Austrian Journal of Statistics
2024, Volume 53, 81–98.
http://www.ajs.or.at/
doi:10.17713/ajs.v53i2.1710

Forecasting Time-varying Value–at–Risk and Expected Shortfall Dependence: A Markov-switching Generalized Autoregressive Score Copula Approach

Katleho Makatjane
University of Botswana
Gaborone Botswana

Abstract

The importance of accurately forecasting extreme financial losses and their effects on the institutions involved in a given financial market has been highlighted by recent financial catastrophes. The flexibility with which econometric models can take into account the highly non-linear and asymmetric dependence in financial returns is a critical component of their capacity to forecast extreme events. Therefore, this study aims to forecast time-varying Value-at-Risk and expected shortfall dependence as a predictive density-based regime changes over time. To achieve this, a non-stationary Markov-switching generalized Autoregressive score model nested with copula is estimated using expectation–maximization (EM) algorithm. Extending this non-stationary model is quite challenging, as it requires specifications not only on how the usual parameters change over time but also those with mass distribution components. Dynamics of the estimated autoregressive score allowed the copula parameters to respond rapidly to time-varying key systemic parameters and risk. This is because regime changes are allowed to oscillated between high and low regimes. This is a clear indication of a regime shift in the parameters of an estimated model. Using the minimum score combining, six extreme value distributions are combined to the estimated MS(2)-GAS(1)-copula model and assessed the performance of each combined model 5 days and 30 days forecasting of value-at-risk and expected shortfall. The results of the forecasting performance indicated that the MS(2)-GAS(1)-GPD is the best model to model and forecast Value-at-risk and expected shortfall for the Botswana stock market. This is a promising technique for stochastic modeling of time-varying Value-at-Risk and Expected Shortfall. In addition, a foundation is provided for future researchers to conduct studies on emerging markets. These results are also important for risk managers and investors.

Keywords: data mining, expectation maximization, Markov-switching, stock market, time-varying.

1. Introduction

Recent financial catastrophes have highlighted the necessity of precisely predicting extreme

financial losses and their effects on the institution's financial stability and, more broadly, on the security of broader economy. Major financial crises, like the Global Financial Crisis (GFC) of 2007–2008, the European Sovereign Debt Crisis (ESDC) of 2010–2011, and 2019–2021 COVID-19, typically affect the entire economy, causing severe downturns and recessions. In fact, during severe crisis episodes, banks and other financial institutions frequently fail, and their failures can set off the failure of other non–financial institutions via balance sheet and liquidity channels, endangering the stability of real economy, see for instance Karaś and Szczepaniak (2017). The ability of econometric models to predict such extreme events strongly relies on their flexibility to model highly non–linear and asymmetric dependence structure of financial returns (McNeil, Frey, and Embrechts 2015). The simple linear correlation, despite being widely used, is unable to capture crucial tail behavior of a joint probability distribution. Thus, in a multivariate environment, especially in light of recent crises episodes, modeling the tail dependence and asymmetric dependence between pairs of assets has been becoming increasingly important. Going beyond the multivariate elliptical assumption for a joint distribution of asset returns is typically required when using a measure of dependence other than linear correlation. The copula methodology enables the modeling of a wide range of dependence structures in this regard (Durante and Sempi 2015). The conditional correlation between asset returns rises during times of financial instability as a result of exposure to common shocks that affect all market participants; despite the fact that the modeling challenge of the joint co-movement of stock returns was already present in Bollerslev (1988). Moreover, structural breaks are observed into dependence structure, which is more evident during crisis periods and other infrequent events. As regards dependence breaks, Markov switching (MS) models have been proven to effectively capture non–smooth evolution of volatility and correlations dynamics. Unlike Degiannakis, Dent, and Floros (2014), who used Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity (FIGARCH) framework to forecast Value–at–Risk (VaR) and Expected shortfall (ES), a regime dependent dynamic of copula parameters from a score driven framework is utilized in this study. Parameters of copula dependence are allowed to depend on the realizations of a first order Markovian process with a specific generalized Autoregressive score (GAS) dynamic of Creal, Koopman, and Lucas (2013) and Harvey (2013) in each regime; while retaining an appropriate arbitrary specification conditional distribution dynamics. In this case, the natural volatility is allowed by substituting an appropriate process for the generalized inverse Gaussian (GIG).

The main contribution of this study is the prediction, and classification of regime shifts using MS–GAS Copula because there have been a tremendous number of recent contributions pertaining to the modeling of stochastic parameters. Creal *et al.* (2013) put forth a group of observation-driven time series models and updated the parameters of the model over time using a scaled likelihood function score. This authors' approach offers a streamlined and cogent framework for adding time-varying parameters to a large class of non-linear models. However, the MS—GAS approach used in this study takes real-time prediction systems into account, which increases forecasting accuracy and makes it easier to spot sudden changes, especially when dealing with economic conditions that are constantly changing over time. Most researchers do not consider this factor when modeling regime shifts and time-varying parameters. The remarkable ability of the GAS to filters and approximate complex non-linear data-generating processes in a straightforward and efficient manner is particularly useful in the case of dependency modeling using copulas discussed here. In particular, the use of a result-oriented model helps in situations where it is not clear how to update the dynamics of the parameters of an Archimedean link function. It is important to note that, estimating this path dependent model is a difficult task because an exact likelihood computation is practically impossible. Due to this challenge, estimation techniques are either simplified or did not depend on a likelihood function. In contrast to the study of Bazzi, Blasques, Koopman, and Lucas (2017), this study adopts the expectation maximization (EM) approach to estimate the model; and, the EM captures uncertainty better than maximum likelihood estimation (MLE). However, there is currently no way to obtain the maximum likelihood

estimator without first altering the model. The asymptotic variance—covariance matrix and maximum likelihood estimator of the Markov-switching-GAS model are computed using a novel method that combines the Monte Carlo expectation-maximization (MCEM) algorithm and importance sampling in this study.

Another related contribution of this paper is the implementation and evaluation of time-varying VaR and ES risk measures. The appealing characteristic of the time-varying VaR and ES measures is that, it inherits the flexibility of a dynamic switching copula framework here developed. The copula approach naturally adapts to environments characterized by different kind of upper and lower tail dependence, enabling VaR and ES as an effective measure of extreme co–movements among financial variables. The literature on co–movement risk measures has proliferated during the last few years, see, Bernardi, Gayraud, and Petrella (2015), Sordo, Suárez-Llorens, and Bello (2015) and Castro and Ferrari (2014) among others who have provided an extensive and up to date survey of systemic risk measures that have been proposed.

2. Markov-switching generalized autoregressive score

Let $Y_t = (Y_{1,t}, \cdots, Y_{d,t}) \in \Re$ be a d–dimensional stochastic vector and let $y_{1,t-1} = (y_1, \cdots, y_{t-n})$ be the past history up to time $t-1$ of a non–stationary stochastic process $\{Y_S, S > 0\}$. Assume also that, there exists a first–order ergodic Markov chain $\{S_s, S > 0\}$ defined on a discrete space $\Omega = \{1, \cdots, L\}$ with transition probability matrix $\Pr\{p_{\ell,k}\}$, where $p_{\ell,k} = \Pr(S_t = k \mid S_{t-1} = \ell)$, $\forall \ell, k \in \Omega$ is the probability that state k is visited at time t given that at time $t-1$, in the chain, was in state ℓ and initial vector of probabilities is $\delta = (\delta_1, \cdots, \delta_L)^T$, $\delta_1 = \Pr(S_1 = \ell)$. That is, the probability of being in state $\ell = \{1, \cdots, L\}$ at time 1. Then, let $U_{i,t} = F(Y_{i,t}; \vartheta_i)$ denote the (conditional) probability integral transformation (PIT) of $Y_{i,t}$. The assumption is that

$$U_t \mid (Y_{1,t-1} = y_{1,t-1}), S_t \sim C\left(\bullet k_t^{\ell} \Psi^{\ell}\right) \tag{1}$$

where, $U_t = (U_{1,t}, \cdots, U_{d,t})$ and $C\left(\bullet k_t^{\ell} \Psi^{\ell}\right)$ is the copula distribution that is conditional to regime $S_t = \ell$.

After marginalizing the latent state S_t in model 1, the distribution of U_t is recovered conditionally on the past observations $Y_{1;t-1} = y_{1;t-1}$. Bernardi and Catania (2019) declared that this is a mixture of component-specific predictive cumulative distributions and it is defined as

$$F(U_t \mid Y_{1;t-1} = y_{1;t-1}, S_t) = \sum_{\ell=1}^{L} \pi_{t|t-1}^{(\ell)} C\left(U_t; k_t^{\ell}, \Psi^{\ell}\right), \tag{2}$$

with the following mixing weights

$$\pi_{t|t-1}^{(\ell)} = \sum_{m=1}^{L} P_{m,\ell} \Pr(S_t = m \mid y_{1;t}), \ell = 1, \cdots, L, \tag{3}$$

where, $\Pr(S_t = m \mid y_{1;t})$ is the filtered probability at time t for state m, which is evaluated as

$$\Pr(S_t \mid y_{1;t}) = \frac{h(y_t \mid y_{t-1}, S_t = m)\, \pi_{t|t-1}^{(m)}}{h(y_t \mid y_{1;t-1})} \tag{4}$$

with $h(y_t \mid y_{t-1}, S_t = m)$ being the density of y_t that is conditionally on the Markov chain $S_t = m$ and $h(y_t \mid y_{1;t-1})$ being the density of y_t after having marginalized out S_t. Exploiting the copula structure, $h(y_t \mid y_{1;t-1})$ can be represented as

$$h(y_t \mid y_{1;t-1}) = \prod_{i=1}^{d} f(y_{1;t} \mid y_{i,1;t-1}) \sum_{\ell=1}^{L} \pi_{t|t-1}^{(\ell)} C(u_t \mid y_{1;t-1}, S_t = \ell) \tag{5}$$

where $C(\bullet \mid y_{t-1}, S_t = \ell)$ denotes the copula density that is conditional to $S_t = \ell$ and, it analogously

$$h(y_t \mid y_{t-1}, S_t = m) = \prod_{i=1}^{d} f(y_{i,t} \mid y_{i,1;t-1}) C(u_t \mid y_{1;t-1}, S_t = m), \tag{6}$$

such that $\Pr(S_t = m \mid y_{1;t})$ simplifies to

$$\Pr(S_t = m \mid y_{1;t}) = \frac{C(u_t \mid y_{1;t-1}, S_t = m)\, \pi_{t|t-1}^{(m)}}{CC(u_t \mid y_{1;t-1})}. \tag{7}$$

$\Pr(S_t = m \mid y_{1;t})$ and consequently $\pi_{t|t-1}^{(m)}$ for $m = 1, \cdots, L$ do not depend from the following marginal density $f(y_{i,t} \mid y_{i,1;t-1}, i = 1, \cdots, d)$ but only on the copula density given by

$$C(u_t \mid y_{1;t-1}) = \sum_{\ell=1}^{L} \pi_{t|t-1}^{(\ell)} C(u_t \mid y_{1;t-1}, S_t = \ell), \tag{8}$$

and $C(u_t \mid y_{1;t-1}, S_t = \ell)$ for $\ell = 1, 2, \cdot, L$. It is worth noting that the copula specification is parameterized in terms of k_t^ℓ and Ψ^ℓ for $\ell = 1, 2, 3, \cdots, L$. In Markov-switching generalized Autoregressive score model, k_t^ℓ allows an AR (herein referenced Autoregressive) where the score of the conditional distribution of the data defined in model 5 acts as a forcing variable. Firstly, $\lambda \Re^n \to D_k$ is defined as an absolutely continuous deterministic vector valued function that maps $\Re^n$ into the natural parameter space D_k. A possible choice for λ is the hyper-spherical coordinate transformation introduced by Catania (2021), where a similar scheme is used. The process of k_t^ℓ for $\ell = 1, \cdots, L$ is defined as

$$\begin{aligned} k_t^\ell &= \lambda k_t^{\sim \ell} \\ k_t^{\sim \ell} &= \omega^\ell + \alpha + S_{t-1}^{\ell\ \ell} + \beta k_{t-1}^{\ell \sim \ell}, \\ S_t^{\sim \ell} &= \Gamma^\ell \bigtriangledown_t^{\sim} \left(u_t; k_t^{\sim \ell}, \Psi^\ell\right) \end{aligned},$$

and it is the score of the density defined in model 5 with respect to $k_t^{\sim}$ that is evaluated in u_t, $\bigtriangledown^{\sim}\left(u_t; {k_t}^\ell \Psi^\ell\right)$, but, pre–multiplied by a positive definite scaling matrix, Γ_t^ℓ. Therefore, Creal *et al.* (2013) suggested that

$$\Gamma_t^\ell = \left[I^{\sim}\left(k_t^{\sim \ell} \Psi^\ell\right)\right]^{-\xi} \tag{9}$$

where $I^{\sim}\left(k_t^{\sim \ell} \Psi^\ell\right)$ is the information matrix of $k_t^{\sim \ell}$ and ξ is a scalar that is chosen in $\{0, 1\}$ or $\{2, 1\}$. In this way, if $\xi = 1$, the score is scaled by its variance-covariance matrix. But, if $\xi = \frac{1}{2}$, it is then scaled by its square root; and, if $\xi = 0$, there is no scaling such that $\Gamma_t^\ell = I_n$, where I_n is the identity matrix (Bernardi and Catania 2019). Note that, if the model is correctly specified, $\mathsf{E}_{t-1}[S_t^{\sim}] = 0, \forall t$. That is, $\{S_z^{\sim}, Z > 0\}$ is a martingale difference sequence with $S_t^{\sim} = \left(S_t^{\sim \ell}, \ell = 1, \cdots, L\right)$. Furthermore, if $\xi = \frac{1}{2}$, the additional conditional moment $\mathsf{E}_{t-1}\left(S_t^{\sim} \times S_t^{\sim T} = I_n\right)$ holds. If $\left(I_n - \beta^\ell\right)^{-1}$ exists, then, the third component of process 2 can be written as

$$K_t^{\sim \ell} = \left(I_n - \beta^\ell\right)^{-1} \omega^\ell + \left(I_n - \beta^\ell B\right)^{-1} A S_{t-1}^{\ell \sim \ell}, \tag{10}$$

with B being a back–shift operator such that $B^k x_1 = x_{t-k}, k \in \mathbf{Z}$ which implies that $\mathsf{E}\left(K_t^{\sim \ell}\right) = \left(I_n - \beta^\ell\right)^{-1} \omega^\ell$. Applying chain rule in model 10, the score and the Fisher information matrix of $K_t^{\sim \ell}$ are now recovered by

$$\bigtriangledown^{\sim}\left(u_t; k_t^{\sim\ell}, \Psi_\ell\right) = J\left(k_t^{\sim\ell}\right) \bigtriangledown \left(k_t^{\ell}, \Psi_\ell\right) \tag{11}$$

$$I^{\sim}\left(k_t^{\sim\ell}\Psi^{\ell}\right) = J\left(k_t^{\sim\ell}\right)^{T} I\left(k_t^{\ell}\Psi^{\ell}\right) J\left(K_t^{\sim\ell}\right), \tag{12}$$

where $\bigtriangledown^{\sim}\left(u_t; k_t^{\sim\ell}, \Psi_\ell\right)$ and $I\left(k_t^{\ell}\Psi^{\ell}\right)$ are the score and a Fisher information matrix with respect to the original vector of parameters $k_t^{\sim\ell}$. These latter quantities are then defined as

$$\bigtriangledown\left(u_t k_t^{\sim\ell}\Psi^{\ell}\right) = \frac{\partial \ln h\left(y_t \mid y_{1;t-1}\right)}{\partial k_t^{\ell}} = \frac{\partial \ln C\left(y_t \mid y_{t-1}\right)}{\partial k_t^{\ell}} \tag{13}$$

$$I(k_t^{\ell}\Psi^{\ell})\mathsf{E}_{t-1}\left[\bigtriangledown_t\left(u_t; k_t^{\ell}, \Psi^{\ell}\right) \times \bigtriangledown_t\left(u_t; k_t^{\ell}, \Psi^{\ell}\right)^{T}\right], \tag{14}$$

and

$$J\left(k_t^{\sim\ell}\right) = \frac{\partial \lambda\left(k_t^{\sim\ell}\right)}{\partial k_t^{\sim\ell}} = \begin{pmatrix} \frac{\partial \lambda_1\left(k_1^{\sim}\right)}{\partial k_1^{\sim\ell}} \mid k_{1,t}^{\sim\ell} & \cdots & \frac{\partial \lambda_1\left(k_1^{\sim}\right)}{\partial k_n^{\sim\ell}} \mid k_{n,t}^{\sim\ell} \\ \vdots & \ddots & \vdots \\ \frac{\partial \lambda_1\left(k_1^{\sim}\right)}{\partial k_1^{\sim\ell}} \mid k_{1,t}^{\sim\ell} & \cdots & \frac{\partial \lambda_1\left(k_1^{\sim}\right)}{\partial k_n^{\sim\ell}} \mid k_{n,t}^{\sim\ell} \end{pmatrix} \tag{15}$$

denotes the Jacobian of the mapping function $\lambda(\bullet)$ that is evaluated at $\left(k_t^{\sim\ell}\right) = \left(k_{1,t}^{\sim\ell} \ldots, k_{1,t}^{\sim\ell}\right)^{T}$. An interesting feature of the MS-GAS model nested with copulas, which also differentiates it from the usual MS models with a dynamic state-dependent conditional density, is that, the MS-GAS embeds a natural mechanism which suggests updating the parameters in each state of the chain according to the posterior probability of being in that state; that is, for state ℓ, $\Pr\left(S_t = \ell \mid y_{1;t}\right)$. Indeed, the score $\bigtriangledown\left(u_t; k_t^{\ell}, \Psi^{\ell}\right)$ is then given by

$$\bigtriangledown\left(u_t; k_t^{\ell}, \Psi^{\ell}\right) = \frac{\pi_{t|t-1}^{(\ell)} C\left(u_t \mid y_{1;t-1}, S_t = \ell\right)}{C\left(u_t \mid y_{1;t-1}\right)} \bigtriangledown^{(\ell)}\left(u_t; k_t^{\ell}, \Psi^{\ell}\right) = \Pr\left(S_t = \ell \mid y_{1;t}\right) \bigtriangledown^{(\ell)}\left(u_t; k_t^{\ell}, \Psi^{\ell}\right) \tag{16}$$

where, $\bigtriangledown\left(u_t; k_t^{\ell}, \Psi^{\ell}\right) = \frac{\partial \ln C(u_t \mid y_{1;t-1}, S_t = \ell)}{\partial k_t^{\ell}}$, is a score of the i^{th} regime copula distribution. Rewriting the third portion of process 2 and fixing for simplicity of $\xi = 0$, then

$$k_t^{\sim\ell} = \omega^{\ell} + J\left(k_{t-1}^{\sim\ell}\right) A^{\ell} \Pr\left(S_{t-1} = \ell \mid y_{1;t-1}\right) \bigtriangledown^{(\ell)}\left(u_{t-1}; k_{t-1}^{\ell}, \Psi^{\ell}\right) + \beta k_{t-1}^{\sim\ell}, \tag{17}$$

is obtained. It is now evident how the information provided by u_{t-1} is shared across the Markovian regimes according to the following posterior probabilities $\Pr\left(S_{t-1} = \ell \mid y_{1;t-1}\right)$ for $\ell = 1, 2, 3, \cdots, L$. In this way, if a state has not been visited by the circuit, the dynamic system will not automatically update the parameters until the state is revisited, that is, until the relevant information about the state is updated. Catania (2021) and Bazzi *et al.* (2017) made similar arguments in related contexts where the conditional distribution of the data is still a mixture of distributions.

2.1. A novel approach to estimate MS—GAS model

A two-step procedure described in the work of Maaziz and Kharfouchi (2018) is followed for estimating the parameters of a model. The first step consists of estimating parameters that include the following marginals $(\vartheta_1, \cdots, \vartheta_n)$, followed by a second step where the parameters governing the evolution of the dependence $\left(\left\{\omega^{\ell}, A^{\ell}, \beta^{\ell}, \Psi^{\ell}\right\}_{\ell=1}^{L}, Q\right)$ are jointly estimated along with the latent Markovian states $\left(\{S_t = \ell, t > 0\}_{\ell=1}^{L}\right)$. Nicolas (2022) has demonstrated

the asymptotical consistency of the proposed two-step procedure of Blazsek, Escribano, and Licht (2020) for conditional copulas. The EM algorithm of Figueroa-Zúñiga, Niklitschek-Soto, Leiva, and Liu (2022) is modified in this study to estimate the dependence parameters subject to both the GAS dynamic and the Markovian structure. Assuming that parameters of the marginal distribution ϑ_i have been consistently estimated by maximum likelihood (see, for example Asheri, Hosseini, and Araabi (2021)), then, the EM algorithm is presented in this study by replacing the pseudo-observation u_t by $\hat{u_t} = \left(F\left(y_{1,t};\hat{\vartheta}_1\right),\cdots,F\left(y_{d,t};\hat{\vartheta}_d\right)\right)^T$ where, $\hat{\vartheta}$ is the maximum likelihood estimate of $\vartheta_1 \forall i = 1,2,3,\cdots,d$. To apply an EM algorithm, the vector of observations $y_{1;T}$ where T is the sample size, is regarded as being incomplete. Following the implementation described in work of Şimşek and Topaloglu (2018), two missing data structures are now introduced. Let $Z_t = (Z_{t,1},\cdots,Z_{t,L})$ and $ZZ_t = (ZZ_{t,1,1},\cdots,ZZ_{t,L,L})^T$ to be defined as

$$Z_{t,\ell}\begin{cases} 1 & \text{if } S_t = \ell; \\ 0 & otherwise. \end{cases}$$

and

$$ZZ_{t,\ell,k}\begin{cases} 1 & \text{if } S_{t-1} = \ell, S_t = k; \\ 0 & otherwise. \end{cases}$$

The class membership is unknown and conveniently treated as the value taken by a latent multinomial variable with one trial and L classes, where the temporal evolution of class membership is driven by a hidden Markov chain S_t for $t = 1,\cdots,T$. This is similar to latent class approaches. Augmenting the observations $y_{1;T}$ with the following latent variables $Z_t, ZZ_t, t = 1,\cdots,T$, this replaces the log-likelihood function with a complete–data log–likelihood which now becomes

$$\begin{aligned} &\log L(\Xi) = \textstyle\sum_{\ell=1}^{L} Z'_{1,\ell}\log(\delta_1) + \sum_{\ell=1}^{L}\sum_{k=1}^{L}\sum_{t=1}^{T} ZZ_{t,\ell,k}\log(q_\ell,k) \\ &+ \textstyle\sum_{\ell=1}^{L}\sum_{t=1}^{T} Z_{t,\ell}\log C\left(\hat{u_t};\Xi^\ell\right) + \sum_{\ell=1}^{L}\sum_{t=1}^{T} Z_{t,\ell}\log C\left(\hat{u_t};\Xi^\ell\right) \\ &+ \textstyle\sum_{t=1}^{T}\sum_{\ell=1}^{d}\log f_i\left(y_{i;t};\hat{\vartheta}_i\right) \end{aligned}$$

where, $\Xi = \left\{\Xi^\ell\right\}_{\ell=1}^{L}$ with $\Xi^\ell = \left(\omega^{\ell'},\left(\overrightarrow{A}^\ell\right)',\left(\overrightarrow{\beta}^\ell\right),\Psi^{\ell'}\right)'$ being a vector containing parameters of the GAS dynamics for the copula dependence parameters k_t^ℓ and Ψ^ℓ for $\ell = 1,\cdots,L$. According to Ahmad and Bladt (2022), the EM algorithm alternates between carrying out a desired (E) step, which develops a function for the desire of the log-likelihood evaluated using the current assessment for the parameters, and an intensification (M) step, which processes parameters by boosting the expected log-likelihood found in the E–step. The distribution of the latent variables in the accompanying E–step is then selected using these parameter-estimates. On the $(m+1)^{th}$ iteration, the EM algorithm proceeds as follows:

Algorithm 1: EM algorithm for Markov-switching GAS Nested Copula model

Input: Observations $\{\hat{u}_0,\cdots,\hat{u}_T\}$, number of regimes K;
Output: Parameter estimates $\hat{Q}$
Then;
E–step: $Q^{(t+1)} = \arg\max_Q F\left(Q,\theta^t\right)$ given the observed data and the current parameters. Then, estimate

$$Q\left(\Xi\Xi^{(m)}\right) = EP^{(Z_t|y_{1;T}\Xi m)}\left[\log L(\Xi);y_{1;T}\right]. \tag{18}$$

M–Step: chose $\Xi^{(m+1)}$ such that the preceding expected values are maximized with respect to Ξ to obtain

$$\Xi^{(m+1)} = \arg\max_{\Xi} Q\left(\Xi,\Xi^{(m)}\right). \tag{19}$$

The E-step requires the computation of the so called the Q–function which calculates the conditional expectation of the complete-data log-likelihood given the observations and the current estimate of the parameter vector $\Xi^{(m)}$. Therefore, the following representation of the function Q is obtained

$$Q\left[\Xi, \Xi^{(m)}\right] = EP^{(Z_t|y_{1;T}\Xi^m)} \propto \sum_{\ell=1}^{L} \hat{Z}_{1,\ell} \log(\delta) + \sum_{\ell=1}^{L}\sum_{k=1}^{L}\sum_{t=1}^{T} \hat{Z}\hat{Z}_{t,\ell,k} \log(q_{\ell,k}) + \sum_{\ell=1}^{L}\sum_{t=1}^{T} Z_{t,\ell} \log C\left(\hat{u}_t; \Xi^{\ell}\right), \tag{20}$$

$\hat{Z}_{t,\ell} = \Pr\left(S_t = \ell \mid y_{1;T}, \Xi^{(m)}\right)$, $\widehat{ZZ}_{t,\ell,k} = \Pr\left(S_t = \ell, S_t = k \mid y_{1:T}\Xi^{(m)}\right)$ for $\ell, k = 1, \cdots, L$ and $\forall t = 1, \cdots, T$ denote the current smoothed probabilities of the states evaluated using the well-known forward-filtering backward-smoothing (FFBS) algorithm. The M–step in model 19 maximizes the function $Q\left(\Xi, \Xi^{(m)}\right)$ with respect to Ξ to determine the next set of parameters $\Xi^{(m+1)}$. The updated estimates of the hidden Markov model (HMM); that is, the elements of the transition probability matrix of the hidden Markov chain $\mathbb{Q}$ are available in closed form as

$$\hat{\delta}_{\ell}^{(m+1)} = \hat{Z}_{1,\ell}$$

$$\hat{q}_{\ell,k}^{(m+1)} = \frac{\sum_{t=2}^{T} \widehat{ZZ}_{t,\ell,k}}{\sum_{\ell=1}^{L}\sum_{t=1}^{T} \widehat{ZZ}_{t,\ell,k}},$$

for $\ell = 1, \ldots, L$ while the parameters Ξ can be obtained as the function of the following optimization problem

$$\Xi^{(m+1)} = \arg\max \sum_{\ell=1}^{L}\sum_{t=1}^{T} \hat{Z}_{t,\ell} \log C\left(\hat{u}_t; \Xi^{\ell}\right). \tag{21}$$

Since the last optimization step provides parameter updates that increase the log-likelihood function, convergence of the algorithm with the ML estimator is guaranteed. After running the direct maximum likelihood estimator of the hidden Markov model, the MS—GAS dependence and standard errors of the marginal parameters can be estimated numerically. However, this procedure does not take into account the estimation errors of the boundary parameters. Therefore, the fixed bootstrap procedure of Shang (2018) can be used as an alternative to the above procedure. In the empirical application, this study take the former approach.

2.2. Time–varying risk measures

Risk measures are statistical measures that are historical predictors of investment risk and volatility, and they are also major components in modern portfolio theory (MPT). The MPT is a standard financial and academic methodology for assessing the performance of a stock or a stock fund as compared to its benchmark index. The economic relevance of this topic follows partly from the Basel Accords (currently the Basel III Accords), which impose that banks and financial institutions have to meet capital requirements, and must rely on state-of-the-art risk systems. In particular, they must assess uncertainty about future values of their portfolios and estimate the extent and the likelihood of potential losses using a risk measure. Nowadays, VaR and ES risk measures are the standards (Ardia, Boudt, and Catania 2018).

The estimation of the VaR and ES thus requires first to accurately estimate the conditional distribution of future portfolios' or assets' returns. Formally, assuming a continuous cumulative density function (cdf) with time-varying parameters $\theta_t \in \Re^d$ d and additional static parameters $\psi \in \Re^q$, $F(\cdot; \theta_t, \psi)$, for the logarithmic returns at time t, $r_t \in \Re$, $VaR_t(\alpha)$ is given by

$$VaR_t(\alpha) \equiv F^{-1}(\alpha; \theta_t, \psi), \tag{22}$$

where, $F^{-1}(\cdot)$ denotes the inverse of the cdf, that is, the quantile function. It follows that $VaR_t(\alpha)$ is nothing more than the $\alpha-$quantile of the return distribution at time t.[1] The ES metric measures the expected loss after a violation of VaR level and it is defined as

$$ES_t(\alpha) = \frac{1}{\alpha}\int_{\alpha}^{VaR_t} z dF(z, \theta_t, \psi). \tag{23}$$

It follows that a crucial point for a correct VaR and ES assessment is the determination of $F(\cdot)$ and its parameters θ_t and ψ. For a general overview of existing methods, the reader is referred to Nieto and Ruiz (2016). In this study, this is illustrated on how it can be achieved using the framework of Markov-switching Generalized Autoregressive Score model introduced by Bernardi and Catania (2019) and Blazsek, Ho, and Liu (2018). The GAS models are also referred to as score driven (SD) and dynamic conditional score (DCS) models and have been used extensively for financial risk management purposes. For a comparison between the accuracy of VaR and ES estimates obtained by the GAS approach against alternative volatility models, see Ardia *et al.* (2018), Gao and Zhou (2016) and Bernardi and Catania (2016) among others.

2.3. Evaluating downside risk forecasts and backtesting risk measures

The recursive prediction method is usually used to backtest the appropriateness of a statistical model, and to perform model comparisons in terms of VaR and ES predictions (Ardia *et al.* 2018). The purpose of a backtesting analysis is to verify the accuracy of the prediction by separating the estimation window and the evaluation period. The goal of a model comparison analysis is usually to sort the models according to a loss function. For this purpose, the full sample of returns T is split into an in-sample period of length S and an out-of-sample period of length H. First, parameters of the model are estimated over the in-sample period, then the h–step forward forecast of the distribution of returns at time $S + h$ is generated along with the associated VaR and ES measures. These steps are repeated, recursively expanding the period in the sample with new observations until the end of the series T is reached. If previous observations are eliminated during the data amplification phase, a moving window is considered; otherwise an expanding window is considered. In this study, a standard approach is followed in the day–to–day risk management of a typical trading desk and use the floating window setup with $h = 1$.

Once a series of VaR predictions is available, forecasts adequacy is assecssed through backtesting procedures. VaR backtesting procedures usually check the correct coverage of the unconditional and conditional left-tail of the log-returns distribution. Correct unconditional coverage (UC) was first considered by Kupiec (1995), while correct conditional coverage (CC) by Christoffersen (1998). The main difference between UC and CC concerns the distribution that one focuses onto. For instance, UC considers correct coverage of the left-tail of the unconditional log-return distribution $f(r_t)$ while CC deals with the conditional density $f(r_t \mid \tau_{t-1})$. From an inferential perspective, UC looks at a ratio between the number of realized VaR violations that are observed from the data and expected number of VaR violations implied by the chosen risk level α during the forecast period; that is, αH. To investigate CC, Christoffersen (1998) proposed a test on the series of VaR exceedance $\{d_t, t = 1, \cdots, s + H\}$ where $d_t \equiv \{r_t < VaR_t(\alpha)\}$ or $d_t \equiv \{r_t < ES_t(\alpha)\}$, which is usually referred to as the hitting series. Specifically, if a correct conditional coverage is achieved by the model, VaR exceedances should be independently distributed over time. For more readings on backtesting VaR and ES, the reader is referred to Bayer and Dimitriadis (2022).

Nevertheless, Escanciano and Olmo (2010) showed that the use of standard unconditional and independence backtesting procedures can be misleading, because they do not take into account uncertainty associated with parameter estimation. They quantify this risk in a very

[1] Sometimes VaR is defined with respect to the loss variable $\ell_t = -r_t$. All the arguments of this paper can be easily adapted to this case.

general class of dynamic parametric VaR models and propose a correction of the standard backtest(s) that takes it into account. This authors showed that one of the main determinants of the corrected asymptotic variance is the forecasting scheme used to generate VaR forecasts, i.e., whether one uses recursive, rolling, or fixed parameter estimates. The backtesting methodologies described above focus only on the number of VaR and ES exceptions, and totally disregard their magnitudes. Hence, Nieto and Ruiz (2016) criticizes the statistics proposed by Christoffersen (1998) because they are two-tailed, and, as a consequence, can reject a risk model for being over conservative. As it was mentioned above, risk models can also be rejected for being over conservative because this is not desirable for financial institutions. Alternatively, the tail risk statistic is proposed and is defined as

$$TR = -\frac{1}{T}\sum_{i=1}^{T}\left(R_t - \alpha\right) I\left(R_t - \alpha\right). \tag{24}$$

The TR statistic in model 24 tells risk managers the size of the aggregate tail losses that a portfolio may incur over the period considered. The asymptotic distribution of the TR statistic is derived under the assumption of normal returns. The Dynamic Quantile (DQ) test by Engle and Manganelli (2004) assesses the joint hypothesis that $\mathsf{E}\left[d_t\right] = \alpha$ and the hit variables are independently distributed. The implementation of the test involves the de-meaned process $Hit_t^{\alpha} \equiv d_t - \alpha$. Under correct model specification, unconditionally and conditionally Hit_t^{α} has zero mean and is serially uncorrelated. The DQ test is then the traditional Wald test of the joint nullity of all coefficients in the following linear regressions

$$Hit_t^{\alpha} = \begin{cases} \delta_0 + \sum_{i=1}^{L}\delta_1 Hit_{t-1}^{\alpha} + \delta_{L+1} VaR_{t-1}\left(\alpha\right) + \varepsilon_t \\ \delta_0 + \sum_{i=1}^{L}\delta_1 Hit_{t-1}^{\alpha} + \delta_{L+1} ES_{t-1}\left(\alpha\right) + \varepsilon_t. \end{cases}$$

Under the null hypothesis of correct unconditional and conditional coverage, the Wald test statistic is asymptotically a chi-square distributed with $L+2$ degrees of freedom. Engle and Manganelli (2004) set $L = 4$ lags, which has become the standard choice.

With the Kupiec likelihood ratio test, the assumption is that the frequency follows the following binomial theorem

$$\Pr(x) = \binom{n}{x} \cdot p^x q^{n-x}.$$

Hence, the Kupiec LR statistic is given by

$$LR_{UC} = 2\ln\left[\frac{\left(1-p\right)^{n-x} p^x}{\left(\left(1-\frac{S_n}{n}\right)\right)^{n-x}\left(\frac{S_n}{n}\right)^x}\right] \sim \chi_1^2 \tag{25}$$

where $\frac{S_n}{n}$ is the rate of violation observed under the null hypothesis. Reject the null hypothesis if the calculated probability value is less than the observed significance level at 1%, 5% and 10% and conclude that the model is not correct, implying that the risk computed from the risk model either both the VaR and ES is unreliable. This means it can give false signals to risk managers.

The Christoffersen test is defined as a new backtesting tool that is based on the length of time between VaR violations. The main insight is that where the VaR model is properly indicated for coverage, p, the expected duration of infringements should be a constant $1/\pi$ days. The fact that a test aims at establishing the dependence of violation, the notation η_{ij} is used to present the number of days where the j^{th} condition occurred given that the i^{th} condition previously occurred. The Christoffersen test statistic can be represented by

$$LR_{CC} = -2\ln\left[\frac{\left(1-\pi\right)\eta_{11} + \eta_{12}\pi\eta_{21} + \eta_{22}}{\left(1-\pi_1\right)\eta_{11}\pi_1\eta_{21}\left(1-\pi_2\right)\eta_{12}\pi_2\eta_{22}}\right] \tag{26}$$

where, π_{ij} presents a violation probability that occurs conditionally on state i at time $t-1$. Reject the null hypothesis if the calculated probability value is less than the observed 1%, 5% and 10% significance level and conclude that the estimated risk measure is not a good measure for a specified risk.

The Basel committee on banking supervision required internal banking or financial market backtesting procedures in 2007. Financial institutions need to regulate their capital requirements and this is done by backtesting in order to cover market risk due to their trading activities. The capital requirement size by Mwamba, Hammoudeh, and Gupta (2014) is computed as

$$Cap = Fact \times \max\left[VaR_t(0.01), \frac{1}{60}\sum_{i=0}^{59} VaR_t(0.01)\right], \tag{27}$$

where Fact is the multiplication factor reported in Table 1. In other words, the required capital is equal to the multiplication factor that is multiplied by the highest value between today's 99% VaR and the mean of the last 60 days' 99% VaR. Subsequently, a good market risk model will produce four or fewer exceptions during a trading-day period. Should the number of exceptions fall into the green zone, as shown in Table1, that is an indication that the market risk model is probably good. If it falls into the yellow zone, there is uncertainty about the correctness of a market risk model.

Table 1: The Basel III zones and the exceptions

Zone	Confidence Level			Multiplicative Factor
	90%	95%	99%	
Green	0-32	0-17	0-4	3
Yellow	33-43	18-26	5-9	3.40, 3.50, 3.65, 3.75 or 3.85
Red	≥ 44	≥ 27	≥ 10	4

Source: Best (2021)

3. Results and data analysis

A five-day financial time series exchange Botswana stock exchange/Domestic Company index (FTSE/BSE-DCI) for the period of January 2010 to June 2023 is fitted with the $MS(k) - GAS(p)$ in this empirical analysis. The use of a five day(s) frequency is based on the fact that, on weekends and holidays, stock markets are closed. Therefore, Botswana stock exchange is not an exception. No trading on these days. Hence the recorded market prices are from Monday(s)–Friday(s). To avoid any exchange rates fluctuations, Makatjane and Moroke (2021) suggested that the index should be kept in its original currency; hence, the FTSE/BSE-DIC used in this study is kept in its BWP (i.e Botswana Pula) currency. As shown in Table 2, the distribution of FTSE/BSE-DCI returns is asymmetric and has fat tails. This suggests that the $MS(k) - GAS(p)$ model follows a skewed distribution in terms of the results of skewness; while normality tests used all rejects the null hypothesis of normality concluding that the returns on FTSE/BSE-DCI is not normally distributed. By assuming asymmetry for the regime differences in this situation, the $MS(k) - GAS(p)$ model would perform more accurately in predictions than the conventional generalized Autoregressive conditional Heteroscedasticity (GARCH) or GAS models.

3.1. Markov-switching generalized autoregressive score framework

To begin the main analysis, the $Ms(k) - GAS(p)$ is fitted to FTSE/BSE-DCI returns. It is assumed that every regime shift in the parameters is valid and that, during the sample period of the study, an model would be correct 95% to 99% of the time; otherwise, regime

Table 2: Descriptive statistics

	Skewness	Kurtosis	J-B	S-W	K-S
FTSE/BSE–DCI	-1.399	17.065	32930.9422 $***$	0.89808 $***$	0.4846 $***$

Codes:$***0.001, **0.01, *0.05, N$

classifiers would be useless. To address unit root and number of regime shift in the non–linear time series, a modified Wald test of Güriş (2019) is applied on the basis of non–linear unit root test by Kapetanios, Shin, and Shell and a likelihood ratio test (LR) of Kasahara and Shimotsu (2018) is also utilized for testing number of regime shifts. The FTSE/BSE-DCI does not contain any non–linear unit roots, therefore, a stationary non–linear time series is achieved. This is in accordance with the results of proposed modified Wald test reported in Table 3. In addition, the null hypothesis of 1 regime shift is rejected and that of $k > 2$ is also rejected concluded that there is at least 2 regime shifts in BSE-DCI index.

Table 3: Modified Wald and likelihood ratio tests

	Wald statistic	df	p-value	Likelihood ratio statistic	df	p-value
FTSE/BSE-DCI	16.341	2	0.001	22.002	2	0.003

Moreover, regime dependence is for the following parameters $\{\alpha, \beta\}$ and $\{\gamma\}$ to derive a Markov–Switching behavior of dependence structure. The results strengthen and provided strong evidence of regime dependence. At both extremes, there is corroboration of state dependence. Because these parameters are significant, as reported in Table 4, this is a clear indication of Markovian structure where returns on FTSE/BSE-DCI display significant switches in both the unconditional mean and persistence of the dependence parameter. $\hat{\delta}$ represent a posterior unconditional volatility mean that has a 96.8% acceptance rate. This implies that the mean is clustered around the rate at which FTSE/BSE-DCI regime switches from low to high regime. This resulted in stable probabilities of 23% in regime 1 and 77% in regime 2. The location parameter in unconditional parameters, which denotes an asymmetric and stable distribution, further supports this. The volatility of a study series is therefore said to be centered on the mean.

Table 4: Parameter estimates

Coefficient	Regime 1	Regime 2
$\hat{\mu}$	0.016(0.025) $***$	0.003(0.0016)$***$
$\hat{\alpha}$	0.341(0.004)$***$	0.235(0.0155)$***$
$\hat{\beta}$	0.9432(0.055)$***$	0.987(0.008)$***$
$\hat{\gamma}$	0.623(0.011)$***$	0.112(0.011)$***$
$\hat{\delta}$	2.254(0.034)$***$	4.3334(0.035)$***$
	Unconditional Parameter Estimates	
$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
-2.0036	0.372	9.236
	Markov Probabilities	
P_{11}		P_{22}
0.624		0.888

Codes:$***0.001, **0.01, *0.05, N$, values in (.) are standard errors

Assessing the goodness of fit (GoF), the mean, kurtosis, skewness and Jarque–Bera for normality test are performed on the residuals of a fitted model. It is worth noting that these results as reported in Table 5 confirms that residuals of a fitted MS(2)-GAS(1) are not normally distributed, confirming the preliminary results reported in Table 2. Additionally, it is discovered that these residuals follow heavy tailed distributions due to observed kurtosis that

is above three and they are also negatively skewed as the reported by a measure of skewness which is -3.0895.

Table 5: Residual analysis

	Mean	Skewness	Kurtosis	JB Test
Residuals	0.006066	-3.0895	54.670	6758.5∗ ∗ ∗

Codes:∗ ∗ ∗0.001, ∗ ∗ 0.01, ∗0.05, N

3.2. Forecasting performance of value-at-risk and expected shortfall using heavy-tailed distributions

Using Minimum score combining proposed by Trucíos and Taylor (2020), the following distributions: the generalized extreme value (GEV) distribution, the generalized Pareto distribution (GPD), the generalized Logistic distribution (GLD), the log–Pearson-3,a Phased Bi-Exponential, and finally Wakeby, are combined with an estimated MS(2)-GAS(1). Based on the estimates of combined models, an out-of-sample analysis evaluates the ability of regime-switching methods coupled with these fat-tailed distributions for VaR and ES forecasting. Volatility predictions and extremes are performed for five, and thirty days steps-ahead forecasts. An out-of-sample data concern the last 5% of FTSE/BSE-DCI log-returns series, using a 168 log-returns sample for the rolling-window estimation with an iterated strategy. This method is suggested by Maciel (2021). The parameters of combined model(s) are updated at every observation for the out-of-sample data, producing the following Value–at–Risk and expected shortfall forecasts: h_{t+5} and h_{t+30}.

Table 6 to Table 8 present the combined forecasts VaR and ES estimates, number of VaR and ES violations, tail risk test, dynamic quantile test, Kupiec Likelihood ratio test, Christoffersen test and Basel III zones results for the different model combinations on the BSE index. All VaR, and ES computation are performed at the common 95%, and 99%, for short positions of trade. These results are vital in the determination of VaR and ES forecasting adequacy for different distributions that are coupled with MS-GAS copula model discussed in section 2. Additionally, it allows comparison across these distributions that are not necessarily fitted over the same part of the data. In this case, the GEV distribution and GPD are compared with other distributions. This empirical analysis is different from that of Huang, Huang, and Chinhamu (2014) who employed both short and long-term trading positions. For the GEV distribution, a block size of 5 is used and this produced weekly maxima. While the GPD is estimated at a threshold of 95% quantile locating the 5% of observations as exceedances. The negative tails are fitted using the relation $M_n = \min\{X_1, \cdots, X_n\} = -M_n = -[\max[X_1, \cdots, X_n]]$. It should be noted that, as one might anticipate given the leptokurtic nature of these data series, normality is rejected almost everywhere. Therefore, the normal distribution assumption frequently results in underestimates of VaR, ES and an excess of violations. The standard Student's t-distribution, which is frequently advised, is a better candidate for VaR and ES estimation, but it is still (in most cases) comparably weaker to the heavy-tailed distributions discussed here (Huang *et al.* 2014).

For BSE-DCI (Table 6) there is a varying pattern of VaR and ES estimates. According to TR, DQ, Kupiec and Christoffersen test(s), the best model(s) at 95% and 99% VaR level is MS-GAS-GPD; the best model at 95% and 99% ES is MS-GAS-GEV. It can then be observed that the clustering of VaR and ES violations start to occur at lesser extreme VaR and ES levels (lower p-values for the Christoffersen test). This may be simply caused by an increase in the number of observations that exceed these risk measures' estimates. Similar observations can be found in the study of Huang *et al.* (2014).

The values in Table 8 show the number of times in backtests, which are used in 5 days and 30 days by VaR and ES in all three zones. The test is ranked as follows (1) a green zone, if in a window of 5 and 30 trade days the number of exceptions reported is 0 to 4, (2) a yellow area,

Table 6: VaR and ES Estimates using heavy-tailed

	VaR			ES		
Distribution	95%	99%	99.9%	95%	99%	99.9%
MS-GAS-GEV	0.775	0.625	0.775	0.075	0.725	0.775
MS-GAS-Logistic	0.419	0.072	0.419	0.519	0.272	0.439
MS-GAS-GPD	0.727	0.781	0.270	0.566	0.657	0.523
MS-GAS-Log-Pearson3	0.572	0.077	0.573	0.105	0.145	0.502
MS-GAS-Phased Bi-Exponential	0.195	0.195	0.582	0.195	0.195	0.582
MS-GAS-Wakeby	0.310	0.891	0.372	0.070	0.874	0.592

Table 7: Backtesting five-step and thirty-step ahead value-at-risk and expected shortfall

	Value-at-Risk 5 days					
Model	Proportion	Violations	TR Test	DQ Test	Kupiec	Christoffersen
MS-GAS-GEV	95%	10	0.062	0.9025	0.8781	0.0058
MS-GAS-Gen.Logistic	95%	26	0.052	0.5305	0.6871	0.0120
MS-GAS-GPD	95%	8	0.6582	0.9525	0.6961	0.0079
MS-GAS-Log.Pearson 3	95%	19	0.0032	0.9872	0.8890	0.1820
MS-GAS-Bi-Exponential	95%	15	0.0062	0.9385	0.8102	0.1890
MS-GAS-Wakeby	95%	25	0.0083	0.9193	0.1387	0.3420
	Expected Shortfall 5 days					
MS-GAS-GEV	95%	14	0.5964	0.2648	0.5195	0.0391
MS-GAS-Gen.Logistic	95%	23	0.4108	0.3937	0.5281	0.0126
MS-GAS-GPD	95%	19	0.3885	0.3991	0.089	0.4084
MS-GAS-Log.Pearson 3	95%	17	0.096	0.3909	0.3989	0.3895
MS-GAS-Phased-Bi-Exponential	95%	20	0.088	0.3049	0.2146	0.2012
MS-GAS-Wakeby	95%	26	0.0108	0.2048	0.2123	0.3047
	Value-at-Risk 30 days					
Model	Proportion	Violations	TR Test	DQ Test	Kupiec	Christoffersen
MS-GAS-GEV	99%	5	0.082	0.9955	0.8611	0.0120
MS-GAS-Gen.Logistic	99%	8	0.0182	0.8381	0.2291	0.0145
MS-GAS-GPD	99%	3	0.7872	0.1258	0.4342	0.0984
MS-GAS-Log.Pearson 3	99%	10	0.0232	0.5139	0.2799	0.0420
MS-GAS-Phased-Bi-Exponential	99%	9	0.0083	0.8504	0.9306	0.0580
MS-GAS-Wakeby	95%	7	0.8772	0.1255	0.8433	0.0596
	Expected Shortfall 30 days					
MS-GAS-GEV	99%	3	0.9552	0.9785	0.963	0.0242
MS-GAS-Gen.Logistic	99%	7	0.3768	0.6796	0.4963	0.0937
MS-GAS-GPD	99%	5	0.5934	0.8897	0.8963	0.0508
MS-GAS-Log.Pearson 3	99%	4	0.7761	0.3879	0.8661	0.0420
MS-GAS-Phased-Bi-Exponential	99%	6	0.9682	0.7768	0.9661	0.0720
MS-GAS-Wakeby	99%	8	0.9123	0.9508	0.9188	0.0108

if there are 5-9 exceptions, and a red zone, if there are more than 9 exceptions. Neither of the risk measures avoids a regulatory penalty, with all of them slipping into the yellow zone, which imposes a penalty between 0.40 and 0.85 times the VaR for calculation of the market risk charge (Bee and Trapin 2018).

Interestingly, the hybrids with 5 and 30 days of data performed uniformly throughout all the risk measures. This gives an average penalty that ranges from 4.3% to 28%. Generally, all the risk measures showed that both the MS-GAS-GEV and MS-GA-GPD distributions are good estimator models for extreme losses for FTSE/BSE-DCI stock returns. At 95% and 99% confidence level, both VaR and ES from MS-GAS-GEV and MS-GAS-GPD did not produce the most worrying results. For example, MS-GAS-GEV only produced zero exceptions out of 30 observations. While MS-GAS-GPD produced zero exception out of 5 observations.

Pigildin (2009) noted there is only a very small likelihood, which is less than 0.01%, that is an accurate model with a correct coverage of 99% would produce as much as 10 or more exceptions. Hence, the current study found that the risk models had estimated extreme tail losses with adequate accuracy for all FTSE/BSE-DCI. It is further worth noting that for five and thirty days penalties, both the MS-GAS-GEV and MS-GAS-GPD, produce less penalties. For the MS-GAS-GEV at five days VaR forecasting, only 8.7% penalty is observed while for ES gave 4.3%. The same interpretation can be made for thirty days and for other distributions.

Table 8: Basel III zones backtest results

Distribution	Zone	5 days VaR	5 days ES	30 days VaR	30 days ES
MS-GAS-GEV	Green Zone	3	3	5	6
	Yellow Zone	2	2	1	3
	Red Zone	2	1	0	0
	penalty	0.087	0.043	0.055	0.054
MS-GAS-Logistic	Green Zone	5	3	4	4
	Yellow Zone	3	1	1	3
	Red Zone	1	1	1	0
	penalty	0.172	0.176	0.251	0.175
MS-GAS-GPD	Green Zone	3	4	5	7
	Yellow Zone	1	2	3	2
	Red Zone	0	0	1	1
	penalty	0.027	0.024	0.024	0.079
MS-GAS-Log-Pearson3	Green Zone	5	6	5	6
	Yellow Zone	2	2	3	1
	Red Zone	2	0	0	1
	penalty	0.179	0.249	0.248	0.289
MS-GAS-Phased Bi-Exponential	Green Zone	5	4	5	6
	Yellow Zone	1	1	2	4
	Red Zone	1	0	1	1
	penalty	0.314	0.766	0.234	0.276
MS-GAS-Wakeby	Green Zone	5	4	5	6
	Yellow Zone	1	1	2	4
	Red Zone	1	0	1	1
	penalty	0.241	0.266	0.234	0.276

3.3. Comparing the performance of the fitted models

The purpose of this section is to determine which model best mimics the data and produces fewer forecasts for both VaR and ES. This helps in assisting the maximum dispatching of Botswana financial sector. Four error metrics are used this study. These metrics are used to measure the performance of each model and results are summarized in Table **??**. Some tentative conclusions are drawn from this Table. The results show that combining point forecasts from different MS(2)-GAS(1) with heavy tailed distributions improves the performance of the model in predicting and forecasting VaR and ES. In Table **??**, the MS(2)-GAS(1)-GPD is the best model. The key performance indicators for this combined forecasting model are far below the MS(2)-GAS(1)-GEV.

Table 9: Key performance indicators for the fitted models

	5 days Forecasting Performance				30 Days Forecasting Performance			
Model	RMSE	MAE	MAPE	MPE	RMSE	MAE	MAPE	MPE
M(2)S-GAS(1)-GEV	16.0352	12.468	1.47832	1.67576	9.2380	0.9023	1.247	1.002
MS-GAS-GPD	13.4014	10.0257	1.27781	1.39646	8.0174	0.7930	0.798	0.9141

4. Discussion and conclusion

Under the regulation of Basel Accords, risk managers of financial institutions need to rely on state-of-the-art methodologies for predicting and evaluating their downside risk. This becomes a complex issue when one has to incorporate regime shift because this requires a substantial amount of planning before throwing machine learning algorithms at it. Nonetheless, it is also an application of data science and machine learning for the good, which makes sure that stochastic parameters are correctly modeled and identified. However, mitigation of financial hazard and management planning plays a critical role in regime shifts with dependence on time-varying parameters in financial society and, the economic world. Due to the asymmetry and leptokurtic nature of financial time series, non-linear models had become a useful tool in modeling and forecasting time-varying VaR and ES. Nevertheless, the statistical procedures of this study lie in forecasting the stochastic time-varying VaR and ES of FTSE/BSE–DCI by employing a MS-GAS that is nested with copulas and heavy tailed distributions. There are few studies that aim at this approach; and to current knowledge, this is the first study to use such a model to show characterizing performance of stochastic time-varying VaR and ES under regime shifts and extreme value distributions. Performing this analysis, explicitly accounted for the presence of different Markovian regimes as well as a smooth within-regime dynamic evolution of the dependence parameters. Specifically, exploiting the recent advantages for score driven processes, the state dependent copula parameters are then allowed to be updated using the scaled score of the conditional copula distribution. This choice helped in achieving a greater level of flexibility in the context of dynamic copula models, and it also introduced extreme stochastic behavior for the class of VaR and ES models in a natural and effective way. A vibrant advantage of the estimated MS(2)–GAS(1) model is that it exploit a full likelihood information and taking a local density score step as a driving mechanism, time-varying parameters become non-stationary and form a clear signal of extreme distribution behavior which empirical properties of this study revealed the same behavior with the residuals of MS(2)–GA(1)-copula following heavy tailed distributions.

The results obtained in this study showed that the MS-GAS-GEV and MS-GAS-GPD, are the best models to forecast 5 days and 30 days VaR and ES; with MS-GAS-GPD out performing the counter part MS-GAS-GEV. Because the model gave less forecasting performance when assessed by RMSE, MAE, MAPE and MPE. Hence the conclusion that MS(2)-GAS(1)-GPD nested with copulas is the best model that mimics the FTSE/BSE-DIC data. This is a promising technique for regime shift susceptibility detection and extreme time-varying risk. For further research, Heavy-GAS-tF model subject to regime switch ought to be utilized and estimate deeper correlation dynamics of the heavy tails of the GAS model. It would be interesting to see what sort of results that can be found if a more sophisticated machine learning procedures are used to model stochastic time-varying parameters as a regime shift process through the use of Markov chain Monte Carlo and bootstrapping of the credible confidence interval for the estimated time-varying parameters. A probabilistic description and modeling of extreme peak time-varying loads using Poison point process is another area that requires future research. This approach helps in estimating the frequency of occurrence of peak time–varying. A sensitivity analysis with respect to daily time-variation performed and the development of a two–stage stochastic integer recourse models with the objective of optimizing parameter distribution is an interesting future research direction with stock market data. This will be studied elsewhere.

References

Ahmad J, Bladt M (2022). "Aggregate Markov Models in Life Insurance: Estimation via the EM Algorithm." *arXiv preprint arXiv:2212.10661*. URL https://www.zhuanzhi.ai/paper/8ec856f314f0342a780cad27f02739f2.

Ardia D, Boudt K, Catania L (2018). "Downside Risk Evaluation With the R Package GAS." *R Journal*, **10**(2), 410–421. doi:10.32614/rj-2018-064.

Asheri H, Hosseini R, Araabi BN (2021). "A New EM Algorithm for Flexibly Tied GMMs with Large Number of Components." *Pattern Recognition*, **114**, 107836. doi:10.1016/j.patcog.2021.107836.

Bayer S, Dimitriadis T (2022). "Regression-based Expected Shortfall Backtesting." *Journal of Financial Econometrics*, **20**(3), 437–471. doi:10.1093/jjfinec/nbaa013.

Bazzi M, Blasques F, Koopman SJ, Lucas A (2017). "Time-varying Transition Probabilities for Markov Regime-switching Models." *Journal of Time Series Analysis*, **38**(3), 458–478. doi:10.1111/jtsa.12211.

Bee M, Trapin L (2018). "Estimating and Forecasting Conditional Risk Measures with Extreme Value Theory: A Review." *Risks*, **6**(2), 45. doi:10.3390/risks6020045.

Bernardi M, Catania L (2016). "Comparison of Value-at-Risk Models Using the MCS Approach." *Computational Statistics*, **31**(2), 579–608. ISSN 1613-9658. doi:10.1007/s00180-016-0646-6. URL https://doi.org/10.1007/s00180-016-0646-6.

Bernardi M, Catania L (2019). "Switching Generalized Autoregressive Score Copula Models with Application to Systemic Risk." *Journal of Applied Econometrics*, **34**(1), 43–65. doi:10.1002/jae.2650.

Bernardi M, Gayraud G, Petrella L (2015). "Bayesian Tail Risk Interdependence Using Quantile Regression." *Bayesian Analysis*, **10**(3), 553–603. doi:10.1214/14-ba911.

Best S (2021). "Minimum Capital Requirements for Market Risk: An Overview and Critical Analysis of the Standardized Approaches under Basel III." *wifin Working Paper 10/2021*, RheinMain University of Applied Sciences, Wiesbaden. doi:10419/266230.

Blazsek S, Escribano A, Licht A (2020). "Nonlinear Common Trends for the Global Crude Oil Market: Markov-switching Score-driven Models of the Multivariate t-Distribution." *Uc3m working papers. economics*, Universidad Carlos III de Madrid. Departamento de Economía. doi:10016/30346.

Blazsek S, Ho HC, Liu SP (2018). "Score-driven Markov-Switching Egarch Models: An Application to Systematic Risk Analysis." *Applied Economics*, **50**(56), 6047–6060. doi:10.1080/00036846.2018.1488073.

Bollerslev T (1988). "A Capital Asset Pricing Model with Time-Varying Covariances." *The Journal of Political Economy*, **96**(1), 116–131. doi:10.1086/261527.

Castro C, Ferrari S (2014). "Measuring and Testing for the Systemically Important Financial Institutions." *Journal of Empirical Finance*, **25**, 1–14. doi:10.1016/j.jempfin.2013.10.009.

Catania L (2021). "Dynamic Adaptive Mixture Models with an Application to Volatility and Risk." *Journal of Financial Econometrics*, **19**(4), 531–564. doi:10.1093/jjfinec/nbz018.

Christoffersen P (1998). "Evaluating Interval Forecasts." *International Economic Review*, **39**(4), 841–862. doi:10.2307/2527341.

Creal D, Koopman J, Lucas A (2013). "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics*, **28**(5), 777–795. doi:10.1002/jae.1279.

Degiannakis S, Dent P, Floros C (2014). "A Monte Carlo Simulation Approach to Forecasting Multi-Period Value-at-Risk and Expected Shortfall Using the FIGARCH-Skt Specification." *The Manchester School*, **82**(1), 71–102. doi:10.2139/ssrn.3259844.

Durante F, Sempi C (2015). *Principles of Copula Theory*. CRC press. URL https://www.taylorandfrancis.com.

Engle RF, Manganelli S (2004). "Caviar: Conditional Autoregressive Value-at-Risk by Regression Quantiles." *Journal of Business and Economic Statistics*, **22**(4), 367–381. doi:10.1198/073500104000000370.

Escanciano JC, Olmo J (2010). "Backtesting Parametric Value-at-Risk with Estimation Risk." *Journal of Business and Economic Statistics*, **28**(1), 36–51. doi:10.1198/jbes.2009.07063.

Figueroa-Zúñiga J, Niklitschek-Soto S, Leiva V, Liu S (2022). "Modeling Heavy-Tailed Bounded Data by the Trapezoidal Beta Distribution with Applications." *REVSTAT-Statistical Journal*, **20**(3), 387–404. doi:10.57805/revstat.v20i3.380.

Gao CT, Zhou XH (2016). "Forecasting VaR and Es Using Dynamic Conditional Score Models and Skew Student Distribution." *Economic Modelling*, **53**, 216–223. doi:10.1016/j.econmod.2015.12.004.

Güriş B (2019). "A New Nonlinear Unit Root Test with Fourier Function." *Communications in Statistics-Simulation and Computation*, **48**(10), 3056–3062. doi:10.1080/03610918.2018.1473591.

Harvey AC (2013). "Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series." *Cambridge University Press.*, **52**.

Huang CS, Huang CK, Chinhamu K (2014). "Assessing the Relative Performance of Heavy-Tailed Distributions: Empirical Evidence from the Johannesburg Stock Exchange." *Journal of Applied Business Research (JABR)*, **30**(4), 1263–1286. doi:10.19030/jabr.v30i4.8675.

Karaś M, Szczepaniak W (2017). "Measuring Systemic Risk with Covar Using a Stock Market Data Based Approach." In K Jajuga, LT Orlowski, K Staehr (eds.), *Contemporary Trends and Challenges in Finance*, pp. 135–143. Springer International Publishing, Cham. doi:10.1007/978-3-319-54885-2_13.

Kasahara H, Shimotsu K (2018). "Testing the Number of Regimes in Markov Regime Switching Models." *arXiv preprint arXiv:1801.06862*. doi:10.48550/arXiv.1801.06862.

Kupiec PH (1995). *Techniques for Verifying the Accuracy of Risk Measurement Models*, volume 95. Division of Research and Statistics, Division of Monetary Affairs, Federal. URL https://ssrn.com/abstract=7065.

Maaziz M, Kharfouchi S (2018). "Parameter Estimation of Markov Switching Bilinear Model Using the (em) Algorithm." *Journal of Statistical Planning and Inference*, **192**, 35–44. doi:10.1016/j.jspi.2017.07.002.

Maciel L (2021). "Cryptocurrencies Value-at-Risk and Expected Shortfall: Do Regime-Switching Volatility Models Improve Forecasting?" *International Journal of Finance and Economics*, **26**(3), 4840–4855. doi:10.1002/ijfe.2043.

Makatjane K, Moroke N (2021). "Predicting Extreme Daily Regime Shifts in Financial Time Series Exchange/Johannesburg Stock Exchange—All Share Index." *International Journal of Financial Studies*, **9**(2), 18. doi:10.3390/ijfs9020018.

McNeil AJ, Frey R, Embrechts P (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition*. Princeton University Press. URL hppts://www.pupress.princeton.edu/titles/8056.html.

Mwamba MJW, Hammoudeh S, Gupta R (2014). "Financial Tail Risks and the Shapes of the Extreme Value Distribution: A Comparison between Conventional and Sharia-Compliant Stock Indexes." *Working Papers 201480*, Univeristy of Pretoria. URL https://EconPapers.repec.org/RePEc:pre:wpaper:201480.

Nicolas MLD (2022). "Pycop: A Python Package for Dependence Modeling with Copulas." *Zenodo Software Package*, **70**, 7030034. doi:10.5281/zenodo.7030034.

Nieto MR, Ruiz E (2016). "Frontiers in VaR Forecasting and Backtesting." *International Journal of Forecasting*, **32**(2), 475–501. doi:10.1016/j.ijforecast.2015.08.003.

Pigildin D (2009). "Value-at-Risk Estimation and Extreme Risk Spillover between Oil and Natural Gas Markets." *PD Economics*. https://www.etd.ceu.edu/2009/pigildin_dmitriy.pdf.

Shang HL (2018). "Bootstrap Methods for Stationary Functional Time Series." *Statistics and Computing*, **28**(1), 1–10. doi:10.1007/s11222-016-9712-8.

Şimşek AS, Topaloglu H (2018). "An Expectation-Maximization Algorithm to Estimate the Parameters of the Markov Chain Choice Model." *Operations Research*, **66**(3), 748–760. doi:10.1287/opre.2017.1692.

Sordo MA, Suárez-Llorens A, Bello AJ (2015). "Comparison of Conditional Distributions in Portfolios of Dependent Risks." *Insurance: Mathematics and Economics*, **61**, 62–69. doi:10.1016/j.insmatheco.2014.11.008.

Trucíos C, Taylor JW (2020). "Forecasting Value-at-Risk and Expected Shortfall of Cryptocurrencies Using Combinations Based on Jump-Robust and Regime-Switching Models." *Available at SSRN 3751435*. doi:10.2139/ssrn.3751435.

Affiliation:

Katleho Makatjane
University of Botswana
Department of Statistics
Gaborone, Botswana
E-mail: makatjanek@ub.ac.bw

http://www.ajs.or.at/
http://www.osg.or.at/
Submitted: 2023-04-19
Accepted: 2023-08-16