

Received September 15, 2020, accepted September 20, 2020, date of publication September 29, 2020, date of current version October 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027631

# The Use of Mutual Information to Improve Value-at-Risk Forecasts for Exchange Rates

ALBERT ANTWI<sup>1,2</sup>, KWABENA A. KYEI<sup>2</sup>, AND RYAN S. GILL<sup>3</sup>

<sup>1</sup>Department of Mathematical Sciences, Sol Plaatje University, Kimberley 8300, South Africa

<sup>2</sup>Department of Statistics, University of Venda, Thohoyandou 0950, South Africa

<sup>3</sup>Department of Mathematics, University of Louisville, Louisville, KY 40292, USA

Corresponding author: Albert Antwi (albert.aantwi@outlook.com)

This work was supported in part by the Amalgamated Bank of South African (ABSA) and in part by the University of Venda with Grant number SMNS/17/STA/06.

**ABSTRACT** In this paper, we show a simple but novel approach in an attempt to improve value-at-risk forecasts. We use mutually dependent covariate returns to create exogenous break variables and jointly use the variables to augment GARCH models to account for time-variations and breaks in the unconditional volatility processes simultaneously. A study of hypothetical mutual dependencies between volatility and the covariates is first carried out to investigate the levels of the shared mutual information among the variables before using the augmented models to forecast 1% and 5% value-at-risks. The results provide evidence of some substantial exchange of information between volatility and the lagged exogenous covariates. In addition, the results show that the estimated augmented models have lower volatility persistence, reduced information leakages, and improved explanatory powers. Furthermore, there is evidence that our approach leads to fewer violations, improved 1% value-at-risk forecasts, and optimal daily capital requirements for all the models. There is, however evidence of relative superiority of the majority of the models for the 5% value-at-risks forecasts from our approach, although they have relatively higher failure rates. Based on these results, we recommend the incorporation of our approach to existing risk modeling frameworks. It is believed that such models may lead to fewer bank failures, expose banks to optimal market risks, and assist them in computing optimal regulatory capital requirements and minimize penalties from regulators.

**INDEX TERMS** Exogenous break, mutual information, value-at-risk, volatility.

## NOMENCLATURE

ARMA	Autoregressive moving average
ADmean	Mean Absolute Deviation
ADmax	Maximum Absolute Deviation
BCMI	Biased Corrected MI
EGARCH	Exponential GARCH
GARCH	Generalized Autoregressive Conditional Heteroscedasticity.
GJRGARCH	Glosten-Jagannathan-Runkle GARCH
MAE	Mean Absolute Error
MCS	Model Confidence set
MI	Mutual information
MSE	Mean Square Error
NAGARCH	Nonlinear Asymmetric GARCH
QMLE	Quasi-Maximum Likelihood Estimation
RMSE	Root Mean Square Error

SGARCH	Standard GARCH
TGARCH	Threshold GARCH
VaR	Value-at-risk

## I. INTRODUCTION

Risk measurement is one of the most important tasks in financial risk management for banks, corporate treasuries and portfolio management firms as well as other financial institutions and practitioners. In financial institutions, risk estimates are used to compute capital requirements - the amount of capital that must be added to a position to make its risk acceptable to regulators [21]. It is also employed in decision-making regarding hedging of assets and portfolio optimization. One of the most commonly used risk measures among financial institutions is value-at-risk (VaR) with the underlying asset's volatility as an input. Inaccurate volatility forecasts may lead to underestimation or overestimation of the actual VaR forecast and financial institutions may lose the opportunity cost or would not be able to recover

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Benedetto<sup>1b</sup>.

losses at crisis periods [52]. Since VaR computation requires volatility input, VaR accuracy is dependent on accurate volatility forecasts. Accurate volatility forecasts yield optimal VaR forecasts and less VaR violations, thus VaR forecasting assessment provides an indirect assessment of the predictive abilities of competing volatility models [15].

A common approach for computing the volatility input of VaR models is the use of GARCH models [5], however, findings from competing GARCH models demonstrate seemingly poor volatility forecasts [16]. This problem has been partly attributed to the models' failure to account for breaks and time-variations in the unconditional volatility [2], [43]. Models, which fail to account for breaks in the unconditional variance lead to sizable upward biases in the degree of persistence in the estimated GARCH models. The fitted models thus may fail to track changes in the unconditional variance and produce systematically underestimated or overestimated volatility forecasts on average, over long horizons [43].

Risk premium theory links the returns of an asset to its volatility [13] and asset returns move in tandem. Taking into account of exchange rate co-movements and the volatility-returns relationship of an asset, there is a possibility of an implied reverse relationship between volatility and inter-market<sup>1</sup> covariate returns. Due to the perceived reverse relationship and the fact that the inter-market covariate returns are exposed to similar uncertainties in the market, the covariates may be adequate proxies for uncertainties in the markets which could also be used to construct break variables. Several attempts in literature have addressed the problem of structural breaks and time-varying unconditional volatility, the use of inter-market covariates to proxy market uncertainty, and its break variables to account for time-variations and breaks in the unconditional volatility of GARCH models, however, have not been given much attention, thus a gap in literature exist. The paper is an attempt to narrow this gap.

In narrowing this gap, the proxies (inter-market), and the break variables (constructed from the proxies) are used to augment GARCH models to account for changes, and breaks in the unconditional volatilities. Empirical studies suggest that changes in the levels of uncertainties in the market affect the unconditional volatility of assets [2], [3], thus this approach has empirical support. One advantage of the approach is that, due to the high levels of co-movements between assets, the proxies and the break variables may share substantial levels of mutual information with volatility, thus it is anticipated to yield improved volatility and value-at-risk forecasts. The approach does not require structural modifications of existing models, thus its implementation is simple and could be easily integrated into various volatility and value-at-risk modeling frameworks.

<sup>1</sup>Inter-market covariates are assets traded on the same market platform with the same delivery date. For example, USD/ZAR, and EUR/ZAR rates are considered inter-market covariate when they are traded on the same market platform (such as the inter-bank forex market) and the same delivery date (such as daily).

A pre-study of the hypothetical mutual dependencies between pre-estimated volatilities and the exogenous variables (proxies together with their respective breaks) is undertaken to investigate the levels of the dependencies and their possible significance. The Jackknife-bias corrected Kernel density estimation approach is used in computing the mutual information while Pearson's method is used in computing the strength of the linear dependency. Simple t-tests are used to test the significance of the linear dependencies. In assessing the individual accuracies of the VaR estimates, the conditional coverage test of [17] and the unconditional coverage test of [35] are used. Competing models, which pass both tests, are then ranked using the MCS procedure of [28]. We also employ several other model comparison tools to assist in selecting the best models. To assess the usefulness of the estimated models to banking institutions, we conduct capital requirement analysis based on the regulations laid down by the Basel accord II and III accords.

The paper is a contribution to literature on value-at-risk forecasting for emerging markets. It would also provide further evidence in support of theoretical and empirical studies, which advocate that structural breaks have potentially important implications for estimated GARCH models and value-at-risk forecasts. Volatility and returns are linked by risk premium, hence, evidence of mutual dependencies between the covariates and volatility will indirectly support the traditional theory of return co-movements. For the rest of the paper, the concept of VaR is discussed in section two. In section three, we formally establish the theoretical link between inter-market covariates and volatility as well as the construction of the exogenous break variables. Data and methodologies are presented in section four. We present the results and discussions in section five and summarize the paper in section six.

## II. THE CONCEPT OF VALUE-AT-RISK AND METHODS OF ESTIMATION

Value-at-risk is concerned with the possibilities of losses associated with a portfolio, at a given time. It is a downside risk metric, which measures the risk as to whether the actual return will fall below or above the expected returns. In simple terms, it is the uncertainty about the magnitude of the differences in returns and the expected returns. It is the preferred risk metric by many experts because of its perceived superiority in backtesting the estimated losses [45]. Technically, value-at-risk is defined as the maximum portfolio loss at a given confidence level,  $\alpha$  within a time interval [54].

Consider a scenario where there is a portfolio of risky assets held over a fixed time in the horizon  $\Delta$ . If the loss distribution associated with this portfolio has distribution function  $F_L(l) = P(L \leq l)$ , the maximum possible loss  $\inf \{l \in R : FL(l) = 1\}$  evaluates the level of risk associated with holding the portfolio over the horizon  $\Delta$ . This scenario leads to the technical definition of VaR below adapted from [45].

*Definition 1:* Given a confidence level  $\alpha \in (0, 1)$ , the VaR of a portfolio is given by the smallest number  $l$  such that the probability that the loss  $L$  exceeds  $l$  is no larger than  $(1 - \alpha)$ , that is:

$$\begin{aligned} \text{VaR}_\alpha &= \inf \{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} \\ &= \inf \{l \in \mathbb{R} : F_L(l) > \alpha\} \end{aligned} \quad (1)$$

In probabilistic terms, value-at-risk is a quantile of the loss distribution. Given losses  $L$ , the generalized inverse  $F \leftarrow$  is called the quantile function of  $L$  such that:

$$\text{VaR}_\alpha(L) = q_\alpha(L) := F \leftarrow(L) = \inf \{l \in \mathbb{R} : F_L(l) \geq \alpha\} \quad (2)$$

Alternatively, value-at-risk can be constructed from the probabilistic function of the underlying returns, which is given in definition 2 [45].

*Definition 2:* Consider the returns of an asset  $r_t$  with the change in the value of the asset over the next  $k$  periods defined by  $\Delta r_t = \Delta V(k) = r(t+k) - r(t)$ . The value of VaR over time horizon  $k$  associated with the left tail probability  $\alpha$  of the returns' distribution is defined as:

$$\alpha = P[\Delta V(k) \leq \text{VaR}_t] = P[\Delta r \leq \text{VaR}_t] \quad (3)$$

Value-at-risk is a function of time and the left tail quantile of the distribution with probability  $\alpha$ . Throughout this paper, any computations and assessments relating to VaR models are based on definition (2).

### III. THEORETICALLY LINKING VOLATILITY TO RETURNS OF INTER-MARKET COVARIATES

In linking volatility to the returns of inter-market covariates, we make the following assumptions:

1. Exchange rate returns are second-order stationary processes.
2. Exchange rate returns are co-integrated.
3. The relationship between two or more exchange rate returns evolve via an ARMA process.
4. The expectation of the square of the ARMA process in assumption (3) is analogous to an exogenous GARCH process.

The plausibility of assumption (1) is derived from empirical studies such as [53] among others. Assumption (2) can be indirectly inferred from the empirical evidence of exchange rate cointegration, which can be found in studies like [20], [34]. Assumption 3 can be deduced from Wold's representative theorem in conjunction with assumption 2. Finally, assumption 4 can be statistically derived from assumption 3.

Consider the endogenous exchange rate returns  $R_t$  and a set of exogenous returns  $r_{j,t}$ , where  $j = 1, 2, \dots, k$  and  $k$  is the total number of currency pairs. If  $R_t$  and  $r_{j,t}$  are *second-order stationary cointegrated processes*, then, there exist integers  $p, q$  and real coefficients  $\lambda_1, \lambda_2 \dots \lambda_k, \varphi_1, \varphi_2 \dots \varphi_p$  and

$\theta_1, \theta_2 \dots \theta_q$  such that by assumption 3 (adapted from [54]):

$$R_t = \varphi_0 + \sum_{j=1}^k \lambda_j r_{j,t} + \sum_{i=1}^p \varphi_i R_{t-i} - \sum_{r=1}^q \theta_r a_{t-r} + a_t \quad (4)$$

where  $R_{t-i}$  is the autoregressive component,  $a_{t-r}$  is the moving average term, and  $a_t$  is the mean corrected return (innovations). By squaring both sides of (4), it can be shown that:

$$R_t^2 = \varphi^2 + \sum_{i=1}^p \varphi_{ii} R_{t-i}^2 + \sum_{r=1}^q \theta_{rr} a_{t-r}^2 + \sum_{k=1}^m \lambda_k r_{k,t} + a_t \quad (5)$$

For stationary asset returns, the unpredictability of the direction of the returns suggest that the autocorrelation function up to lag 1 is  $\rho = 0$ . Applying the law of total variance and taking expectation with respect to the information set  $\mathcal{F}_{t-1}$ , the volatility  $h_t$  can be expressed as

$$h_t = \delta + \sum_{r=1}^p \psi_r a_{t-r}^2 + \sum_{i=1}^q \varphi_i h_{t-i} + \sum_{k=1}^m \lambda_k E_{t-1} [r_{k,t}] \quad (6)$$

It can clearly be seen that equation 6 is analogous to the exogenous version of [14], where the exogenous term is represented by the last set of terms. These terms move in tandem with volatility by the indirect implications of the asset-return co-movement theory, thus, their absolute values are used to proxy the degree of uncertainties in the exchange rate market in this paper. The use of the absolute returns of the exogenous terms as proxies instead of the original values are due to the fact that market uncertainty is a positive measure. If one considers the positive returns from the  $k$  exogenous exchange rates with a sample of size  $N$ , then, the lag-one break variables  $B_{j,t-1}$  are constructed using the definition below:

$$B_{j,t-1} = \begin{cases} 1, & \text{if } |r_{j,t-i-1}| < |r_{t-i}| \\ 0, & \text{if } |r_{j,t-i-1}| > |r_{t-i}| \end{cases} \quad \text{for } i = 1, 2, \dots, N \quad \text{and } j = 1, 2, \dots, k \quad (7)$$

### IV. DATA AND METHODOLOGY

#### A. DATA

We use daily inter-bank closing spot-ask prices from the rand forex market to illustrate the empirical significance of our approach. The rand forex market was chosen because being an emerging market and ranked as the 18th world's most traded currency [8], it has not received much attention in terms of value-at-risk forecasting in comparison to other emerging and developed markets. The period of the data spans from July 6, 2011, to June 28, 2016. The data were sourced from OANDA (<https://www1.oanda.com>) and the currency pairs include the Swedish Kroner (SEK), Norwegian Kroner (NOK), Botswana Pula (BWP), Brazilian Real (BRL), Israeli Shekel (ILS), Indian Rupees (INR) and Malawian Kwacha (MWK). Due to the data download restrictions during the time of sourcing the data, fifteen currency pairs including the most traded currency pairs on the rand forex market such as the USD/ZAR, GBP/ZAR,

and EUR/ZAR among others were selected. Due, however, to problems such as autocorrelations and serial correlations of some pairs and the fact that some pairs from developing and emerging economies have not received substantial attention in literature when it comes to volatility and value-at-risk forecasting, only ten currency pairs were selected for the empirical exercise. The daily returns used in the study are continuously compounded as indicated in equation (8).

$$R_t = \ln(P_t) - \ln(P_{t-1}) \quad \text{and} \quad r_t = \ln(p_t) - \ln(p_{t-1}) \quad (8)$$

An upper case letter denotes an endogenous return while lower case letter denotes an exogenous return.

## B. MUTUAL INFORMATION

Information theory attempts to study communication systems. It can be employed in the study of statistical dependency between random variables via mutual information (MI). Mutual information measures the reduction in the uncertainty about a random variable conditioned on the knowledge of another variable, thus it is more closely related to the concept of entropy (a measure of the uncertainty of a random variable or the expected amount of information contained in a variable) introduced by [49]. There are several ways of measuring dependency [58] although, MI is by far the best statistic in several ways [46]. Unlike linear correlation, MI is more general in the sense that it contains all information about the dependency of variables including linear and non-linear and it is very effective in measuring any kind of relationship [18]. MI also has a straightforward interpretation, grounded in information theory and insensitive to the size of data sets. Mutual information is a function of the expected amount of information contained in a variable, called entropy [47]. Given a pair of continuous random variables  $(X, Y)$  with values over the space  $\chi \times y$  the entropy of  $X$  denoted  $H(X)$  is defined by [18] as

$$H(X) = E[-\log(f(X))] = - \int_{\chi} f(x_i) \log_b f(x_i) dx \quad (9)$$

where  $b$  is the base of the logarithm,  $f(x_i)$  is the probability density function of  $X$ ; and  $X = x_1$ . If  $y = y_1$ , the joint entropy of  $X$  and  $Y$  is defined as

$$H(X, Y) = - \int_{\chi \cdot y} f(x_i, y_i) \log_b f(x_i, y_i) dx dy \quad (10)$$

The mutual information between  $X$  and  $Y$  [18]:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \quad (11)$$

Mutual information between bivariate variables can either be zero (corresponding to independence) and positive (corresponding to dependency). When  $MI(X, Y) = 0$ , observing  $Y$  tells us nothing about  $X$ , thus the variables are independent. MI is a positive unbounded measure i.e.  $MI \in [0, \infty)$  with unit in bits (base 2) or nats (base  $e$ ).

Among the commonly used estimation approaches for mutual information are the probability density-based methods such as the Burg's maximum entropy method (Burg's MEM), the kernel density estimation (KDE), and the nearest-neighbor approach. For instance, [9], [10] used the Burg's MEM to model the flow of information between financial variables. Reference [42] also developed a new KDE for application to large high-dimensional datasets frequently used in genomic experiments. In as much the Burg's MEM of [9], [10] and the KDE of [42] are non-parametric estimators and are suitable for high-dimensional datasets; however, they differ in three major aspects.

In the first instance, the new Kernel density estimation is based on Shannon's definition of entropy and joint entropy. The probability density function is estimated by filtering the data with a kernel, which is then normalized with an integral of one, which is usually symmetric and localized. The Burg's MEM on the other hand is based on the extrapolation of the autocorrelation function of variables by using the entropy rate definition after which the estimation of the power spectral density is estimated by the Fourier transformation of the extended autocorrelation function.

Secondly, the new Kernel density estimation is purposely designed to model static relationship between variables while the Burg's MEM is designed to model dynamic relationship and to predict future entropy of a time series by exploiting the unknown but predicted autocovariance function of the future time interval.

Thirdly, the Burg's MEM relies on the assumption of second-order stationary. If this hypothesis is not satisfied, the series can be partitioned into smaller epochs, which are approximately stationary, or the series can be represented by alternative functions instead of the usual sine and cosine functions. Simulation studies, however, have shown that the Burg's approach to spectral analysis is robust in the presence of non-stationarity. Unlike the Burg's MEM, the KDE does not rely on stationarity assumption, although, it relies heavily on the choice of the tuning parameters, thus the corresponding estimators may be very unstable or seriously biased. This problem is addressed by using the Jackknife-bias corrected algorithm. The Jackknife-bias corrected Kernel density estimator automates the bandwidth selection such that the optimal bandwidth is estimated. This helps to reduce the bias at the boundary region and thus improve the efficiency of estimation [58]. Unlike the randomized resampling approaches for correcting bias in Kernel estimation, the Jackknife approach is deterministic in the sense that it gives the same result when re-applied to any given data. In addition, by restricting the resampling to a specific group of  $n$  subsamples, substantial computational costs can be avoided. The approach puts an upper limit on the number of subsamples and the relationships between Jackknife repetitions can be exploited to avoid redundant computations [42].

Lastly, unlike the Burg's MEM, the Kernel density estimation models the distribution of a continuous variable as a mixture of conditional distributions for each level of a

categorical variable, thus it is suitable for the estimation of mutual information between a mixture of continuous and discrete variables. Due to this, KDE can be used to model the relationship between a mixture of discrete or categorical and continuous variables without the need for variable transformation.

In comparison to other density-based estimation approaches such as the mirrored KDE, ensemble KDE, copula-based generalized nearest-neighbor graphs and the mixed generalized nearest-neighbor graphs, the Jackknife-biased corrected KDE is more computationally efficient. In addition, the procedure is completely data-driven and there is no need for a predetermined tuning parameter. It does not necessitate boundary correction and yet it retains the same estimation efficiency because the boundary biases are eliminated automatically. Furthermore, the estimates are numerically stable. Due to these advantages over existing methods, the Jackknife-biased corrected KDE is employed in this paper.

**C. VALUE-AT-RISK ESTIMATION**

There are several metrics for computing risk but in this paper, we use the VaR metric. VaR metric corresponds to an amount that could be lost at some pre-selected probability. It also measures the risk of the risk factors and the risk factor sensitivities. The metric applies to all activities and types of risks in financial institutions and it can be compared across different markets at different exposures. The metric can also be measured at any level, from a single trade or portfolio case up to a single enterprise-wide metric covering all the risks in the firm. It can be used to find the total VaR of a very large portfolio in aggregated form or to isolate component risks corresponding to different types of risk factors in disaggregated form. The metric, moreover, accounts for the dependencies between the component of assets or portfolios [1]. The methods used in computing VaR includes the historical simulation method [33]. One of the advantages of this approach is that it determines the joint probability distribution of the market variables and avoids the need for cash-flow mapping; however, it is computationally slow and does not easily allow volatility updating schemes to be used [30]. Notwithstanding the disadvantage of the historical simulation method, the study use this approach to estimate and forecast VaR, because the conditional volatility models used here are built on historical returns. The GARCH framework is used to specify the volatility input. By definition [32] VaR is given by

$$VaR_t^D = Z_\alpha^D \sqrt{\hat{h}_t} + \mu \tag{12}$$

where  $VaR_t^D$  denotes the VaR estimate based on volatility model with appropriate distribution  $D$ ,  $\hat{h}_t$  is the forecasted volatility at time  $t$  and  $\mu$  is the estimated mean from the volatility model. In considering the returns series  $R_t$  we estimate the mean,  $\mu$  and the volatility  $h_t$  using the

ARMA-GARCH representation [54],

$$R_t = \varphi_0 + \sum_{j=1}^k \lambda_j r_{jt-1} + \sum_{i=1}^p \varphi_i R_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{13}$$

where  $\varphi_0, \varphi_i$  and  $\theta_j$  are parameters,  $R_{t-i}$  is the  $i^{th}$  lag autoregressive term with corresponding order  $p$ ,  $\varepsilon_{t-j}$  is the  $j^{th}$  lag moving average component with order  $q$  and  $\varepsilon_t$  term is the mean corrected return (innovations). The  $\eta_t$  is a sequence of *i.i.d* random variables and  $h_t$  is the conditional variance. In this paper, we consider the conditional volatilities for GJRGARCH, NAGARCH, SGARCH, TGARCH, and EGARCH. The break variable in (9) and the absolute values of the exogenous covariates in (8) are passed to the unconditional volatilities of the GARCH models to augment the models. The augmented form of the standard GARCH model of [14] is defined by adding the market uncertainty proxy identified in equation (6) and the break variables from equation (7), hence,

$$h_t = \alpha_0 + \Psi_t + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \tag{14}$$

where  $\Psi = \sum_{m=1}^k \theta_m B_{t-1} + \sum_{l=1}^k \lambda_l |r_{l,t-1}|$ . To ensure that  $h_t > 0$  the parameters are conditioned, such that  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$  and  $\beta_i \geq 0$ . The necessary and sufficient condition required for the existence of the second moment of the returns is that  $\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1$ . The GARCH model is popular because of its simplicity in modeling very complex GARCH processes. The GJR version [25] is formulated as

$$h_t = \alpha_0 + \Psi_t + \sum_{i=1}^q \left( \alpha_i a_{t-i}^2 - \gamma_i I_{t-i} a_{t-i}^2 \right) + \sum_{j=1}^p \beta_j h_{t-j} \tag{15}$$

where  $I_{t-j}$  is an indicator function, which takes the value of 1 if  $a \leq 0$  or 0 otherwise. To ensure that the variance,  $h_t > 0$ , the parameters are constrained such that  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_i \geq 0$  and  $\alpha_i + \gamma_i \geq 0$ . The  $\gamma_i$  parameter provides information about asymmetric effects. If  $\gamma_i = 0$ , there is no volatility asymmetry, if  $\gamma_i > 0$  negative shocks will increase volatility more than positive shocks of the same magnitude and if  $\gamma_i < 0$ , positive shocks increase volatility more than negative shocks. The persistence parameter is  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j + \sum_{i=1}^q \gamma_i \kappa$ , where  $\kappa$  is the expected value of the standardized residuals  $z_t$  below zero (effectively the probability of being below zero) defined by  $\kappa = E \left[ I_{t-j} z_{t-j}^2 \right] = \int_{-\infty}^0 f(z, 0, 1, \dots) dz$ . The exponential GARCH version [41] is formulated as

$$\ln(h_t) = \alpha_0 + \Psi_t + \sum_{j=1}^q \alpha_j \ln(h_{t-j}) \tag{16}$$

where  $H_{t-1} = \sum_{n=1}^n \gamma_n \left( \frac{\varepsilon_{t-n}}{\sqrt{h_{t-n}}} \right) + \sum_{i=1}^p \beta_i \left( \left| \frac{\varepsilon_{t-i}}{\sqrt{h_{t-n}}} \right| - E \left| \frac{\varepsilon_{t-i}}{\sqrt{h_{t-n}}} \right| \right)$ ,  $\alpha_i$ ,  $\beta_j$  and  $\gamma_n$  are real constants with  $p$  and  $q$  being the respective orders. To ensure the existence of a second moment,  $\alpha$  is conditioned such that  $\sum_{j=1}^q \alpha_j < 1$ . The main advantage of the EGARCH over the standard GARCH and GJRARCH is that the EGARCH model does not require any artificially imposed non-negativity constraints on the model parameters. The threshold GARCH version [57] of the augmented model is

$$h_t = \omega + \Psi_t + \sum_{i=1}^q (\alpha_i + \gamma_i I_{t-i}) \alpha_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \quad (17)$$

where all notations are the same as the GJRARCH. The nonlinear asymmetric GARCH specification [19] is formulated as:

$$h_t = \omega + \Psi_t + \sum_{i=1}^q \alpha_i h_{t-i} (z_{t-i} - \delta_i)^2 + \sum_{j=1}^p \beta_j h_{t-j}^2 \quad (18)$$

where  $\delta_i$  is a parameter, which controls shift (asymmetry for small shocks) in the news-impact curve. The existence of the second moment requires that  $\sum_{i=1}^q \alpha_i (1 + \delta_i^2) + \sum_{j=1}^p \beta_j < 1$ . Due to numerical difficulties encountered by the Quasi Maximum Likelihood Estimation (QMLE) method, we use variance targeting to alleviate the degree of these difficulties [22].

### D. BACKTESTING METHODS FOR EVALUATING VAR ESTIMATES

Backtesting techniques require the simulation of VaR models on past returns and the predicted losses from VaR calculations are then compared to the actual realized losses at a given time horizon. The comparison identifies periods where the portfolio losses are greater than the expected VaR. If the expected return is less than the estimated VaR, a violation or exception occurs, thus backtesting techniques are used to systematically count the number of these violations or exceptions and compare them to acceptable rates at pre-selected confidence intervals. Consider the realization of asset returns over a fixed time interval  $R_{t,t+1}$  with VaR estimated at time  $t$  and probability of  $\alpha$  defined by  $VaR_t(\alpha)$ . The hit function as defined in [32] is

$$I_{t+1}(\alpha) = \begin{cases} 1 & \text{if } R_{t,t+1} \leq VaR_t(\alpha) \text{ when violation occurs} \\ 0 & \text{if } R_{t,t+1} > VaR_t(\alpha) \text{ when no violation occurs} \end{cases} \quad (19)$$

If the loss at day  $(t + 1)$  is larger than the predicted VaR estimate, the hit sequence returns 1 or 0 otherwise. Reference [17] explains that the hit sequence of VaR estimates needs to pass the tests based on the unconditional coverage and independence properties before they can be deemed accurate. Under the *unconditional coverage property*, the probability of loss at day  $(t + 1)$  being larger

than the predicted VaR estimate should be exactly  $(1 - \alpha)$  or equivalently, the probability of loss at day  $(t + 1)$  being smaller than the predicted VaR estimate should be exactly  $\alpha$ . Mathematically a VaR model has *correct unconditional coverage* if:

$$P(I_{t+1}(\alpha) = 1) = E[I_{t+1}] = 1 - \alpha \quad (20)$$

It also has correct conditional coverage if:

$$P_t(I_{t+1}(\alpha) = 1) = E_t[I_{t+1}] = 1 - \alpha \quad (21)$$

It should be noted that *correct unconditional coverage* is implied by *correct conditional coverage* but not vice versa. If violations are more frequent than  $\alpha$ , the VaR model systematically underestimates the actual VaR and if violations are less frequent than  $\alpha$ , the model overestimates the expected VaR [32]. The *independence property* requires that for any  $i < j$  hits,  $I_{t+i}(\alpha)$  and  $I_{t+j}(\alpha)$  are independent if  $i \neq j$ . In other words, when the exceedances are not clustered over time, VaR estimates are said to be independent. Accurate VaR estimates, therefore, have current violations at time  $(t + i)$ , which is independent of violations at previous time  $(t + j)$  or has non-clustered exceedance over time. The hit sequences from VaR estimates,  $I_{t+1}(\alpha) \sim i.i.d. \text{ Bernoulli}(1 - \alpha)$  with success probability  $(1 - \alpha)$ , thus, in testing for VaR violation we are interested in the nulls:

$$\begin{aligned} H_0: & E[I_{t+1}] = \pi = 1 - \alpha \\ H_0: & E_t[I_{t+1}] = \pi = 1 - \alpha \end{aligned} \quad (22)$$

$\pi$  is the sample average. The first null corresponds to the unconditional coverage test while the last one corresponds to the conditional coverage test. In this paper, the unconditional coverage Kupiec test [35] and the Christoffersen's interval forecast test [17] are used.

### E. MODEL COMPARISON TOOL

Due to large set of models available for use by financial institutions, model comparison has become an integral part of the model-building process. The Model Set Confidence (MSC) test of [27] is convenient for comparisons where there is no natural benchmark. This advantage of MCS's procedure makes it more suitable for use in this study because the reference models in this study are not natural benchmarks. Even though the MCS procedure does not require a natural benchmark model as in the case of multiple comparison procedure with controls, the procedure can still rank the models in order of superiority after selecting the superior set of models. One main advantage of this procedure is that after a set of superior models have been selected, the models can be aggregated and used to forecast future volatility levels, predict future levels of observations, conditioned on past information [11] or forecast value at risk levels [12].

## V. EMPIRICAL RESULTS

### A. DESCRIPTIVE ANALYSIS

Descriptive statistics are reported in Table 1 while time plots of the log-returns of the entire length of the sample are

TABLE 1. Descriptive statistics.

Currency	Mean	Std dev	Skewednes	Kurtosis	Jarque Bera	Minimum	Maximu
BRL	2.6E-05	0.0071	0.1463	6.5496	96.4566	-0.0420	0.0510
BWP	0.0002	0.0064	-0.1615	10.187	3923.26	-0.0586	0.0387
ILS	0.0004	0.0075	0.2256	42.716	119564	-0.1000	0.1000
INR	0.0002	0.0065	6.7305	6.7305	1074.07	-0.0412	0.0326
MWK	-0.0004	0.0203	-1.2720	22.633	29703.8	-0.2327	0.1112
NOK	0.0002	0.0059	0.3745	7.8875	1853.03	-0.0448	0.0399
SEK	0.0003	0.0060	0.1792	6.4483	910.980	-0.0438	0.0350

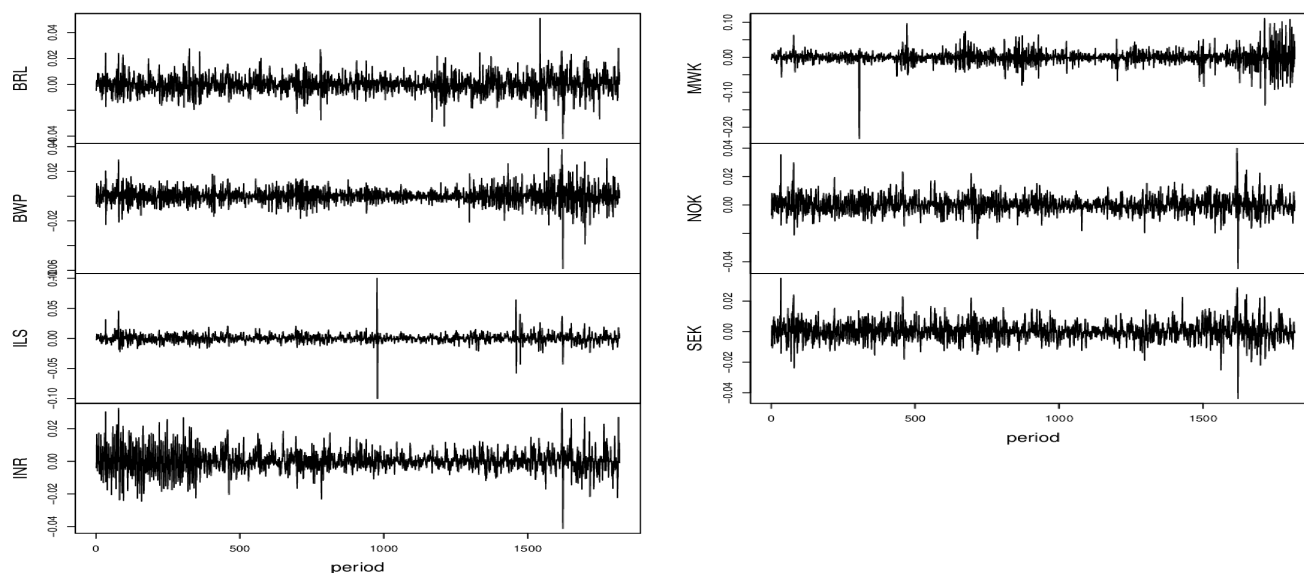


FIGURE 1. Time plots of returns for the sample period July 6, 2011, to June 28, 2016.

displayed in Figure 1. A look at Figure 1 reveals that high returns are more likely to be followed by another high returns. This is an indication of the presence of asymmetry in the log returns. The Figure also depicts time-varying volatility clustering of the log-returns. These observations confirm the suitability of GARCH models for the data. In Table 1, relatively extreme negative returns appear to be more pronounced than extreme positive returns for the currency pairs: BWP/ZAR and NOK/ZAR. In addition, it is observed that all the returns are slightly and positively skewed except for BWP/ZAR and MWK/ZAR pairs, which are negatively skewed. The excess kurtosis for all the returns is large and far from zero and the Jacque-Bera tests statistics are very large. This is an indication of non-normality of the return with associated heavy tails, thus heavy-tailed distributions may be appropriate in modeling the volatility of the returns.

**B. ESTIMATION OF HYPOTHETICAL MUTUAL INFORMATION**

One of the main reasons we advocate the use of inter-market covariates and their break variables to model time-variations and breaks in the unconditional volatility of GARCH models is their perceived mutual dependence with volatility, thus in

this section, we investigate the levels and the significance of these dependencies. The latent nature of volatility implies that actual shared information cannot be computed, thus the mutual information computed in this paper are hypothetical in a broader sense. Since we intend to compute value-at-risk for the currency pairs: MWK/ZAR, BWP/ZAR, BRL/ZAR, ILS/ZAR, and SEK/ZAR- it is more appropriate to use pre-estimated volatilities for these pairs using similar specifications (refer to Table 4), as that will be employed in estimating the value-at-risks. Individual estimation of the volatilities of the above currency pairs is time-consuming, thus the simultaneous estimation approach implemented in rugarch package [24] is utilized.

The GARCH estimation procedure used for computing the value-at-risks in this paper is based on a moving window, thus simple lag-one moving averages of the variables are computed. The lagging is required because we are interested in using the variables for forecasting. Since MI is a positive unbounded measure, to be able to compare two MI estimates we normalize the estimates using equation (23)-adapted from [18].

$$\text{Normalized BCMI} = \frac{BCMI(X, Y)}{\sqrt{H(X)H(Y)}} \tag{23}$$

TABLE 2. Estimated BCMI and normalized BCMI between volatilities and returns.

Volatility	MWK/ZAR	BWP/ZAR	BRL/ZAR	ILS/ZAR	SEK/ZAR	INR/ZAR	NOK/ZAR
MWK/ZAR		0.743 (0.323) <b>-0.009*</b>	0.600 (0.28) <b>0.114</b>	0.820 (0.35) <b>-0.059</b>	0.917 (0.39) <b>-0.077</b>	0.868 (0.37) <b>-0.228</b>	0.869 (0.37) <b>-0.124</b>
BWP/ZAR	0.155 (0.06) <b>0.127</b>		0.086 (0.04) <b>0.200 L</b>	0.142 (0.06) <b>0.056</b>	0.127 (0.05) <b>0.084</b>	0.142 (0.06) <b>-0.160</b>	0.128 (0.05) <b>0.022*</b>
BRL/ZAR	0.581 (0.23) <b>0.138</b>	0.506 (0.21) NL <b>-0.051</b>		0.637 (0.26) NL <b>-0.031*</b>	0.660 (0.27) <b>-0.064</b>	0.634 (0.26) <b>-0.193</b>	0.644 (0.26) <b>-0.096</b>
ILS/ZAR	0.321 (0.15) NL <b>0.055</b>	0.318 (0.16) <b>0.075</b>	0.310 (0.16) <b>0.118</b>		0.415 (0.20) <b>0.109</b>	0.486 (0.23) <b>-0.006*</b>	0.420 (0.20) <b>0.087</b>
SEK/ZAR	0.313 (0.14) <b>-0.101</b>	0.281 (0.13) <b>0.247</b>	0.241 (0.12) <b>0.260</b>	0.362 (0.16) <b>0.141</b>		0.386 (0.17) <b>0.114</b>	0.326 (0.15) <b>0.233</b>

Normalized BCMI estimates are reported in brackets while the strength of linear dependencies is in bold values. All estimated information measures are measured in bits. Asterisks indicate that the correlation coefficient is not significant at 5%.

TABLE 3. Estimated BCMI and normalized BCMI between volatilities and break variables.

Metric	MWK/ZAR	BWP/ZAR	BRL/ZAR	ILS/ZAR	SEK/ZAR	INR/ZAR	NOK/ZAR
MWK/ZAR	0.666 (0.29) <b>-0.036*</b>	0.434 (0.19) <b>-0.070</b>	0.627 (0.28) <b>0.081</b>	0.490 (0.21) <b>-0.067</b>	0.686 (0.26) <b>0.082</b>	0.627 (0.24) <b>-0.061</b>	0.548 (0.23) <b>-0.036*</b>
BWP/ZAR	0.122 (0.05) <b>-0.025*</b>	0.163 (0.07) <b>0.141</b>	0.133 (0.06) <b>0.029*</b>	0.170 (0.07) <b>0.058</b>	0.244 (0.08) <b>0.063</b>	0.236 (0.08) <b>0.062</b>	0.132 (0.05) <b>0.060</b>
BRL/ZAR	0.550 (0.23) <b>-0.069</b>	0.391 (0.17) <b>0.035*</b>	0.475 (0.21) <b>0.036*</b>	0.375 (0.15) <b>-0.030*</b>	0.534 (0.19) <b>0.069</b>	0.489 (0.18) <b>-0.044*</b>	0.396 (0.16) <b>-0.024*</b>
ILS/ZAR	0.295 (0.15) <b>0.045*</b>	0.202 (0.10) <b>0.114</b>	0.282 (0.14) <b>0.011*</b>	0.235 (0.11) <b>0.055</b>	0.348 (0.15) <b>0.042*</b>	0.251 (0.11) <b>0.069</b>	0.259 (0.12) <b>0.073</b>
SEK/ZAR	0.285 (0.13) <b>0.066</b>	0.214 (0.10) <b>0.180</b>	0.249 (0.12) <b>-0.043*</b>	0.196 (0.09) <b>0.186</b>	0.355 (0.14) <b>-0.011*</b>	0.254 (0.11) <b>0.147</b>	0.265 (0.12) <b>0.035*</b>

Normalized BCMI estimates are reported in brackets while the strength of linear dependencies is in bold values. All estimated information measures are measured in bits. MWK/ZAR, BWP/ZAR, BRL/ZAR, ILS/ZAR, and SEK/ZAR are endogenous break variables while INR/ZAR and NOK/ZAR are exogenous. Asterisks indicate that the correlation coefficient is not significant at 5%.

The empirical results of the mutual dependency estimation are reported in Table 2, 3, and Figure 2. From Table 2, it is observed that the returns of SEK/ZAR have the highest percentage of exchanged information with the volatilities of both MWK/ZAR and BRL/ZAR pairs. Between the two volatilities, the information exchange is highest with MWK/ZAR suggesting that returns of SEK/ZAR may have a better chance of predicting the volatility of MWK/ZAR than the volatility of BRL/ZAR. Similarly, returns of INR/ZAR may have a better chance of predicting the volatility of ILS/ZAR than the volatility of SEK/ZAR. Concerning the volatility of BWP/ZAR, the returns of MWK/ZAR stand a better chance of predicting the volatility of BWP/ZAR than the other returns although there is only 5.9% shared information.

In Table 3, the endogenous break variables for MWK/ZAR and SEK/ZAR respectively share more percentages of mutual information with their respective volatilities than the corresponding exogenous break variables. In relation to the volatility of MWK/ZAR, the non-normalized shared information with the exogenous SEK/ZAR break variable is, however, slightly higher than that of the endogenous break variable. In respect of the volatility of BWP/ZAR, the break variables for SEK/ZAR and INR/ZAR have the highest percentage of exchanged information with the volatility. A similar statement can be made for the break variables for MWK/ZAR, and SEK/ZAR with respect to the volatility of ILS/ZAR although the non-normalized exchanged information is relatively higher for the break variable of SEK/ZAR. Finally, the break variable for MWK/ZAR exchanges the highest



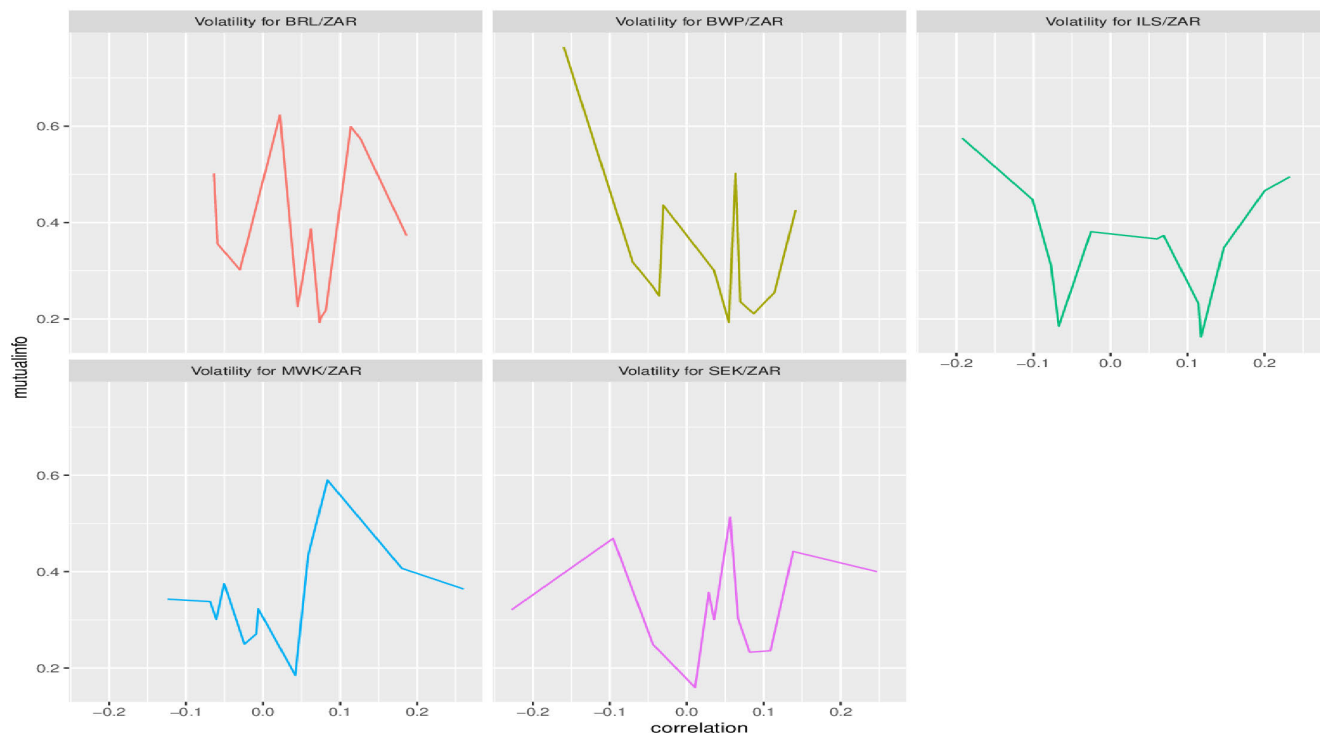


FIGURE 2. Correlation versus mutual information plots.

percentage of information with the volatility of BRL/ZAR. In general, among the variable pairings considered in the paper, the exogenous break variables tend to exchange relatively higher mutual information with volatility in comparison to the endogenous break variables. This suggests that break variables constructed from exogenous returns have higher likelihoods of volatility predictive abilities, hence, are more likely to be adequate than the endogenous break variables in accounting for breaks in the unconditional volatilities of exchange rates.

There are quite a lot of variable pairings with insignificant linear dependencies although their respective BCMI are substantial. For example, the strength of the linear dependency between the returns of BWP/ZAR and the volatility of MWK/ZAR is abysmally low and insignificant although the normalized BCMI is about 32%. This suggests that, it is very unlikely that the two variables will move linearly together, thus the nature of the relationship between the two variables is more likely to be non-linear. This observation is consistent with the rationale behind the concept of mutual information in measuring dependency [18]. Based on the foregoing deduction, the nature of the relationships between the variable pairings in Tables 2 and 3 (indicated by cells with asterisks) are more likely to be non-linear, hence non-linear volatility models may be more appropriate for estimating volatility of the exchange rate returns.

When the joint distribution of paired variables is a bivariate normal, there is an exact logarithmic relation between mutual

information and linear correlation coefficient  $\rho$  which is defined by [23]

$$MI = -0.5 \log(1 - \rho^2) \tag{24}$$

It is observed from Figure 2 that, the relationship between MI and correlation coefficient does not assume the form of equation (24) but rather it evolves randomly. This implies that the joint distributions of the paired variables are not bivariate normal, hence, since the individual variables are characterized by heavy-tailed distributions (see the data description section); the bivariate joint distributions may be a mixture of heavy-tailed distributions. In addition, the relationship suggest that increased levels of mutual information are not associated with increased strength of linear dependencies, thus there may be low levels of shared information between variables, although, the strength of the linear dependency structure may be comparatively higher.

In conclusion, the levels of exchange mutual information between volatilities and their respective lagged covariates provides substantial evidence of the predictive abilities of the covariates and these may be adequate in modeling variations and breaks in the unconditional volatilities of GARCH models for exchange rates. The results indirectly support the findings of returns dependencies from [40].

C. VOLATILITY AND VALUE-AT-RISK ESTIMATIONS

In this and the next section, the VaR estimation procedures discussed in the methodology section are applied to data from

TABLE 4. Specifications of estimated VaR model.

	Model	Returns	Proxy	Dummy	GARCH	ARMA	Target	Distribution
Augmented models	VaR-GJRGARCH	MWK	BRL	BRL	(1,1)	(5,5)	YES	NIG
	VaR-NAGARCH	BWP	BRL	BRL	(3,0)	(2,1)	YES	JSU
	VaR-SGARCH	BRL	MWK	MWK	(2,1)	(5,4)	YES	JSU
	VaR-TGARCH	ILS	MWK	BRL	(2,2)	(1,1)	YES	NIG
	VaR-EGARCH	SEK	INR, NOK	INR	(1,1)	(1,1)	NO	SGED
Benchmark models	VaR-GJRGARCH	MWK	None	None	(1,1)	(5,5)	YES	NIG
	VaR-NAGARCH	BWP	None	None	(3,0)	(2,1)	YES	JSU
	VaR-SGARCH	BRL	None	None	(2,1)	(5,4)	YES	JSU
	VaR-TGARCH	ILS	None	None	(2,2)	(1,1)	YES	NIG
	VaR-EGARCH	SEK	None	None	(2,1)	(1,1)	NO	SGED

The variables under the dummy column represent the returns that were used in constructing the dummy or break variable

the inter-bank rand forex market. The data for estimation consists of 1819 days of closing returns from the seven selected currency pairs, which are divided, into two sub-samples. The first sample consisting of 1379 returns from 2011/11/07 to 2015/04/15 is used for the in-sample estimation, while the second sample consisting of the remaining 440 returns from 2015/04/16 to 2016/06/28 is used to backtest the VaR estimates. The VaR estimates were computed for the single asset portfolios namely – MWK/ZAR, BWP/ZAR, BRL/ZAR, ILS/ZAR, and SEK/ZAR using the summarized specifications in Table 4. Hereafter, the augmented VaR models are represented as VaR<sub>A</sub> while the benchmark (reference or non-augmented) models are represented as VaR<sub>B</sub>. For each model, one-step-ahead VaR forecasts at 5% and 1% confidence levels (corresponding to the risk metrics methodology and the requirements of the Basel II accord) were obtained. The rolling window with a re-fitting interval of 100 days was used. From the empirical characteristics of the data under the descriptive analysis section, the return data were found to be heavy-tailed with high peaks, thus the innovations from the standard VaR-GARCH, VaR-NAGARCH, and VaR-GJRGARCH models are modeled with the Jonson’s SU reparametrized distribution (JSU). The skewed generalized error distribution (SGED) is used to model the innovations from the VaR-EGARCH model while the normal inverse Gaussian distribution (NIG) is used to model the innovations from the VaR-TGARCH model.

1) IN-SAMPLE VOLATILITY ESTIMATION

In this section, we discuss the in-sampling performances of the volatility models. The empirical results are reported in Table 5. With the exception of the GJRGARCH model, at least one of the regressors has a significant impact on volatility. For instance, the break variables have significant effects on the volatilities of BWP/ZAR, SEK/ZAR, and BRL/ZAR while proxy 1 has significant effects on the volatilities of ILS/ZAR, and BRL/ZAR. These observations suggest that the levels of uncertainties surrounding the exchange rate markets indeed have significant impacts on the estimated volatility. Furthermore, they suggest that extreme events as captured by the break variables have significant impacts on

the estimated volatility. The aggregated impacts of all the regressors are seen in the improved statistical efficiencies of the estimated models as discussed below.

The log-likelihood estimates from the unrestricted models (augmented) are significantly higher than the values from the restricted versions (non-augmented). Consequently, the absolute AIC values for the unrestricted models are higher than the values for the restricted versions. The results suggest that, the incorporation of the mutual information into the volatility modelling process has the tendency to reduce information leakages in GARCH models. Furthermore, it is observed that, shocks persist more in the non-augmented models than they do in the augmented models. Consequently, it takes a shorter time for half of the shocks in the augmented models to decay than it takes in the non-augmented models. The augmented models also yielded lower estimated unconditional variances. The adjusted R-squares from the Mincer-Zarnowitz regression suggest that the explanatory powers of the augmented models are higher than the non-augmented models. By implication, the consistent lower MAE and RMSE values observed from the augmented models indicate that the models have relatively superior predictive abilities over the non-augmented versions. The volatility persistence results as well as the predictive accuracies are in full support of the results from [43].

2) OUT-OF-SAMPLE VOLATILITY ESTIMATION

Summary statistics of the forecasted volatility components of the VaR models and their performance metrics are reported in Table 6. It is evident from the Table that the augmented models have larger log-likelihoods and these are significant in three of the models. These suggest a possibility that the augmentation may have controlled significant information leakages from the forecasted models. In addition, the volatility persistence of the augmented models is consistently smaller, thus they may be suitable for modeling series that exhibit IGARCH effects. These results are consistent with the in-sample observations. In terms of the RMSE and MAE values, the non-augmented models for SGARCH and GJRGARCH are relatively superior to the augmented versions. Similarly, the augmented models for NAGARCH

TABLE 5. Summary of in-sample estimation.

Metric	NAGARCH	TGARCH	EGARCH	SGARCH	GJRGARCH
Unconditional Var.	<b>2.534E-05</b>	<b>1.901E-05</b>	<b>2.361E-05</b>	<b>4.093E-05</b>	<b>2.765E-05</b>
	2.536E-05	1.902E-05	2.517E-05	4.096E-05	2.765E-05
Half-life	<b>3.0897</b>	<b>5.5910</b>	<b>1.0260</b>	<b>2.0306</b>	<b>7.5656</b>
	9.1355	7.2276	27.804	2.0036	8.9594
Persistence	<b>0.7990</b>	<b>0.8834</b>	<b>0.5089</b>	<b>0.7108</b>	<b>0.9125</b>
	0.9269	0.9086	0.9754	0.7076	0.9256
AIC	<b>-7.8937</b>	<b>-7.6131</b>	<b>-7.8078</b>	<b>-7.3821</b>	<b>-5.9145</b>
	-7.8905	-7.6085	-7.7376	-7.3797	-5.9148
Log Likelihood	<b>5452.8*</b>	<b>5258.4*</b>	<b>5391.6</b>	<b>5103.3*</b>	<b>4093.5*</b>
	5460.7	5253.2	5340.2	5099.6	4091.3
MAE	<b>2.717E-05</b>	<b>5.603E-05</b>	<b>2.708E-05</b>	<b>4.463E-05</b>	<b>33.385</b>
	2.778E-05	5.657E-05	2.878E-05	4.468E-05	33.326
RMSE	<b>5.075E-05</b>	<b>5.241E-04</b>	<b>4.928E-05</b>	<b>8.130E-05</b>	<b>182.83</b>
	5.090E-05	5.302E-04	5.107E-05	8.164E-05	183.65
Adj.MZ-Rsq.	<b>0.0712</b>	<b>0.0377</b>	<b>0.1351</b>	<b>0.0110</b>	<b>0.0698</b>
	0.0626	0.0339	0.0340	0.0090	0.0626

Regressor	NAGARCH	TGARCH	EGARCH	SGARCH	GJRGARCH
Proxy1	0.000522	0.000755*	47.45561*	0.000157	<0.00000
	0.000308	0.019259	14.80736	0.000112	0.001406
Proxy 2			4.488539		
			14.988733		
Break	0.000006*	0.000000	1.098860*	0.000009*	0.000033
	<0.00000	0.019259	0.115273	0.000001	0.000022

Metrics for the augmented models are in bold while values with an asterisk indicate the significance of the likelihood ratio tests at 5%. Adj.MZ-Rsq.is the adjusted R-square values from the Mincer-Zarnowitz regression. Here it is solely used to measure the explanatory powers of the estimated models. The Mincer-Zarnowitz regression models were estimated using ordinary least square with robust standard errors due to the heteroscedasticity of the volatility series.

TABLE 6. Out-of-sample 1-day ahead rolling forecasts estimation.

Metric	NAGARCH	TGARCH	EGARCH	SGARCH	GJRGARCH
LogLik	<b>5311.435*</b>	<b>5222.287*</b>	<b>5351.331*</b>	<b>5014.461</b>	<b>3842.161</b>
	5308.214	5217.694	5287.627	5014.218	3842.031
RMSE	<b>8.209E-05</b>	<b>8.457E-05</b>	<b>4.985E-05</b>	<b>7.530E-05</b>	<b>7.697E-04</b>
	8.224E-05	8.472E-05	4.943E-05	7.474E-05	7.675E-04
MAE	<b>6.338E-03</b>	<b>5.899E-03</b>	<b>5.074E-03</b>	<b>6.204E-03</b>	<b>1.800E-02</b>
	6.342E-03	5.900E-03	5.098E-03	6.147E-03	1.799E-02
Persistence	<b>0.9503</b>	<b>0.7833</b>	<b>0.4690</b>	<b>0.9610</b>	<b>0.9541</b>
	0.9587	0.82691	0.9820	0.9780	0.9548

Metrics for the augmented models are in bold while values with asterisk indicate significance of the likelihood ratio tests at 5%. In computing the persistence parameter for the rolling GJRGARCH and TGARCH models, the distributions were assumed approximately symmetric.

and TGARCH are relatively superior to the non-augmented versions, however, neither versions of the EGARCH model is superior to each other. This is because the augmented version is superior in terms of the MAE while the non-augmented version is superior in terms of the MSE, thus there is no clear superiority of any of the versions. Based on the overall observations, we cannot generalize the superiority nor the inferiority of the augmented models in forecasting volatility. The results are in full agreement with [42] in terms of volatility persistence but in partial contrast to the same study in terms of the accuracy of the forecasts.

### 3) BACKTESTING VALUE-AT-RISK MODELS

Figures 3 to 5 in the Appendix display the plots of realized returns, out-of-sample VaRs predictions, and VaR violations. Inspections of the plots give impressions that there were very few VaR violations in all the models (represented by the red dots) for the 99% quantile but relatively more violations for the 95% quantile. There is pre-indication that all the models may respond early to changing market conditions. This is due to the fact that the evolutions of the VaR estimates seem to be non-clustering, however, albeit useful, the visual inspections do not constitute proper backtest analysis, hence,

TABLE 7. Coverage and independence tests.

99% VaR estimation							
Models	violations (VaR)	Unconditional		Serial Independence		Conditional	
		LRUC	Reject	LRind	Reject	LRCC	Reject
VaR <sub>A</sub> (GJR)	4	0.038	No	0.073	No	0.111	No
VaR <sub>R</sub> (GJR)	4	0.038	No	0.073	No	0.111	No
VaR <sub>A</sub> (NA)	3	0.507	No	0.041	No	0.548	No
VaR <sub>R</sub> (NA)	3	0.507	No	0.041	No	0.548	No
VaR <sub>A</sub> (S)	10	5.292	No	0.466	No	5.758	No
VaR <sub>R</sub> (S)	10	5.292	No	0.466	No	5.758	No
VaR <sub>A</sub> (T)	10	5.292	No	0.466	No	5.758	No
VaR <sub>R</sub> (T)	11	7.059	Yes	0.565	No	7.624	No
VaR <sub>A</sub> (E)	6	0.528	No	0.166	No	0.694	No
VaR <sub>R</sub> (E)	9	3.730	No	0.377	No	4.107	No
95% VaR estimation							
VaR <sub>A</sub> (GJR)	34	5.949	Yes	5.716	Yes	11.665	Yes
VaR <sub>R</sub> (GJR)	34	5.949	Yes	5.716	Yes	11.665	Yes
VaR <sub>A</sub> (NA)	23	0.047	No	0.041	No	0.088	No
VaR <sub>R</sub> (NA)	21	0.049	No	2.110	No	2.159	No
VaR <sub>A</sub> (S)	32	4.222	Yes	1.093	No	5.315	No
VaR <sub>R</sub> (S)	30	2.763	No	0.746	No	3.509	No
VaR <sub>A</sub> (T)	27	1.119	No	0.074	No	1.193	No
VaR <sub>R</sub> (T)	26	0.725	No	0.144	No	0.869	No
VaR <sub>A</sub> (E)	26	0.725	No	3.276	No	4.001	No
VaR <sub>R</sub> (E)	25	0.413	No	3.021	No	3.434	No

Correct conditional correct unconditional coverage is implied by correct conditional coverage but not vice versa

we proceed to discuss the Kupeic’s and the Christoffersen’s tests reported in Table 7. The critical values corresponding to 1% and 5% Chi-Squares with one degree of freedom are 6.635 and 3.841 respectively. Similarly, the corresponding two degrees of freedom critical values are 9.21 and 5.99 respectively. A good VaR estimate must pass both the unconditional coverage and the independence tests, thus our interest is in failing to reject the nulls of these tests. The nulls are rejected when the test statistics are higher than the corresponding critical values.

It is noted from the Table that the test statistics for all the models are below the corresponding critical values except for the VaR (GJR) and VaR<sub>A</sub>(S) models for the 5% VaR estimates, therefore, the nulls for correct exceedances and independence of failures are not rejected for the models with smaller test statistics. These signify that the respective models passed the tests at 1% and 5% confidence levels.

Observations from Table 7 indicate that all the models for the 99% VaR estimates, with the exception of the VaR<sub>B</sub>(T) have LRUC test statistics, which are less than the 6.635 critical value, thus we failed to reject the nulls of correct exceedances of VaR violations from the models. Similarly, we fail to reject the nulls of correct exceedances of VaR violations from the 95% estimated VaR models, with the exception of the VaR<sub>A</sub>(GJR), VaR<sub>B</sub>(GJR) and VaR<sub>A</sub>(S) models. These results suggest that all the models are significantly accurate and acceptable for making risk decisions, however, we cannot categorically make this conclusion. This is due to the fact that, the unconditional test does not account for time-varying volatility, in the sense that it ignores the time losses which

occur, thus, the test may fail to reject a model that produces clustered VaR violations [45]. To address this problem, we use the conditional coverage test. Observations from Table 7 indicate that the nulls of correct exceedances of VaR violations and independence are not rejected for all the models except the 95% VaR(GJR) models. The observed failure rates of all the models except for the 95% VaR(GJR) models, hence, are not significantly different from the corresponding expected rates. We can, therefore, categorically conclude that with the exception of the 95% VaR(GJR) models, the VaR estimates from all the models are significantly accurate and would respond early to changing market conditions, without clustering over time.

#### 4) COMPARING VALUE-AT-RISK MODELS

In comparing the superiority of the estimated VaR models, the MSC procedure alongside with the number of VaR violations, VaR violation failure rates and the ADmean and ADmax of VaR violating returns contemplated in [39] are used. In theory, the number of expected violations with 95% and 99% confidence levels for 1-step ahead forecasts are 22 and 4.4 (5% of 440 and 1% of 440) respectively. Models with violations closer to the expected violations tend to have less failure rates. VaR estimates only an upper bound on the losses that occur with a given frequency; thus, we do not know anything about the sizes of the potential losses, which is of much interest to financial risk practitioners. To address this theoretical drawback of VaR, an alternative risk measure, the expected shortfall (ES), is used. Given an integrable loss  $L$

TABLE 8. Model comparison metrics.

Model	99% VaR and ES estimations								%Voting
	VaR Mean Loss	Mean VaR	Mean ES	MSE (ES)	VaR Rank	Failure rate VaR	AD of VaR violations		
							Maximum	Mean	
VaR <sub>A</sub> (GJR)	0.059416	-0.06035	-0.08227	0.00923	1	0.009	0.034237	0.017118	<b>100</b>
VaR <sub>B</sub> (GJR)	0.059729	-0.06067	-0.08244	0.00926	2	0.009	0.034296	0.017148	00
VaR <sub>A</sub> (NA)	0.019871	-0.01977	-0.02484	0.00077	1	0.007	0.001414	0.000943	<b>100</b>
VaR <sub>B</sub> (NA)	0.020608	-0.02052	-0.02609	0.00084	0	0.007	0.002098	0.001399	00
VaR <sub>A</sub> (S)	0.019770	-0.01949	-0.02613	0.00082	1	0.023	0.010186	0.003008	<b>57</b>
VaR <sub>B</sub> (S)	0.019772	-0.01951	-0.02594	0.00081	2	0.023	0.012205	0.002970	43
VaR <sub>A</sub> (T)	0.017405	-0.01682	-0.02198	0.00065	1	0.023	0.020258	0.008171	<b>88</b>
VaR <sub>B</sub> (T)	0.017642	-0.01707	-0.02214	0.00066	0	0.025	0.022355	0.007944	12
VaR <sub>A</sub> (E)	0.014298	-0.01373	-0.01677	0.00042	1	0.014	0.005688	0.002783	<b>88</b>
VaR <sub>B</sub> (E)	0.014340	-0.01379	-0.01776	0.00040	2	0.020	0.007131	0.002809	12
95% VaR and ES estimations									
VaR <sub>A</sub> (GJR)	0.059419	-0.06035	-0.05080	0.00387	1	0.077	0.034237	0.017118	<b>86</b>
VaR <sub>B</sub> (GJR)	0.059729	-0.06067	-0.01646	0.00389	2	0.077	0.034296	0.017148	14
VaR <sub>A</sub> (NA)	0.018306	-0.01818	-0.01727	0.00038	1	0.052	0.001414	0.000943	<b>88</b>
VaR <sub>B</sub> (NA)	0.020608	-0.02052	-0.01706	0.00041	0	0.048	0.002098	0.001399	12
VaR <sub>A</sub> (S)	0.019770	-0.01949	-0.01694	0.00040	1	0.073	0.010186	0.003008	50
VaR <sub>B</sub> (S)	0.019772	-0.01951	-0.01457	0.00039	2	0.068	0.012205	0.002970	50
VaR <sub>A</sub> (T)	0.017405	-0.01682	-0.01468	0.00033	1	0.061	0.020258	0.008171	<b>88</b>
VaR <sub>B</sub> (T)	0.017642	-0.01707	-0.01224	0.00034	0	0.059	0.022355	0.007944	12
VaR <sub>A</sub> (E)	0.014298	-0.01373	-0.01291	0.00026	1	0.059	0.005688	0.002783	<b>75</b>
VaR <sub>B</sub> (E)	0.014340	-0.01379	-0.08227	0.00024	2	0.057	0.007131	0.002809	25

GJR, NA, S, T and E denotes GJRGARCH, NAGARCH, standard GARCH, TGARCH and EGARCH respectively. Zero rank indicates that a model was eliminated by the MSC procedure at 99% confidence level based on the asymmetric mean loss. When there is a tie, the particular metric is excluded when computing the percentage votes.

with  $E(|L|) < \infty$  having a continuous distribution function  $F_L$  and a confidence level  $\alpha \in (0, 1)$ , the expected shortfall is defined as [45]

$$ES_\alpha = E(L|L \geq VaR_\alpha) = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_u(L) du \quad (25)$$

In addition to the VaR estimates, we report the mean ES values as well as the MSE values for comparison purposes. A model with smaller mean VaR, mean VaR loss, mean ES and MSE of the expected shortfall, less failure rate, superior MSC rank, as well as a minimum ADmax and ADmean, is preferred. It may not be possible for a model to be superior in terms of all the eight metrics, hence, a voting criterion based on these metrics is introduced to help in selecting the model with overall superior ability (i.e. a model with majority of the votes is adjudged superior).

The results of the comparison metrics are reported in Tables 8. A look at the MCS values reveals that, all the augmented models were selected into the superior set of models at both 1% and 5% confidence levels and they are consistently ranked number one. The VaR<sub>B</sub>(NA) and the VaR<sub>B</sub>(T) models were eliminated by the procedure, thus in terms of this test; the corresponding augmented models have superior predictive abilities.

On the average, the augmented models yield lower failure rates in comparison to the referenced models at 1% confidence level but not at 5%, which is in agreement with the

results from [36], hence, the 1% confidence level VaR models may lead to fewer bank failures. In addition, the augmented models tend to have lower MSE values from the expected shortfall estimates and lower mean VaR losses, hence; the augmented models have superior predictive abilities in comparison to the reference models, however there are exceptions to this generalization as seen in the VaR(S) and VaR(E) models. It is further observed that, the mean absolute VaR estimates from the augmented models are consistently lower than the estimates from the reference models. Furthermore, the mean ES for all the augmented models are lower than the values for the corresponding reference models, except for the VaR(S) and VaR(E) models, therefore, on average, it is anticipated that the use of these augmented models may lead to lower bank costs. The augmented models also tend to have lower mean absolute deviations for VaR violating estimates but produce large maximum absolute deviations in some instances.

The overall predictive abilities of the models based on the voting patterns indicate that, the augmented models are relatively superior to the reference models for all the 1% VaR estimates. The same conclusion cannot be made for the 5% VaR estimates. This is because; there is a split of votes among the VaR(S) models. It is worth noting that, although the VaR<sub>A</sub>(GJR) model was decisively adjudged superior to the VaR<sub>B</sub>(GJR), neither can be used in making risk decisions because they all failed the independence and the unconditional coverage tests. In general, the available

evidence suggests superiority of the proposed method in forecasting VaR which agrees with the results from [36], [43], and [44].

It seems on the surface that models with more accurate volatility forecasts do not necessarily yield better VaR forecasts when we compare the superiorities of the volatility models in Table 6 to the superiorities of the VaR models, however, this may not be fully true. The use of squared returns as a proxy for actual volatility tend to over-exaggerate true volatility, thus they lead to inflated MSE with distorted forecasts [4]. VaR forecast evaluation measures are therefore more appropriate in assessing volatility forecast. In this regard, based on the superiority of the augmented VaR models for the 1% VaR estimates we can conclude that our approach of modeling changes and breaks in the unconditional volatilities of exchange rates yielded improved volatility forecast. On the other hand, we cannot make such generalization for the 5% VaR estimates although majority of the augmented models outperformed the non-augmented versions.

5) CAPITAL REQUIREMENT ANALYSIS

Although our approach yielded superior VaR and ES estimates for all models at 1% and the majority at 5%, they are of no practical importance to risk practitioners especially, the banks when they are unable to use the estimates to compute acceptable regulatory capital requirements (as set out by the Basel II Accord). The capital requirements are used to control and monitor market risk exposure of financial institutions. Furthermore, they act as a buffer for adverse market conditions. Overestimation of value-at-risk forces institutions to hold significant amounts of capital and lose opportunity costs, while underestimation overexposes institutions to market risks and losses in their balance sheets that cannot be recovered at crisis periods [15]. This may lead to repercussions on their positions, on the market. Basel II accord allows banks to use their internal models to compute VaR estimates. Reference [39], however, emphasizes that, banks have the responsibility to demonstrate the accuracy of their models sufficiently through backtest analysis based on the number of VaR violations. In addition to this, the Base II accord has instituted penalty zone (see Table 9) to penalize bad models in terms of a multiplicative factor  $k$ , based on the VaR estimates over the last 250 business days. Based on the penalty zones, the capital requirement is defined by [7]

$$Capital\ requirement_t = \max \left\{ -VaR_{t-1}, - (3 + k) \overline{VaR}_{60} \right\} \tag{26}$$

where  $\overline{VaR}_{60}$  is the average VaR over the last 60 business days. In line with the quantitative standards prescribed by the Basel Committee, we focus the backtest analysis on the 99% quantile VaR estimates.

Table 10 reports the mean daily capital requirements (MDCR) for the out-of-sample period. A result that immediately emerges from the Table is that the augmented models consistently produced lower MDCR in comparison

TABLE 9. Penalty thresholds.

Zone	250 business days		440 business days
	violations	Increase in k	violations
Green	0-4	0	0-7
Yellow	5	0.4	8-14
	6	0.5	
	7	0.65	
	8	0.75	
	9	0.85	
Red	≥ 40	1	15 or more

The Basel II penalty zones are based on 250 business-trading days. To obtain an equivalent threshold for 440 days, we use the binomial probabilities associated with the considered true levels of coverage (99%) for a sample size of 444 days. The yellow zone begins at the point such that the probability of obtaining at maximum, number of exceptions equals or exceeds 95%. The starting point of the red zone is the one for which the same probability equals or exceeds 99.99%.

TABLE 10. Mean daily capital requirement.

	VaRGJR)	VaRN	VaRS	VaRT	VaRE
MDCR	0.06014	0.0197	0.019	0.016	0.0136
	0.06047	0.0205	0.019	0.017	0.0137
Zone	Green	Green	Yello	Yello	Green
	Green	Green	Yello	Yello	Yellow

to the non-augmented versions. The superior performance of the augmented models in terms of the MDCR coincides with our previous analysis based on the eight model comparison metrics. Furthermore we can observe that, majority of the augmented models avoided the regulatory penalty zone while a few of them slipped into the yellow zone with the associated penalties; however, the converse of this statement is the case for the non-augmented models. These observations suggest that most of the augmented models would easily (no imposed penalty) pass the scrutiny of regulatory bodies unlike the non-augmented versions. The outcomes in the yellow range are plausible for both accurate and inaccurate models; however, the presumption that a model is inaccurate grows as the number of exceptions increases in the range [7]. In Table 7, the expected violations for the VaR<sub>A</sub>(T) model is smaller than that of VaR<sub>B</sub>(T), the VaR<sub>A</sub>(T) models, thus, would face less hurdles (penalties) in passing through the scrutiny of regulatory bodies, unlike the VaR<sub>B</sub>(T). Both versions of the VaR(S) models, however, will face similar challenges before been certified accurate by the regulatory bodies.

VI. CONCLUSION

There are undesirable consequences associated with inaccurate VaR estimations for banks, thus accurate forecasting of value-at-risk forms an integral part of decision-making and long-term stability of financial institutions. In this paper, 1% and 5% value-at-risk were estimated using univariate GARCH models augmented with exogenous variables. Hypothetical mutual dependencies between pre-estimated volatilities and the exogenous variables were computed to investigate the levels of exchanged information. The Jackknife-biased corrected Kernel density estimation was

used in computing the mutual information while Pearson's method was used in computing the strength of the linear dependencies. In assessing the individual accuracies of the VaR estimates, the conditional coverage and the unconditional coverage tests were used. Competing models, which passed both tests, were then ranked using the model set confidence test procedure. Several other model comparison tools were also employed to assist in selecting the best models. We also conducted a capital requirement analysis to assess the usefulness of the models to banking institutions in computing mean daily capital requirements. In the next four paragraphs, we briefly discuss the main contributions of the study in relation to existing studies.

In studying dependencies among the variables (volatility, returns, and dummy or break variables constructed from returns), to the best of our knowledge, this is the first study to employ the concepts of mutual entropy and the Jackknife-biased corrected KDE approach to data from the rand forex market. Hence, this aspect of the study serves as a contribution to literature on the relationships among the dynamics of exchange rate markets of emerging economies from the perspective of mutual information. The mutual information concepts can be used to re-examine and re-affirm the various established forms of theoretical and phenomenological relationships in literature. In addition, due to the insensitivity of mutual information to the size of data sets, it can be used to study the relationships between financial variables with limited or small data samples

In speculative asset markets, asset returns are related to their volatilities via the risk premium theory [13]. This relationship is the underlying basis for computing risk premium and can be used to study the effects of an asset's volatility on its returns. The reverse of this relationship is used to study the concept of volatility asymmetry. Assets move in tandem, hence, a cross-asset returns-volatility relationship (*the relationship between the volatility of one asset and the returns of another asset*) and its reverse may exist. These relationships may be useful in studying the effects of cross-asset risk on asset returns and asymmetric effects of exogenous returns on volatility respectively. The reverse of cross-asset returns-volatility relationship has been useful in studying spillover of volatility asymmetry across different speculative asset markets [56], as well as the improvement of multivariate volatility forecasts [48]. Cross-asset returns-volatility relationship is a theoretical construct, thus empirical evidence of substantial mutual dependencies between volatility and the exogenous returns support the plausibility of this hypothesis.

The study also contributes to the extant literature on VaR estimation. In its unique contribution, it brings on-board, a simple but a novel approach to account for breaks and changes in the unconditional volatility of GARCH-type models. The approaches used in [2], [43], and [44] to model breaks and time-variations involve relatively complicated procedures which are sometimes not easy to incorporated into other modeling frameworks. However, in comparison to these approaches, our methodology is simple and easy

to incorporate into other volatility frameworks such as the stochastic volatility framework. Furthermore, since the break variables used are exogenous unlike [31], they prevents the compounding of bias, which may be introduced by consecutive endogenous outliers in the parameter estimation. In addition, our approach takes into consideration the actual economic state of volatility in constructing the break variables unlike [31]; hence, periods of crises are modeled differently from periods of increased volatility. Again, due to the superiority of the models built on our approach and the fact that the MCS procedure ranked all models built on our approach number one, our approach, thus, provides alternative or complementary tools, which can be used to mitigate risk in financial institutions comprehensively. It is also useful to individual traders and investors who may not have any standard approach of computing financial risk associated with their daily decision-making.

Finally, the forecasting structures of non-time varying GARCH models such as [14] induce a monotonic mean-reversion path on the long-run forecasts, thus the forecasts converge or revert along a monotonic path, which is inconsistent with the underlying stochastic path. This is because the conditional forecasts converge to a monotonic non-time varying long-run variance. The evidence of significant estimated coefficients of the break and the proxy variables as reported in the bottom section of Table 5 indicates a potential source of uncertainties, which can induce time-variations in the long-run variance of GARCH models so that the conditional forecasts revert or converge to a stochastic time-varying long-run variance consistent with realized volatility.

The main findings of the study include the following:

- The joint distributions of all the paired variables used in the hypothetical study are not bivariate normal and since the individual variables are characterized by heavy-tailed distributions, the bivariate joint distributions may be a mixture of heavy-tailed distributions
- The hypothetical study provides evidence of substantial percentages of exchanged information between the lagged exogenous variables suggesting that there is a better chance of predicting exchange rate volatility with these variables.
- In general, the exogenous break variables tend to exchange relatively higher mutual information with volatility in comparison to the endogenous break variables, thus, suggesting that break variables constructed from exogenous returns have higher likelihoods of volatility predictive abilities. They are, hence, more likely to be adequate in accounting for breaks in the unconditional volatilities of exchange rates.
- Our approach led to a significant reduction of information leakages in the in-sample fitted models and perceived reductions of information leakages in the out-of-sample models. In addition, the approach also yielded less persistent volatilities, reduced half-life, and improved in-sample explanatory powers of the models.

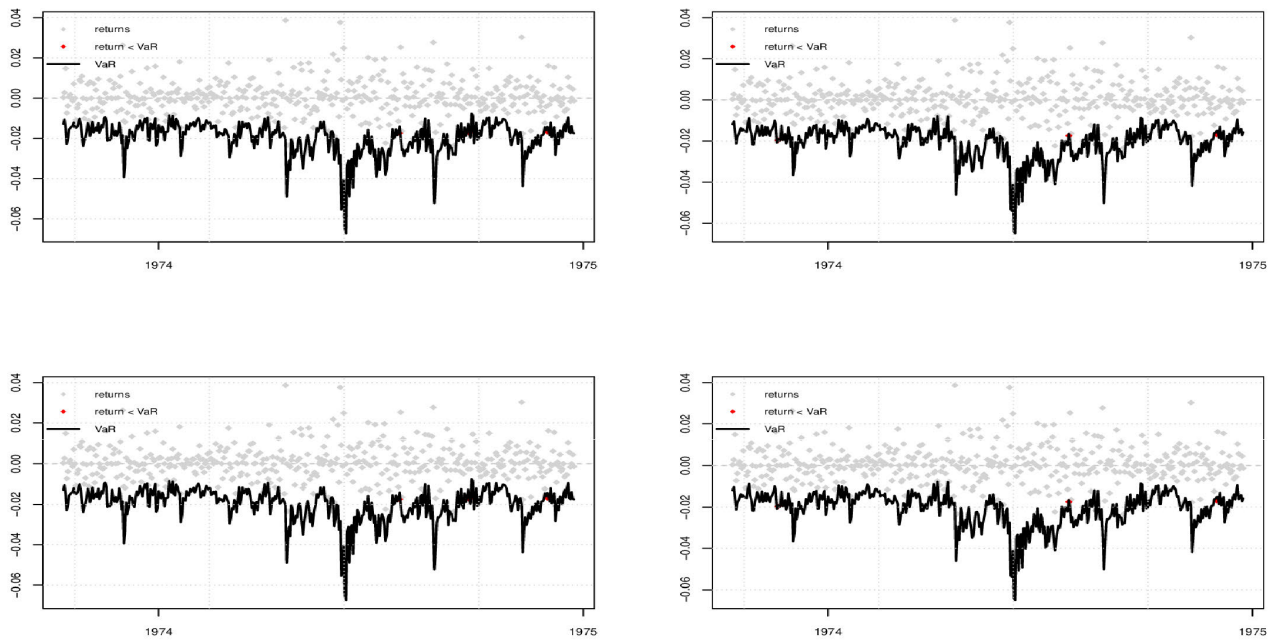


FIGURE 3. VaR plots for EGARCH.

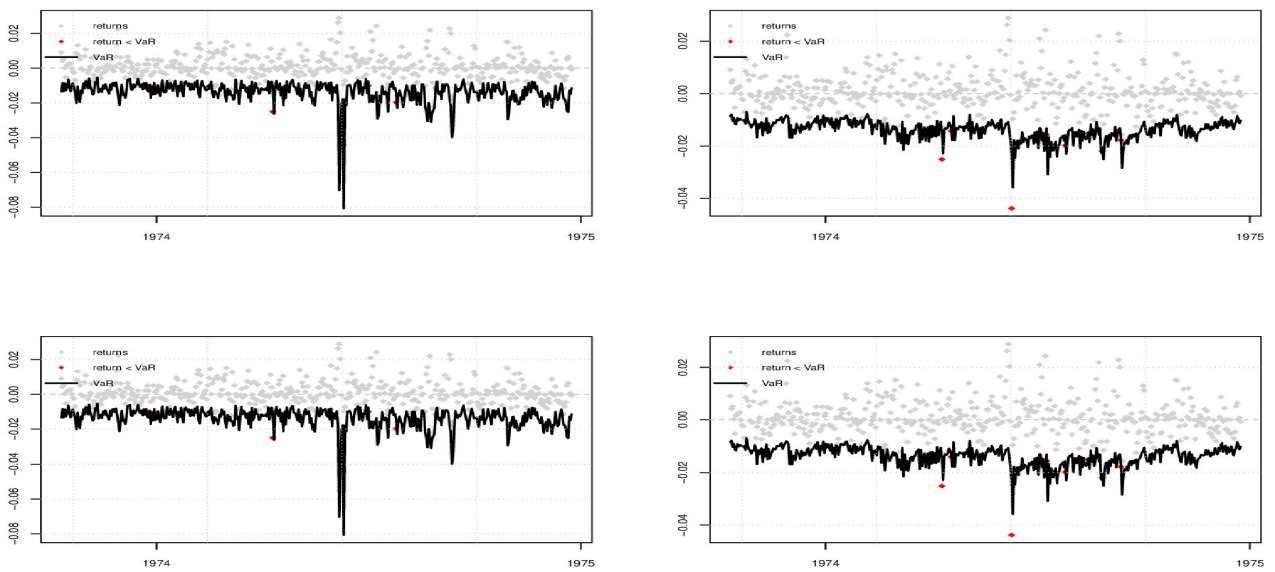


FIGURE 4. VaR plots for NAGARCH.

In addition, there were improvements in the predicted volatilities from all the models; however, the same cannot be said about the forecasted volatilities.

- Our approach yielded better forecasts for all the 1% VaR models and the majority of the 5% VaR models. Accurate volatility is implied by an accurate VaR forecast [15], thus our approach yielded similar superiorities in terms of the volatility forecasts.
- On the usefulness of the VaR estimates in computing daily capital requirements, our approach consistently

produced lower MDCR for all the models. Furthermore, majority of the models built on our approach avoided the regulatory penalty zones while few of them slipped into the yellow zone with relatively less associated penalties.

In conclusion, our approach led to fewer VaR violations, improved 1% value-at-risk forecasts, lower ES forecasts, and optimal daily capital requirements, thus, the models are preferred from regulatory and institutional point of views, because they would lead to optimal bank costs and fewer bank failures. The 5% value-at-risk forecasts for the VaR(NA),



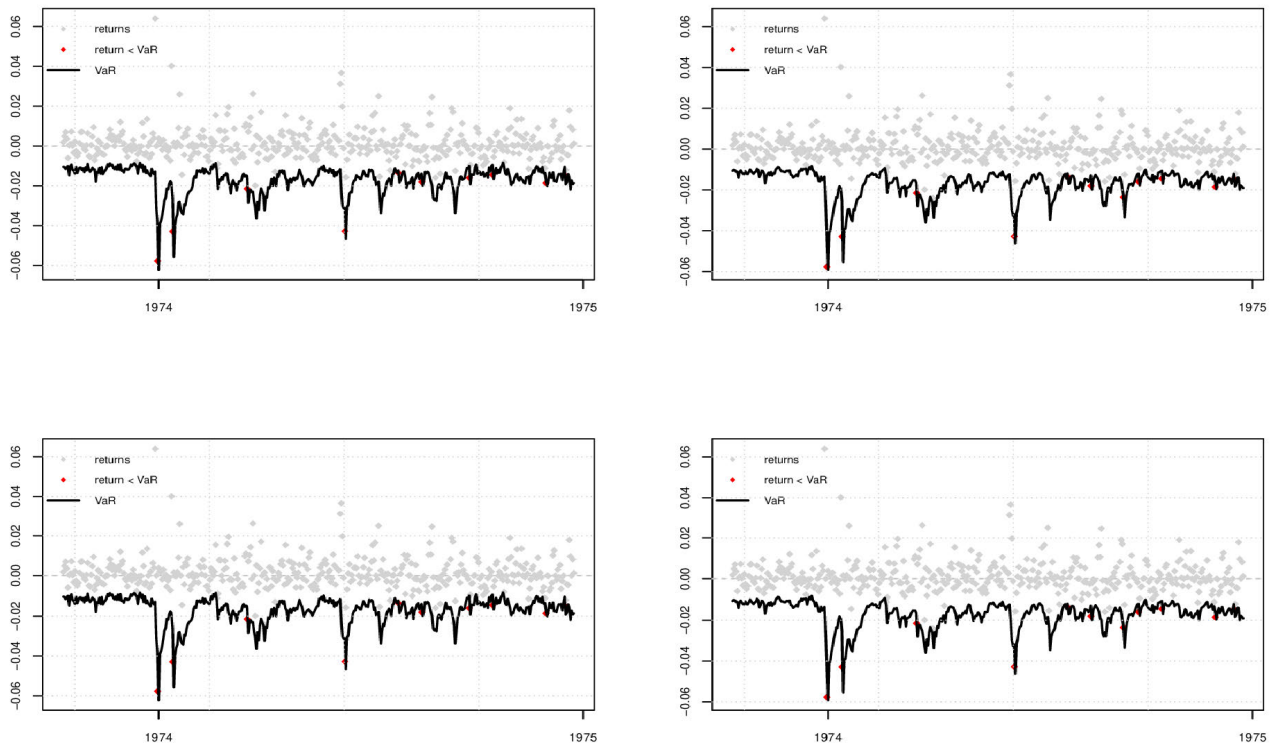


FIGURE 5. VaR plots for TGARCH.

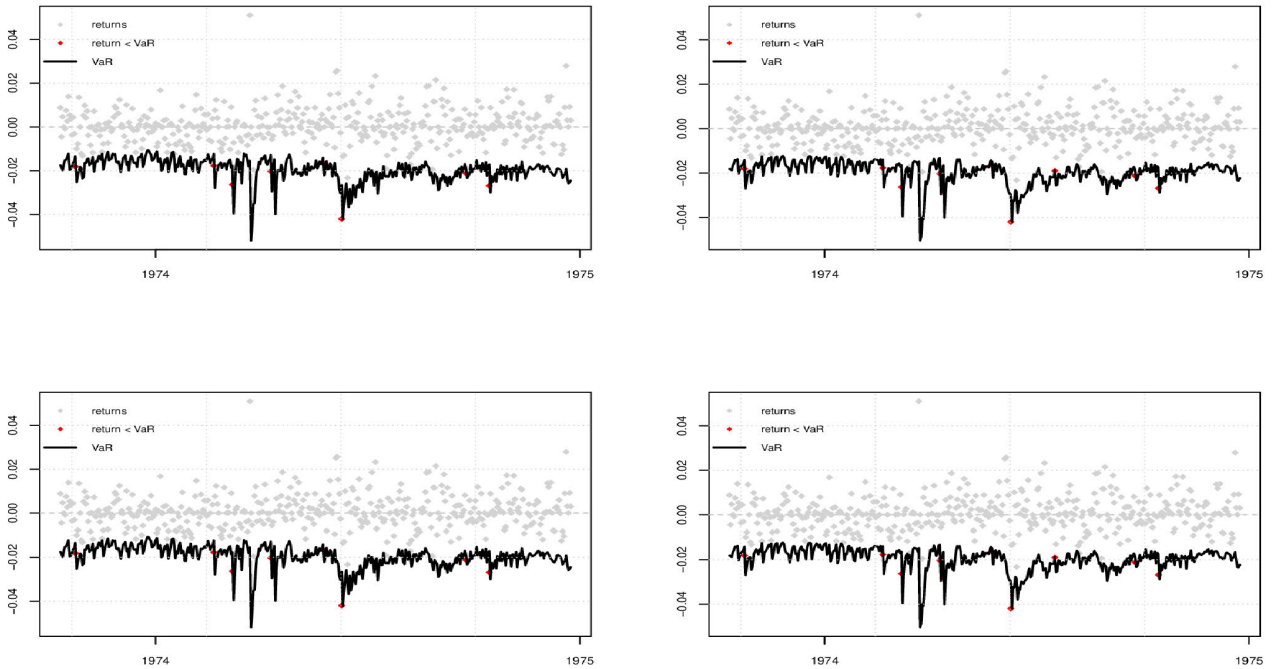
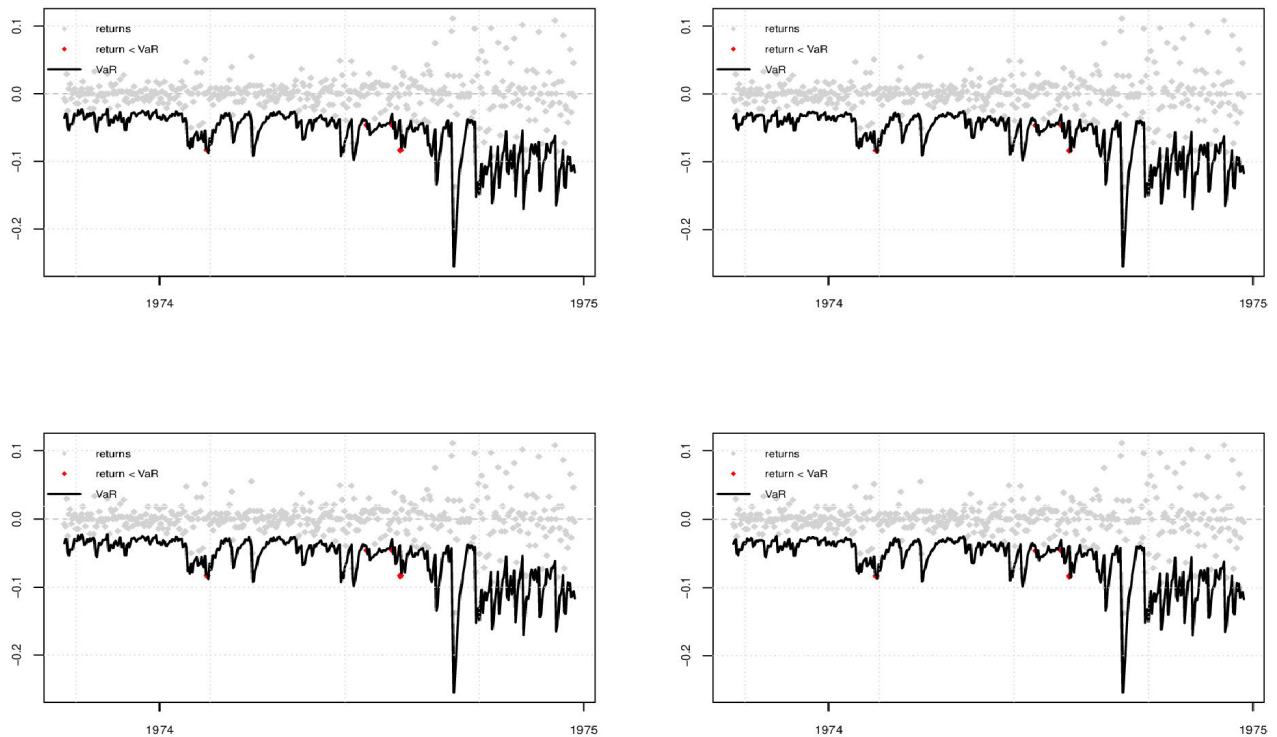


FIGURE 6. VaR plots for SGARCH.

VaR(T) and VaR(E) models, however, may not be preferred from regulatory point of view, although they yielded improved VaR and ES forecasts. This is due to the fact that

the models have relatively higher violations; hence, they may lead to frequent bank failures or severe regulatory penalties. However, from an institutional point of view, they are



**FIGURE 7.** VaR plots for GJR-GARCH.

recommended because of their perceived lower bank costs. It should be noted that our proposed methodology *per se* might not be the cause of the relative higher violations for the 5% VaR estimates. This may be due to the use of inappropriate volatility model specifications and or inadequate market uncertainty proxies and exogenous break variables to estimate the volatility inputs of the VaR model. In a broader sense, the results are in support of studies, which advocate that failure to account for breaks in the unconditional variance lead to sizable upward biases in the degree of persistence in the estimated GARCH models with forecasts that systematically underestimate or overestimate volatility and the subsequent value-at-risk on average, over long horizons [43].

The proposed method depends on substantial levels of mutual dependencies among assets, thus, it is unlikely to yield improved forecasts when there is no substantial evidence of mutual dependencies. In addition, the approach is not parsimonious because it sometimes requires more exogenous covariates and higher-order ARMA-GARCH terms to guarantee optimal parameters that may ensure forecast accuracy. The methodology can be extended to multivariate case, where common exogenous covariates can be incorporated into the volatility processes to improve value-at-risk forecasts. The methodology may also be useful in forecasting value-at-risk for other speculative class of assets, which are known to be mutually dependent. In future, attention would be focussed on applying the methodology to broader exchange rate markets in an attempt to generalize the findings.

## APPENDIX

The first rows of each of the graphs in Appendix 3 to 7 represent 99% VaR models while the second rows are for the 95% VaR models. In each row, the first model represents the augmented model while the last represent the non-augmented versions.

## CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

## ACKNOWLEDGMENT

The authors would like to express their profound gratitude to Amalgamated Bank of South African (ABSA) and the University of Venda for financially supporting the study. They also extend our gratitude to the anonymous reviewers whose constructive criticisms have enriched the contents of this article.

## REFERENCES

- [1] C. Alexander, *Market Risk Analysis: Value-at-Risk Models*, vol. 4. Hoboken, NJ, USA: Wiley, 2008.
- [2] C. Amado and H. Laakkonen, "Modeling time-varying volatility in financial re-turns: Evidence from the bond markets," in *Essays in Nonlinear Time Series Econometrics*. London, U.K.: Oxford Univ. Press, 2014, pp. 139–160.
- [3] C. Amado and T. Teräsvirta, "Modelling changes in the unconditional variance of long stock return series," *J. Empirical Finance*, vol. 25, pp. 15–35, Jan. 2014, doi: [10.1016/j.jempfin.2013.09.003](https://doi.org/10.1016/j.jempfin.2013.09.003).
- [4] T. G. Andersen and T. Bollerslev, "Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies," *J. Finance*, vol. 53, no. 1, pp. 219–265, Feb. 1998.

- [5] N. A. Aridi, C. W. Cheong, and T. S. Hooi, "An estimation of value at risk using GARCH models for the conventional and islamic stock market in Malaysia," *Int. J. Acad. Res. Bus. Social Sci.*, vol. 8, no. 11, pp. 2054–2065, Dec. 2018.
- [6] R. T. Baillie and T. Bollerslev, "Common stochastic trends in a system of exchange rates," *J. Finance*, vol. 44, no. 1, pp. 167–181, Mar. 1989, doi: [10.1111/j.1540-6261.1989.tb02410.x](https://doi.org/10.1111/j.1540-6261.1989.tb02410.x).
- [7] Basel Committee on Banking Supervision. (2011). Revisions to the Basel II Market Risk Framework. Bank for International Settlements. Switzerland. [Online]. Available: <https://www.bis.org/publ/bcbs128.pdf>
- [8] Monetary and Economic Department. (2019). Triennial Central Bank Survey Foreign Exchange Turnover in April 2019. Bank for International Settlements. Switzerland. [Online]. Available: [https://www.bis.org/statistics/rpfx19\\_fx.pdf](https://www.bis.org/statistics/rpfx19_fx.pdf)
- [9] F. Benedetto, L. Mastroeni, G. Quaresima, and P. Vellucci, "Does OVX affect WTI and brent oil spot variance? Evidence from an entropy analysis," *Energy Econ.*, vol. 89, Jun. 2020, Art. no. 104815.
- [10] F. Benedetto, L. Mastroeni, and P. Vellucci, "Modeling the flow of information between financial time-series by an entropy-based approach," *Ann. Oper. Res.*, 2019. [Online]. Available: <https://doi.org/10.1007/s10479-019-03319-7>
- [11] M. Bernardi and L. Catania, "The model confidence set package for R," 2014, *arXiv:1410.8504*. [Online]. Available: <http://arxiv.org/abs/1410.8504>
- [12] M. Bernardi, L. Catania, and L. Petrella, "Are news important to predict large losses?" 2014, *arXiv:1410.6898*. [Online]. Available: <http://arxiv.org/abs/1410.6898>
- [13] F. Black, "Studies of stock price volatility changes," in *Proc. Meeting Bus. Econ. Statist. Sect. Amer. Stat. Assoc.*, 1976, vol. 27, no. 1, pp. 399–418.
- [14] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *J. Econometrics*, vol. 31, no. 1, pp. 307–327, 1986, doi: [10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).
- [15] A. Bucci. (2017). *Forecasting Realized Volatility: A Review*. [Online]. Available: <https://mpr.aub.uni-muenchen.de/83232>
- [16] L. Chen, J. J. Dolado, and J. Gonzalo, "Detecting big structural breaks in large factor models," *J. Econometrics*, vol. 180, no. 1, pp. 30–48, May 2014. [Online]. Available: <http://hdl.handle.net/1814/32939>
- [17] P. F. Christoffersen, "Evaluating interval forecasts," *Int. Econ. Rev.*, vol. 39, no. 4, p. 841, Nov. 1998.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
- [19] R. F. Engle and V. K. Ng, "Measuring and testing the impact of news on volatility," *J. Finance*, vol. 48, no. 5, pp. 1749–1778, Dec. 1993.
- [20] M. Firoj and S. Khanom, "Efficient market hypothesis: Foreign exchange market of Bangladesh," *Int. J. Econ. Financial Issues*, vol. 8, no. 6, pp. 99–103, 2018, doi: [10.32479/ijefi.7097](https://doi.org/10.32479/ijefi.7097).
- [21] C. Francq and J. M. Zakoian, *GARCH Models: Structure, Statistical Inference and Financial Applications*. Hoboken, NJ, USA: Wiley, 2010.
- [22] C. Francq, L. Horváth, and J. M. Zakoian, "Merits and drawbacks of variance targeting in GARCH models," *J. Financial Econometrics*, vol. 9, no. 4, pp. 619–656, Oct. 2011.
- [23] I. M. Gel'fand and A. M. Yaglom, "Calculation of amount of information about a random function contained in another such function," *Amer. Math. Soc. Transl.*, vol. 2, no. 12, pp. 199–246, 1957.
- [24] A. Ghalanos. (2020). *Rugarch: Univariate GARCH Models, R Package Version 1.4-1*. [Online]. Available: <https://cran.rstudio.com/web/packages/rugarch/index.html>
- [25] L. R. Glosten, R. Jagannathan, and D. E. Runkle, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *J. Finance*, vol. 48, no. 5, pp. 1779–1801, Dec. 1993.
- [26] P. R. Hansen, "A test for superior predictive ability," *J. Bus. Econ. Statist.*, vol. 23, no. 1, pp. 365–380, 2005.
- [27] P. R. Hansen and A. Lunde, "A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)?" *J. Appl. Econometrics*, vol. 20, no. 7, pp. 873–889, 2005.
- [28] P. R. Hansen, A. Lunde, and J. M. Nason, "The model confidence set," *Econometrica*, vol. 79, no. 2, pp. 453–497, 2011.
- [29] P. R. Hansen, A. Lunde, and J. M. Nason, "Choosing the best volatility models: The model confidence set approach," *Oxford Bull. Econ. Statist.*, vol. 65, no. 1, pp. 839–861, 2003.
- [30] J. Hull and A. White, "Incorporating volatility updating into the historical simulation method for value-at-risk," *J. Risk*, vol. 1, no. 1, pp. 5–19, Sep. 1998.
- [31] O. Karlsson, "Volatility forecasting under structural breaks," M.S. thesis, Dept. Statist., Uppsala Univ., Uppsala, Sweden, 2016. [Online]. Available: <https://www.diva-portal.org/smash/get/diva2:957411/FULLTEXT01.pdf>
- [32] C. Kosapattarapim. (2013). *Improving Volatility Forecasting of GARCH Models: Applications to Daily Returns of Emerging Stock Market*. [Online]. Available: <https://ro.uow.edu.au/theses/3999/>
- [33] K. Kuester, S. Mittnik, and M. Paoletta, "Value-at-risk prediction: A comparison of alternative strategies," *J. Financial Econometrics*, vol. 4, no. 1, pp. 85–89, 2006.
- [34] M. Kuhl. (2007). *Cointegration in the Foreign Exchange Market and Market Efficiency Since the Introduction of the Euro: Evidence Based on Bivariate Co Integration Analyses*. [Online]. Available: <https://www.econstor.eu/bitstream/10419/31987/1/550537864.pdf>
- [35] P. H. Kupiec, "Techniques for verifying the accuracy of risk measurement models," *J. Derivatives*, vol. 3, no. 2, pp. 73–84, Nov. 1995, doi: [10.3905/jod.1995.407942](https://doi.org/10.3905/jod.1995.407942).
- [36] D. Kumar, "Value-at-risk in the presence of structural breaks using unbiased extreme value volatility estimator," *J. Quant. Econ.*, vol. 18, no. 3, pp. 587–610, Sep. 2020, doi: [10.1007/s40953-020-00197-w](https://doi.org/10.1007/s40953-020-00197-w).
- [37] H. Lee, "The effect of level shift in the unconditional variance on predicting conditional volatility," *J. Econ. Theory Econometrics*, vol. 26, no. 2, pp. 36–56, 2015.
- [38] H. C. Liu and J. C. Hung, "Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 4928–4934, 2010.
- [39] M. McAleer and B. da Veiga, "Single-index and portfolio models for forecasting value-at-risk thresholds," *J. Forecasting*, vol. 27, no. 3, pp. 217–235, 2008, doi: [10.1002/for.1054](https://doi.org/10.1002/for.1054).
- [40] P. O. Junior, A. M. Adam, G. Tweneboah, and K. T. Soo, "Comovement of real exchange rates in the West African Monetary Zone," *Cogent Econ. Finance*, vol. 5, no. 1, Jan. 2017, Art. no. 1351807, doi: [10.1080/23322039.2017.1351807](https://doi.org/10.1080/23322039.2017.1351807).
- [41] D. B. Nelson, "Conditional heteroscedasticity in asset returns: A new approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991.
- [42] C. Pardy, "Mutual information as an exploratory measure for genomic data with discrete and continuous variables," M.S. thesis, Univ. New South Wales, Sydney, NSW, Australia, 2013. [Online]. Available: <http://handle.unsw.edu.au/1959.4/52767>
- [43] D. Rapach, J. Strauss, and M. Wohar, "Forecasting stock return volatility in the presence of structural breaks," *Frontiers Econ. Globalization*, vol. 3, pp. 381–416, 2007, doi: [10.1016/S1574-8715\(07\)00210-2](https://doi.org/10.1016/S1574-8715(07)00210-2).
- [44] L. Reddy, D. Pillai, C.-S. Huang, and C.-K. Huang, "Forecasting stock market volatility in the presence of structural breaks: An application to value-at-risk estimation in South Africa," Tech. Rep., 2017, doi: [10.5281/zenodo.3357652](https://doi.org/10.5281/zenodo.3357652).
- [45] S. Roccioletti, *Backtesting Value at Risk and Expected Shortfall*. Wiesbaden, Germany: Springer, 2015.
- [46] B. C. Ross, "Mutual information between discrete and continuous data sets," in *Proc. Annu. South Afr. Stat. Assoc. Conf.*, Bloemfontein, South Africa, vol. 2017, Dec. 2017, pp. 33–40.
- [47] M. D. Scheuerell. Muti: An R Package for Computing Mutual Information. (Version v1.0.0). Zenodo. [Online]. Available: <https://doi.org/10.5281/zenodo.439391>
- [48] A. A. P. Santos, F. J. Nogales, and E. Ruiz, "Comparing univariate and multivariate models to forecast portfolio value-at-risk," *J. Financial Econometrics*, vol. 11, no. 2, pp. 400–441, Apr. 2013.
- [49] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948, doi: [10.1002/j.1538-7305.1948.tb00917.x](https://doi.org/10.1002/j.1538-7305.1948.tb00917.x).
- [50] H. Singh, N. Misra, V. Hnizdo, A. Fedorowicz, and E. Demchuk, "Nearest neighbor estimates of entropy," *Amer. J. Math. Manage. Sci.*, vol. 23, nos. 3–4, pp. 301–321, Feb. 2003, doi: [10.1080/01966324.2003.10737616](https://doi.org/10.1080/01966324.2003.10737616).
- [51] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, "Entropy and information in neural spike trains," *Phys. Rev. Lett.*, vol. 80, no. 1, pp. 197–200, Jan. 1998, doi: [10.1103/PhysRevLett.80.197](https://doi.org/10.1103/PhysRevLett.80.197).
- [52] J.-B. Su and J.-C. Hung, "The value-at-risk estimate of stock and currency-stock portfolios' returns," *Risks*, vol. 6, no. 4, p. 133, Nov. 2018, doi: [10.3390/risks6040133](https://doi.org/10.3390/risks6040133).
- [53] S. Suess, E. Jondeau, S. Poon, and M. Rockinger, "Financial modeling under non-Gaussian distributions," *Financial Markets Portfolio Manage.*, vol. 22, no. 1, pp. 91–92, 2008.

- [54] R. S. Tsay, *Analysis of Financial Time Series*, 3rd ed. New York, NY, USA: Wiley, 2010.
- [55] H. White, "A reality check for data snooping," *Econometrica*, vol. 68, no. 5, pp. 1097–1126, Sep. 2000.
- [56] L. Yarovaya, J. Brzezczynski, and C. K. M. Lau, "Asymmetry in spillover effects: Evidence for international stock index futures markets," *Int. Rev. Financial Anal.*, vol. 53, pp. 94–111, Oct. 2017, doi: [10.1016/j.irfa.2017.07.007](https://doi.org/10.1016/j.irfa.2017.07.007).
- [57] J.-M. Zakoian, "Threshold heteroskedastic models," *J. Econ. Dyn. Control*, vol. 18, no. 5, pp. 931–955, Sep. 1994.
- [58] X. Zeng, Y. Xia, and H. Tong, "Jackknife approach to the estimation of mutual information," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 40, pp. 9956–9961, Oct. 2018, doi: [10.1073/pnas.1715593115](https://doi.org/10.1073/pnas.1715593115).



**ALBERT ANTWI** received the Diploma degree in mathematics and science education from the University of Cape Coast, Ghana, the B.Sc. degree in mathematics from the Kwame Nkrumah University of Science and Technology, the B.Sc. (Hons.) and M.Sc. degrees in statistics from the University of Venda, South Africa, and the M.Tech. degree in mathematical technology from the Tshwane University of Technology. He is currently pursuing the Ph.D. degree in statistics with the University of Venda. He is also a Lecturer in statistics and operations research with Sol Plaatje University, Kimberley, South Africa. His research interest includes but not limited to statistical applications to finance and economics with emphasis on modeling and forecasting. He is a member of the Local Organizing Committee for the 2020 Second International Multidisciplinary Information Technology and Engineering Conference organized by Sol Plaatje University.

**KWABENA A. KYEI** received the B.Sc. degree in mathematics and statistics from the Kwame Nkrumah University of Science and Technology, the M.Sc. degree in demography from the Universite Catholique de Louvain, LLN Belgium, and the Ph.D. degree in demography from the University of Pretoria, South Africa. As a Statistician/Demographer, he has been involved in research projects in primary health care and population issues. He did his Postdoctoral Research in actuarial/biostatistics from the University of Louisville, Louisville, KY, USA. He has been doing research on econometrics and banking. At the Human Sciences Research Council, he was the Manager of Demographic and Health Survey from 1992 to 1994, a large survey that covered more than 30 000 households countrywide, he is currently an associate editor of some journals and had reviewed articles for some international journals.

**RYAN S. GILL** received the Ph.D. degree in statistics from The University of Texas at Dallas, Richardson, TX, USA, in 2002. He is currently a Professor with the Department of Mathematics, University of Louisville, Louisville, KY, USA, where he has been since 2002. He was an Honorary Fellow with the Department of Mathematics, University of Wisconsin–Madison, in the summer of 2003. His research interests include change point analysis, generalized linear models, differential network analysis, and applications of statistical methods in financial models.

• • •