Journal of Statistical and Econometric Methods, Vol.9, No.1, 2020, 1-26
ISSN: 1792-6602 (print), 1792-6939 (online)
Scientific Press International Limited, 2020

News Classification using Support Vector Machine to Model and Forecast Volatility

Alpha Basweti Kenyatta[1], Antony Ngunyi[2] and Anthony Gichuhi Waititu[3]

Abstract

Volatility modelling and forecasting in the financial market is significant in risk management, monetary policy making, security valuation and portfolio creation. Standard volatility models use historical asset price returns to model and predict volatility. The purpose of this study is to add an exogenous variable to the standard volatility model. The exogenous variables used in this research are the news sentiments from Safaricom news articles extracted from Business daily, a Kenyan news publisher that consistently publishes business news. These news sentiments are the counts of positive and negative articles. Safaricom was chosen due to its huge market capitalization compared to other stocks in Kenya and it also has enough news data points for analysis. The Safaricom news articles were classified into either positive or negative using Support Vector Machine. The volatility model that incorporates news sentiments was formulated and its modelling and forecasting capabilities was compared to some standard volatility models. The empirical results indicate that the news sentiments augmented GARCH model

[1] Department of Mathematics, Pan African University Institute for Basic Sciences, Technology and Innovation, Kenya. E-mail: baswetialpha@gmail.com
[2] Department of Statistics and Actuarial Sciences, Dedan Kimathi University of Technology, Kenya. E-mail: ngunyi.antony@dkut.ac.ke
[3] Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology. E-mail: awaititu@jkuat.ac.ke

Article Info: *Received* : August 7, 2019. *Revised* : September 1, 2019.
Published online : January 5, 2020

performed best in forecasting volatility compared to standard GARCH and E-GARCH models.

Keywords: Support Vector Machine; News Sentiments Augmented GARCH; Modelling, Forecasting

1 Introduction

In the financial market improved predictions can lead to tremendous gains by portfolio managers making better investment decisions. The portfolio managers analyse the financial market behaviour in order to make their buy or sell strategies, here accuracy of the forecasted results impacts the profitability outcome. The financial market behaviour is affected by many factors including historical prices and news in the economy. Standard financial models use historical prices and due to this limitation, the incorporation of behavioural economics in the financial markets is gaining traction [1]. Behavioural finance states that strategies built in the financial market are impacted by social and emotional factors. These social and emotional factors are as a result of news on stocks or news on the general economy. Thus, new information in the financial market shapes investor decisions which in turn impacts the state of the financial market. Hence, volatility models should be modified to capture news.

The financial market behaviour was first explained by the EMH theory which was developed by [2]. The EMH theory states that market prices factor in all available and relevant information in the financial market thus, the market cannot be beaten consistently. EMH is divided into weak EMH which suggests that all past information is included in security prices, semi-strong EMH which suggests that any new public information is instantly factored in security prices and strong EMH which suggests that all information, public and private is reflected in securities prices and no investor has a competitive advantage of the market as a whole. According to EMH theory, securities prices are random walk and thus past information cannot be used to accurately predicted future price movements. EMH theory implies that news variables will not impact financial market models significantly. However, several researchers have disputed the EMH theory empirically and theoretically. Socionomic theories as described by [3] and behavioural economics theory are some of the theories that criticize the EMH theory and suggest that predictions in the financial markets can be made successfully. These theories indicate that investors do not make decisions instantly and are not always rational. Furthermore, the strategies in the financial markets are not merely determined by facts which include the technical and fundamental analysis indicators but are also impacted by moods and emotions which are gotten from news. The technical and fundamental factors are "facts" in the financial market while news acts as an exogenous variable. With a lot of information which has been brought about by the increased adoption of the internet and electronic media, investors are not immediately aware of all these news and thus they do not act accordingly instantly. This information is slowly incorporated to security prices allowing for partial asset price prediction. Partial prediction of security prices has been supported by [4]'s empirical study, which provided evidence that some variables effectively predict asset prices when incorporated to econometric models. These findings dispute the EMH theory.

Over the last few decades, researchers have modelled and analysed financial time series conditional variance temporal behaviour. Most academic research has focused in modelling and forecasting stock return volatility. Thus, models that estimate and forecast conditional volatility of financial time series data have been proposed, with conditional heteroscedastic being the most popular. The well known conditional heteroscedastic models are standard GARCH by [5], E-GARCH and GED-GARCH by [6], GJRGARCH by [7], TGARCH by Zakoian 1994, GRS-GARCH by [8] and SEMIFAR-GARCH by [9] amongst others.

The impact of news in the financial markets has also been studied over the years to determine the relationship between published news and financial model behaviours. For instance, [10] indicated that there was a weak relationship between financial market movements and news and that news announcement patterns do not explain market seasonalities in the day of the week in their study. [11] indicated that conditional stock volatility is impacted differently depending on the type of news. This argument was supported by [12] whose study showed that bad news increased volatility more compared to good news. [13] indicated that there is a strong autocorrelation pattern in news arrival. [14] determined news intensity (stock press releases) as the most effective explanatory variable that can improve GARCH modelling. [15] studied the relationship between securities market and sentiments from investors. [16] confirmed that stronger market reactions are observed when the news in the market is negative compared to when the news is positive.

[17]'s study indicated that GARCH model provides good forecasts in volatility and simple parametrization, but the model does not fully capture the asymmetric effect between asset returns and volatility such as leverage effect. Furthermore, GARCH models are good in estimating model parameters as they give the best in-sample fit for the data, but they often give poor out-of-sample volatility forecasts. Thus, it would be significant to investigate and use exogenous data beside historical asset prices to improve the predictive power of standard GARCH model. [18], [19], [20] incorporated trading volume as an exogenous variable into their volatility models. [21] used interest rate levels as an exogenous variable in volatility modelling and forecasting. These studies showed improvement of the forecasting ability of volatility models when relevant exogenous variables were incorporated to the model structure.

The financial market is affected by new information from different news sources like social media, newscasts, articles and announcements. News can either be expected or unexpected. Expected news involves government announcements while unexpected news is mainly from social media platforms. Both news have been evidenced to significantly impact volatility. [22], [23] used scheduled and unscheduled news as exogenous variables in volatility models. The evidence of the relationship between news and asset price fluctuations has encouraged researchers investigate how news impacts trading strategies made by investors. [24] studied how the stock and bond markets are affected by macro-economic news. [25] showed that news significantly impacted return asset price volatility and its incorporation to the GARCH model improved forecasting. [26] determined the relationship between company specific news and stock price movements. Recently, [27], [28] modified the GARCH model by adding news impact scores as exogenous variables in the model equation and split the news impact scores as positive and negative. [29] indicated that financial market decisions are impacted by news sentiments, thus it would be relevant to incorporate news as an exogenous variable in volatility models to improve the models forecasting capabilities. To apply sentiment data from news to volatility models efficiently and effectively, researchers need to extract relevant news items and correctly classify them. The news sentiments should then be aggregated into daily numbers in order to link the sentiments to daily asset prices, this helps reduce noise in the experimental data.

The use of news sentiments in volatility models has been studied in the developed markets, frontier markets like Kenya have not studied the use of news sentiments in volatility models, yet there are scenarios where news impacted the Kenyan financial market significantly. Examples of scenarios where news significantly affected the Kenyan stock market include the news on the failure by Members of Parliament in Kenya to review the law on interest rate cap in September 2018, this led to a five day decline in the stock prices of majority of the listed banks in Kenya. Additionally, the Safaricom share price declined to 10-month low when taxes were increased on mobile money transfers, internet and telephone services. Thus it would be relevant to study the effect of news sentiments in frontier markets like Kenya. The study develops the news sentiments augmented GARCH and compares its modelling and forecasting capabilities with standard GARCH and E-GARCH.

2 Data

The data for Safaricom, the largest listed company on the NSE was used. The data used in the analysis includes the times series data for Safaricom stocks' closing prices and news sentiments for a five-year period (i.e. 1st January 2014 to 31st December 2018). Safaricom was the preferred company for this study because it has a large market capitalization compared to other stocks at NSE. Additionally, the stock guarantees enough news data points because of its huge impact in almost all sectors of the Kenyan economy. The daily returns defined by r_t were computed as continuously compounded returns;

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

P_t is the closing stock price at time t and P_{t-1} is the closing stock price at time $t-1$. News articles from 2014 to 2018 on Safaricom were extracted from

Business Daily, a news publisher that consistently publishes business news in Kenya.

3 Methodology

3.1 News article classification using SVM

Let the training data points be defined by $x_i \in \mathbb{R}^d$ for $i = 1, .., N$. The training set has two classes, $y_i = 1$ characterizing class 1 (positive) and $y_i = -1$ characterizing class 2 (negative). The two classes are separated by a hyperplane that is defined by;

$$f(x) = x_i \cdot w + b,$$

where b is a bias term and w is an N-dimensional vector. The purpose for SVM is to orient the separating hyperplane such that it is as far as possible from the closest points of both classes. For data that is not fully linearly separable, the training set is defined by;

$$x_i \cdot w + b \geq 1 - \epsilon_i \ for \ y_i = 1 \quad (3.1)$$

$$x_i \cdot w + b \leq -1 + \epsilon_i \ for \ y_i = -1 \quad (3.2)$$

$$\epsilon_i \geq 0 \ for \ i = 1, .., N$$

ϵ_i is a positive slack variable that allows the consideration of misclassified data points. Combining equations (3.1) and (3.2) the result (3.3) is obtained;

$$y_i(x_i \cdot w + b) - 1 + \epsilon_i \geq 0 \quad (3.3)$$

The points that are close to the separating hyperplane are defined by;

$$x_i \cdot w + b = 1 - \epsilon_i \ for \ H_1$$

$$x_i \cdot w + b = -1 + \epsilon_i \ for \ H_2$$

H_1 and H_2 are the hyperplanes where the closest positive and negative data points lie respectively. These two separating hyperplanes are the confidence intervals of the main separating hyperplane. Let d_1 be the distance from the main separating hyperplane to H_1 and d_2 be the distance from the main separating hyperplane to H_2. $d_1 = d_2$ and this distance is called the SVM margin.

The SVM margin needs to be maximized for the separating hyperplane to be as far as possible from the points closest to the separating hyperplane. The SVM margin is given by $\frac{1}{\|w\|}$. Misclassified data points have a retribution that increases with the distance from the margin boundary. Thus to decrease misclassified points, the SVM margin $\frac{1}{\|w\|}$, needs to be maximized. Maximizing the SVM margin given the constraints in (3.3) is equivalent to;

$$\begin{aligned} & min\frac{1}{2} \parallel w \parallel^2 + C\sum_{i=1}^{N} \epsilon_i \\ & s.t.\ y_i(x_i \cdot w + b) - 1 + \epsilon_i \geq 0 \end{aligned} \tag{3.4}$$

C controls the compromise between the SVM margin and ξ_i's, the slack variable, retribution.

To cater for the constraint in (3.4), allocate Lagrange multipliers β and ζ, where $\beta_i, \zeta_i \geq 0$;

$$L = \frac{1}{2} \parallel w \parallel^2 + C\sum_{i=1}^{N} \epsilon_i - \sum_{i=1}^{N} \beta_i [y_i(x_i \cdot w + b) - 1 + \epsilon_i] - \sum_{i=1}^{N} \zeta_i \epsilon_i \tag{3.5}$$

Therefore, w, b and ξ that minimizes (3.5) need to be evaluated, and Lagrange multipliers β and ζ that maximize (3.5). This is done by differentiating equation (3.5) w.r.t w, b, ξ_i, and equating the results to 0;

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \beta_i y_i x_i = 0$$

$$\implies w = \sum_{i=1}^{N} \beta_i y_i x_i \tag{3.6}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} \beta_i y_i = 0 \tag{3.7}$$

$$\frac{\partial L}{\partial \xi_i} = C - \beta_i - \zeta_i = 0$$

$$\implies C = \beta_i + \zeta_i \tag{3.8}$$

Replacing results from (3.6), (3.7) and (3.8) to (3.5), equation (3.9) is obtained. Equation (3.9) depends on β and thus needs to be maximized;

$$\mathcal{L} = -\frac{1}{2}\sum_{i,j}\beta_i\beta_j y_i y_j x_i \cdot x_j - \sum_{i=1}^{N}\xi_i(\beta_i + \zeta_i) + \sum_{i=1}^{N}\beta_i(-1+\xi_i) + \sum_{i=1}^{N}\zeta_i\xi_i \quad (3.9)$$

$$= -\frac{1}{2}\sum_{i,j}\beta_i\beta_j y_i y_j x_i \cdot x_j + \sum_{i=1}^{N}\beta_i \quad (3.10)$$

Equation (3.9) is the dual form of the primary (3.5). Next;

$$maximize \ -\frac{1}{2}\sum_{i,j}\beta_i\beta_j y_i y_j x_i \cdot x_j + \sum_{i=1}^{N}\beta_i \quad (3.11)$$

$$s.t \sum_{i=1}^{N}\beta_i y_i = 0 \ and \ 0 \leq \beta_i \leq C$$

Equation (3.11) is a convex quadratic optimization problem that gives β and from (3.6), w is obtained. Replacing w to (3.3), equation (3.12) is obtained;

$$y_k\left(\sum_{s\in K}\beta_s y_s x_s \cdot x_k + b\right) = 1 - \xi_s \quad (3.12)$$

where K is the set of support vectors indices and is evaluating by finding the indices i for $0 \leq \beta_i \leq C$. Multiplying both sides of (3.12) by y_k and keeping in mind $y_k^2 = 1$;

$$\sum_{s\in K}\beta_s y_s x_s \cdot x_k + b = y_k(1-\xi_s)$$

$$\Longrightarrow b = y_k(1-\xi_s) - \sum_{s\in K}\beta_s y_s x_s \cdot x_k$$

Hence, b and w that characterize the optimal separating hyperplane have been formulated and thus, the Support Vector Machine. The unclassified data set is classified into;

$$\begin{cases} positive \ if \ f(x) > 0, \\ negative \ if \ f(x) < 0. \end{cases}$$

If $f(x) = 0$, x is on the separating hyperplane and is thus not classifiable.

3.2 News sentiments augmented GARCH

Consider a volatility model given by;

$$a_t = \sigma_t \epsilon_t$$

and let $(\mathcal{F}_{t-1})$ be the filtration representing the information set at time $t-1$.

$$Var(a_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$$

Assume ϵ_t is iid (0,1) and it is adapted to information set available at time $t-1$, and σ_t is given by;

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + f(s_t, \omega)$$

for parameters $\alpha_0 > 0$ $\alpha_i, \beta_j \geq 0$ and $\overset{[}{i} = 1]max(p,q)\sum(\alpha_i + \beta_i) < 1$. s_t is an exogenous time series, for which in this case it is the **news sentiments**, $f(s_t, \omega)$ is strictly positive. Thus the news sentiments augmented GARCH (p,q) model is defined as;

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \gamma_1 ln \left(1 + \frac{P_{t-1}}{T_{t-1}}\right) + \gamma_2 ln \left(1 + \frac{N_{t-1}}{T_{t-1}}\right) \quad (3.13)$$

where P_{t-1}, N_{t-1} and T_{t-1} are the number of positive, negative and total news articles at time $t-1$. The positive and negative news articles define the news sentiments. Time $t-1$ is used in the news sentiments to make predictions possible.

Equation (3.13) can be presented as an ARMA (p,q) model by defining $\eta_t = a_t^2 - \sigma_t^2$ as a martingale difference sequence with $E(\eta_t) = 0$ and $cov(\eta_t, \eta_t - j) = 0$

for $j \geq 1$. Moreover, $\sigma_t^2 = a_t^2 - \eta_t$, replacing this in (3.13);

$$\begin{aligned}
a_t^2 - \eta_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \left(a_{t-j}^2 - \eta_{t-j}\right) \\
&\quad + \gamma_1 ln \left(1 + \frac{P_{t-1}}{T_{t-1}}\right) + \gamma_2 ln \left(1 + \frac{N_{t-1}}{T_{t-1}}\right) \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j a_{t-j}^2 - \sum_{j=1}^{q} \beta_j \eta_{t-j} \\
&\quad + \gamma_1 ln \left(1 + \frac{P_{t-1}}{T_{t-1}}\right) + \gamma_2 ln \left(1 + \frac{N_{t-1}}{T_{t-1}}\right) \\
&= \alpha_0 + \sum_{i=1}^{max(p,q)} \left(\alpha_i + \beta_i\right) a_{t-i}^2 - \sum_{j=1}^{q} \beta_j \eta_{t-j} + \\
&\quad \gamma_1 ln \left(1 + \frac{P_{t-1}}{T_{t-1}}\right) + \gamma_2 ln \left(1 + \frac{N_{t-1}}{T_{t-1}}\right)
\end{aligned}$$

Therefore

$$a_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} \left(\alpha_i + \beta_i\right) a_{t-i}^2 + \eta_t - \sum_{j=1}^{q} \beta_j \eta_{t-j} + \gamma_1 ln \left(1 + \frac{P_{t-1}}{T_{t-1}}\right) + \gamma_2 ln \left(1 + \frac{N_{t-1}}{T_{t-1}}\right).$$

3.2.1 Forecasting of News sentiments augmented GARCH

Consider the news sentiments augmented GARCH (1,1) model defined as;

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma_1 ln \left(1 + \frac{P_{t-1}}{T_{t-1}}\right) + \gamma_2 ln \left(1 + \frac{N_{t-1}}{T_{t-1}}\right) \quad (3.14)$$

Assume the forecast origin for (3.14) is k, then the 1-step ahead forecast becomes;

$$\sigma_{k+1}^2 = \alpha_0 + \alpha_1 a_k^2 + \beta_1 \sigma_k^2 + \gamma_1 ln \left(1 + \frac{P_k}{T_k}\right) + \gamma_2 ln \left(1 + \frac{N_k}{T_k}\right) \quad (3.15)$$

At time index k, a_k and σ_k^2 are known. The $1 - step$ ahead forecast can also be written as;

$$\sigma_k^2\left(1\right) = \alpha_0 + \alpha_1 a_k^2 + \beta_1 \sigma_k^2 + \gamma_1 ln \left(1 + \frac{P_k}{T_k}\right) + \gamma_2 ln \left(1 + \frac{N_k}{T_k}\right).$$

3.3 Standard GARCH model

For a log return series r_t, let $a_t = r_t - \mu_t$ be the innovation at time t. Then a_t follows a GARCH (p,q) model if;

$$a_t = \sigma_t \epsilon_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3.16}$$

where $\{\epsilon_t\}$is a sequence of i.i.d random variables with mean 0 and variance 1, $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$, and $\overset{[}{i}= 1]max\,(p,q)\sum(\alpha_i + \beta_i) < 1$. The latter constraint on $(\alpha_i + \beta_i)$ implies that the unconditional variance of a_t is finite, and its conditional variance σ_t^2 evolves over time. ϵ_t is always assumed to be standardized Student-t, standard normal or generalized error distribution. If $q = 0$ (3.16) reduces to an ARCH(p) model.α_i's are the ARCH parameters and β_j's are the GARCH parameters.

To describe GARCH properties, let $\eta_t = a_t^2 - \sigma_t^2 \implies \sigma_t^2 = a_t^2 - \eta_t$. By plugging $\sigma_{t-1}^2 = a_{t-1}^2 - \eta_{t-1}$ into equation (3.16), the GARCH model becomes;

$$a_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)\, a_{t-i}^2 + \eta_t - \sum_{j=1}^{q} \beta_j \eta_{t-j} \tag{3.17}$$

η_t is a martingale difference series $(i.e., E(\eta_t) = 0 \; and \; cov(\eta_t, \eta_t - j) = 0 \; j \geq 1)$. Equation 3.17 is the ARMA form for $\{a_t^2\}$. Thus, a GARCH model is as an application of the ARMA model to $\{a_t^2\}$. The unconditional mean for (3.17) is;

$$E\left(a_t^2\right) = \frac{\alpha_0}{1 - \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)} \tag{3.18}$$

provided the denominator for 3.18 is positive.

3.3.1 Forecasting the standard GARCH Model

Consider the standard GARCH (1,1) model defined as;

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.19}$$

Assume the forecast origin for (3.19) is k, then the 1-step ahead forecast becomes;

$$\sigma_{k+1}^2 = \alpha_0 + \alpha_1 a_k^2 + \beta_1 \sigma_k^2$$

At time index k, a_k and σ_k^2 are known. The $1-step$ ahead forecast can also be written as;

$$\sigma_k^2(1) = \alpha_0 + \alpha_1 a_k^2 + \beta_1 \sigma_k^2.$$

3.4 EGARCH Model

E-GARCH was developed to overcome some weaknesses of GARCH in particular to capture the asymmetric effect of positive and negative returns. The EGARCH (p,q) model can also be represented as;

$$ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{|a_{t-i}| + \gamma_i a_t - i}{\sigma_{t-i}} + \sum_{j=1}^{q} \beta_j ln(\sigma_{t-j}^2)$$

Here, a positive a_{t-i} contributes $\alpha_i(1+\gamma_i)|\epsilon_{t-i}|$ to the log volatility, while a negative a_{t-i} contributes $\alpha_i(1-\gamma_i)|\epsilon_{t-i}|$, where $\epsilon_{t-i} = \frac{a_{t-i}}{\sigma_{t-i}}$. The γ_i parameter signifies the leverage effect of a_{t-i}. If $\gamma_i > 0$ then good news (positive shocks) have a greater effect on volatility compared to bad news (negative shocks).

3.4.1 Forecasting using E-GARCH model

Consider the E-GARCH (1,1) model, assume that the model parameters are known and $\epsilon_t' s$ are standard GAussian, then;

$$\begin{aligned} ln(\sigma_t^2) &= (1-\alpha_1)\alpha_0 + \alpha_1 ln(\sigma_{t-1}^2) + g(\epsilon_{t-1}) \\ g(\epsilon_{t-1}) &= \theta\epsilon_{t-1} + \gamma(|\epsilon_{t-1}| - \sqrt{\frac{2}{\pi}}). \end{aligned} \tag{3.20}$$

Taking exponentials, equation (3.20) becomes;

$$\begin{aligned} \sigma_t^2 &= \sigma_{t-1}^{2\alpha_1} exp[(1-\alpha_1)\alpha_0] exp[g(\epsilon_{t-1})] \\ g(\epsilon_{t-1}) &= \theta\epsilon_{t-1} + \gamma(|\epsilon_{t-1}| - \sqrt{\frac{2}{\pi}}). \end{aligned}$$

Let k be the forecast origin. The 1-step ahead forecast becomes;

$$\sigma_{k+1}^2 = \sigma_k^{2\alpha_1} exp[(1-\alpha_1)\alpha_0] exp[g(\epsilon_k)]$$

where all the quantities at time k are known.

3.5 Model parameter estimation

The conditional variance model parameters have to be estimated to enable volatility prediction of time series data. The MLE method is used on the residuals to estimate the conditional volatility model parameters. The vector of model parameters is $\theta = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)^T$ for the GARCH model, $\theta = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \gamma_1, \gamma_2)^T$ for the news sentiments augmented GARCH and $\theta = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \gamma)^T$ for E-GARCH. Assume θ belongs to the set $\Theta = \left\{((\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)^T)\right\}:\ \alpha_0 \geq 0, \alpha_i > 0, \beta_j > 0$ for the GARCH (p,q) model, denote $\theta^* = (\alpha_0^*, \alpha_1^*, ..., \alpha_p^*, \beta_1^*, ..., \beta_q^*)^T$, the vector of true values of parameters. The goal is to find θ^* that maximizes the maximum likelihood function given an observation sequence $a_0, ..., a_n$ of length n. Defining the sequence $(\tilde{\sigma_1}, ..., \tilde{\sigma_n})$ by recursion;

$$\tilde{\sigma}_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_i \tilde{\sigma}_{t-j}^2 \ 1 \leq t \leq n$$

$a_{1-p}, ..., a_0$ and $\tilde{\sigma}_{1-q}, ..., \tilde{\sigma}_0$ are the initial values for $a's$ and $\sigma's$ respectively. Given the initial values and assuming the residuals are normally distributed, the Gaussian quasi-likelihood function can be written as;

$$L(\theta) = L(\theta; a_1, ..., a_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} exp\left(\frac{-a_t^2}{2\tilde{\sigma}_t^2}\right)$$

The log likelihood function is;

$$\mathcal{L}(\theta) = lnL(\theta) = -\frac{n}{2} ln\,(2\pi) - \frac{1}{2}\sum_{t=1}^{n} ln\tilde{\sigma}_t^2 - \frac{1}{2}\sum_{t=1}^{n}\left(\frac{-a_t^2}{2\tilde{\sigma}_t^2}\right) \tag{3.21}$$

Taking the first derivative of (3.21) with respect to θ, the system of equations cannot be solved analytically. Since (3.21) does not have an analytical solution, numerical methods are used.

3.6 Out-of-sample model performance

The forecasting performance of the employed volatility models is obtained from a sequence of rolling regressions. One-step ahead volatility forecasts were used. The mean absolute error (MAE) (3.22) and root mean squared error (RMSE) (3.23) are used to evaluate the forecasts out-of-sample performance.

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |\hat{\sigma_t}^2 - \sigma_t^2| \tag{3.22}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (\hat{\sigma_t}^2 - \sigma_t^2)} \tag{3.23}$$

T is the number of forecasts made, σ_t^2 is the volatility forecast and $\hat{\sigma_t}^2$ is the proxy for actual volatility. But, since conditional volatility is not observable a proxy for actual volatility needs to be defined. The common proxy for actual volatility is taking the daily squared return series [30].

4 Results and Discussions

4.1 Text classification using SVM

300 articles were classified by the researcher as either positive or negative. 67% of the articles classified by the researcher were positive while 33% were negative. The news articles classified by the researcher were then divided into training and test sets to determine how well the training set can correctly classify the text. The training set was 75% of the positive and negative articles independently. Linear kernel on SVM was used on the training set and this was applied to the test set. The SVM algorithm results on the training set are;

Table 1: Linear kernel on SVM

Support Vector Machine	
SVM type	Class classification
SVM kernel	Linear
Cost parameter	0.05
Number of support vectors	143
Positive support vectors	87
Negative support vectors	56

Table 1 shows that out of the 300 articles 143 of them were used as support vectors. In the 143 articles, 56 and 87 are support vectors for the negative and positive classes respectively. The cost parameter reduces misclassification by controlling the compromise between the SVM margin size and misclassified data. Table 2 shows the results of correctly classified Safaricom articles using the SVM classifier that was applied on the training set to check how well the test set is classified.

Table 2: Prediction using SVM

Confusion Matrix Statistics	
Accuracy	0.8899
95% confidence interval	(0.8417, 0.9274)
Positive	0.9205
Negative	0.8289

Table 2 indicates that 89% of the articles are correctly classified, 92% of them are correctly classified as positive and 82% of them are correctly classified as negative. Now that the SVM classifier gives an accuracy of 89%, it was applied to the unclassified Safaricom news articles. 79% of the unclassified Safaricom articles were categorized as positive while the other 21% were categorized as negative.

4.2 Daily closing prices and returns for Safaricom from 204 to 2018

4.2.1 Descriptive statistics

Table 3 shows the summary statistics for the closing prices and returns.The results indicate that Safaricom closing prices do not have zero mean and the variance is very high, but the daily returns have zero mean and zero variance. The Kurtosis of the returns is > 3, this implies that the data has heavier tails than that of a normally distributed data.

Table 3: Descriptive statistics on Safaricom prices and returns

	Closing prices	Returns
Number of observations	1220	1219
Minimum	10.75	-0.08
Maximum	23.75	0.06
Mean	19.07	0.00
Median	17.30	0.00
Sum	23268.40	0.80
Standard error mean	0.16	0.00
Variance	31.90	0.00
Standard deviation	5.65	0.00
Skewness	0.61	-0.14
Kurtosis	-0.85	3.10

4.2.2 Time series plots

Figure 1 shows the plots on closing prices and returns over time. The time series plot for the closing prices on Safaricom clearly shows that the variance is not constant as it evolves steadily over time, thus it is volatile.

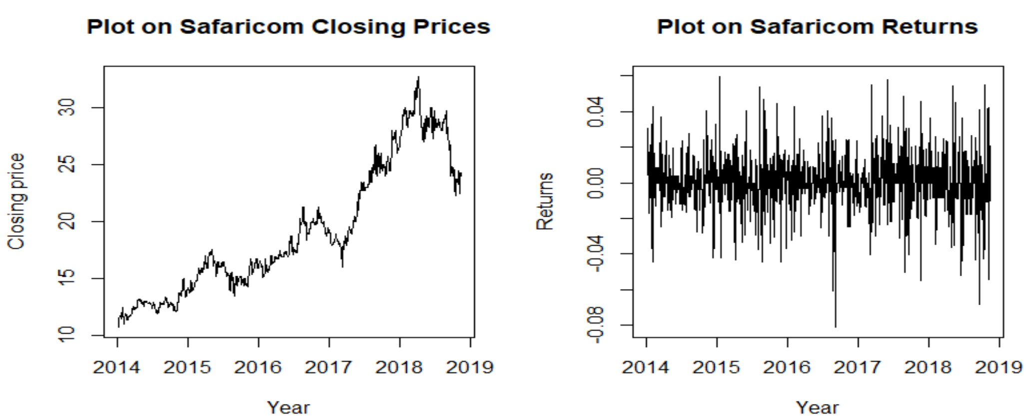

Figure 1: Plots on daily closing prices and returns on Safaricom from 2014-2018

The time series plot on the returns for Safaricom daily is close to zero, this implies they are stationary as indicated by the test on stationarity in (4).

Volatility clustering is also observed, as high spikes are followed by high spikes and low spikes are followed by low spikes.

Table 4: Augmented Dickey-Fuller Stationarity Test

Dickey-Fuller	P-value
-11.83	0.01

Table 4 shows that Safaricom returns are stationary since the $p - value < \alpha = 0.05$.

Table 5: Shapiro-Wilk Normality Test

W	P-value
0.94567	$<$0.01

Table 5 shows that the Safaricom returns do not follow a normal distribution, this is because the $p - value < \alpha = 0.05$.

4.2.3 ARMA modelling

The ACF and PACF plots on returns are plotted to determine the order of the ARMA (p,q) model.

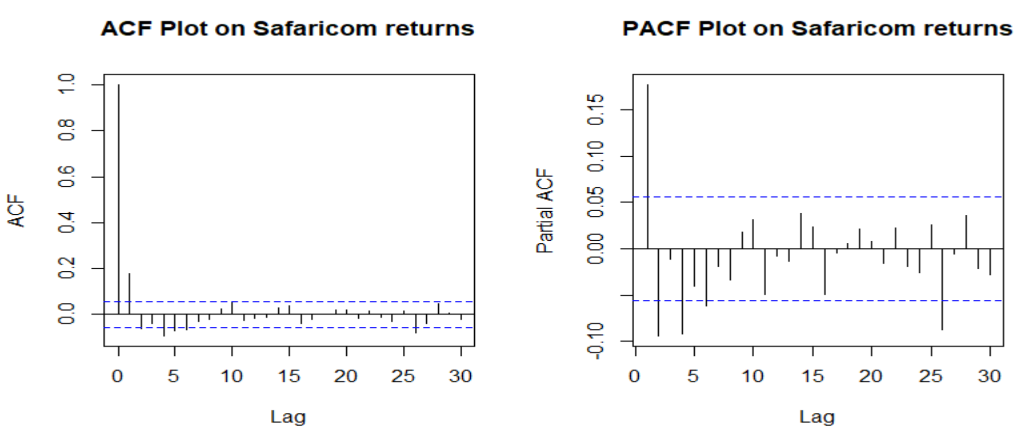

Figure 2: The ACF and PACF plots on returns from 2014-2018

Figure 2 shows that the ACF and PACF plots have significant lags and from this an ARMA (2,2) can be modelled. The fitted model is;

$$r_t = 0.00006 + 0.7832r_{t-1} - 0.0151r_{t-2} + 0.5989a_{t-1} + 0.2122a_{t-2}$$

r_t is the return and a_t is the residual returns.

The ACF and PACF plots on the residuals are plotted to determine if the fitted ARMA (2,2) model is adequate. Figure 3 shows that there are no

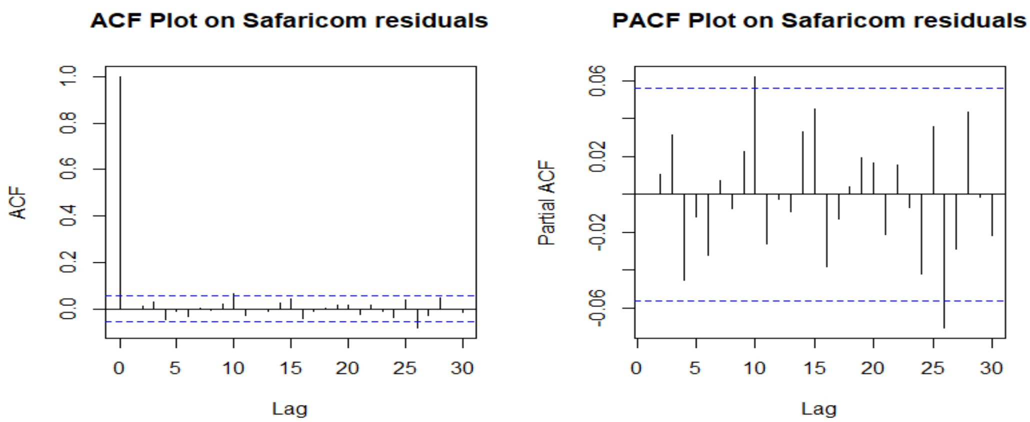

Figure 3: The ACF and PACF plots on residuals from 2014-2018

significant lags in the ACF and PACF plots thus ARMA (2,2) is adequate.

4.2.4 ARCH effects test on residual returns for Safaricom

Table 6: Box-Ljung Test for ARCH effects

χ^2	df	P-value
1.403	1	<0.01

Table 6 indicates that ARCH effects are present in the residuals of the return series since $p - value < \alpha = 0.05$, thus volatility should be modelled.

4.3 News Sentiments Augmented GARCH, Standard GARCH and E-GARCH Models

4.3.1 Identification of the (p,q) models

BIC is used to determine the order of the employed consitional volatility models.

Table 7: BIC of the employed conditional volatility models

Order (p,q)	BIC-News sentiments augmented GARCH	BIC-Standard garch	BIC-EGARCH
(1,1)*	-5.6766	-5.6882	-5.6879
(1,2)	-5.6715	-5.6831	-5.6852
(2,1)	-5.6709	-5.6825	-5.6823
(2,2)	-5.6674	-5.6773	-5.6768

Table 7 shows that models of order (1,1) are the optimal models for in-sample fitting since they have the least BIC values. Also, the standard GARCH (1,1) model provides the best in-sample fit for modelling volatility since it has the least BIC value followed by E-GARCH (1,1) model. The news sentiments augmented GARCH (1,1) performs poorest on in-sample model fitting. This outcome is consistent to literature as studies found the standard GARCH (1,1) to give the best in-sample model fit for volatility [17].

4.3.2 Fitting the conditional volatility models of order (1,1)

Table 8 shows the estimated parameters of the conditional volatility models used in this study using MLE.

Table 8: Estimated parameters for the employed conditional volatility models

Augmented GARCH(1,1)			Standard GARCH(1,1)			E-GARCH(1,1)		
Parameters	Estimate	P-value	Parameters	Estimate	P-value	Parameters	Estimate	P-value
alpha0	0.000022	<0.01	alpha0	0.000019	<0.01	alpha0	-0.854998	<0.01
alpha1	0.123837	<0.01	alpha1	0.123009	0.01	alpha1	-0.023155	0.23
beta1	0.789225	<0.01	beta1	0.793349	<0.01	beta1	0.897327	<0.01
gamma1	-0.000005	<0.01				gamma	0.250673	<0.01
gamma2	0.000000	1.00						

The News sentiments augmented GARCH (1,1) model is;

$$\sigma_t^2 = 0.000022 + 0.123837a_{t-1}^2 + 0.789225\sigma_{t-1}^2 - 0.000005ln\left(1 + \frac{P_{t-1}}{T_{t-1}}\right)$$

The parameters α_0, α_1, β and γ_1 are statistically significant, this is because their $p - values < 0.05$. γ_1 is the coefficient for positive news sentiments and γ_2 is the coefficient for the negative news sentiments. The negative news sentiments are not statistically significant because, the $p - value$ for $\gamma_2 \not< 0.05$. Economically, at time t positive news sentiments from time $t - 1$ on Safaricom reduce its volatility whereas, negative news sentiments at time $t - 1$ do not have an effect on volatility at time t. This is the case since for the studied period from 2014 to 2018, majority of the news articles on Safaricom from Business Daily were positive. This could also be explained by the fact that only news specific to Safaricom was used(company-specific-news), therefore not all news in the Kenyan economy was covered, yet stocks are also affected by general economic news. $\alpha_1 + \beta_1 = 0.913062 < 1$ therefore the unconditional variance for the news sentiments augmented GARCH (1,1) model is stationary.

The Standard GARCH(1,1) model is;

$$\sigma_t^2 = 0.000019 + 0.123009a_{t-1}^2 + 0.793349\sigma_{t-1}^2$$

The parameters α_0, α_1, and β are statistically significant, this is because their $p - values < 0.05$. $\alpha_1 + \beta_1 = 0.916358 < 1$ therefore the unconditional variance for the standard GARCH (1,1) model is stationary.

The fitted model for E-GARCH(1,1) from Table 8 is;

$$ln(\sigma_t^2) = -0.854998 - 0.023155(|\epsilon_{t-1}| + 0.250673\epsilon_{t-1}) + 0.897327ln(\sigma_{t-1}^2)$$

The parameters, α_0, β_1, and γ are statistically significant, this is because their $p - values < 0.05$.

4.3.3 Test for ARCH effects

ARCH effects is tested on the squared residuals to determine if the fitted conditional volatility models fit the data well.

Table 9: Weighted Ljung-Box test for ARCH effects

	News sentiments augmented GARCH(1,1)		Standard GARCH(1,1)		E-GARCH(1,1)	
	Statistic	P-value	Statistic	P-value	Statistic	P-value
Lag[1]	0.01825	0.89	0.0341	0.85	0.2342	0.63
Lag[5]	0.81192	0.90	0.8243	0.90	0.7587	0.91
Lag[9]	3.27949	0.71	3.2921	0.71	3.2507	0.72

The $p - values$ for all the models in Table 9 are greater than zero, this implies that there are no ARCH effects, thus the conditional volatility models employed fit the data well.

4.4 Out-of-sample model comparison

The one-day-ahead forecasts from $15^{th} November, 2018$ to $31^{st} December, 2018$ was constructed. Figure 4 shows the volatility forecasts but performing model cannot be determined from the plot.

Plot on Volatility Forecasts

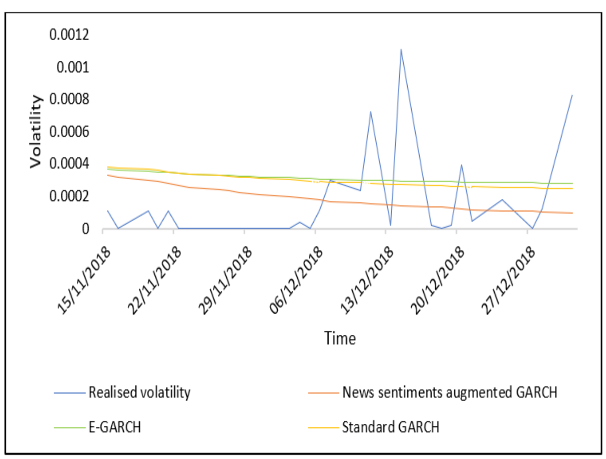

Figure 4: Plot on 30-day-ahead volatility forecasts

The RMSE and MAE are used to determine the best performing conditional volatility model in forecasting volatility.

Table 10: RMSE and MAE for the employed conditional volatility models

	News sentiments augmented GARCH (1,1)*	Standard GARCH (1,1)	E-GARCH (1,1)
RMSE	0.0003026939	0.0003231025	0.0003236592
MAE	0.000229945	0.0002851235	0.0002902799

Table 10 shows the RMSE and MAE results. The News sentiments augmented GARCH model has the least RMSE and MAE of approximately 0.00030 and 0.00023 respectively, thus it outperforms the other conditional volatility models in forecasting volatility. It is also observed that the model that performs best on in-sample modelling does not necessarily perform best in out-of-sample forecasting. This outcome is consistent with literature as studies found the augmented GARCH models gave the best out-of-sample forecast for volatility [17].

5 Conclusion

This study included news sentiments on Safaricom from Business Daily in the conditional variance equation. The positve and negative news sentiments were incorporated into the standard GARCH model separately. This was done to allow the conditional volatility model capture the asymmetric effect of news sentiments. It was observed that positive news sentiments had a significant negative impact on volatility, this is consistent with literature since positive news is expected to reduce volatility. But, negative news sentiments did not have a significant effect on volatility, this could be due to the fact that for the studied period most of the articles were positive, 79% were positive and the other 21% were negative. In-sample and out-of-sample model comparisons were also performed to determine the best performing model. BIC was used to determine the best in-sample model, standard GARCH (1,1) gave the best in-sample fit since it had the least BIC value. Out-of-sample model comparison was performed to determine if the model that performs best on in-sample modelling also performs best in out-of-sample forecasting. The News sentiments augmented GARCH model performed poorly on in-sample modelling but, better in forecasting compared to the other volatility models employed in this study. This study used company specific news, therefore future studies can incorporate general economic news to determine if negative news for the studied period had an effect on Safaricom volatility.

Acknowledgments We would like to thank the African Union, the Pan African University, Institute for Basic Sciences Technology and Innovation together with Japan International Cooperation Agency for supporting this research.

References

[1] Jasmina Smailović, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. Predictive sentiment analysis of tweets: A stock market application. In *Human-computer interaction and knowledge discovery in complex, unstructured, Big Data*, pages 77–88. Springer, 2013.

[2] Burton G Malkiel and Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[3] Robert R Prechter Jr and Wayne D Parker. The financial/economic dichotomy in social behavioral dynamics: the socionomic perspective. *The Journal of Behavioral Finance*, 8(2):84–108, 2007.

[4] Bo Qian and Khaled Rasheed. Stock market prediction with multiple classifiers. *Applied Intelligence*, 26(1):25–33, 2007.

[5] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[6] Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

[7] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.

[8] Stephen F Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62, 1996.

[9] Yuanhua Feng, Jan Beran, and Keming Yu. Modelling financial time series with semifar–garch model. *IMA Journal of Management Mathematics*, 18(4):395–412, 2007.

[10] Mark L Mitchell and J Harold Mulherin. The impact of public information on the stock market. *The Journal of Finance*, 49(3):923–950, 1994.

[11] Torben G Andersen. Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1):169–204, 1996.

[12] Michel Crouhy and Michael Rockinger. Volatility clustering, asymmetry and hysteresis in stock returns: International evidence. *Financial engineering and the Japanese markets*, 4(1):1–35, 1997.

[13] Petko S Kalev, Wai-Man Liu, Peter K Pham, and Elvis Jarnecic. Public information arrival and volatility of intraday stock returns. *Journal of Banking & Finance*, 28(6):1441–1467, 2004.

[14] Jean-Gabriel Cousin and Tanguy de Launois. New intensity and conditional volatility on the french stock market. *Finance*, 27(1):7–60, 2006.

[15] Paul C Tetlock. Does public financial news resolve asymmetric information? *The Review of Financial Studies*, 23(9):3520–3557, 2010.

[16] Ryan Riordan, Andreas Storkenmaier, Martin Wagener, and S Sarah Zhang. Public information arrival: Price discovery and liquidity in electronic limit order markets. *Journal of Banking & Finance*, 37(4):1148–1159, 2013.

[17] Mai Thi Thanh Hien and T Thanh. Modelling and forecasting volatility by garch-type models: the case of vietnam stock exchange. *A dissertation presented in part consideration for the degree of MA. Finance and Investment*, pages 1–97, 2008.

[18] Jingliang Xiao, Robert D Brooks, and Wing-Keung Wong. Garch and volume effects in the australian stock markets. *Annals of Financial Economics*, 5(01):0950005, 2009.

[19] B Ashok, P Subhadeep, H Soham, and D Rahul. Impact of information arrival on volatility of intraday stock returns. Technical report, Working Paper, 2011.

[20] Sergei Sidorov, Paresh Date, Vladimir Balash, et al. Using news analytics data in garch models. *Applied Econometrics*, 29(1):82–96, 2013.

[21] Engle e Patton. What good is a volatility model. *Quantitative Finance*, 1:237–245, 2001.

[22] Kevin P Evans. Intraday jumps and us macroeconomic news announcements. *Journal of Banking & Finance*, 35(10):2511–2527, 2011.

[23] José Gonzalo Rangel. Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking & Finance*, 35(5):1263–1276, 2011.

[24] Bala Arshanapalli, Edmond d'Ouville, Frank Fabozzi, and Lorne Switzer. Macroeconomic news effects on conditional volatilities in the bond and stock markets. *Applied Financial Economics*, 16(5):377–384, 2006.

[25] Leela Mitra, Gautam Mitra, and Dan Dibartolomeo. Equity portfolio risk estimation using market information and sentiment. *Quantitative Finance*, 9(8):887–895, 2009.

[26] PA Hafez. Impact of news sentiment on abnormal stock returns. *White paper, RavenPack*, 2009.

[27] Xiang Yu. *Analysis of new sentiment and its application to finance*. PhD thesis, 2014.

[28] Xiang Yu, Gautam Mitra, Cristiano Arbex-Valle, and Tilman Sayer. An impact measure for news: Its use in daily trading strategies. *Available at SSRN 2702032*, 2015.

[29] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.

[30] Robert Cumby, Stephen Figlewski, and Joel Hasbrouck. Forecasting volatilities and correlations with egarch models. *The Journal of Derivatives*, 1(2):51–63, 1993.