European Journal of Economic and Financial Research
ISSN: 2501-9430
ISSN-L: 2501-9430
Available on-line at: http://www.oapub.org/soc

DOI: 10.46827/ejefr.v8i7.1883 Volume 8 | Issue 7 | 2024

MODELING ELECTRONIC MONEY TRANSACTION FRAUD VOLUMES IN KENYA USING GENERALISED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC MODEL

Ondiwa Simon Oluoch[1],
Ondiwa Micah Oketch[2],
Ondiwa Domnic Odoyo[3],
Martin Sure Ondiwa[4i]
[1]Adjunct Lecturer, Dr.,
Department of Accounting and Finance,
Maseno University,
Kenya
[2]MSc, Financial Engineering,
World Quant University,
New Orleans, Louisiana,
United States of America
[3]Adjunct Lecturer,
Department of Accounting and Finance,
KCA University,
Nairobi, Kenya
[4]Geospatial Engineer,
School of Surveying and Geospatial Science,
The Technical University of Kenya,
Kenya

Abstract:

Electronic transaction fraud has been on the increase in recent times. Though technology advancement is cited as a major milestone in the global business environment, at times, it comes with challenges, such as financial risks. Businesses and individuals have been losing their hard-earned funds through online-related transaction frauds, and the trend continues to increase. Studies reviewed heavily used deep learning and machine learning to investigate the detection of online-related fraudulent activities. The current study, however, deviated from this norm by focusing on modelling electronic transaction fraud volumes using the Generalised Auto-regressive Conditional Heteroscedastic (GARCH) model. The study employed grid search cross-validation parameter optimisation techniques and popular loss functions, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) to find the best-fitting model. The study objective was to

[i] Correspondence: email sondiwa81@gmail.com, micah.oketch.ondiwa@gmail.com, domnicodoyo@gmail.com, surekondiwa@gmail.com

establish a suitable model for forecasting changes (Variations) in the volume of money lost from fraudulent transactions in Kenya. The study is anchored on the fraud triangle model and fraud diamond model. The study revealed that GARCH (1,1) model predicts electronic transaction fraud volume deviations in Kenya; with the prediction, the risk of financial losses can be averted. This study is important to stakeholders such as the public, corporations, regulators such as the Central Bank of Kenya, financial institutions such as commercial banks, and scholars.

JEL: E52, E43, G21, G23, C25, D12, E42

Keywords: modelling, electronic money, GARCH, machine learning, deep learning, fraud detection

1. Introduction

Electronic money transaction (EMT) systems have reshaped the financial landscape globally, offering unprecedented convenience and accessibility, particularly as evidenced in Kenya's robust mobile money ecosystem spearheaded by M-Pesa. This study delved into modelling electronic money transfer fraud volumes using the Generalised Autoregressive Conditional Heteroskedastic Model. EMT Systems have democratised financial transactions by enabling swift, secure, and reliable funds transfers, reducing dependence on traditional banking infrastructure, and enhancing financial inclusion (Regnard-Weinrabe *et al.*, 2024). However, this digital evolution has also ushered in new challenges, prominently electronic money transfer fraud. In Kenya, as in many other regions, common forms of fraud include identity theft, phishing scams, unauthorised transactions, and exploitation of vulnerabilities in mobile money platforms and banking systems (Mustafa, 2024).

Other commonly used methods by fraudsters to defraud their victims are the use of cell phone random calls, short message services, cybercrimes, bank account and mobile banking frauds (KNBS, 2021). Reviewed literature underscores that in a three-month review between September and December 2023, the Kenya Incident Response Team (KE-IRT) detected 52,075 mobile application fraud attempts targeting end users' devices. The fraud attempts constituted an increase of 94.15% compared to the previous three months, July and September 2023 (CAK, 2023). Electronic transactions have changed the business landscape globally, in Africa, East Africa and Kenya. Individual citizens who lost their funds through electronic fraud increased from 8.4% in 2019 to 47.4% in 2021. In Kenya, mobile money transfers rose from 8% in 2009 to 81% in 2021 (CBK, 2021).

According to Ogay (2020), the rapid adoption of EMT services has often outpaced the development of robust cybersecurity measures, leaving systems susceptible to cyber threats such as hacking and data breaches; this requires strategies for both fraud detection and prevention, which are unique and dynamic (Arkhipov *et al.*, 2021). Several factors contribute to the prevalence of EMT fraud in Kenya. Technological vulnerabilities arising from the fast-paced evolution of digital payment systems expose gaps that fraudsters exploit. From Bankole's (2024) point of view, inadequate consumer awareness regarding safe transaction practices and the intricacies of security features contributes significantly to susceptibility to fraud schemes.

Regulatory challenges, including gaps in frameworks and enforcement, further exacerbate vulnerabilities within the digital financial ecosystem (Ge, 2024). To combat these challenges, researchers and practitioners have developed sophisticated modelling approaches. According to Hetzel (2021), machine learning is used to analyse transaction patterns and detect fraud. Besides, analysing behaviour can also be used to monitor user behaviours, identify deviations from typical patterns, and trigger alerts for potential fraud (Burgalassi & Matsumoto, 2024).

Jawad & Naz (2023) assert that collaborative efforts are crucial in combating sophisticated fraud networks. Public-private partnerships facilitate sharing of information among various stakeholders, such as regulators, policymakers and financial institutions, enhancing collective resilience against evolving fraud tactics. The empirical evidence reviewed underscores the efficacy of two-factor authentication and customer education in bolstering cybersecurity awareness and reducing fraud incidents (Sanchez-Castany, 2024).

1.1 Objectives of the Study

1) The main objective of the study was to establish a suitable model for forecasting changes (Variations) in the volume of money lost from fraudulent transactions in Kenya.
2) The specific objective of the study was to establish whether the GARCH model is a suitable model for predicting the volume of money lost from fraudulent electronic transactions in Kenya.

1.2 Hypothesis of the Study

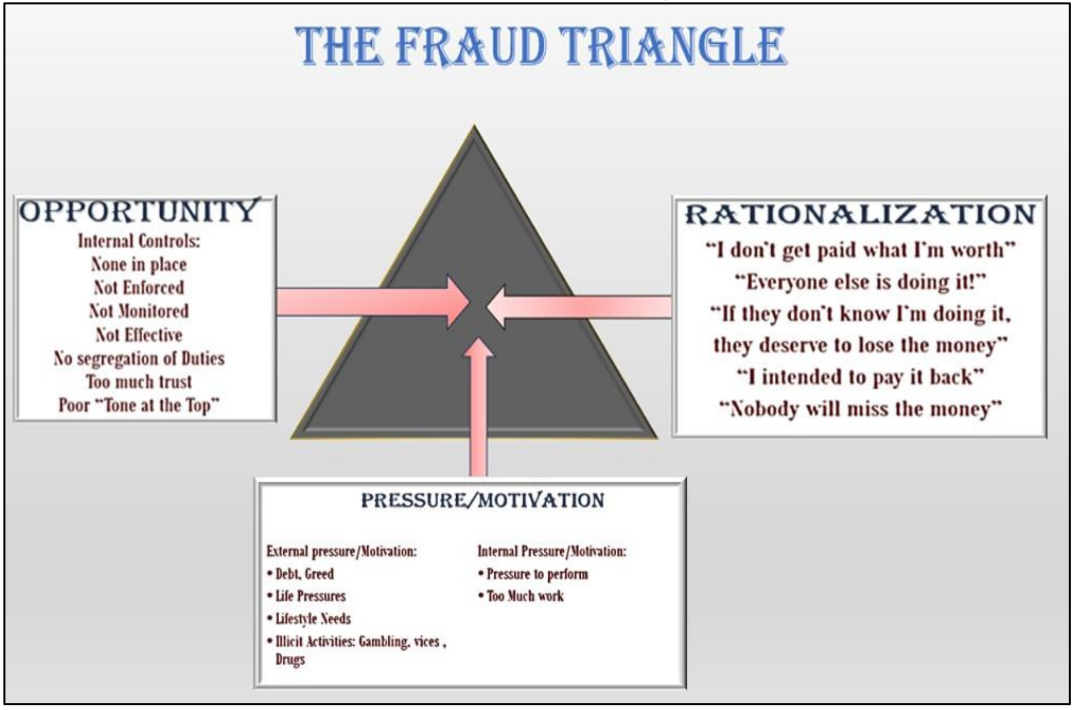
H$_0$1: The GARCH Model is not a suitable model for predicting the volume of money lost from fraudulent electronic transactions in Kenya.

2. Literature Review

2.1 Theoretical Underpinning

Fraud is an old phenomenon in the field of finance and accounting. Fraud can be traced from as early as the 1970s. One of the theories about fraud is the fraud triangle; the theory was proposed by Donald R. more than five decades ago. The theory outlines various conditions that may lead to fraud, which are occupational-related. The three instances are motivation, opportunity, and rationalisation. Opportunity implies that there must be something of value to steal; internal control weakness in oversight is an opportunity to steal. Pressure also motivates one to steal; personal situations and interests that may make someone commit fraud constitute pressure. Finally, rationalisation implies that a person should first realise that the gain from fraud is greater than the possibility of detection; hence, the fraudster would justify the fraud. Cressey (1971) asserts that when people are trusted, they become trust violators, especially when they find themselves in financial problems that they cannot share; they are also aware that such problems can be solved by violating the financial trust position.

Figure 1: Fraud Triangle



THE FRAUD TRIANGLE

OPPORTUNITY
Internal Controls:
None in place
Not Enforced
Not Monitored
Not Effective
No segregation of Duties
Too much trust
Poor "Tone at the Top"

RATIONALIZATION
"I don't get paid what I'm worth"
"Everyone else is doing it!"
"If they don't know I'm doing it, they deserve to lose the money"
"I intended to pay it back"
"Nobody will miss the money"

PRESSURE/MOTIVATION

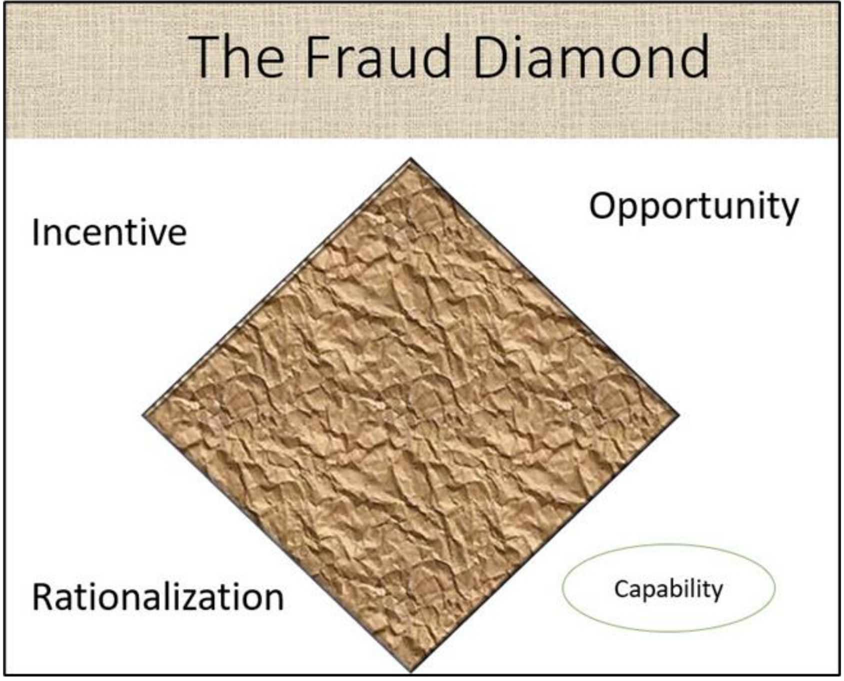
External pressure/Motivation:
- Debt, Greed
- Life Pressures
- Lifestyle Needs
- Illicit Activities: Gambling, vices , Drugs

Internal Pressure/Motivation:
- Pressure to perform
- Too Much work

Source: Cressey (1971).

The second theory guiding this study is the fraud diamond theory, which was proposed in 2004. The proponents of this theory were David Wolfe and Hermanson; the proponents averred that the capacity of fraudsters should be considered while analysing the chances of fraud happening. The fraudster must have required traits, including greed, weakness or character, excessive pride, and dishonesty. The proponents of the theory assert further that only 10% of the population will not commit fraud even if there is an opportunity to do so, 80% of the population may commit fraud in case there is an opportunity to do so and finally, the remaining 10% of the population will always be actively looking for the opportunity to commit fraud.

Figure 2: The Fraud Diamond



Source: Wolfe and Hermanson (2004).

The theories reviewed demonstrate that people are likely to be involved in fraud, given a favourable environment and opportunity favouring fraud. The theories, therefore, guide the study by making the audience understand how deeply fraud and dishonesty are rooted in contemporary society. In a broader sense, in the public sector, for instance, those entrusted with public resources always find themselves with an opportunity to commit fraud. Likewise, the public may occasionally be exposed to opportunities that lead to fraud justifications, leading to huge losses through electronic transactions. Recently, Kenya has witnessed large volumes of electronic transaction fraud. Therefore, these theories help us understand how fraudulent activities originate and justify the need to model electronic money transfer fraud in Kenya.

2.2 Empirical Literature Review

Lokanan (2023) studied mobile money transaction fraud. The study used machine learning to achieve its objective. The study underscored how easily mobile money transactions are common in cross-border transactions that pose money laundering threats. The study sourced data from real-time transactions to identify mobile transaction fraud. The study used ensemble, gradient descent and logistic models; the two models were also compared. Results revealed that logistic regression portrayed superior performance. The reviewed study used machine learning to study mobile money transactions but only concentrated on mobile money transactions, while the present study is wide in scope. Furthermore, the present study used the Generalised Auto-regressive Conditional Heteroscedastic (GARCH) model to assess volatility or deviation of monthly electronic fraud volumes in the Kenyan context. In this way, the results of the current study may be useful in knowing deviations of fraudulent transaction volumes in subsequent months, which in turn may be useful in averting electronic transaction fraud.

Okoye, Onuaha and Akuesodo (2021) studied the performance of commercial banks. The study used inferential statistics and revealed an inverse insignificant relationship between mobile banking fraud and the performance of commercial banks in Nigeria. The study recommended that commercial banks in Nigeria should be held responsible for any fraudulent e-commerce transactions. However, the present research modelled a single parameter to assess the deviation of monthly electronic transaction fraud volumes. Interest in electronic transaction fraud volumes is occasioned by the fact that modelling it may help know the magnitude of exposure to financial risk and action to avert the same applied. Therefore, the two studies used different approaches to solve different problems. The reviewed study can, therefore, not be generalised, especially in the Kenyan context.

Amiri and Hekmat (2021) investigated banking fraud from a customer perspective about categories and frameworks related to detecting and preventing fraud. The study underpinned the importance of mobile transactions to both bank customers and commercial banks. Besides, the study underscored the importance of fraud detection and prevention. The study concluded that anomaly detection, machine learning, and artificial intelligence are some of the key fraud detection approaches that can be explored. The reviewed study used machine learning to study banking fraud, customer perspectives and focused on fraud detection, while the present study used the GARCH model to study the deviation of electronic transaction fraud volumes in Kenya. The electronic transaction fraud volume assessment is imperative because the magnitude of such transactions is essential, especially when assessing the level of financial risk. The two studies are, therefore, not in tandem; hence, the reviewed study may not be generalised when assessing the volume of fraudulent electronic transactions.

Abdel-Rahman and Andriansyah (2023) studied artificial intelligence, fraud detection, compliance, and risk management in the modern banking environment. The researchers underpinned the importance of risk management and reduction of fraud in banking. Therefore, an integration of these aspects with artificial intelligence is imperative. Furthermore, the researchers opined that the traditional compliance-based approach needs to be updated in the modern banking business environment; in most cases, such traditional approaches are not real-time. It is observed that AI can detect fraudulent transactions in real time as they transpire. For instance, when neural networks are trained on historical fraudulent data, they can swiftly detect the possibility of fraudulent transactions. The study suggests using AI in commercial bank’s real-time fraud detection. The reviewed study used AI to detect fraud in the banking environment, while the present study used GARCH to model electronic transaction fraud volumes. Besides, the empirical study reviewed only focused on risk management in the modern banking environment, while the present study focused on the entire volume of electronic transaction frauds in Kenya and modelled its volatility for purposes of decision-making and proving useful information in averting exposure to financial risk by the players and stakeholders.

Gabudeanu *et al.* (2021) observed that the financial market literature has focused on data collection and prevention of financial banking-related frauds; consequently, legislation is also in tandem with this direction. However, the problematic question of data protection and privacy concerns comes into play. The study gives insight into complying with the legislation and detecting fraudulent financial banking activities under the law. The reviewed study only concentrated on financial banking fraud detection, while the present study assessed electronic transaction fraud volumes; the present study is broad. Besides, the study focused on fraud detection, while the present study modelled monthly electronic transaction fraud volumes using the GARCH model. Modelling electronic transaction fraud volume is essential since it assists in knowing the deviations of fraud volumes so that actions can be taken that fit the magnitude of fraud volumes to avert exposure to financial risks.

Marakalala (2023) studied biometric-based solutions and mobile fraud. The reviewed study was occasioned by the presence of many cybercriminals in the online space who constantly cause havoc in South Africa. The study concludes that forensic investigators should be preventive rather than reactive. Besides, the study concludes that investigation should be done progressively; law enforcement must also be done proactively. The study also outlines the best methods to use in controlling mobile fraud transactions, which include multi-factor authentication, computer emails and text alerts, online activity logging and behavioural analysis, multi-channel fraud and suspicious activity monitoring solutions, educating employees and consumers, constant cleaning and monitoring malware regularly, using secure access through https and managing online credentials wisely.

Marakalala (2023) assessed biometric solutions in fraud prevention, while the current study used the GARCH model. The GARCH model is suitable for the current study since the current study focused on monthly fraud volumes of electronic transactions; the GARCH model is suitable for assessing the volatility of monthly electronic transaction fraud volumes. Fraud prevention using biometric solutions is essential. However, it is also important to know and predict the magnitude of fraudulent activities to prevent the chances of such fraud occurring. The reviewed study was done in South Africa; the business environment in South Africa, legislation, and the electronic transaction volumes in South Africa and Kenya are different; hence, generalisation of the reviewed study may not be possible.

Shrivastava and Johari (2022) studied Neural Networks, mobile banking transactions, and fraud detection. The study revealed that the unauthorised use of mobile payment through credit cards as the most common banking fraud. Furthermore, intelligent fraud detection remains unexplored. The study asserted that fraudsters invented new methods of stealing money. The current study, however, used the GARCH model to model the monthly electronic transaction fraud volumes. Shrivastava and Johari (2022) focused on fraud detection, while the current study focused on modelling monthly electronic transaction fraud volumes and their volatility. Fraud detection is imperative; however, it is also essential to know the electronic transaction fraud volumes to ascertain the financial risk or exposure level so that a way forward is found. The reviewed study, just like other previous studies, focused on fraud detection as opposed to the assessment of monthly volatility of electronic transaction fraud volumes. The present study deviates from the reviewed studies by assessing the volatility of fraud volumes using the GARCH model. Fraud detection is imperative. However, it is also essential to know the movements of monthly deviations of electronic fraud volumes to take appropriate actions to prevent fraud and reduce financial risk.

Phiri, Lavhengwa and Segooa (2024) studied online banking fraud. The study areas were South Africa and Spain. The study underscored the importance of online banking by asserting that online banking allows for efficient and convenient access to banking services. The study employed a qualitative research paradigm. The results of this review revealed insufficient fraud detection experts in South Africa. However, in Spain, the study revealed that there is no law regulating online banking, which may pose a high risk. The study concludes that banks should hire fraud detection experts, and online regulations should also be in place to tame online banking fraud cases. The present study used the GARCH model, a more rigorous quantitative analysis approach. While the reviewed study used a qualitative approach using primary data collection approaches, qualitative approaches based on opinions are not the best in assessing a problem involving absolute numbers. Besides, the reviewed study was done in South Africa and Spain; despite these countries being more developed, the legislative and business environment differs from Kenya; hence, the reviewed study may not be generalised. Therefore, modelling electronic transaction fraud volumes in Kenya is essential.

Ahmad *et al.* (2021) studied e-banking frauds. These scholars underpinned that the emergence of technology in the banking industry has hastened fraud. The study employed a systematic literature review. Results revealed advanced security techniques that can be used to prevent fraud cases. The reviewed study focused on e-banking fraud. At the same time, the present study focused on electronic transaction fraud volumes. Electronic transaction fraud is broad in scope as compared to e-banking fraud. Besides, the study used a systematic literature review methodology. In contrast, the current study employed a quantitative design using the GARCH model to assess Kenya's monthly electronic transaction fraud volumes. The prevention study is, therefore, considered more rigorous than the reviewed study.

In conclusion, effectively modelling electronic money transfer fraud in Kenya necessitates a multifaceted approach integrating technological advancements, robust regulatory frameworks, and comprehensive user education initiatives (Bache, 2024). Continuous research and adaptation are essential to anticipate and mitigate emerging fraud trends, ensuring the continued integrity and security of digital financial systems in Kenya and beyond.

3. Material and Methods

3.1 Research Philosophy

Research philosophy has been defined differently by scholars; Creswell (2013) has referred to it as a worldview; Mertens (2010) and Cohen, Manion and Marrison (2000) referred to it as a paradigm; Crotty (1998) calls it epistemologies and ontologies; and Neuman (2009) says it is broadly conceived research methodologies. The current study is guided by positivist research philosophy. Positivism research philosophy views the social world objectively. This approach holds that knowledge develops by investigating social reality by observing facts.

3.2 Types of Data

This study used secondary data sourced from the Central Bank of Kenya's official website. The data set constitutes monthly data for the period ranging from 2010 to 2024; the data is reported in absolute volumes (millions) and was used in their absolute form for statistical analysis. To meet the statistical format for analysis and diagnostic tests, the data is transformed into their natural logarithms and differences.

3.3 Model Specification

It is observed from the previous tests that there are changing variances across the observations (heteroskedasticity). Since the changes in the volume of fraud are conceptually linked with the variance and standard deviation, modelling how the standard deviation of fraud volume changes over time is essential. A generalised auto-regressive conditional heteroscedastic GARCH (p, q) model is used; Auto-regressive conditional heteroscedastic {ARCH(p)} is used to model the variation in the conditional variance of the volume of fraudulent electronic transactions.

3.3.1 Background of the ARCH Process

The ARCH process is for modelling volatility clustering. Volatility clustering happens where periods of low deviations are followed by periods of low deviations, and periods of high deviations are followed by periods of high deviations (Engle, 1982). In the case of this study, the ARCH process is used in a case where high changes follow high changes in the amount of money lost from fraudulent electronic transactions and low changes in the amount of money lost from fraudulent transactions are followed by low changes in the amount of money lost from fraudulent electronic transactions. In the ARCH Process, variation at the current time is conditional or dependent on the changes in the previous times. The general ARCH (q) model is:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \quad (1)$$

Where:

q = the number of squared error/ past periods for modelling the conditional variance,

σ_t^2 = conditional Variance at time t,
α_0 = positive constant (time, t =0),
$\alpha_1, \alpha_2, \ldots, \alpha_q$ = positive coefficient of the past squared error at lags i = 1,2……q,
ε_{t-i}^2 = the error terms or residuals at lag i = 1,2……q.

3.3.2 The Framework: The GARCH Process

The GARCH process extends the ARCH process by adding the conditional variance to the squared error terms. Capturing the past conditional variance makes GARCH flexible and suitable for modelling time-varying volatilities. The general GARCH (p,q) equation is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \qquad (2)$$

p = the number of lagged squared error terms,
q = the number of lagged conditional variances,
σ_t^2= conditional variance at time t,
α_0 = positive constant (time, t = 0),
α_1, α_2, …, α_p = positive coefficients of the past squared errors at lags i = 1, 2, …, p,
β_i, β_i, …, β_i = positive coefficients of the past squared error at lags i=1, 2, …, q,
ε_{t-i}^2 = the positive coefficients of the past conditional variances at lags i = 1, 2, …, p.

The constant α_0 is the baseline volatility when there are no ARCH and conditional variances.

The ARCH Process captures the effect of the past squared residuals or shocks in the current conditional variances. The last part models the current conditional variance based on the past conditional variances and allows the model to capture volatility persistence. An example of the GARCH (1,1) model equation is shown below: -

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

3.3.3 Model Evaluation and Selection

Using different combinations of p and q, the best combination gave the best GARCH model considering the mean Absolute Error (MAE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Based on the four evaluation criteria, the best model is the one with the least MAE, RMSE, BIC and AIC. The MAE measures the difference between predicted and actual or observed variances. The equation of MAE is given by:

$$\text{MAE} = \frac{\sum_{t=1}^{n} |\bar{\sigma}^2 - \sigma^2|}{n}$$

Where:
$\bar{\sigma}^2$ = the predicted conditional variance,
σ^2 = the observed/ true conditional variance,
n = the number of data points.

The root squared mean error is the square root of the mean absolute error, hence:

$$\text{RSME} = \sqrt{\frac{\sum_{t=1}^{n} |\bar{\sigma}^2 - \sigma^2|}{n}}$$

Akaike Information Criterion (AIC) measures the relative amount of information lost from the prediction. The Akaike of the model is given by: -

$$AIC = 2k - 2ln(\bar{L}),$$

K = number of parameters in the model,
L = maximised likelihood function estimated from the model parameters.

The Bayesian Information Criterion (BIC) measures the goodness of fit of the model while penalising the complexity of the model. BIC is given by: -

$$BIC = kln(n) - 2ln(\bar{L})$$

Where:

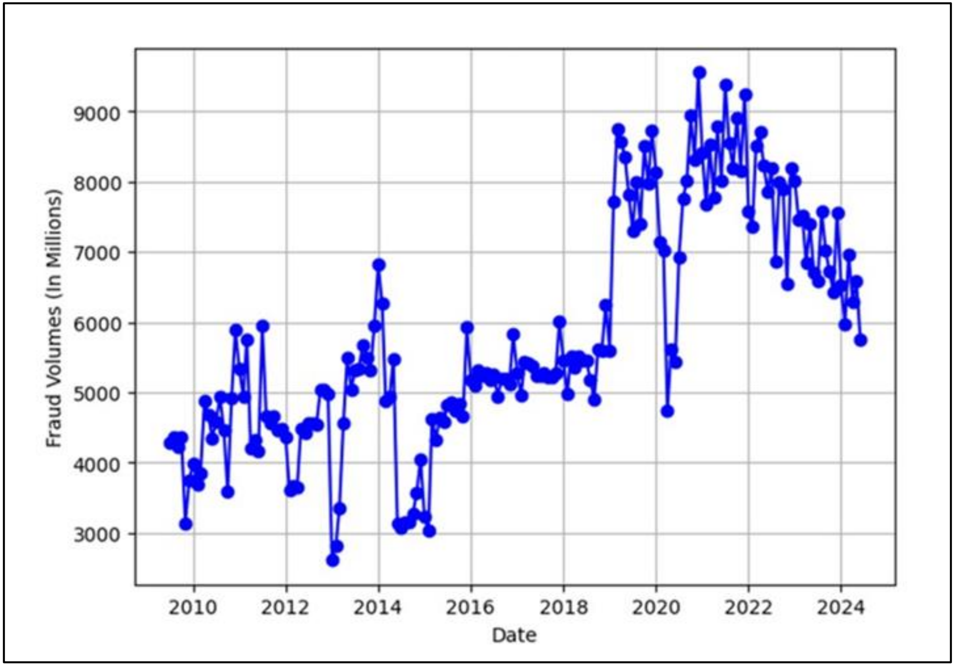
n = number of data points,
K = the number of parameters in the model,
L = maximised likelihood function estimated from the model parameters.

4. Results

4.1 Exploratory Data Analysis: Original Series

Figure 3 is a line plot showing the trends in fraud volume between the years 2010 and 2024. The figure indicates a significant rise in money lost in online fraudulent transactions in Kenya during this period. This rise in fraud volumes is attributed to technological advancements and innovations such as online banking and mobile money transfer, which were significantly adopted in Kenya. Similarly, figure 3 reveals that there was a surge in fraud volumes between 2020 and 2023, which is the period covering the peak of the COVID-19 Pandemic. The pandemic came with regulatory measures to the spread, including mandatory cashless transactions and a waiver of mobile and online banking transaction costs, resulting in an overall increase in the volume of online transactions.

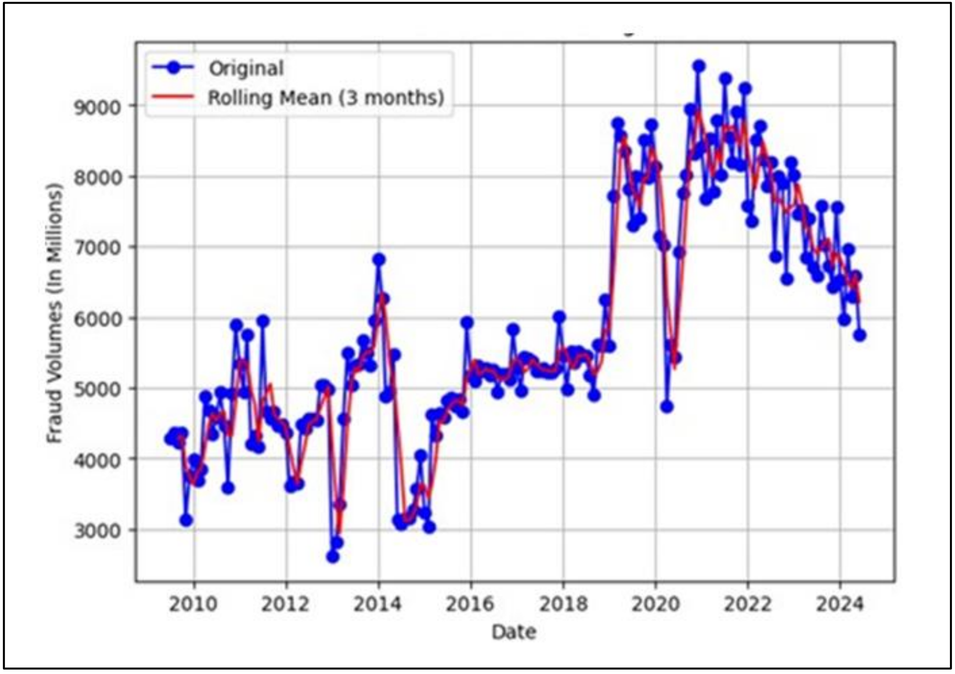
Figure 3: Plot Showing fraud Volume over Time



Source: Analysed Research Data.

Original fraud volumes and the rolling mean (moving averages) were compared to smooth out short-term fluctuations and highlight long-term trends without being distracted by the noise, as shown in Figure 4 below.

Figure 4: Fraud Volume with Rolling Mean

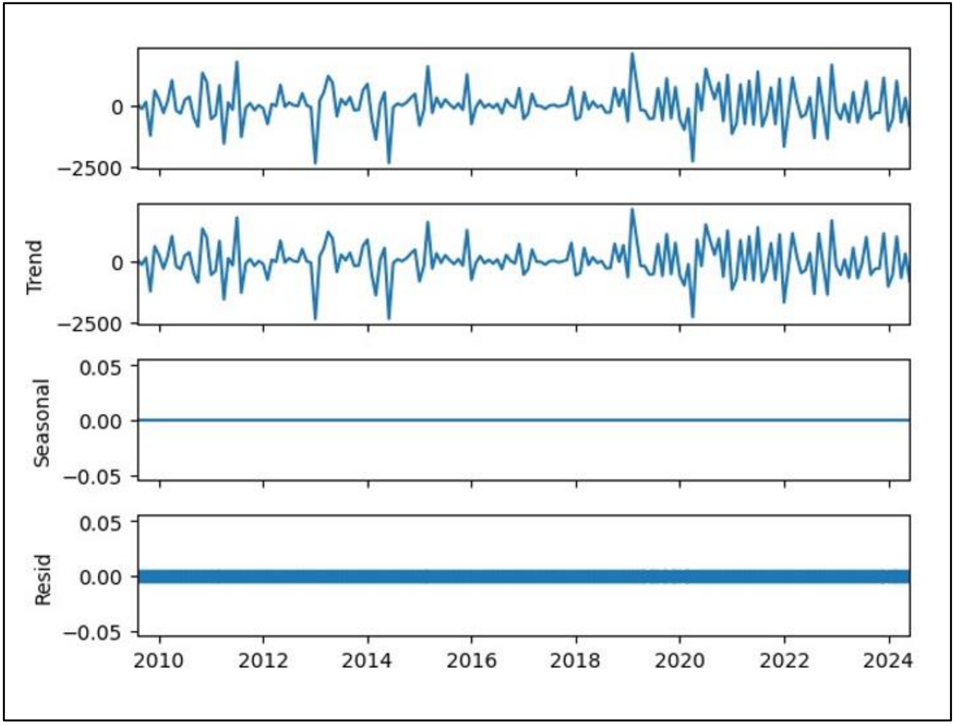


Source: Analysed Research Data.

The 3-month rolling means is obtained by averaging 3-month data to obtain a single data point, basically resulting in 4 data points in a year instead of 12. The idea behind using the rolling means technique is to establish whether there is a noticeable difference between the short-term and longer-term trends by comparing the original series and the rolling means series. Overall, it is observed that the rolling means have a relatively consistent trend as the original time-series plot. Finally, an overall rise in the volume of fraudulent transactions between 2010 and 2024, with a surge between 2020 and 2023, is witnessed.

The time series decomposition technique is used to investigate whether the variations in the volume of fraudulent transactions are affected by systematic trends such as seasonal variations or random trends. Figure 5 provides the decomposition plots, including the trend, seasonal component, and residual component. It is observed that there were no significant upward or downward variations in residuals of the volume of fraudulent transactions over time. Similarly, the seasonal plots show a nearly flat graph, suggesting that there are no significant seasonal patterns in the volume of fraudulent transactions. The residuals and seasonal plot suggest that fraud is not influenced by specific seasons over time. Overall, it is observed that fraudulent transactions are likely influenced by random factors rather than seasonal or systematic factors. Therefore, a modelling approach for the volume of fraudulent transactions should consider short-term periods rather than long-term periods.

Figure 5: Fraud Volumes in Millions



Source: Analysed Research Data.

4.2. Diagnostic Tests: Distribution Assumptions of a GARCH Model

4.2.1 Test 1: Stationarity

One of the key distribution assumptions of a GARCH model is stationarity, which implies that the series should have a constant mean and variance to capture long-term trends.

Augmented Dickey-Fuller Test is used to test for stationarity of the fraudulent transactions series, given a null hypothesis that the series has a unit root or is not stationary. Results in Table 1 demonstrate that the ADF statistic (-1.2224) is greater than the critical values at the 1% (-3.4706), 5% (-2.8792), and 10% (-2.5762) significance levels, so it does not fall within the rejection region. The p-value is higher than all three levels, so it does not fall within the rejection region. The p-value is also higher than all three levels of significance (p-value =0.6632>0.01, 0.05, 0.1). Therefore, the null hypothesis that the series has a unit root is not rejected. There is not enough evidence to conclude that the volume of fraud is stationary. Therefore, it is concluded that the series is likely to have a unit root and may be stationary.

Table 1: ADF Test

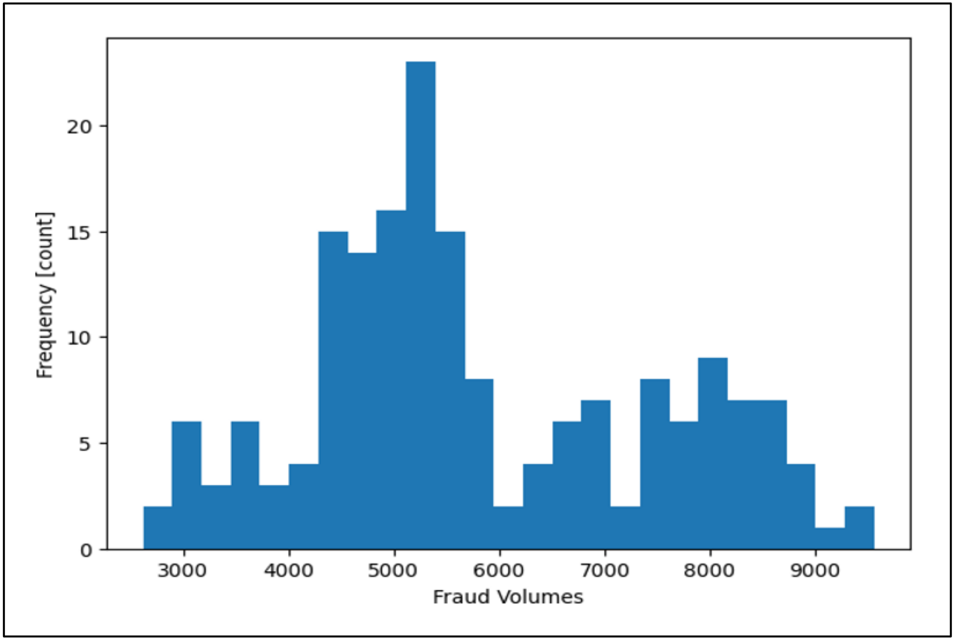
	Value
ADF Statistic	-1.2224
p-value	0.6632
Critical Values: 1%	-3.4706
Critical Values: 5%	-2.8792
Critical Values: 10%	-2.5762

Source: Analysed Research Data.

4.2.2 Test 2: Conditional Normality

The other key distribution assumption of a GARCH model is the conditional normality of the error terms, which implies that the series is normally distributed given the past information. A histogram is used to test for the normality of the fraud volume series. Figure 6 demonstrates that the fraud volume follows a non-normal distribution, with the observations skewed to both the left and right of the graph.

Figure 6: Fraud Volumes in Millions

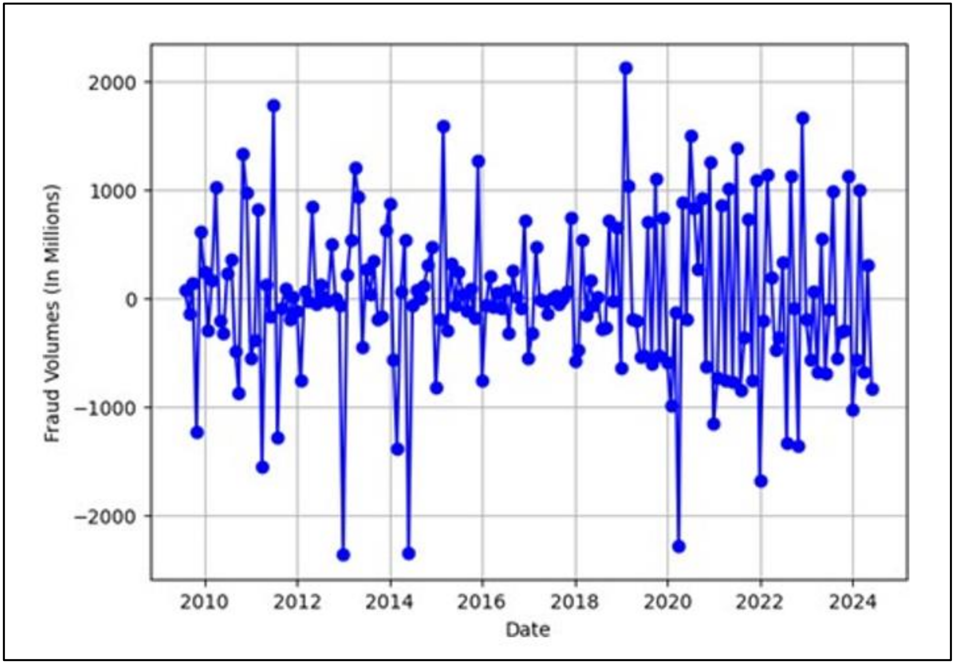


Source: Analysed Research Data.

4.3 Exploratory Data Analysis: Transformed Series

From the previous tests, it is established that the fraud volume series is non-stationary. To attain stationarity, the data is first transformed into their natural logarithms. The transformation does not make the data stationary. Therefore, the first differences were used for analysis, as shown in Figure 7 below. It is seen that after obtaining the first difference, the series is now stationary and meets the stationary assumption of a GARCH model.

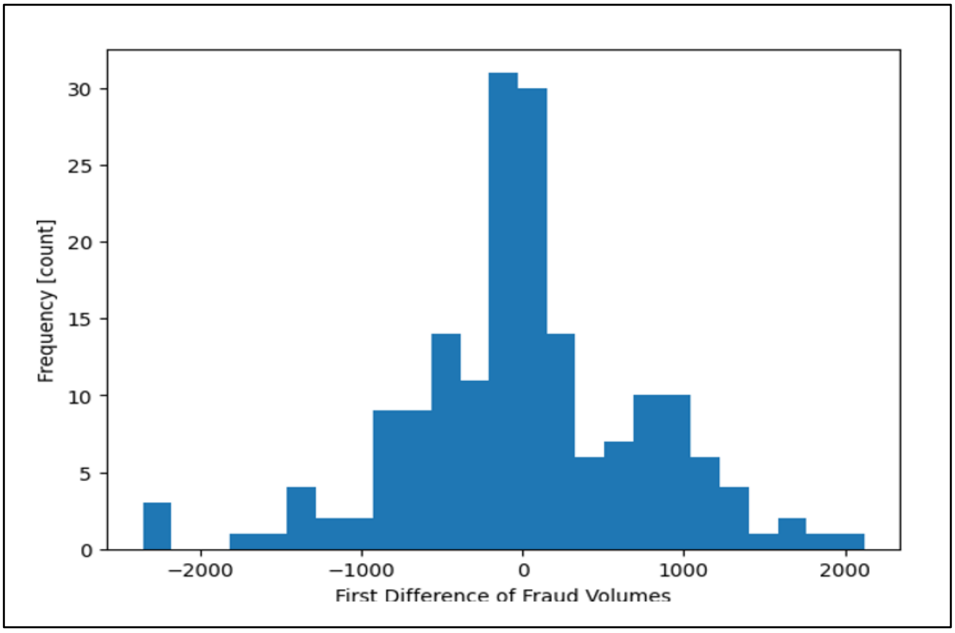
Figure 7: First Difference of Fraud Volume Over Time



Source: Analysed Research Data.

Besides fulfilling the stationarity condition, the first difference series also meets the conditional normality assumption. The histogram of the first difference series in Figure 8 below demonstrates a nearly bell-shaped distribution, which is typical of normal distribution. Therefore, the conclusion is that the conditional mean and variances of first differences, given the past observations, follow a normal distribution.

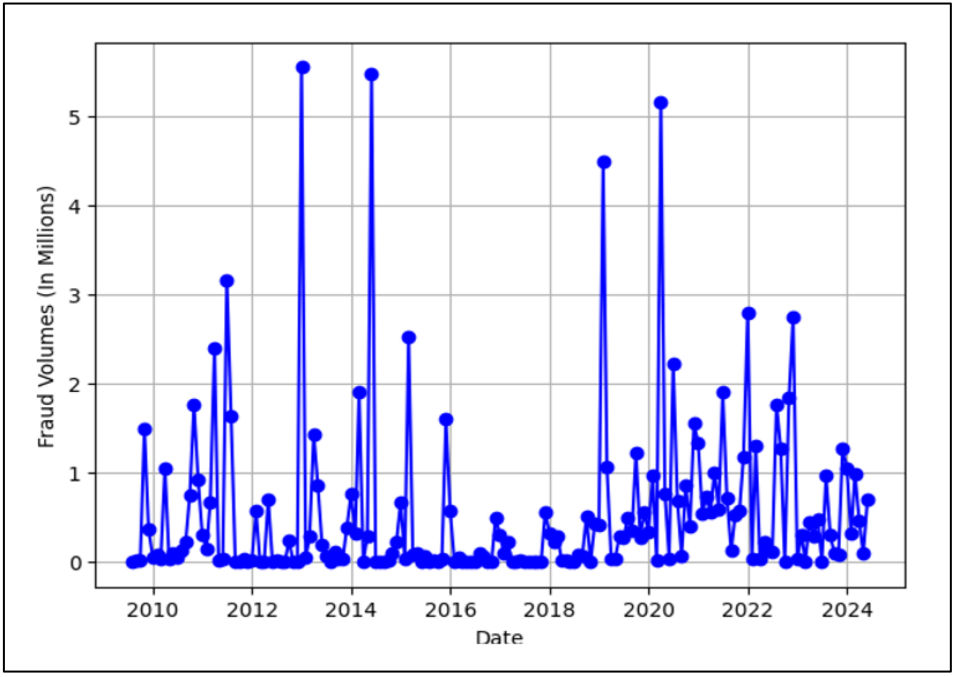
Figure 8: Normal Distribution of First Difference of Fraud Volumes (In Millions)



Source: Analysed Research Data

Figure 9 shows that the standard deviation goes up when the first difference of the fraud volumes changes drastically, either up or down. For instance, there was a big increase between 2014 and 2015, when there were several months of large negative differences.

Figure 9: Distribution of First Difference of Fraud Volumes (in Millions)



Source: Analysed Research Data.

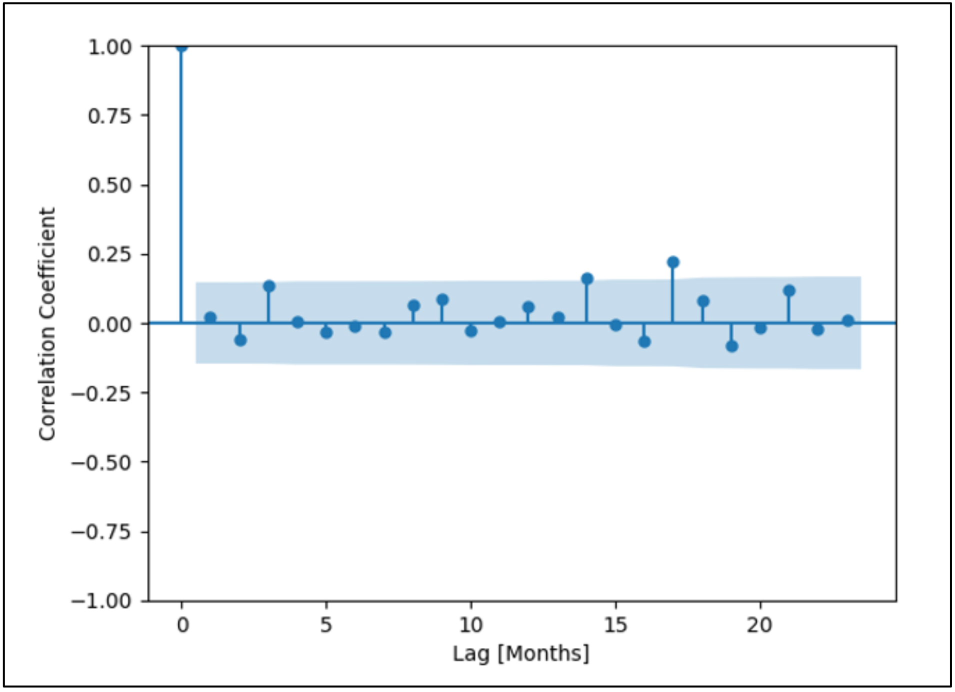
Furthermore, the standard deviation will decrease in 2025, when there will be only small day-to-day changes in electronic fraud transaction volumes. Therefore, the first difference values are used to check whether a high standard deviation in a particular month is associated with a high standard deviation in the following month. However, high standard deviations are caused by large changes in the first difference of the fraud volumes, which can be either positive or negative. Positive and negative numbers can be assessed together without the numbers cancelling each other by taking the absolute value of the numbers and then using them to calculate performance metrics like mean absolute error. The other solution, which is more common in this context, is to square all the values.

4.4 Parameter Selection and Model Evaluation

4.4.1 ACF and PACF- Criteria

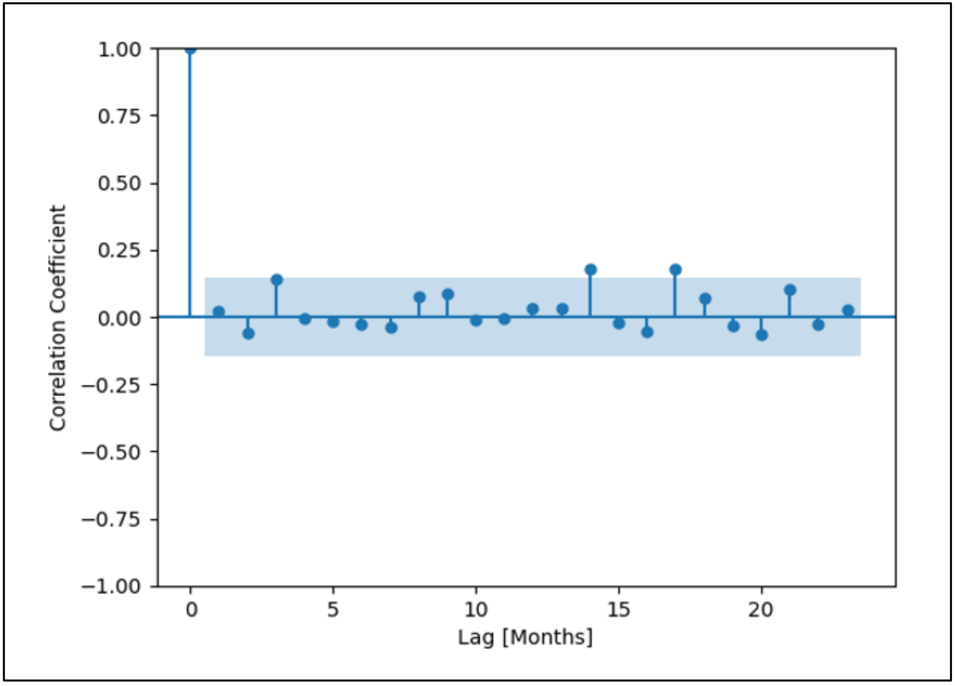
From Figure 10, periods of high and low differences and high difference months tend to cluster together. This is a perfect situation to use a GARCH model. From the ACF and PACF plots in Figures 10 and 11, the autocorrelation coefficient significantly differs from zero at lag 1. This indicates that the squared differences at time t are significantly correlated with those at time t-1, suggesting that a GARCH (1,1) model would be appropriate for this series of data. The performance of the GARCH models is ranked by evaluating the model’s ability to minimise the mean absolute error (MAE). The models with lower MAE indicate better accuracy and performance.

Figure 10: ACF of Squared Differences



Source: Analysed Research Data.

Figure 11: PACF of Squared Differences



Source: Analysed Research Data.

4.4.2 RMSE and MAE Criteria

Table 2 below reveals that GARCH (1,1) achieved the lowest mean absolute error (1044.65), outperforming the other variations of the GARCH model, suggesting that GARCH (1,1) provides a more accurate prediction than other variations of GARCH Models. In addition, it is concluded that as the complexity of the GARCH models increases, the mean absolute error also increases, suggesting that adding parameters to the GARCH model does not necessarily improve the accuracy of the predictions.

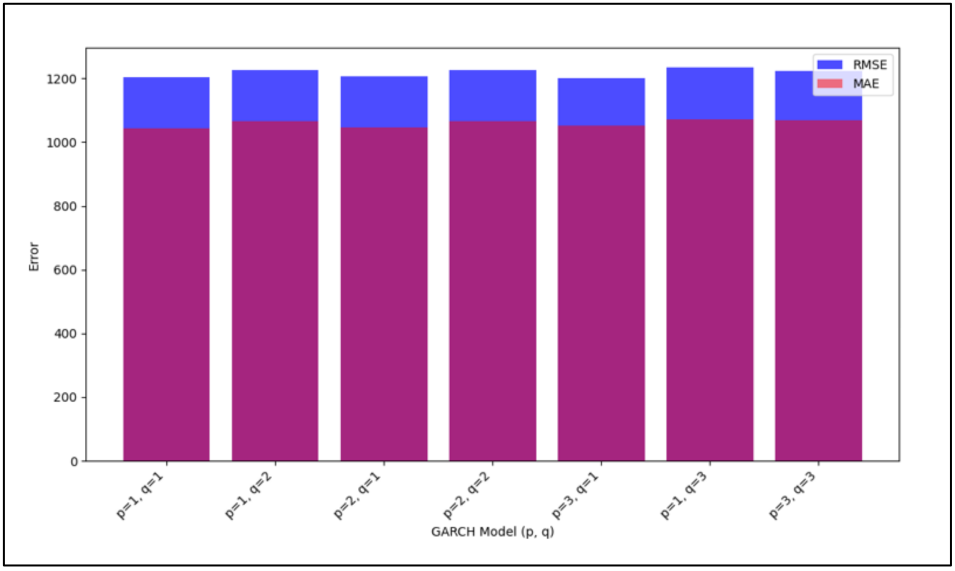
In addition, the mean absolute error criteria and the root mean square error (RMSE) criteria penalise larger deviations or errors. In terms of the root mean square error criteria, the GARCH (3,1) model performs better with the lowest RMSE of 1200.86, suggesting a slightly better performance for larger deviations compared to the GARCH (1,1) performance is marginal, suggesting that adding complexity to the model may not be warranted given the increased computational complexity of the model.

Table 2: RMSE and MAE Criteria

GARCH (p, q)	RMSE	(MAE)
GARCH (1, 1)	1203.8960	1044.6548
GARCH (3, 1)	1200.8647	1052.2963
GARCH (2, 1)	1205.8699	1046.8183
GARCH (3, 3)	1222.6808	1068.0970
GARCH (2, 2)	1227.0326	1065.6065
GARCH (1, 2)	1227.0345	1068.0970
GARCH (1,3)	1233.8812	1071.2249

Source: Analysed Research Data.

Figure 12: RMSE and MAE criteria



Source: Analysed Research Data.

4.4.3 AIC and BIC Criteria

Finally, the performance of models using Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC) is evaluated. AIC is a model-fitting technique for evaluating model overfitting based on in-sample fit and estimating the likelihood of a model predicting or estimating future values. On the contrary, BIC measures the trade-off between the model fit and complexity of the model. Using the two criteria, GARCH (1,1) is the best-performing model with an AIC of 2290.52 and BIC of 2302.38. The results suggest that increasing the parameters of the GARCH models gives a higher value of AIC and BIC, suggesting possible overfitting without substantial improvement in the quality of the model fit.

Table 3: AIC and BIC Criteria

GARCH (p, q)	AIC	BIC
GARCH (1, 1)	2290.5238	2302.3751
GARCH (3, 1)	2290.6007	2308.3778
GARCH (1, 2)	2291.3719	2306.1861
GARCH (2, 1)	2292.4992	2307.3134
GARCH (1, 3)	2293.0327	2310.8098
GARCH (2, 2)	2293.371	2311.1490
GARCH (3, 3)	2293.9650	2317.6678

Source: Analysed Research Data.

Based on the four evaluation criteria, the best model is the GARCH (1,1) model. The GARCH (1,1) model is the optimal model because it offers lower errors (MAE and RMSE) while maintaining parsimony and handling overfitting. Therefore, GARCH (1,1) is optimal for capturing the volatility structure in the data.

4.5 Model Fitting and Performance: GARCH (1,1) Model

Tables 4, 5, 6 and 7 contain the summary results of the GARCH (1,1) model. Table 4 specifies the results of a constant mean GARCH model when the assumption is that there is no volatility in the mean of the dependent variable, which is the average of the first difference in the fraud volumes. Table 6 specifies the model parameters for the GARCH (1,1) model.

Overall, the model parameters from Table 6 are:

α = 5.0492e+04
0
α = 0.0340
1
β = 0.8701

From the parameters, the resulting GARCH (1,1) is as follows:

$$\sigma_t^2 = (5.0492e^{04}) + 0.0340\varepsilon_{t-1}^{2} + 0.8701_1\sigma_{t-1}^2$$

Table 4: Constant Mean-GARCH model Summary

	Value
Dep. Variable	Fraud Volumes (In Millions)
Mean Model	Constant Mean
Vol Model	GARCH
Distribution	Normal
Method	Maximum Likelihood
R-squared	0.000

Source: Analysed data.

Table 5: Constant Mean-GARCH Model Results

Adj. R-squared	0.000
Log-Likelihood	-1141.26
AIC	2290.52
BIC	2302.38
No. observations	143
Df Residuals	142
Df Model	1

Source: Analysed Data.

Table 6: Mean Model Results

	Coef	**Std err**	**T**	**p>ǀtǀ**	**95% comf.int**
Mu	22.3699	57.098	0.392	0.695	[-89.540, 1.343e=02]

Table 7: Volatility Model Results

	Coef	**Std err**	**T**	**p>|t|**	**95% comf.int**
Omega	5.0492e+04	7.042e+04	0.717	0.473	[-8.752e+04, 1.885e+05]
alpha[1]	0.0340	5.524e-02	0.616	0.538	[-7.422e-02, 0.142]
beta[2]	0.8701	0.113	7.715	1.207e-14	[0.649, 1.091]

Source: Analysed Data.

4.5.1 Summary of Key Findings

Deviations in monthly fraud volumes follow similar probabilistic properties as return on investments. Therefore, it assumes popular econometric models (GARCH) for modelling returns to find a suitable model for detecting fraud. These properties include stationarity of the first difference values and autocorrelation of the error terms. Using the grid-search cross-validation technique together with the four-loss function (MAE, RMSE, AIC, and BIC), it is established that the deviations in the volume of fraudulent transactions follow a GARCH (1,1) model.

4.6 Discussions

The primary objective of this study was to establish a suitable model for forecasting changes (variations) in the volume of money lost from fraudulent transactions in Kenya. Although significant studies have been done on the topic, most of the studies focused on using machine learning to find suitable models for fraud detection and monitoring. This study is peculiar in that instead of using machine learning to develop fraud detection and monitoring models, the study employed a different approach, using machine learning to forecast the deviations in the volume of money lost from fraudulent activities. The other unique approach is that it is assumed that changes and variations in the volume of fraudulent transactions follow similar probabilistic properties as the returns on investments. Therefore, the study attempts to model the variations in fraud volume using similar econometric models and approaches used in finance for modelling returns. With the focus of finding a suitable autoregressive model, the study narrowed down to finding an optimal GARCH model for forecasting variations in fraud volumes. The study employed grid search cross-validation parameter optimisation techniques and popular loss functions, including MAE, RMSE, AIC, and BIC, to find the best-fitting model.

5. Recommendations

The GARCH (1,1) model implies that the expected change in the volume of money lost from fraudulent transactions is short-term and is significant up to the previous month. For example, a spike in the volume of fraud today would highly likely suggest that the next month might be the same. Therefore, if short-term fraud detection policies are in place, the anti-fraud handlers can take necessary actions to prepare for the coming month or keenly monitor the trend in the short term. Anti-fraud handlers can also use such results to audit the system and find the possible causes of spikes (high fraud volume). In addition, the existing anti-fraud systems may be improved by providing an alternative to fraud handling. Policymakers can use this model or a similar approach to implement machine learning techniques for real-term monitoring of the deviations in the volume of fraudulent transactions. Besides establishing long-term fraud prevention policies, policymakers should also consider short-term policies since significant changes can take place in short periods, like one month, enhancing real-time fraud detection and prevention mechanisms to cover the expected short-term changes in the amount of money lost from fraudulent transactions.

6. Conclusions

The study used only the GARCH model approach; improvements of future research would be to use multiple econometric models and make a comparison of the performance of the models. For the grid-search cross-validation for parameter optimisation with pre-set hyperparameters (p and q), an improvement for future research would be to use a random grid search to find the optimal values of p and q. Therefore, based on the specific objective of this study, it is concluded that GARCH (1,1) can be used to predict electronic transaction volumes in Kenya.

Conflict of Interest Statement

The authors declare no conflicts of interest.

About the Authors

Dr. Ondiwa Simon is an MSME consultant, and he is currently working with DT Global International Development EA Limited. Previously, he worked as an adjunct lecturer in the School of Business and Economics at Maseno University in Kenya. He has also been a collaborator in research by The University of Maryland, USA, and Maseno University, assessing the current status of the cassava value chain and food security in Kenya and the Democratic Republic of Congo. Dr. Ondiwa holds a PhD in Business Administration, a Finance Major, an MBA, a Finance Major and a Bachelor of Business Administration Finance. His research interest is the operations of financial institutions and markets.

Mr. Micah Oketch Ondiwa holds a Bachelor of Science in Actuarial Science from Jomo Kenyatta University of Agriculture and Technology and a Master of Science in Financial Engineering from WorldQuant University, New Orleans, Louisiana, United States of America. Mr. Micah is a professional data analyst with top-notch knowledge of deep learning and machine learning coupled with software engineering. Currently, Mr. Micah works with The Kenya National Highway Authority. With over seven years of experience, Mr. Micah's research interest is in software engineering and machine learning.

Mr. Domnic Odoyo holds a bachelor's degree in business administration with IT, specialising in accounting. Furthermore, he is a professional accountant in Kenya. He is a registered member of the Institute of Certified Public Accountants of Kenya (ICPAK). Dominic is also an adjunct lecturer at KCA University in Kenya. Domnic's interest is auditing and taxation. He has cumulatively over eight years of industry and academic experience.

Mr. Martin Sure holds a Bachelor of Science in Land Survey from The Technical University of Kenya. Mr. Martin is, therefore, a land surveyor and software developer. Besides, Mr. Martin's interest is in agriculture. His work in agriculture was recognised and rewarded by anzisha price, the African Leadership Academy, YALI-East Africa and the Global Student Entrepreneurs Award.

References

Abdel-Rahman, L. A., & Andriansyah, Y. (2023). The Role of Artificial Intelligence in Modern Banking: An Exploration of AI-Driven Approaches for Enhanced Fraud Prevention, Risk Management, and Regulatory Compliance. *RCBA 6(1)* https://orcid.org/0000-0003-3394-5222, 110-132.

Adedoyin, A. (2018). Predicting Fraud in Mobile Money Transfer. *Unpublished PhD Thesis at The University of Brighton*.

Ahmad, I., Iqbal, S., Jamil, S., & Kamran, M. (2021). A Systematic Literature Review of E-Banking Frauds: Scenario and Security Techniques. *Lingustica Antyerpiesia* 2, 3509-3517.

Amiri, M., & Hekmat, S. (2021). Banking Fraud: A Consumer-side Overview of Categories and Frameworks of Detection and Prevention. *Journal of Applied Intelligent Systems and Information Sciences 2(2)* www.journal.research.fanap.com, 58-68.

Arkhipov, A., Arkhipova, N., & Karminsky, A. (2021). Regulation of Financial Risks in Emerging Markets: Past, Present, and Future. In Risk Assessment and Financial Regulation in Emerging Market's Banking. *Trends and Prospects* https://link.springer.com/chapter/10.1007/978-3-030-69748-8_2.

Bache, I. W. (2024). *Financial Infrastructure Report.* Oslo: Norges Bank, Norway https://norges-bank.brage.unit.no/norges-bank xmlui/bitstream/handle/11250/3132845/fi_2024_eng_web_1245.pdf?sequence=1.

Bankole, O. A. (2024). Implementation of Posti's Mail and Parcel Logistics System for the Modernization and Efficiency Enhancement of Nigerian Postal Services. *Bachelor's Thesis at Jamk University of Applied Sciences,* https://www.theseus.fi/bitstream/handle/10024/863436/Bankole_Omotoyosi.pdf?sequence=2&isAllowed=y.

Burgalassi, D., & Matsumoto, T. (2024). Demographic change in cities: Trends, challenges and insights from G7 economies. *OECDilibrary* https://www.oecd-ilibrary.org/content/paper/f2aec988-en.

CBK. (2024). *The Central Bank of Kenya Official Website.* Nairobi: Central Bank of Kenya https://www.centralbank.go.ke/national-payments-system/payment-cards/value-of-transactions-ksh-millions/.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education.* London: Routledge Falmer.

Cressey, D. R. (1950). The Criminal Violation of Financial Trust. *American Sociological* Review https://doi.org/doi:10.2307/2086606, 738-743.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches (3rd ed).* Thousand Oaks: CA: Sage.

Crotty, M. (1998). *The foundation of social research: Meaning and perspective in the research process.* Thousand Oaks: CA: Sage.

Gabudeanu, L., Brici, I., Mare, C., Mihai, I. C., & Scheau, M. C. (2021). Privacy Intrusiveness in Financial- Banking Fraud Detection. *Risks 9: 104.*

Ge, X. (2024). Design of an Enterprise Financial Information Management System Based on Virtual Reality for Hypertext System Interactive Devices. *Computer-Aided Design & Applications 21(S17)* https://www.cad-journal.net/files/vol_21/CAD_21(S17)_2024_270-288.pdf, 270-288.

Hetzel, M. (2021). How Technological Frames Transform: The case of the global microgrid industry. *Doctoral Dissertation* https://iris.luiss.it/bitstream/11385/183510/5/Finance%20for%20Growth.pdf.

Jawad, M., & Naz, M. (2023). Financial Technological Innovation, Sustainable Operations, and Efficiency: A study of SMBs in Times of Crisis. *Journal of the Knowledge Economy* https://link.springer.com/article/10.1007/s13132-023-01574-5, 1-22.

KNBS. (2021). *FinAccess Household Survey.* Nairobi: Kenya National Bureau of Statistics.

Lokanan, M. E. (2023). Predicting mobile money transaction fraud using machine learning algorithms. *Wiley.*

Marakalala, M. C. (2023). Forensic Intelligence: The Effectiveness of the Biometric-based Solution to Combat Mobile Fraud. *International Journal of Sustainable Development* http://www.ssrn.com/link/OIDA-Intl-Journal-Sustainable-Dev.html, 20-30.

Mertens, D. M. (2010). *Research and evaluation in education and Psychology: Integrating diversity with quantitative, qualitative, and mixed methods (3rd ed).* Thousand Oaks: CA: Sage.

Mustafa, J. A. (2024). Integrating Financial Literacy, Regulatory Technology, and Decentralized Finance: A New Paradigm in Fintech Evolution. *Investment Management and Financial Innovations, Business Perspectives 21(2)* https://www.businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/20066/IMFI_2024_02_Mustafa.pdf, 213-226.

Neuman, W. L. (2009). *Social Research Methods: Qualitative and Quantitative approaches (7th ed).* Pearson Education Inc: Boston.

Ogay, M. (2020). Capital markets development in the emerging market. https://tesi.univpm.it/bitstream/20.500.12075/1386/2/Ogay%20Marina_Capital%20markets%20development%20in%20the%20emerging%20markets_17.06.2021.pdf

Okoye, N. J., Onuoha, O. C., & Akuesodo, O. E. (2021). E-Commerce and Mobile Banking Fraud: Examining the Nexus on the Performance of Deposit Money Banks in Nigeria. *IOSR Journal of Economics and Finance 12(3)*, 42-60.

Phiri, J., Lavhengwa, T., & Segooa, M. A. (2024). Online banking fraud detection: A comparative study of cases from South Africa and Spain. *South Africa Journal of Information Management*, 1-8.

Regnard-Weinrabe, B., Kikki, J., & Finlayson-Brown, J. (2024). Adapting to a Challenging Payments landscape. *Elgaronline* https://doi.org/10.4337/9781035314751.00012, 24-74.

Sanchez-Castany, R. (2024). Industry Insights About Translation Technologies: Current Needs and Future Trends. In New Translation Technology: Applications and Pedagogy. *Springer Nature* https://link.springer.com/chapter/10.1007/978-981-97-2958-6_6, 99-119.

Shrivastava, S., & Johari, P. K. (2022). Convolutional Neural Network Approach for Mobile Banking Fraudulent Transaction to Detect Financial Frauds. *International Journal of Engineering Technology and Management Sciences*, 30-37.

Wolfe, D. T., & Hermanson, D. R. (2004). The Fraud Diamond: Considering the Four Elements of Fraud. *The CPA Journal 74(12)*, 38-42.

Zekos, G. I. (2021). E-Globalization and Digital Economy. Economics and Law of Artificial Intelligence: Finance, Economic Impacts, Risk Management and Governance. *Springer* https://link.springer.com/chapter/10.1007/978-3-030-64254-9_2, 13-66.

Appendix 1: Link for Data Analysis and Codes

https://github.com/micahondiwa/ml_dl-in-finance/tree/main/0x00-fraud_volumes_detection_with_garch