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.000562    0.004523 0.124140  0.90121
## omega   0.000005    0.000360 0.014406  0.98851
## alpha1  0.049820    0.277642 0.179438  0.85759
## beta1   0.867023    0.121712 7.123591  0.00000
## vxreg1  0.000000    0.001045 0.000000  1.00000
##
## LogLikelihood : 2389.767
##
## Information Criteria
## ------------------------------------
##
## Akaike       -2.3131
## Bayes        -2.2994
## Shibata      -2.3131
## Hannan-Quinn -2.3081
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.003213  0.9548
## Lag[2*(p+q)+(p+q)-1][2]  1.250368  0.4237
## Lag[4*(p+q)+(p+q)-1][5]  2.272937  0.5555
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
```

```
## ------------------------------------
##                         statistic p-value
## Lag[1]                  0.0003515   0.985
## Lag[2*(p+q)+(p+q)-1][5] 0.0077962   1.000
## Lag[4*(p+q)+(p+q)-1][9] 0.0132847   1.000
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]  0.005667 0.500 2.000  0.9400
## ARCH Lag[5]  0.013659 1.440 1.667  0.9992
## ARCH Lag[7]  0.015437 2.315 1.543  1.0000
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  114.2244
## Individual Statistics:
## mu       0.1498
## omega    0.7113
## alpha1   0.7033
## beta1    0.2333
## vxreg1 112.7109
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:         1.28 1.47 1.88
## Individual Statistic:    0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.8098 0.4182
## Negative Sign Bias  0.2818 0.7781
## Positive Sign Bias  0.2648 0.7912
## Joint Effect        0.6681 0.8807
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20      4984            0
## 2    30      6297            0
## 3    40      7148            0
## 4    50      8348            0
##
##
## Elapsed time : 0.096946
```

```
#MODEL SELECTION
#OBJECTIVE 1: Causality between trading volume and stock price movement

# Load necessary libraries
library(tidyverse)
```

```r
library(lubridate)
library(vars)

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: strucchange

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

##
## Attaching package: 'strucchange'

## The following object is masked from 'package:stringr':
##
##     boundary

## Loading required package: urca

## Loading required package: lmtest

library(dplyr) # Explicitly load dplyr


# Load the datasets
prices <- read.csv("C:/Users/munir/Desktop/nse prices.csv")
volumes <- read.csv("C:/Users/munir/Desktop/nse volumes.csv")

# Convert date columns to appropriate formats
prices$DATE <- dmy(prices$DATE)
volumes$MONTH <- my(volumes$MONTH)

# Clean the VOLUME column in the volumes dataset
volumes$VOLUME <- as.numeric(gsub(",", "", trimws(volumes$VOLUME)))

# Merge datasets on ISSUER and ensure date alignment
merged_data <- merge(prices, volumes, by = "ISSUER")
merged_data <- merged_data %>%
  mutate(
    DATE = as.Date(DATE),
    MONTH = floor_date(MONTH, "month")
  ) %>%
```

```
  filter(DATE >= as.Date("2012-01-01") & DATE <= as.Date("2018-12-31"))

# Calculate log returns for each ISSUER
merged_data <- merged_data %>%
  group_by(ISSUER) %>%
  arrange(DATE) %>%
  mutate(
    LOG_PRICE = log(VWAP),
    RETURNS = c(NA, diff(LOG_PRICE))
  ) %>%
  ungroup()


# Analyze causality relationship year by year
results <- list()

for (year in 2012:2018) {
  # Filter data for the year
  yearly_data <- merged_data %>%
    filter(lubridate::year(DATE) == year) %>%
    drop_na(RETURNS, VOLUME) # Drop rows with NA values in RETURNS and VOL
UME

  # Ensure data is sufficient for analysis
  if (nrow(yearly_data) > 10) {
    # Select relevant columns for the VAR model
    var_data <- yearly_data %>%
      dplyr::select(RETURNS, VOLUME) # Use dplyr::select() explicitly

    # Fit VAR model
    var_model <- VAR(var_data, p = 2, type = "const")

    # Perform Granger causality test
    granger_results <- list(
      "Volume causes Returns" = causality(var_model, cause = "VOLUME"),
      "Returns cause Volume" = causality(var_model, cause = "RETURNS")
    )

    # Store results
    results[[as.character(year)]] <- granger_results
  } else {
    results[[as.character(year)]] <- "Insufficient data for analysis"
  }
}

# Display results
results

## $`2012`
## $`2012`$`Volume causes Returns`
## $`2012`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
```

```
## data:  VAR object var_model
## F-Test = 0.016872, df1 = 2, df2 = 1886404, p-value = 0.9833
##
##
## $`2012`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 0.63252, df = 1, p-value = 0.4264
##
##
##
## $`2012`$`Returns cause Volume`
## $`2012`$`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 0.30919, df1 = 2, df2 = 1886404, p-value = 0.734
##
##
## $`2012`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 0.63252, df = 1, p-value = 0.4264
##
##
##
##
## $`2013`
## $`2013`$`Volume causes Returns`
## $`2013`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
## data:  VAR object var_model
## F-Test = 0.53708, df1 = 2, df2 = 1979714, p-value = 0.5845
##
##
## $`2013`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 0.50339, df = 1, p-value = 0.478
##
##
##
## $`2013`$`Returns cause Volume`
## $`2013`$`Returns cause Volume`$Granger
##
```

```
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 1.0553, df1 = 2, df2 = 1979714, p-value = 0.3481
##
##
## $`2013`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 0.50339, df = 1, p-value = 0.478
##
##
##
##
## $`2014`
## $`2014`$`Volume causes Returns`
## $`2014`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
## data:  VAR object var_model
## F-Test = 0.021004, df1 = 2, df2 = 2123466, p-value = 0.9792
##
##
## $`2014`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 0.1024, df = 1, p-value = 0.749
##
##
##
## $`2014`$`Returns cause Volume`
## $`2014`$`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 0.10056, df1 = 2, df2 = 2123466, p-value = 0.9043
##
##
## $`2014`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 0.1024, df = 1, p-value = 0.749
##
##
##
##
```

```
## $`2015`
## $`2015`$`Volume causes Returns`
## $`2015`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
## data:  VAR object var_model
## F-Test = 0.08343, df1 = 2, df2 = 2117588, p-value = 0.92
##
##
## $`2015`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 0.94148, df = 1, p-value = 0.3319
##
##
##
## $`2015`$`Returns cause Volume`
## $`2015`$`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 0.42744, df1 = 2, df2 = 2117588, p-value = 0.6522
##
##
## $`2015`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 0.94148, df = 1, p-value = 0.3319
##
##
##
##
## $`2016`
## $`2016`$`Volume causes Returns`
## $`2016`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
## data:  VAR object var_model
## F-Test = 0.20353, df1 = 2, df2 = 2053558, p-value = 0.8158
##
##
## $`2016`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 1.2054, df = 1, p-value = 0.2722
```

```
##
##
##
## $`2016`$`Returns cause Volume`
## $`2016`$`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 0.34882, df1 = 2, df2 = 2053558, p-value = 0.7055
##
##
## $`2016`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 1.2054, df = 1, p-value = 0.2722
##
##
##
##
## $`2017`
## $`2017`$`Volume causes Returns`
## $`2017`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
## data:  VAR object var_model
## F-Test = 2.5498, df1 = 2, df2 = 1996030, p-value = 0.0781
##
##
## $`2017`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 1.4511, df = 1, p-value = 0.2283
##
##
##
## $`2017`$`Returns cause Volume`
## $`2017`$`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 0.1557, df1 = 2, df2 = 1996030, p-value = 0.8558
##
##
## $`2017`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
```

```
## data:  VAR object var_model
## Chi-squared = 1.4511, df = 1, p-value = 0.2283
##
##
##
##
## $`2018`
## $`2018`$`Volume causes Returns`
## $`2018`$`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
## data:  VAR object var_model
## F-Test = 0.38649, df1 = 2, df2 = 2058926, p-value = 0.6794
##
##
## $`2018`$`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 0.95579, df = 1, p-value = 0.3282
##
##
##
## $`2018`$`Returns cause Volume`
## $`2018`$`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 0.14991, df1 = 2, df2 = 2058926, p-value = 0.8608
##
##
## $`2018`$`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 0.95579, df = 1, p-value = 0.3282

#general analysis for the causality relationship

# Load necessary libraries
library(tidyverse)
library(lubridate)
library(vars)

# Load the datasets
prices <- read.csv("C:/Users/munir/Desktop/nse prices.csv")
volumes <- read.csv("C:/Users/munir/Desktop/nse volumes.csv")

# Convert date columns to appropriate formats
prices$DATE <- dmy(prices$DATE)
```

```r
volumes$MONTH <- my(volumes$MONTH)

# Clean the VOLUME column in the volumes dataset
volumes$VOLUME <- as.numeric(gsub(",", "", trimws(volumes$VOLUME)))

# Merge datasets on ISSUER and align dates
merged_data <- merge(prices, volumes, by = "ISSUER")
merged_data <- merged_data %>%
  mutate(
    DATE = as.Date(DATE),
    MONTH = floor_date(MONTH, "month")
  ) %>%
  filter(DATE >= as.Date("2012-01-01") & DATE <= as.Date("2018-12-31"))

# Calculate log returns for each ISSUER
merged_data <- merged_data %>%
  group_by(ISSUER) %>%
  arrange(DATE) %>%
  mutate(
    LOG_PRICE = log(VWAP),
    RETURNS = c(NA, diff(LOG_PRICE))
  ) %>%
  ungroup()

# Drop rows with NA values in RETURNS and VOLUME
final_data <- merged_data %>%
  drop_na(RETURNS, VOLUME)

# Ensure there is sufficient data for the analysis
if (nrow(final_data) > 10) {
  # Select relevant columns for the VAR model
  var_data <- final_data %>%
    dplyr::select(RETURNS, VOLUME)

  # Fit the VAR model
  var_model <- VAR(var_data, p = 2, type = "const")

  # Perform Granger causality test
  granger_results <- list(
    "Volume causes Returns" = causality(var_model, cause = "VOLUME"),
    "Returns cause Volume" = causality(var_model, cause = "RETURNS")
  )

  # Display results
  print(granger_results)
} else {
  print("Insufficient data for analysis")
}

## $`Volume causes Returns`
## $`Volume causes Returns`$Granger
##
##  Granger causality H0: VOLUME do not Granger-cause RETURNS
##
```

```
## data:  VAR object var_model
## F-Test = 0.06537, df1 = 2, df2 = 14215770, p-value = 0.9367
##
##
## $`Volume causes Returns`$Instant
##
##  H0: No instantaneous causality between: VOLUME and RETURNS
##
## data:  VAR object var_model
## Chi-squared = 2.4443, df = 1, p-value = 0.1179
##
##
##
## $`Returns cause Volume`
## $`Returns cause Volume`$Granger
##
##  Granger causality H0: RETURNS do not Granger-cause VOLUME
##
## data:  VAR object var_model
## F-Test = 1.4781, df1 = 2, df2 = 14215770, p-value = 0.2281
##
##
## $`Returns cause Volume`$Instant
##
##  H0: No instantaneous causality between: RETURNS and VOLUME
##
## data:  VAR object var_model
## Chi-squared = 2.4443, df = 1, p-value = 0.1179
```

```
# Load necessary libraries
library(tidyverse)
library(lubridate)
library(rugarch)
# Load the datasets
prices <- read.csv("C:/Users/munir/Desktop/nse prices.csv")
volumes <- read.csv("C:/Users/munir/Desktop/nse volumes.csv")

# Preprocess the 'prices' dataset
prices$DATE <- dmy(prices$DATE)  # Convert DATE to Date format
prices <- prices %>%
  arrange(ISSUER, DATE) %>%
  group_by(ISSUER) %>%
  mutate(Return = c(NA, diff(log(VWAP))))  # Calculate log returns

# Preprocess the 'volumes' dataset
volumes$VOLUME <- as.numeric(gsub(",", "", trimws(volumes$VOLUME)))  # Cle
an and convert VOLUME to numeric
volumes$MONTH <- my(volumes$MONTH)  # Convert MONTH to Date format

# Merge the datasets
merged_data <- left_join(prices, volumes, by = c("ISSUER" = "ISSUER", "DAT
E" = "MONTH"))

# Filter out rows with missing data
```

```r
merged_data <- merged_data %>% filter(!is.na(Return) & !is.na(VOLUME))

# Add a YEAR column to facilitate year-wise analysis
merged_data <- merged_data %>% mutate(YEAR = year(DATE))

# Define a function to fit the GARCH-X(1,1) model for a given year
fit_garchx_for_year <- function(data, year) {
  cat("Fitting GARCH-X(1,1) for year:", year, "\n")
  yearly_data <- data %>% filter(YEAR == year)

  if (nrow(yearly_data) == 0) {
    cat("No data available for year", year, "\n")
    return(NULL)
  }

  # Specify the GARCH-X(1,1) model with volume as an external regressor in
the variance equation
  spec <- ugarchspec(
    variance.model = list(
      model = "sGARCH",  # Standard GARCH model
      garchOrder = c(1, 1),  # GARCH(1,1)
      external.regressors = matrix(yearly_data$VOLUME)  # Include trading
volume as an external regressor
    ),
    mean.model = list(
      armaOrder = c(0, 0),  # No ARMA terms in the mean equation
      include.mean = TRUE  # Include a mean term
    ),
    distribution.model = "norm"  # Assume normal distribution for residual
s
  )

  # Fit the model
  fit <- tryCatch({
    ugarchfit(spec = spec, data = yearly_data$Return)
  }, error = function(e) {
    cat("Error fitting model for year", year, ":", e$message, "\n")
    return(NULL)
  })

  return(fit)
}

# Loop through years 2012 to 2018 and fit the GARCH-X(1,1) model
results <- list()
for (year in 2012:2018) {
  fit <- fit_garchx_for_year(merged_data, year)
  if (!is.null(fit)) {
    results[[as.character(year)]] <- fit
    print(fit)  # Print the results for each year
  }
}
```

```
## Fitting GARCH-X(1,1) for year: 2012
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error    t value Pr(>|t|)
## mu      0.002864    0.003642   0.786220 0.431738
## omega   0.000004    0.000021   0.180222 0.856978
## alpha1  0.049938    0.024621   2.028262 0.042533
## beta1   0.868002    0.007770 111.714615 0.000000
## vxreg1  0.000000    0.000000   0.000477 0.999620
##
## Robust Standard Errors:
##         Estimate  Std. Error   t value Pr(>|t|)
## mu      0.002864    0.002738  1.045751  0.29568
## omega   0.000004    0.000059  0.066084  0.94731
## alpha1  0.049938    0.016491  3.028194  0.00246
## beta1   0.868002    0.060032 14.459114  0.00000
## vxreg1  0.000000    0.000142  0.000001  1.00000
##
## LogLikelihood : 288.8294
##
## Information Criteria
## ------------------------------------
##
## Akaike       -2.4156
## Bayes        -2.3420
## Shibata      -2.4164
## Hannan-Quinn -2.3859
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.005395  0.9414
## Lag[2*(p+q)+(p+q)-1][2]  0.045828  0.9596
## Lag[4*(p+q)+(p+q)-1][5]  0.078236  0.9988
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.006403  0.9362
## Lag[2*(p+q)+(p+q)-1][5]  0.020905  0.9999
## Lag[4*(p+q)+(p+q)-1][9]  0.035424  1.0000
```

```
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##              Statistic Shape Scale P-Value
## ARCH Lag[3]  0.007238 0.500 2.000  0.9322
## ARCH Lag[5]  0.017538 1.440 1.667  0.9989
## ARCH Lag[7]  0.026806 2.315 1.543  1.0000
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  60.5522
## Individual Statistics:
## mu      0.07497
## omega   0.18594
## alpha1  0.38938
## beta1   0.19490
## vxreg1 45.77912
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias          0.41934 0.6754
## Negative Sign Bias 0.45163 0.6520
## Positive Sign Bias 0.01094 0.9913
## Joint Effect       0.72600 0.8671
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     648.0   3.695e-125
## 2    30     797.4   4.079e-149
## 3    40     766.7   7.724e-136
## 4    50     781.0   1.763e-132
##
##
## Elapsed time : 0.6795769
##
## Fitting GARCH-X(1,1) for year: 2013
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
```

```
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error   t value Pr(>|t|)
## mu      0.001762    0.001575  1.118294 0.263441
## omega   0.000001    0.000003  0.229070 0.818814
## alpha1  0.049981    0.017184  2.908534 0.003631
## beta1   0.866340    0.026957 32.138197 0.000000
## vxreg1  0.000000    0.000000  0.000033 0.999973
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.001762    0.002062 0.854430  0.39287
## omega   0.000001    0.000006 0.105734  0.91579
## alpha1  0.049981    0.070296 0.711005  0.47708
## beta1   0.866340    0.128933 6.719279  0.00000
## vxreg1  0.000000    0.000012 0.000001  1.00000
##
## LogLikelihood : 570.8517
##
## Information Criteria
## ------------------------------------
##
## Akaike       -4.1454
## Bayes        -4.0793
## Shibata      -4.1461
## Hannan-Quinn -4.1189
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.7427  0.3888
## Lag[2*(p+q)+(p+q)-1][2]    1.7966  0.2989
## Lag[4*(p+q)+(p+q)-1][5]    5.6146  0.1107
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                      2.578 0.10839
## Lag[2*(p+q)+(p+q)-1][5]     6.004 0.08977
## Lag[4*(p+q)+(p+q)-1][9]     6.665 0.22883
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.0432 0.500 2.000  0.8353
## ARCH Lag[5]    0.7426 1.440 1.667  0.8105
## ARCH Lag[7]    0.8312 2.315 1.543  0.9394
##
## Nyblom stability test
## ------------------------------------
```

```
## Joint Statistic:  30.4347
## Individual Statistics:
## mu      0.04693
## omega   0.80743
## alpha1  1.04841
## beta1   0.47111
## vxreg1 29.74640
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.3034 0.7618
## Negative Sign Bias  0.6004 0.5487
## Positive Sign Bias  0.2351 0.8143
## Joint Effect        0.5025 0.9184
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     244.3    4.427e-41
## 2    30     230.2    3.388e-33
## 3    40     318.8    1.351e-45
## 4    50     441.5    1.464e-64
##
##
## Elapsed time : 0.286829
##
## Fitting GARCH-X(1,1) for year: 2014
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error   t value Pr(>|t|)
## mu      0.006991    0.011425   0.61191 0.540594
## omega   0.000027    0.000137   0.19478 0.845562
## alpha1  0.049634    0.018086   2.74437 0.006063
## beta1   0.873263    0.006286 138.91261 0.000000
## vxreg1  0.000000    0.000000   0.00392 0.996873
##
## Robust Standard Errors:
```

```
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.006991    0.221705 0.031534  0.97484
## omega   0.000027    0.157113 0.000170  0.99986
## alpha1  0.049634    1.410568 0.035188  0.97193
## beta1   0.873263    4.190512 0.208390  0.83492
## vxreg1  0.000000    0.012951 0.000000  1.00000
##
## LogLikelihood : 17.7021
##
## Information Criteria
## ------------------------------------
##
## Akaike       -0.081164
## Bayes        -0.021320
## Shibata      -0.081663
## Hannan-Quinn -0.057249
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.1046  0.7464
## Lag[2*(p+q)+(p+q)-1][2]    0.1171  0.9079
## Lag[4*(p+q)+(p+q)-1][5]    0.1291  0.9969
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.005256  0.9422
## Lag[2*(p+q)+(p+q)-1][5]  0.016982  0.9999
## Lag[4*(p+q)+(p+q)-1][9]  0.029162  1.0000
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]  0.006027 0.500 2.000  0.9381
## ARCH Lag[5]  0.014554 1.440 1.667  0.9992
## ARCH Lag[7]  0.021357 2.315 1.543  1.0000
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  55.5415
## Individual Statistics:
## mu      0.3370
## omega   0.1316
## alpha1  0.9504
## beta1   0.2080
## vxreg1 14.2314
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
```

```
##
## Sign Bias Test
## ------------------------------------
##                      t-value   prob sig
## Sign Bias             0.5841 0.5596
## Negative Sign Bias  0.2242 0.8227
## Positive Sign Bias  0.0381 0.9696
## Joint Effect        0.3515 0.9501
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20      1880            0
## 2    30      2267            0
## 3    40      2355            0
## 4    50      2470            0
##
##
## Elapsed time : 0.1378851
##
## Fitting GARCH-X(1,1) for year: 2015
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error   t value Pr(>|t|)
## mu     -0.008297    0.002173 -3.818391 0.000134
## omega   0.000001    0.000004  0.244859 0.806566
## alpha1  0.049966    0.012874  3.881207 0.000104
## beta1   0.866591    0.015720 55.127917 0.000000
## vxreg1  0.000000    0.000000  0.000077 0.999939
##
## Robust Standard Errors:
##         Estimate  Std. Error   t value Pr(>|t|)
## mu     -0.008297    0.004441 -1.868371  0.06171
## omega   0.000001    0.000008  0.111151  0.91150
## alpha1  0.049966    0.044151  1.131710  0.25776
## beta1   0.866591    0.065064 13.319072  0.00000
## vxreg1  0.000000    0.000016  0.000001  1.00000
##
## LogLikelihood : 502.8606
##
## Information Criteria
## ------------------------------------
```

```
##
## Akaike       -3.7293
## Bayes        -3.6621
## Shibata      -3.7300
## Hannan-Quinn -3.7023
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                      4.229 0.03974
## Lag[2*(p+q)+(p+q)-1][2]     4.295 0.06283
## Lag[4*(p+q)+(p+q)-1][5]     5.017 0.15175
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.004467  0.9467
## Lag[2*(p+q)+(p+q)-1][5]  0.221987  0.9909
## Lag[4*(p+q)+(p+q)-1][9]  0.328201  0.9996
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.1597 0.500 2.000  0.6894
## ARCH Lag[5]    0.2407 1.440 1.667  0.9556
## ARCH Lag[7]    0.2802 2.315 1.543  0.9938
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  47.8275
## Individual Statistics:
## mu      0.1294
## omega   0.9513
## alpha1  0.4958
## beta1   0.7690
## vxreg1 46.7153
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.5603 0.5757
## Negative Sign Bias  0.3350 0.7379
## Positive Sign Bias  0.4839 0.6288
## Joint Effect        0.4053 0.9392
##
##
## Adjusted Pearson Goodness-of-Fit Test:
```

```
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     169.5    3.593e-26
## 2    30     202.1    7.475e-28
## 3    40     213.7    5.726e-26
## 4    50     224.9    2.255e-24
##
##
## Elapsed time : 0.05036402
##
## Fitting GARCH-X(1,1) for year: 2016
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error   t value Pr(>|t|)
## mu      0.000316    0.001545  0.204674  0.83783
## omega   0.000001    0.000002  0.631994  0.52739
## alpha1  0.049957    0.007853  6.361435  0.00000
## beta1   0.866671    0.017956 48.266770  0.00000
## vxreg1  0.000000    0.000000  0.000029  0.99998
##
## Robust Standard Errors:
##         Estimate  Std. Error   t value Pr(>|t|)
## mu      0.000316    0.004009  0.078866  0.93714
## omega   0.000001    0.000007  0.131871  0.89509
## alpha1  0.049957    0.030526  1.636538  0.10173
## beta1   0.866671    0.084173 10.296348  0.00000
## vxreg1  0.000000    0.000017  0.000000  1.00000
##
## LogLikelihood : 749.2016
##
## Information Criteria
## ------------------------------------
##
## Akaike       -3.6126
## Bayes        -3.5638
## Shibata      -3.6129
## Hannan-Quinn -3.5933
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.0042  0.9483
## Lag[2*(p+q)+(p+q)-1][2]    0.1307  0.8986
```

```
## Lag[4*(p+q)+(p+q)-1][5]    2.5688  0.4914
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                  0.0002403  0.9876
## Lag[2*(p+q)+(p+q)-1][5] 1.7377901  0.6815
## Lag[4*(p+q)+(p+q)-1][9] 3.4451861  0.6838
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]   0.03365 0.500 2.000  0.8545
## ARCH Lag[5]   2.98532 1.440 1.667  0.2919
## ARCH Lag[7]   3.72518 2.315 1.543  0.3879
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  36.8548
## Individual Statistics:
## mu      0.2981
## omega   4.3143
## alpha1  6.4154
## beta1   5.9512
## vxreg1 33.7616
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.1525 0.8789
## Negative Sign Bias  0.4345 0.6641
## Positive Sign Bias  0.7342 0.4632
## Joint Effect        1.2698 0.7363
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     410.7    2.580e-75
## 2    30     656.1   1.440e-119
## 3    40     891.7   9.068e-162
## 4    50    1137.8   3.982e-206
##
##
## Elapsed time : 0.1638479
##
## Fitting GARCH-X(1,1) for year: 2017
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error   t value Pr(>|t|)
## mu     -0.003920    0.002014 -1.946273 0.051622
## omega   0.000001    0.000003  0.330465 0.741048
## alpha1  0.049954    0.012274  4.069942 0.000047
## beta1   0.865153    0.016180 53.472023 0.000000
## vxreg1  0.000000    0.000000  0.000088 0.999930
##
## Robust Standard Errors:
##         Estimate  Std. Error   t value Pr(>|t|)
## mu     -0.003920    0.002492 -1.572825 0.115759
## omega   0.000001    0.000007  0.161436 0.871750
## alpha1  0.049954    0.027028  1.848254 0.064566
## beta1   0.865153    0.043545 19.868187 0.000000
## vxreg1  0.000000    0.000012  0.000001 0.999999
##
## LogLikelihood : 572.8625
##
## Information Criteria
## ------------------------------------
##
## Akaike       -3.6285
## Bayes        -3.5687
## Shibata      -3.6290
## Hannan-Quinn -3.6046
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.3199 0.57170
## Lag[2*(p+q)+(p+q)-1][2]    0.5600 0.66611
## Lag[4*(p+q)+(p+q)-1][5]    5.8535 0.09736
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic  p-value
## Lag[1]                   0.009306 0.923149
## Lag[2*(p+q)+(p+q)-1][5]  6.278197 0.077304
## Lag[4*(p+q)+(p+q)-1][9] 14.992037 0.003593
## d.o.f=2
```

```
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale   P-Value
## ARCH Lag[3]      3.88 0.500 2.000 0.0488697
## ARCH Lag[5]     15.20 1.440 1.667 0.0003410
## ARCH Lag[7]     17.51 2.315 1.543 0.0002723
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  50.0934
## Individual Statistics:
## mu      0.4956
## omega   1.1546
## alpha1  1.1957
## beta1   1.0807
## vxreg1 48.0698
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.8971 0.3704
## Negative Sign Bias  0.8343 0.4048
## Positive Sign Bias  0.2970 0.7667
## Joint Effect        0.9482 0.8138
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     129.7    1.627e-18
## 2    30     221.9    1.296e-31
## 3    40     243.0    2.624e-31
## 4    50     262.2    6.385e-31
##
##
## Elapsed time : 0.8221531
##
## Fitting GARCH-X(1,1) for year: 2018
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
```

```
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.005512    0.002054 2.683579 0.007284
## omega   0.000001    0.000004 0.248627 0.803649
## alpha1  0.049969    0.062148 0.804030 0.421380
## beta1   0.864409    0.105200 8.216847 0.000000
## vxreg1  0.000000    0.000000 0.000041 0.999967
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.005512    0.014294 0.385603  0.69979
## omega   0.000001    0.000063 0.014136  0.98872
## alpha1  0.049969    0.792688 0.063037  0.94974
## beta1   0.864409    1.360156 0.635522  0.52509
## vxreg1  0.000000    0.000013 0.000001  1.00000
##
## LogLikelihood : 506.8707
##
## Information Criteria
## ------------------------------------
##
## Akaike       -4.0311
## Bayes        -3.9605
## Shibata      -4.0319
## Hannan-Quinn -4.0027
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                  1.989e-05  0.9964
## Lag[2*(p+q)+(p+q)-1][2] 1.277e+00  0.4164
## Lag[4*(p+q)+(p+q)-1][5] 2.264e+00  0.5576
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic  p-value
## Lag[1]                    0.04481 0.832363
## Lag[2*(p+q)+(p+q)-1][5]  11.07239 0.004866
## Lag[4*(p+q)+(p+q)-1][9]  16.35843 0.001666
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale   P-Value
## ARCH Lag[3]     14.85 0.500 2.000 1.162e-04
## ARCH Lag[5]     18.87 1.440 1.667 4.111e-05
## ARCH Lag[7]     19.63 2.315 1.543 7.943e-05
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  20.2074
```

```
## Individual Statistics:
## mu      0.1909
## omega   0.4356
## alpha1  0.6489
## beta1   0.6150
## vxreg1 19.3750
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.7503 0.4538
## Negative Sign Bias  1.1739 0.2416
## Positive Sign Bias  0.6951 0.4877
## Joint Effect        1.8782 0.5981
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     163.9    4.353e-25
## 2    30     193.0    3.748e-26
## 3    40     208.7    4.673e-25
## 4    50     213.0    2.447e-22
##
##
## Elapsed time : 1.123082
```

```r
# Summary of results
cat("GARCH-X(1,1) model fitting completed for years 2012 to 2018.\n")
```

```
## GARCH-X(1,1) model fitting completed for years 2012 to 2018.
```