Scientific Research Publishing

Forecasting Value-at-Risk of Financial Markets under the Global Pandemic of COVID-19 Using Conditional Extreme Value Theory

Cyprian Omari*, Simon Mundia, Immaculate Ngina

Department of Statistics and Actuarial Science, Dedan Kimathi University of Technology, Nyeri, Kenya
Email: *cyomari@dkut.ac.ke, simon.maina@dkut.ac.ke, nginakyalo2015@gmail.com

Abstract

The recent global pandemic of coronavirus (COVID-19) has had an enormous impact on the financial markets across the world. It has created an unprecedented level of risk uncertainty, prompting investors to impetuously dispose of their assets leading to significant losses over a very short period. In this paper, the conditional heteroscedastic models and extreme value theory are combined to examine the extreme tail behaviour of stock indices from major economies over the period before and during the COVID-19 pandemic outbreak. Daily returns data of stock market indices from twelve different countries are used in this study. The paper implements a dynamic method for forecasting a one-day ahead Value at Risk. As a first step, a comprehensive in-sample volatility modelling is implemented with skewed Student's-t distribution assumption and their goodness of fit is determined using information selection criteria. In the second step, the VaR quantiles are estimated with the help of conditional Extreme Value Theory framework and then used to estimate the out-of-sample VaR forecasts. Backtesting results suggest that the conditional EVT based models consistently produce a better 1-day VaR performance compared with conditional models with asymmetric probability distributions for return innovations and maybe a better option in the estimation of VaR. This emphasizes the importance of modelling extreme events in stock markets using conditional extreme value theory and shows that the ability of the model to capture volatility clustering accurately is not sufficient for a correct assessment of risk in these markets.

Keywords

Asymmetry, Backtesting, COVID-19, GARCH-EVT, Skew-Student-t
Distribution, Conditional Extremes, Heavy Tails, Value-at-Risk

1. Introduction

Measurement of market risk that arises from movements in stock prices, interest rates, exchange rates and commodity prices is a focal point in the practice of financial risk management. This measurement relies heavily on the use of statistical financial models. These models attempt to capture the stylized facts that determine price fluctuations and sensitivities in financial markets. The recent COVID-19 pandemic fears, uncertainties and related confinement measures accelerated an unprecedented contraction in economic activities. The global financial markets reacted to the pandemic with a plunge in stock market indices and extreme volatility in equity prices due to panic sell-outs of equities out of fear. For example, on 12th March 2020 Dow Jones plunged 10%, experiencing its sharpest decline since the Black Monday crash in 1987. Also the S&P 500 index plunged 9% to close in the bear market territory, thus officially ending the bull market that began in 2009 during the throes of the financial crisis. The Asia-Pacific stock markets closed down (with the Nikkei 225 of the Tokyo Stock Exchange also falling to more than 20% below its 52-week high) and European stock markets closed 11% lower in their worst one-day decline in history on coronavirus fears[1]. At the same time, the economic turmoil associated with the COVID-19 pandemic has also had wide-ranging and severe impacts upon other sectors of the financial markets, including bond and commodity (including crude oil and gold) markets. The collapse of crude oil prices was one of the biggest price shocks the energy market has ever experienced since the first oil shock of 1973. The effects upon markets are part of the coronavirus recession and among the many economic impacts of the pandemic.

One of the most established and widely used standard measures of exposure to market risk is the Value at Risk (VaR). It calculates the worst loss that might be expected of an asset or portfolio of assets at a given confidence level over a given period under normal market conditions. It gives a fixed probability (or confidence level) that any losses suffered by an asset or portfolio over the holding period will be less than the limit established by VaR. VaR can be derived as a quantile of an unconditional distribution of financial returns, but it is preferable to model VaR as the conditional quantile so that it captures the time-varying volatility inherent to financial markets [1]. The popularity of VaR as a risk measure can be attributed to its ability to provide an aggregate measure of risk; a single number that is related to the maximum loss that might be incurred on a position, at a given confidence level.

To estimate market risk measures, several methodologies have been developed; the non-parametric approach (for example, historical simulation), the fully parametric approach (for example, based on an econometric model) and the semi-parametric method (for example, extreme value theory, filtered historical simulation and CAViaR method). However, over the last decade, conventional VaR models have been subject to massive criticism, as they failed to predict the

---

[1]https://www.cnbc.com/2020/03/12/stock-market-today-live.html.

huge repetitive losses that devastated financial markets during the 2007-2008 global financial crisis. These VaR models normally assume that the asset returns follow a normal distribution and thus ignoring the fat-tailed properties of actual returns. This underestimates the likelihood of extreme price movements. Thus, VaR may ignore important information regarding the tails of the underlying distributions. On the other hand, the concept of VaR as a risk measure has one disadvantage for measuring extreme price movements. The tail risk of VaR emerges since it measures only a single quantile of the return distributions and disregards any loss beyond the VaR level. Hence, when financial markets experience volatile behaviour and extreme price movements, the conventional methods are not capable of appropriately measuring the risk. Since VaR estimations are only related to the tails of a probability distribution, techniques from Extreme Value Theory (EVT) may be particularly effective. Therefore, emphasis and focus are now placed on adequate modelling of extreme quantiles for the conditional distribution of financial returns rather than the entire distribution. EVT has been proved to be successful in VaR estimation [2].

The Extreme Value Theory approach focuses on limiting the distribution of extreme returns observed over a long period, which is essentially independent of the distribution of the returns themselves. There two main models for Extreme Value Theory; the Block Maxima (BM) and the Peaks-Over-Threshold (POT) model. In the POT model, extreme values above a high threshold are analysed using a Generalized Pareto Distribution (GPD). A difficulty finding the optimal threshold for GPD fitting is encountered in this method. Threshold choice involves balancing bias and variance. The threshold must be sufficiently high to ensure that asymptotic underlying the GPD approximations is reliable, thus reducing bias. However, the reduced sample size for high thresholds increases the variance of the parameter estimates [3].

EVT provides a theoretical framework and practical foundation for statistical models describing the tail behaviour of extreme observations. Unlike the GARCH-family models, EVT models do not consider the entire conditional distribution of financial returns. Instead, they focus directly on the tails of the sample distribution in order to account for the heavy tails. Therefore, EVT potentially performs better than other approaches in terms of predicting unexpected extreme price changes, especially in volatile financial markets. However, applying EVT directly to return series is inappropriate as the current volatility background is not taken into consideration and the models deal with independently and identically distributed (i.i.d.) random variables. To overcome this shortcoming, McNeil and Frey [4] introduced a two-stage hybrid approach that combines the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model and Extreme Value Theory (EVT), referred to as Conditional Extreme Value Theory (CEVT) model or GARCH-EVT model, which addresses the fat tails phenomenon and stochastic volatility. The GARCH-EVT model captures the important stylized facts exhibited by most financial time series, such as stochastic volatility,

volatility clustering and leptokurtosis of conditional return distributions, and quickly adapts to recent market movements.

In econometrics and finance, implementing risk measurement methodology based on the theory of extremes is an important area of research. To this day, many researchers have investigated the estimation of Value at Risk (VaR) and Conditional VaR (CVaR), with the help of Extreme Value Theory. Most studies have VaR as their primary measure of interest. McNeil and Frey [4] showed that the application of combined GARCH and EVT results in a more accurate estimation of Value at Risk as compared with EVT methods and GARCH-type models. Several researchers have used the McNeil and Frey [4] approach in estimating market risk. Fernandez [5] showed that EVT outperforms a GARCH model with normal innovations by far and that it provides similar results to a GARCH model with Student-t innovations, as long as these innovations arise from an asymmetric and fat-tailed distribution. Gencay and Selcuk [6] showed that at the 99th and higher quantiles the Generalized Pareto Distribution model is superior to five other methods used in the study in terms of VaR forecasting. [7] [8] [9] and [10] among others, have demonstrated that the methods for estimating VaR based on modelling extreme observations measure financial risk more accurately compared to the conventional approaches.

In recent studies in finance VaR is be estimated more accurately using the conditional-EVT approach especially in modelling the distribution of extreme events and estimating extreme tail risks than the conventional models. Moreover, in some other studies, the conditional Value-at-Risk has been used as the risk measure, and researchers have shown that theoretically and empirically, using EVT contributes to a more precise estimation of the Value-at-Risk. Researchers have also compared the Extreme Value Theory models with other conventional methods, such as Historical Simulation (HS), Filtered Historical Simulation (FHS) and the GARCH models, in the estimation of Value at Risk, and have shown that conditional EVT models perform better.

Among the many studies on estimating the Value-at-Risk on the financial market with the conditional-EVT model, is the work of Soltane *et al.* [11] that combines Extreme Value Theory (EVT) and GARCH model to estimate VaR for the Tunisian Stock Market. They observe that GARCH-EVT-based VaR approach appears more effective and realistic than conventional methods. Singh *et al.* [12] applied univariate extreme value theory to model extreme market risk for the ASX-All Ordinaries (Australian) index and the S&P-500 (USA) Index. Results from backtesting showed that conditional-EVT based dynamic approach outperforms GARCH(1, 1) model and RiskMetrics in estimating VaR forecasts. Karmakar and Shukla [13] investigated the relative performance of Value-at-Risk (VaR) models using daily share price index data from six different countries across Asia, Europe and the United States. The empirical results showed the superior performance of Conditional EVT model in estimating and forecasting VaR measures compared with other competing models. Totić and Božović [14]

modelled the tail behaviour of returns using conditional Extreme Value Theory and results from backtesting showed that a conditional EVT-based model provides a more reliable VaR forecasts than the alternative models in all the six markets. Zhang and Zhang [15] also showed that the exponential GARCH (EGARCH) model with Generalized Error Distribution (GED) combined with the EVT approach does very well in predicting a critical loss for precious metal markets. Omari *et al.* [16] used the conditional extreme value theory to estimate Value at Risk of daily currency exchange rates and their results show that the conditional EVT-based model provides a more accurate out-of-sample VaR forecasts in estimating the currency tail risk. Stoyanov *et al.* [17] also investigated the out-of-sample behaviour of 41 equity market indices using the GARCH-EVT model and empirical results show that GARCH-EVT model performed better in estimating Value-at-Risk and Expected shortfall at 1% tail probability. Tabasi *et al.* [18] showed that the GARCH-EVT model outperforms the simple GARCH model with Student's *t* and normal distributions for residuals.

The application of GARCH-EVT model in empirical research requires the selection of an appropriate threshold level which separates tails of distribution from its middle part. The choice of threshold level is ambiguous, but critical in the estimation of Generalized Pareto distribution's parameters and the corresponding accuracy of value-at-risk. The standard practice is to adopt as low a threshold as possible, but there is a trade-off between variance and bias. If a threshold is too low, the asymptotic basis of the model is violated leading to high bias. However, too high a threshold generates insufficient excesses with which the model is estimated, leading to high variance. Most authors preferred to select a threshold as a fixed quantile of the data set, instead of determining a threshold value at each step, especially when they use a moving window of observation to find out-of-sample VaR estimates. [19] [20] and [21] chose either the 90th or 95th quantiles of the loss distribution as a threshold. In contrast, [6] [11] [22] used a less conservative but fixed thresholds.

Recent advances in the financial econometrics and applications of extreme value theory to finance have led to the development of several models. These models are used in estimating and forecasting tail-related quantities of the asset returns' conditional distribution. Some of the recent developments include; Huang *et al.* [23] who proposed a new approach to extreme value modelling for the forecasting of Value-at-Risk (VaR). In particular, both the block maxima and the peaks-over-threshold methods are generalized to exchangeable random sequences that cater for the dependencies such as serial autocorrelations of financial returns observed empirically. The results of VaR forecasts show that the conditional GARCH-EVT model performs better compared to the unconditional extreme value theory (EVT) approach. [1] proposed a new self-exciting probability peaks-over-threshold (SEP-POT) model for forecasting extreme loss probability and the value at risk. The results from backtesting of SEP-POT value at risk

(VaR) forecasts on seven stock indices favoured the SEP-POT model as an alternative for forecasting extreme quantiles of financial returns. There are several reviews on applications of conditional EVT in finance for instance [20] [24]. Echaust and Just [25] used four different optimal tail selection algorithms, that is, the path stability method, the automated Eye-Ball method, the minimization of asymptotic mean squared error method and the distance metric method with a mean absolute penalty function, to estimate out-of-sample Value at Risk (VaR) forecasts and compare them to the fixed threshold approach.

In this study, the conditional VaR at a 1-day horizon is estimated based on conditional Extreme Value Theory (conditional-EVT) approach and conventional GARCH-type models assuming asymmetric innovations distributions. We take into account volatility clustering and leverage effects in return volatility by using the GARCH, EGARCH, GJRGARCH, CSGARCH and APARCH models under different probability distributions assumed for the standardized innovations: Gaussian, Student-*t*, skewed Student-*t* and generalized error distribution. The two-step procedure of [4] fits a generalized Pareto distribution to the extreme values of the standardized residuals generated by an AR(1)-EGARCH(1, 1) model. Then, we compare the out-of-sample one-step-ahead value at risk (VaR) forecasts the performance of all these models before and during the COVID-2019 pandemic period using daily data. For VaR evaluation, the most widely used backtesting procedures, Unconditional Coverage (*UC*) and Conditional Coverage (*CC*) tests are used. The empirical analysis is based on the daily log-returns of twelve international stock market indices for the period between January 2006-July 2020 (that is, S&P 500 (US; SPX), FTSE 100 (UK; FTSE), DAX 30 (Germany; GDAXI), CAC 40 (France; FCHI), SMI (Switzerland; SMI), Euro Stoxx 50 (Europe; STOXX 50), S&P/TSX Composite (Canada; GSPTSE), NIKKEI 225 (Japan; N225), KOSPI 200 (South Korea; KS11), Hang Seng (Hong Kong; HSI), Shanghai Composite (China, SSE), Sensex (India, BSESN)). The daily log returns for the equity market were calculated from the adjusted daily closing prices downloaded from https://markets.businessinsider.com/indices.

This work provides an empirical study of conditional extreme value theory and contributes to the literature on the estimation of the tail risk of stock markets in four ways. First, several GARCH-type volatility specifications in an EVT model to take into account volatility clustering and asymmetric returns are used. Secondly, conditional EVT models that incorporate conditional models with asymmetric probability distributions used in the financial literature to calculate VaR are compared. Thirdly, VaR over the 1-day horizon for market risk management is calculated. Finally, we focus on the accuracy of our risk models for VaR estimation during pre-pandemic and during pandemic periods as well as using different significance levels. The empirical results indicate that the conditional EVT based models consistently produce a better 1-day VaR performance compared with conditional models with asymmetric probability distributions for return innovations and maybe a better option in the estimation of VaR.

The rest of the paper is organized as follows. Section 2 provides the metho-

dological details. It presents the estimation of the GARCH models with the selected innovations distribution assumptions, conditional GARCH-EVT modelling framework including the Peaks Over Threshold (POT) model, the Quasi Maximum Likelihood (QML) and describes the tail selection problem in this model, VaR estimation and backtesting procedures. Section 3 presents data and some preliminary summary descriptive statistics. Section 4 presents estimation results and empirical results from backtesting, while Section 5 gives a conclusion the paper.

2. Methodology

2.1. Value at Risk

Value-at-Risk (VaR) is a popular approach to measuring market risk. It is defined as the maximum loss that will be incurred on an asset with a given level of confidence over a specified period under normal market conditions. Given some confidence level $p \in (0,1)$, the VaR at a confidence level p is given by the smallest number l such that the probability that the expected loss L exceeds l is no larger than $(1-p)$. That is,

$$\text{VaR}_p = \inf \left\{ l \in \mathbb{R} : P(L > l) \leq 1-p \right\} = \inf \left\{ l \in \mathbb{R} : F_L(l) \geq p \right\} \tag{1}$$

Let r_t denote the return on assets at time t. The one-day-ahead Value-at-Risk (VaR) for holding a long trading position at p level of significance, denoted as $\text{VaR}_{t+1}^p(r_t)$ is defined as

$$P\left(r_{t+1} \leq \text{VaR}_{t+1}^p \mid \mathcal{F}_t \right) = p \tag{2}$$

where $\mathcal{F}_t$ is the information set available at time t. In this definition VaR is the pth conditional quantile of the return distribution. For a short trading position VaR is the $(1-p)^{\text{th}}$ conditional quantile of the return distribution.

$$P\left(r_{t+1} > \text{VaR}_{t+1}^p \mid \mathcal{F}_t \right) = 1-p \tag{3}$$

2.2. The GARCH Model

The Generalized Autoregressive Conditional Heteroskedasticity models are the most commonly used in the literature for modelling volatility and estimating Value-at-Risk. Let $r_t \in R$ be the percentage log-returns of the financial asset (stock index) of interest at time t. Then

$$r_t = \mu_t + \sigma_t z_t, \tag{4}$$

where μ_t denotes the conditional mean, σ_t a conditional volatility process and z_t is a zero-mean white noise. The mean component of daily log returns r_t is assumed to be represented by an AR(1) model. The GARCH-type models are used in modelling conditional volatility dynamics in log-returns of financial time series. There are several representations of common GARCH-type models but we consider the ones that follow the above specification in Equation (4); however, in each case, the volatility process σ_t is different. For brevity, all of the models will be restricted to a maximum order of one. In addition, for each

GARCH-type model, the innovation process ε_t is allowed to follow either the normal distribution or Student's-t distribution or skew Student's-t distribution or skew generalized error distribution.

The standard GARCH(1, 1) model introduced by Bollerslev [26] is given as:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{5}$$

where $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $(\alpha_1 + \beta_1) < 1$. The GARCH(1, 1) model which is the most commonly used in financial literature and the main feature of the model is that it captures volatility clustering in the data. The "persistence" parameter (which accounts for the amount of volatility clustering captured by the model) for this model is $\alpha_1 + \beta_1$. The parameter restrictions are necessary for the model in Equation (5) to be weakly stationary and the unconditional variance is given by $\omega / (1 - (\alpha + \beta))$, thus higher order moments exist.

Since financial returns tend to display leverage effects, which is the negative correlation between returns and its volatility, the asymmetric GARCH models are introduced to address the problem.

The exponential GARCH (EGARCH) model of Nelson [27] is defined as:

$$\log_e \sigma_t^2 = \omega + \alpha_1 z_{t-1} + \gamma_1 \left(|z_{t-1}| - E|z_{t-1}| \right) + \beta_1 \log_e \left(\sigma_{t-1}^2 \right) \tag{6}$$

where the coefficient α_1 captures the sign effect, and $\gamma_1 > 0$ the size of leverage effect. The persistence parameter for this model is β_1.

The Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model of Glosten *et al.* [28] models positive and negative shocks on the conditional variance asymmetrically via the use of the indicator function *I*. The GJR-GARCH(1, 1) model is given as:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 I_{t-1} \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{7}$$

where γ_1 now represents the "leverage" term. The indicator function *I* takes on value of 1 for $\varepsilon_{t-1} \leq 0$ and 0 otherwise. The persistence depends on the parameter γ_1, through $\alpha_1 + \beta_1 + \gamma_1 \kappa$, where κ denotes the expected value of the standardized residuals.

The asymmetric power ARCH (APARCH) model of Ding *et al.* [29] allows for both leverage and the Taylor effect, named after Taylor [30] who observed that the sample autocorrelation of absolute returns was usually larger than that of squared returns. The APARCH(1, 1) model can be expressed as:

$$\sigma_t^\delta = \omega + \alpha_1 \left(|z_{t-1}| - \gamma_1 z_{t-1} \right)^\delta + \beta_1 \sigma_{t-1}^\delta, \tag{8}$$

where $\delta \in R^+$, is a Box-Cox transformation of σ_t, and $-1 < \gamma_1 < 1$ is the coefficient in the leverage term. The persistence parameter is equal to $\beta_1 + \alpha_1 \kappa_1$, where κ_1 is the expected value of the standardized residuals under the Box-Cox transformation of the term, which includes the leverage parameter γ_1.

The component standard GARCH (CS-GARCH) model of Engle and Lee [31] decomposes the component of the conditional variance so as to investigate the long and short-run movements of volatility. Let q_t represent the permanent

component of the conditional variance, the component model can be written as

$$\sigma_t^2 = q_t + \alpha_1\left(z_{t-1}^2 - q_{t-1}\right) + \beta_1\left(\sigma_{t-1}^2 - q_{t-1}\right) \tag{9}$$

$$q_t = \alpha_0 + \rho q_{t-1} + \phi\left(z_{t-1} - \sigma_{t-1}^2\right)$$

where effectively the intercept of the GARCH model is now time-varying following first order autoregressive type dynamics.

For a better fit of the GARCH models, the standardised Student's-t distribution, skewed Student's-t distribution and Generalized Error Distributions (GED) are instead of the normal distribution, since returns exhibit fat tails and skewness. The standardized Student's t-distribution is given by:

$$f\left(\varepsilon_t, \nu\right) = \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{\pi(\nu-2)}}\left[1 + \frac{\varepsilon_t^2}{\nu-2}\right]^{-\left(\frac{\nu+1}{2}\right)} \tag{10}$$

with degrees of freedom parameter $\nu > 0$, controlling the thickness of the tail, $\Gamma(z) = \int_0^\infty x^{z-1}\mathrm{e}^{-x}\mathrm{d}x$.

An alternative distribution for modelling skewed and heavy-tailed data is the skewed Student's t-distribution proposed by Hansen [32]. The distribution as in Zhu and Galbraith [33] is given by

$$f\left(x; \mu, \alpha, \nu, \mu, \sigma\right) = \begin{cases} \dfrac{1}{\sigma}K(\nu)\left[1 + \dfrac{1}{\nu}\left(\dfrac{x-\mu}{2\alpha\sigma}\right)^2\right]^{-(\nu+1)/2}, & x \le \mu \\[4mm] \dfrac{1}{\sigma}K(\nu)\left[1 + \dfrac{1}{\nu}\left(\dfrac{x-\mu}{2(1-\alpha)\sigma}\right)^2\right]^{-(\nu+1)/2}, & x > \mu \end{cases} \tag{11}$$

This parametrization of the distribution is equivalent to those of [32] and [34].

The Generalised Error Distribution (GED) is given by

$$f\left(x, \mu, \alpha, \beta\right) = \frac{\beta}{2\alpha\Gamma\left(\dfrac{1}{\beta}\right)}\mathrm{e}^{-\left(\frac{|x-\mu|}{\alpha}\right)^\beta} \tag{12}$$

Using the Quasi-Maximum Likelihood (QML), the parameters $\left(\mu, \omega, \alpha, \gamma, \beta\right)$ may be estimated simultaneously by maximizing the log likelihood. The log-likelihood function is obtained under the assumption that the random error term follows the standardized Student's t-distribution is given by

$$\begin{aligned} l_t(\boldsymbol{\theta}) = {}& \log\Gamma\left(\frac{\nu+1}{2}\right) - \frac{1}{2}\log\left(\pi(\nu-2)\right)\log\Gamma\left(\frac{\nu}{2}\right) \\ & - \frac{\nu+1}{2}\log\left(1 + \frac{\varepsilon_t^2}{\nu-2}\right) - \frac{1}{2}\log\sigma_t^2 \end{aligned} \tag{13}$$

where $\boldsymbol{\theta} = \left(\mu, \omega, \alpha, \gamma, \beta\right)$ is the unknown parameters in GARCH-type models

to be estimated. Solving the first-order conditions of the log-likelihood function with respect to the parameters $\omega, \alpha_1, \gamma_1, \beta_1$, a specified optimal GARCH-type model is obtained. The standardized residuals ε_t of the fitted GARCH-type model can also be extracted. Next, the forecasts of the conditional mean μ_{t+1} and variance σ_{t+1} are obtained using the estimated parameters from QML above. To this extent, one-step ahead forecasts of the conditional variance of returns are recursively obtained as $\hat{\sigma}_{t+1}^2 = E\left(\sigma_{t+1}^2 \mid F_t\right)$.

The one-step-ahead conditional variance $\hat{\sigma}_{t+1}^2$ forecast for the GARCH(1, 1), EGARCH(1, 1) GJR-GARCH(1, 1), APARCH(1, 1) and CS-GARCH(1, 1) respectively, is:

$$
\begin{aligned}
\hat{\sigma}_{t+1}^2 &= \hat{\omega} + \hat{\alpha}_1 \varepsilon_t^2 + \hat{\beta}_1 \sigma_t^2 \\
\log_e\left(\hat{\sigma}_{t+1}^2\right) &= \hat{\omega} + \hat{\alpha}_1\left[\hat{\gamma}_1 z_t + \gamma_2\left(|z_t| - E|z_t|\right)\right] + \hat{\beta}_1 \log_e\left(\sigma_t^2\right) \\
\hat{\sigma}_{t+1}^2 &= \hat{\omega} + \hat{\alpha}_1 \varepsilon_t^2 + \hat{\beta}_1 \sigma_t^2 + \hat{\gamma}_1 I_t \varepsilon_t^2 \\
\hat{\sigma}_{t+1}^\delta &= \hat{\omega} + \hat{\beta}_1 \sigma_t^\delta + \hat{\alpha}_1\left(|\varepsilon_t| - \hat{\gamma}\varepsilon_t\right)^\delta \\
\hat{\sigma}_{t+1}^2 &= q_{t+1} + \hat{\alpha}_1\left(\varepsilon_t^2 - q_t\right) + \hat{\beta}_1\left(\sigma_t^2 - q_t\right)
\end{aligned}
\tag{14}
$$

The VaR forecast for the GARCH-type models rely on the one-day-ahead conditional variance forecast, σ_{t+1}^2 of the volatility model. For each GARCH-type model, under the assumption of different error distribution, the one-day-ahead VaR forecast at $p\%$ confidence level is obtained as:

$$
\widehat{\text{VaR}}_{t+1}^p = \hat{\mu}_{t+1} + F^{-1}(p)\hat{\sigma}_{t+1}
\tag{15}
$$

where $F^{-1}(p)$ is p^{th} quantile of the cumulative distribution function of the innovations distribution.

2.3. Modelling Tails Using Extreme Value Theory

Extreme Value Theory primarily focuses on analysing the asymptotic behaviour of extreme values of a random variable. The theory provides robust statistical tools for estimating only extreme values distribution instead of the whole distribution. There are two main approaches in applying EVT; the Block Maxima (BM) model and the Peaks-Over-Threshold (POT) model. The approaches rely on different references to determine the extreme values. The BM model selects the maximum value given a specified period or block while the POT model focusses on the observations exceeding some pre-specified high threshold. Modelling the maximum of a block of random variables is considered wasteful if other data on extreme values are available. Therefore, a more efficient approach to modelling extreme events is to focus not only the largest (maximum) events, but also on all events greater than some large preset threshold. This is the Peaks Over Threshold (POT) modelling. The POT models are generally considered to be more appropriate in practical applications, due to their efficient use of data at the extreme values.

In this study, the POT approach to model extreme events is adopted. The POT method specifies the observations above the chosen threshold as extreme values

and focusses on the "exceedance" part to estimate parameters of the tail distribution rather than the entire data set. Let $X_1, X_2, \cdots, X_n$ be a sequence of of independent and identically distributed (i.i.d.) random variables, with common distribution function F. The POT model approach focuses on estimating the distribution function F_u of values of x above a high threshold u. The distribution of excesses over a high threshold u is defined as:

$$F_u(y) = \Pr(X - u \le |X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad 0 < y < x_0 - u, \qquad (16)$$

where $x_0 \le \infty$ is the right endpoint of F.

As in Balkema and de Haan [35] and Pickands [36], for a large class of underlying distributions functions F the conditional excess distribution function $F_u(y)$, for a large u, is well approximated by $F_u(y) \approx G(y)$ with $u \to \infty$. That is

$$\lim_{u \to x_F} \sup_{0 \le y \le x_F - u} \left| F_u(y) - G_{\xi, \sigma(u)}(y) \right| = 0 \qquad (17)$$

where $G_{\xi, \sigma(u)}(y)$, the Generalized Pareto Distribution (GPD), given by

$$G_{\xi, \sigma(u)}(y) = \begin{cases} 1 - \left(1 + \xi \dfrac{y}{\sigma(u)}\right)^{-\frac{1}{\xi}} & \text{if } \xi \ne 0, \\[2mm] 1 - \exp\left(-\dfrac{y}{\sigma(u)}\right) & \text{if } \xi = 0, \end{cases} \qquad (18)$$

where $y \ge 0$ for $\xi \ge 0$ and $0 \le y \le -\dfrac{\sigma(u)}{\xi}$ for $\xi < 0$. The distribution has a scale, σ and shape parameter ξ. It subsumes a number of other specific distributions under its parametrization. When $\xi > 0$, then G is a parameterized version of a heavy tailed ordinary Pareto distribution; when $\xi = 0$ corresponds to the light tailed exponential distribution and when $\xi < 0$ we have a short tailed Pareto type II distribution. In financial risk management, $\xi > 0$ is generally chosen as the most relevant for analysis purposes since the GPD tends to describe heavy tails. Estimates of the parameters ξ and $\sigma(u)$ can be obtained using the method of maximum likelihood [37]. For $\xi = -0.5$, Hosking and Wallis [38] present evidence that maximum likelihood regularity conditions are fulfilled and the maximum likelihood estimates are asymptotically normally distributed.

By setting $x = u + y$, an approximation of $F(x)$, for $x > u$, can be obtained from Equation (16):

$$F(x) = (1 - F(u)) G_{\xi, \sigma(u)}(y) + F(u), \qquad (19)$$

The function $F(u)$ can be estimated non-parametrically using the empirical c.d.f:

$$\hat{F}(u) = \frac{n - N_u}{n} \qquad (20)$$

where N_u represents the number of exceedances over the threshold u and n is the sample. By substituting Equations (18) and (20) into Equation (19), an estimate for $F(x)$ is obtained as follows:

$$\hat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\left(\frac{x - \hat{u}}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}},\qquad (21)$$

where $\hat{\xi}$ and $\hat{\sigma}$ are estimates of ξ and σ, respectively, which can be estimated by the method of maximum likelihood.

For $p > F(u)$, x_p can be obtained from Equation (21) by solving for x;

$$x_p = \hat{u} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(\frac{n}{N_u}(1 - p)\right)^{-\hat{\xi}} - 1\right]\qquad (22)$$

where u is a threshold, $\hat{\sigma}$ is the estimated scale parameter, $\hat{\xi}$ is the estimated shape parameter.

One of the challenging problems in practical application of POT-method is setting the appropriate threshold. Single threshold selection involves a bias-variance trade-off. An excessively low threshold may violate the asymptotic underlying the GPD approximation and, consequently, increase the bias. Conversely, an excessively high threshold may involve a smaller sample size and generate few excesses, leading to high variance in the parameter estimations. It is thus of importance of finding a good balance in setting the threshold to find a suitable balance between the variance and the bias of the model. In this paper, a quantile rule using an upper threshold of 10% (the 90th percentile) for setting the threshold value is adopted. This is a common practice.

2.4. Conditional Extreme Value Theory Model

The GARCH-EVT model introduced by McNeil and Frey [4] is used to estimate Value at Risk by extending the EVT framework to dependent data. To utilize EVT, an important assumption is for the data to be independently and identically distributed (i.i.d.). The EVT is used to model the tails of standardized residues e_t obtained from the GARCH-type model. First, the GARCH model is fitted to the financial return series to filter the serial autocorrelation and obtain close to independently and identically distributed standardized residuals. Subsequently, the standardized residuals are fitted using the POT-EVT framework. The GARCH-EVT approach is summarized as follows:

Fit a suitable GARCH-type model to the return data by quasi maximum likelihood. That is, maximize the log-likelihood function of the sample assuming the standardized Student's t-distributed innovations. Estimate μ_{t+1} and σ_{t+1} from the fitted model and extract the standardized residuals z_t.

Consider the standardized residuals computed in Step 1 to be realizations of a white noise process. Apply EVT to model the tails of the innovations and estimate x_q for a given probability q.

Hence, the standardized residuals can be computed as a white noise process

$z_t = \hat{\varepsilon}_t / \hat{\sigma}_t$. Given the 1-step forecasts $\hat{\mu}_{t+1}$, $\hat{\sigma}_{t+1}$ and standardized residuals series, the conditional value-at-risk, $\widehat{\text{VaR}}_{t+1}^p$ can be estimated as follows as:

$$\widehat{\text{VaR}}_{t+1}^p = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{x}_q \tag{23}$$

with $\hat{x}_q$ obtained from Equation (22).

VaR_{t+1}^p is computed during the out-of-sample period along with the parameter estimates by using the previous in-sample observations n returns. That is, VaR_{t+1}^p with t in the set $n, n+1, \cdots, m-1$ is calculated with returns $r_{t-n+1}, \cdots, r_t$. This implementation is rolled forward for each day, which effectively captures time-varying characteristics. Besides, the 90th percentile of the return distribution is set as the threshold, so k equals 10% of the daily observations. Since the backtesting period is relatively long, the threshold value is set at the 90th quantile in order to simplify the procedures. The advantage of this combination lies in its ability to capture conditional heteroscedasticity in the data through the GARCH framework, while at the same time modelling the extreme tail behaviour through the EVT method.

2.5. Backtesting the VaR Models

The adequacy of models used for estimating VaR forecasts can be statistically tested using the backtesting procedure. This procedure consists of comparing the out-of-sample VaR estimates with actual realized loss in the next period. An accurate VaR_{t+1}^p measure guarantees that the actual return will only be worse than the VaR forecasts $p \times 100$ of the time. Given a time series of past ex-ante VaR forecasts and past ex-post returns, one can define the "hit sequence" (also referred to as indicator function) of VaR violations as:

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t+1} < -\text{VaR}_{t+1}^p \\ 0 & \text{if } r_{t+1} \geq -\text{VaR}_{t+1}^p \end{cases} \tag{24}$$

The hit sequence returns a value of 1 on day $t+1$ if the ex-post loss on that day exceeds the VaR number predicted in advance for that day and value zero otherwise. When performing backtesting on VaR models, a hit sequence is created across T days indicating when the past violations occurred. For a VaR model to be accurate in its predictions, then the average hit ratio or the failure rate over the full sample should be equal α for the $(1-\alpha)^{\text{th}}$ quantile VaR (that is, for 95% VaR, $\alpha = 5\%$). The violations involve counting the number of actual realized returns that exceed the VaR forecast and comparing this number with the expected number of violations. As expected, the closer the hit ratio is to the expected value, the better the forecasts of the risk model. If the hit ratio is greater than the expectation, then the model underestimates the risk; with a hit ratio smaller than the expected value, the model overestimates risk.

In this study, the commonly used backtesting procedures; the Kupiec's proportion of failures test (also known as the unconditional coverage test) and Christoffersen's test (also known as the conditional coverage test) are used.

2.5.1. Unconditional Coverage Test

According to Kupiec [39] test, the interest is to check if the proportion of violations obtained from VaR models, call it $\hat{p}$, is significantly different from the expected proportion, p. This is called the unconditional coverage hypothesis. Assuming that the probability of obtaining an exceedance is constant, the number of VaR violations by actual returns, $T_1 = \sum_{t=1}^{T} I_t$ follows a binomial distribution $b(T_1, p)$, where T represents the total number of observations. An accurate $\text{VaR}(p)$ measure should produce an unconditional coverage $\hat{p} = T_1/T$ equal to p percent. The unconditional coverage test has a null hypothesis that the probability of failure for each trial ($\hat{p}$) should be equals to (p), that is, $H_0 : \hat{p} = p$.

The likelihood ratio statistic,

$$LR_{UC} = 2\left(\log\left(\hat{p}^{T_1}\left(1-\hat{p}\right)^{T-T_1}\right) - \log\left(p^{T_1}\left(1-p\right)^{T-T_1}\right)\right) \tag{25}$$

is used to perform this test. When the null hypothesis is true, the statistic has an asymptotic Chi-square distribution with one degree of freedom. The advantage of this test is that it assesses the adequacy of the model taking into account either too large and too small number of exceedances. A good model used for VaR estimation should also be characterized by independence of exceedances.

2.5.2. Conditional Coverage Test

Christoffersen [40] proposed a conditional coverage test procedure that jointly examines the correct unconditional coverage and serial independence. The procedure is a joint test of these two properties and the corresponding test statistic is the sum of the individual test statistics for the properties; that is, $LR_{CC} = LR_{UC} + LR_{IND}$ when conditioned on the first observation. The LR_{IND} denotes the likelihood ratio statistic that tests whether exceptions are independent, and the LR_{UC} is defined in the previous subsection. When the model accurately estimates VaR, then a present exceedance should not depend on whether or not an exceedance occurred on the previous day.

According to this test, the hit sequence is assumed to be dependent over time and that it can be described as a first-order Markov sequence with a transition probability matrix given by

$$\Pi = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix} \tag{26}$$

where $\pi_{ij} = \Pr\left(I_t = j \mid I_{t+1} = i\right)$. These transition probabilities simply mean that conditional on today being a non-violation (that is, $I_t = 0$), then the probability of tomorrow being a violation (that is, $I_{t+1} = 1$) is π_{01}. The probability of tomorrow being a violation given today is also a violation is: $\pi_{11} = \Pr\left(I_t = 1 \text{ and } I_{t+1} = 1\right)$. The Markov chain reflects the existence of an order one memory in the sequence of exceedances. In the case of the hit sequence being independent over time, then the probability of a violation tomorrow does not depend on today being a violation or not. In this case, the null hypothesis in

the independence test is

$$H_0 : \pi_{01} = \pi_{11} = \pi \tag{27}$$

Ultimately, one is interested in simultaneously testing if the VaR violations are independent and the average number of violations is correct. The conditional coverage test jointly examines whether the percentage of exceptions is statistically equal to the one expected ($\hat{p} = p$) and the serial independence of the exception indicator. In this test, the null hypothesis takes the form:

$$H_0 : \pi_{01} = \pi_{11} = p \tag{28}$$

Thus, under the null hypothesis of the expected proportion of exceptions equals p and the failure process is independent, the appropriate likelihood ratio test statistic is of the form:

$$LR_{CC} = 2\left(\log\left(\hat{\pi}_{01}^{T_{01}} \left(1 - \hat{\pi}_{01}\right)^{T_{00}} \hat{\pi}_{11}^{T_{11}} \left(1 - \hat{\pi}_{11}\right)^{T_{10}} \right) - \log\left(p^{T_{01}+T_{11}} \left(1-p\right)^{T_{00}+T_{10}} \right) \right) \tag{29}$$

where $\hat{\pi}_{ij} = T_{ij} / (T_{i0} + T_{i1})$, T_{ij} denote the number of days when condition j occurred assuming that condition i occurred on the previous day (1 if exceedance occurs, 0 if no exceedance occurs).

Under the null hypothesis the likelihood ratio statistic, LR_{CC}, has an asymptotically Chi-square distribution, with two degree of freedom. The Christoffersen's test enables the use to test both coverage and independence hypotheses at the same time. Moreover, it checks if the VaR model fails a test of both hypotheses combined. This approach makes us enable to test each hypothesis separately, and therefore establish where the model failure arises.

3. Data

In this study, twelve major international stock indices in the world are analysed. The set include: S&P500 (US; SPX), FTSE 100 (UK; FTSE), DAX 30 (Germany; GDAXI), CAC 40 (France; FCHI), Swiss Market Index (Switzerland; SMI), Euro Stoxx 50 (Europe; STOXX 50), S&P/TSX Composite (Canada; GSPTSE), Nikkei 225 (Japan; N225), KOSPI 200 (South Korea; KS11), Hang Seng (Hong Kong; HSI), Shanghai Composite (China, SSE) and Sensex (India, BSESN). The motivation for selection of these stock indices is to examine the reliability of the proposed VaR forecast models for the major world stock indices in periods of financial distress and turmoil. Therefore, the data covers the period for 1st January 2006 to 31st July 2020, covering the 2008 global financial crisis, the 2011 European financial crisis and the current COVID-19 pandemic period. Each price series is expressed in the local currency. The daily percentage log-returns on assets over the sample period are used. Daily log-returns are computed as 100 times the difference of the log prices, *i.e.* $r_t = 100 \log\left(P_t / P_{t-1}\right)\%$, where P_t is the adjusted closing price (value) on day *t*. The daily log returns for the equity market were calculated from the adjusted daily closing prices downloaded from https://markets.businessinsider.com/indices.

Figure 1 presents the time plots of the log-return series which shows evidence of volatility clustering and extreme price movements in the returns. From the
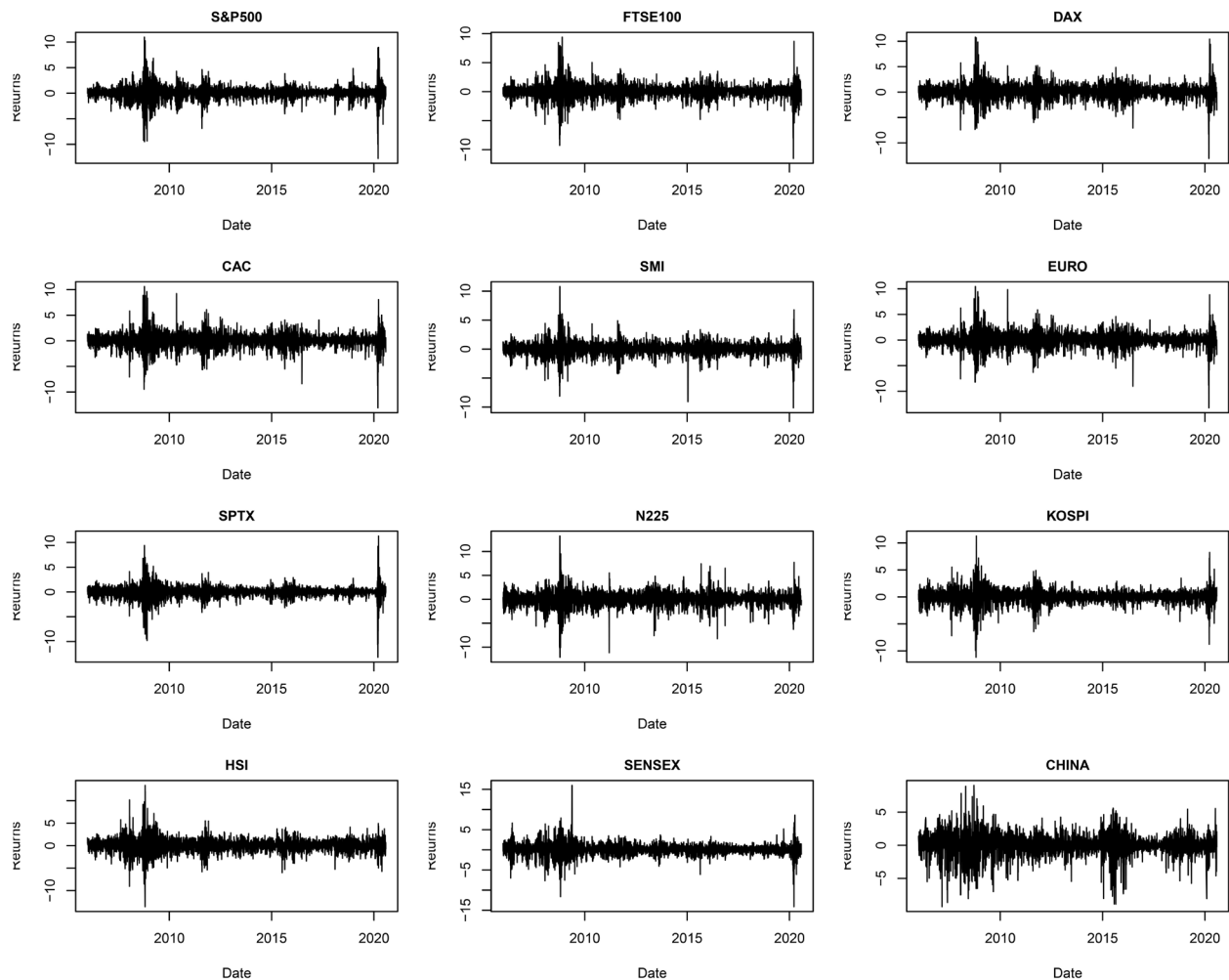
Figure 1. Daily log-return plots of the twelve major stock market indices for the period starting from January 1, 2006 to July 31, 2020.

figure, we can also see the effects of the 2008 global financial crisis, the 2011 European financial crisis as well as the 2019 COVID pandemic shocks in March 2020. All the ten stock market indices display similar patterns of volatility clustering dynamics over time and extreme price jumps. Table 1 shows the summary statistics and statistical test results computed over the in-sample, out-of-sample and full sample periods for all stock market indices considered in this paper. All the stock market indices record a positive mean close to zero except for CAC40, EURO and N225 in the in-sample, FTSE in the out-of-sample and EURO again in the full sample that have a negative mean. The log-return series for each stock market index are far from being normally distributed as indicated by their negative skewness and high excess-kurtosis. The Jarque-Bera normality test also confirms that all stock market indices are non-normally distributed. The Augmented Dickey-Fuller (ADF) results further show that all series are stationary. The Ljung-Box Q statistic tests the null hypothesis of no serial correlation and is calculated using up to 5 lags. A significant Q statistic for returns implies that we reject the null hypothesis of no serial correlation in returns, while a

significant Q statistic for the squared return series implies that the null hypothesis of homoscedastic returns is rejected. From the results, it is observed that the Ljung-Box Q statistics are significant for most returns as well as squared return series. Thus, Ljung-Box tests confirm presence of serial correlation in squared returns series. Again, the null hypothesis of no ARCH effects is rejected, by the Lagrange multiplier test for AutoRegressive Conditional Heteroscedasticity (ARCH-LM test) thus confirming presence of ARCH effects in all return series. The presence of serial correlation supports the need to filter the heteroscedasticity in all series using an appropriate conditional heteroscedastic model.

Table 1. Descriptive summary statistics and statistical tests results for the log-returns (%) of twelve stock market indices for in-sample, out-of-sample and full sample period.

Statistics	S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
In-sample												
No. of Obs.	2670	2713	2695	2727	2656	2740	2659	2567	2604	2592	2607	2540
Min	−9.4695	−9.2656	−7.4335	−9.4715	−9.0704	−9.1009	−9.7879	−12.1110	−11.1720	−13.5820	−11.6044	−9.2562
Max	10.957	9.3843	10.7975	10.5946	10.7876	10.4376	9.3703	13.2346	11.2844	13.4068	15.9899	9.0343
Mean	0.0203	0.0071	0.0252	−0.0024	0.0018	−0.0058	0.0093	−0.0003	0.0138	0.0123	0.0417	0.0351
Std.Dev	1.2900	1.2425	1.453	1.5007	1.1940	1.5025	1.1913	1.6284	1.3302	1.6257	1.5438	1.8063
Skew	−0.3334	−0.1345	−0.0281	−0.0269	−0.2237	−0.0678	−0.6611	−0.4551	−0.5644	0.0301	0.0913	−0.6087
Kurt.	10.3369	7.2749	5.5163	5.7007	7.8009	5.3376	9.9723	7.1498	8.6935	8.4008	8.3019	3.5603
Out-of-sample												
No. of Obs.	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Min	−12.765	−11.512	−13.055	−13.098	−10.1339	−13.2405	−13.1761	−8.2529	−8.7669	−5.7202	−14.1017	−8.0391
Max	8.9683	8.6668	10.4143	8.0561	6.7805	8.8343	11.29446	7.7314	8.2513	4.9249	8.5947	5.5543
Mean	0.0397	−0.0141	0.0142	0.0085	0.02314	0.0042	0.0107	0.0319	0.01304	0.0184	0.0305	0.0133
Std.Dev	1.2682	1.0703	1.2578	1.1951	0.9772	1.1813	1.1368	1.2266	1.0725	1.1344	1.1918	1.0822
Skew	−1.1713	−1.5436	−0.7376	−1.7609	−1.2942	−1.7337	−2.1783	−0.3285	−0.3119	−0.4915	−1.9551	−0.7165
Kurt.	23.2205	22.1706	22.0029	21.6635	17.3195	23.7462	49.344	7.6226	13.0891	2.7614	30.1745	6.8474
Full-sample												
No. of Obs.	3670	3713	3695	3727	3656	3740	3659	3567	3604	3592	3607	3540
Min	−12.765	−11.512	−13.055	−13.098	−10.134	−13.240	−13.176	−12.111	−11.172	−13.582	−14.102	−9.256
Max	10.957	9.384	10.797	10.595	10.788	10.438	11.294	13.235	11.284	13.407	15.990	9.034
Mean	0.0256	0.0014	0.222	0.00055	0.00766	−0.00311	0.00967	0.00873	0.014	0.014	0.0386	0.0289
Std.Dev	1.284	1.198	1.403	1.425	1.138	1.424	1.176	1.5263	1.264	1.505	1.455	1.634
Skew	−0.554	−0.40275	−0.1654	−0.3035	−0.413	−0.325	−1.036	−0.45176	−0.528	−0.033	−0.2172	−0.6419
Kurt.	13.6775	9.999	8.5298	8.098	9.482	7.966	19.380	7.711	9.683	8.719	11.516	4.549
JB	28832.83*	15590.73*	11234.93*	10256.82*	13818.48*	9968.14*	57985.92*	8972.53*	14269.23*	11394.75*	19987.47*	3301.71*
ADF	−15.026*	−15.526*	−15.031*	−15.424*	−15.847*	−15.386*	−14.92*	−14.825*	−15.107*	−14.665*	−13.937*	−13.437*
$Q(5)$	86.857*	28.647*	5.465	18.218*	46.115*	18.982*	34.64*	9.874	10.152	7.793	10.179	19.988*
$Q^2(5)$	2281.50*	1342.60*	680.77*	1026.9*	1644.9*	946.48*	2425*	1675.9*	1685.1*	1859.4*	747.07*	514.49*
LM(5)	1086.8*	722.86*	438.11*	579.30*	754.93*	554.94*	994.98*	788.80*	838.20*	851.46*	430.42*	305.83*

JB is the test statistic of Jarque-Bera test, ADF is the test statistic of Augmented Dickey-Fuller test, $Q(5)$ is the test statistic of Ljung-Box test on returns series, $Q^2(5)$ is the test statistic of Ljung-Box test on squared returns series and LM is the test statistic of Lagrange multiplier test for autoregressive conditional heteroscedasticity (ARCH-LM test), with * stands for significance at 0.01%.

4. Empirical Results

4.1. In-Sample Analysis

First an in-sample analysis is considered, where the GARCH-type models are fitted to the in-sample data. An approximately 70% log-returns (in percent) are used for the estimation and run the backtest over 1000 (about 4 years) out-of-sample log returns for the period from August, 01, 2016 to July, 31, 2020 (the full data set starts on January, 01, 2006). Each model is estimated on a rolling window basis and both the density and one-step-ahead log-returns forecasts are obtained. The model parameters are updated every 20th (monthly) observations. This frequency was selected in order to speed up the computations. Similar results were obtained for a subset of stocks when the parameters were updated every day. This is also in line with the observations of [41], who noted that in the context of GARCH models, that the performance of VaR forecasts is not affected significantly when moving a daily updating frequency to a weekly or monthly updating frequency. It is important to note that, while the parameters are updated every 20 observations, the density and downside risk measures are computed every day.

As we are interested in the volatility dynamics, as a first step we de-mean the stock indices return series and remove autoregressive effects in the data using a first order autoregressive model, AR(1)-filter and estimate the models on the residuals. As noted, the log-returns are skewed and leptokurtic. Thus, to account for the excess kurtosis, skewness and the dynamics of fluctuations typical in the financial time series data, we consider different error distributions including Student's-*t* distribution, skewed Student's-*t* distribution and Generalized error distribution (GED). The standard AR(1)-GARCH(1, 1) model specification is fitted to all return series with the different error distributions. Table 2 presents the AIC and BIC values for fitted AR(1)-GARCH(1, 1) model with different error distributions. Overall the skewed Student's *t* distribution and Generalized error distributions fits well majority of stock market indices returns. The skewed Student's-*t* distribution accounts for the excess skewness and kurtosis typical of financial time series data.

Table 2. Criterion for selecting the appropriate error distribution for the log returns.

	S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
Dist.						*Student-t distribution*						
AIC	2.749	2.861	3.269	3.329	2.792	3.337	2.636	3.492	2.963	3.371	3.279	3.696
BIC	2.762	2.874	3.282	3.342	2.805	3.350	2.649	3.506	2.977	3.385	3.293	3.709
						skewed Student-t distribution						
AIC	2.741	2.855	3.264	3.325	2.785	3.334	2.610	3.482	2.955	3.369	**3.279**	3.690
BIC	2.756	2.870	3.279	**3.339**	**2.801**	3.349	**2.626**	**3.498**	2.971	3.385	3.294	3.707
						Generalized error distribution						
AIC	2.739	2.856	3.262	3.328	2.795	3.334	2.635	3.489	2.957	3.367	3.282	3.687
BIC	**2.752**	**2.869**	**3.275**	3.341	2.808	**3.347**	2.649	3.502	**2.970**	**3.380**	3.295	**3.701**

For brevity, the AR(1)-SGARCH(1, 1), AR(1)-EGARCH(1, 1), AR(1)-GJRGARCH(1, 1), AR(1)-APARCH(1, 1) and AR(1)-CSGARCH(1, 1) models are used to filter conditional volatilities in all return series and estimate the out-of-sample VaR forecasts. Table 3 reports the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values of all the GARCH-type models considered in the study with error distributions; skewed Student's *t*-distribution and generalized error distribution. The AR(1)-EGARCH(1, 1) model with skewed Student's *t* distribution is selected as the optimal GARCH specification with the smallest AIC and BIC values for modelling volatility dynamics in all stock indices returns. According to the results reported in Table 4, the expected daily index return is not statistically different from zero for all stock return indices. However, the index for FTSE, CAC, SMI and EURO give a positive return. The estimated mean return for S&P, CAC, N225 and BSESN are statistically significant at 5% level. Volatility persistence in the return series for each of the stock indices is described by both α_1 and β_1 terms as reported in Panel B of Table 4. The parameter α_1 represents the lagged squared residuals, while β_1 is the lagged conditional variance term in the EGARCH model. Volatility is said to be persistent if the sum of the two volatility terms is close to unity, less persistent if less than unity and explosive if greater than unity. The coefficients for all the stock market indices are significant at 5% level of significance in favour of the EGARCH(1, 1) model. The implication of these results is that, all the stock indices returns show no evidence of long-memory in their respective return series. This means that shocks to volatility tend to decay quickly, implying that positive volatility do not have a strong predictive power on current volatility. These results are, however, conditional on the model specification and the distribution assumption made in the estimation. A number of empirical research in stock market returns distribution points to significant leverage effect, where higher volatility tends to follow negative returns. Asymmetries in the distribution of returns may arise as a result of shocks due to systemic risk factors that affect the cross-section of returns, or because of country-specific shocks. The results reveals significant positive leverage effect for all return series. Similarly, the parameters of skew and shape are statistically significant for all the stock indices series.

Panel C of Table 4 reports Ljung-Box test results for standardized residual series and squared standardized residual series as well as Lagrange multiplier tests for autoregressive conditional heteroscedasticity (ARCH-LM test) on the residuals. The Ljung Box results on standardized residuals up to 5 lags are significant for all stock residuals except for SMI and SSE at 5% level. For the squared residuals the results are also significant for all stock indices except for KOSPI and HSI. The ARCH-LM test confirms that no ARCH effects are present in the standardized residuals of most stock indices except for KOSPI, HSI and BSESN. Therefore the AR(1)-EGARCH(1, 1) model sufficiently filters the serial autocorrelation and conditional volatility dynamics present in stock indices returns effectively producing standardized residuals that are closer to being independently

and identically distributed (i.i.d.) compared to the original log-return series. However, the fitted GARCH-type models fails to capture extreme observations experienced in the stock markets.

In the next step, the GARCH-EVT model is utilized in estimating VaR forecasts. The standardized residuals from the fitted AR(1)-EGARCH(1, 1) model are approximately i.i.d. which is a standard requirement for extreme value theory to be applied in modelling extreme observations. The Peak over threshold (POT) approach is used to model the tail behaviour of standardized residuals of stock market indices returns. A threshold value is set at 90% quantile of the in-sample observations to estimate the parameters of the generalized Pareto distribution (GPD). Table 5 reports threshold values, number of exceedances and parameter estimates of the fitted GPD with their corresponding standard errors enclosed in brackets. The shape parameter (ξ) is positive and significantly different from zero except for S&P500, FTSE, S&PTX, KOSPI, HSI and SSE indicating heavy-tailed distributions and a finite variance. This also implies that tail distributions of stock market indices belong to Frechet class which is heavy-tailed. The scale parameter is positive and significant for all the stock market indices.

Table 3. AIC and BIC values for the fitted GARCH-type models for stock market indices returns.

		S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
Dist.						*Skewed Student-t distribution*							
GARCH	AIC	2.741	2.855	3.264	3.325	2.785	3.334	2.610	3.482	2.955	3.369	3.279	3.690
	BIC	2.756	2.870	3.279	3.3398	2.801	3.349	2.626	3.498	2.971	3.385	3.294	3.707
EGARCH	AIC	2.684	2.806	3.2198	3.261	2.734	3.269	2.577	3.455	2.933	3.356	3.257	3.691
	BIC	2.702	2.824	3.237	3.278	2.751	3.286	2.595	3.474	2.951	3.375	3.275	3.709
GJRGARCH	AIC	2.694	2.809	3.228	3.277	2.744	3.287	2.585	3.463	2.933	3.357	3.256	3.691
	BIC	2.712	2.826	3.246	3.295	2.762	3.304	2.603	3.481	2.951	3.375	3.274	3.709
CSGARCH	AIC	2.712	2.852	3.265	3.325	2.784	3.335	2.611	3.483	2.955	3.370	3.280	3.692
	BIC	2.761	2.872	3.285	3.345	2.804	3.354	2.631	3.503	2.976	3.390	3.301	3.713
APARCH	AIC	2.678	2.801	3.217	3.259	2.728	3.268	2.575	3.454	2.931	3.353	3.254	3.691
	BIC	2.698	2.821	3.237	3.279	2.748	3.288	2.595	3.474	2.951	3.374	3.275	3.712
Dist.						*Generalized error distribution*							
GARCH	AIC	2.739	2.856	3.262	3.328	2.795	3.334	2.635	3.489	2.957	3.367	3.282	3.687
	BIC	2.752	2.869	3.275	3.341	2.908	3.347	2.649	3.502	2.970	3.380	3.295	3.701
EGARCH	AIC	2.699	2.815	3.228	3.275	2.754	3.280	2.612	3.465	2.939	3.354	3.266	3.689
	BIC	2.715	2.830	3.243	3.291	2.7699	3.295	2.627	3.481	2.955	3.370	3.281	3.705
GJRGARCH	AIC	2.704	2.816	3.234	3.288	2.764	3.294	2.617	3.471	2.939	3.355	3.265	3.688
	BIC	2.7195	2.831	3.249	3.304	2.997	3.309	2.633	3.487	2.955	3.370	3.281	3.704
CSGARCH	AIC	2.7395	2.853	3.263	3.328	2.795	3.335	2.636	3.489	2.957	3.367	3.283	3.689
	BIC	2.757	2.871	3.280	3.346	2.812	3.352	2.654	3.507	2.975	3.385	3.301	3.707
APARCH	AIC	2.693	2.810	3.225	3.272	2.749	3.278	2.610	3.465	2.937	3.351	3.262	3.688
	BIC	2.711	2.828	3.242	3.2898	2.767	3.296	2.627	3.483	2.955	3.369	3.281	3.707

Table 4. AR(1)-EGARCH(1, 1)-sstd estimation results for stock market indices returns.

	S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
					Panel A: Conditional mean equation							
μ	−0.0202	0.0057	−0.0175	0.0113	0.0087	0.0185	−0.0089	−0.0092	−0.0012	−0.0166	−0.0332	−0.0397
	(0.0746)	(0.6844)	(0.2696)	(0.6102)	(0.6107)	(0.2493)	(0.5036)	(0.7154)	(0.9478)	(0.4585)	(0.0929)	(0.1003)
ϕ_1	−0.0635*	−0.0299	0.0112	−0.0440*	0.0294	−0.0352	0.0233	−0.0402*	−0.0077	0.0075	0.0789*	0.0088
		(0.1059)	(0.6389)		(0.1384)	(0.0741)	(0.2212)		(0.6947)	(0.6997)		(0.4965)
					Panel B: Conditional variance equation							
ω	−0.0026	0.0008	0.0123*	0.0137*	−0.0018	0.0134*	−0.0035	0.0232*	0.0043	0.0094*	0.0097*	0.0109*
	(0.5199)	(0.8349)			(0.7086)		(0.2152)		(0.1809)			
α_1	0.1954*	0.1584*	0.1372*	0.1864*	0.1833*	0.1850*	0.1237*	0.1162*	0.0961*	0.0699*	0.1129*	0.0026*
β_1	0.9769*	0.9750*	0.9739*	0.9693*	0.9628*	0.9718*	0.9866*	0.9639*	0.9823*	0.9846*	0.9791*	0.9900*
γ_1	0.1253*	0.1425*	0.1263*	0.1089*	0.1526*	0.1075*	0.1076*	0.1752*	0.1490*	0.1507*	0.1756*	0.1438*
skew	1.2052*	1.1541*	1.1540*	1.1578*	1.1741*	1.1365*	1.3209*	1.1510*	1.1489*	1.0716*	1.0619*	1.1051*
shape	6.5448*	10.3353*	24.8099*	9.0280*	8.8667*	8.7654*	13.7407*	12.4688*	7.8757*	8.9537*	6.8498*	4.9935*
					Panel C: Diagnostics and heteroscedastic tests							
Loglik	−3575.08	−3798.49	−4330.73	−4438.05	−3622.05	−4470.67	−3418.13	−4426.92	−3810.68	−4341.88	−4237.70	−4679.94
AIC	2.684	2.806	3.2198	3.2608	2.7335	3.269	2.577	3.455	2.9329	3.3564	3.2572	3.6913
BIC	2.702	2.824	3.2373	3.2781	2.7512	3.286	2.595	3.474	2.951	3.3745	3.275	3.709
$Q(5)$	3.6089	2.0227	2.1944	4.3065	7.4562	3.9546	2.5234	3.1462	2.857	2.198	2.1103	7.330
	(0.2947)	(0.7048)	(0.6548)	(0.1807)	(0.0127)	(0.2328)	(0.5597)	(0.3954)	(0.4682)	(0.6537)	(0.6794)	(0.0143)
$Q^2(5)$	5.807	3.958	3.628	1.875	1.8946	3.399	4.502	0.9711	7.141	9.119	4.6195	2.1849
	(0.0998)	(0.2594)	(0.3043)	(0.6483)	(0.6435)	(0.3390)	(0.1978)	(0.8662)	(0.0479)	(0.0155)	(0.1863)	(0.5754)
LM(5)	0.5185	4.880	1.633	0.3841	4.947	0.7143	1.3063	1.6715	8.2317	11.725	6.985	2.570
	(0.8781)	(0.1095)	(0.5582)	(0.9172)	(0.1057)	(0.8191)	(0.6447)	(0.5484)	(0.0177)	(0.0025)	(0.0351)	(0.3585)

*Significant at 5% level of significance and *p*-values are shown in parentheses.

Table 5. AR(1)-EGARCH(1, 1)-EVT estimation results for stock market indices returns.

	S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
Left Tail												
No. of Obs.	2670	2713	2695	2727	2656	2740	2659	2567	2604	2592	2607	2540
u	1.2657	1.2655	1.2709	1.2649	1.2489	1.2483	1.3653	1.3123	1.237	1.2689	1.1906	1.176
No. of Exceed.	267	272	270	273	266	274	266	257	261	260	261	254
ξ	−0.01147	−0.0590	0.00149	0.02926	0.06821	0.03189	−0.0257	0.0313	−0.1476	−0.0941	0.0500	−0.0343
(*s.e*)	(0.05305)	(0.04643)	(0.05714)	(0.02923)	(0.06821)	(0.03189)	(0.05504)	(0.04934)	(0.05165)	(0.05304)	(0.06573)	(0.07201)
β	0.72964	0.68182	0.64025	0.62709	0.60549	0.62645	0.63118	0.58458	0.77753	0.63272	0.60508	0.76105
(*s.e*)	(0.05909)	(0.05201)	(0.05345)	(0.05050)	(0.04818)	(0.05083)	(0.05199)	(0.04648)	(0.06228)	(0.05151)	(0.05460)	(0.07267)

4.2. Out-of-Sample Analysis

We now turn to an out-of-sample analysis where we compare the ability of the conditional EVT and GARCH-type models to correctly forecast the one-day ahead Value-at-Risk (VaR). We use out-of-sample data for backtesting; thus we have an in-sample of the return observations for the rolling window estimation procedure, containing the 2008 global financial crisis period, and we run the backtest over 1000 out-of-sample observations for a period starting from 1st June 2017, to 31st July 2020. VaR forecasts are also estimated following a rolling-window approach. The out-of-sample data is further divided into blocks of 500 and 1000 trading days to observe how the models behave for both shorter and longer periods of observation. To test the ability of our models to capture the true VaR, we compare the realization of the returns with the one-day ahead VaR forecasts at 95% and 99% risk levels. To that aim, we adopt the *UC* test of [39] and the *CC* test of [40] to evaluate the accuracy of each of the 5 models considered in terms of predicting accurate VaR forecasts at the 5% and 1% levels for all daily returns on the 11 stock indices. Table 6 presents the out-of-sample VaR Violations and *p*-values of the Unconditional Coverage (*UC*) test and Table 7 presents the one-day ahead VaR backtesting results computed using 1000 and 500 out-of-sample observations at the 5% and 1% risk levels.

Table 6 summaries the violation ratio percentages of the underlying VaR models and the *p*-values corresponding to the unconditional coverage (*UC*) tests for the 500 and 1000 blocks at 5% and 1% significance levels. The expected out-of-sample VaR violations for the 1000 window is 50 at 95% and 10 for 99% confidence level while for the 500 window is 25 at 95% and 5% for 99% confidence level respectively. Based on the proximity of the actual violation ratio to the expected violation ratio, the GARCH-EVT VaR model perfoms better than the standard GARCH(1, 1), EGARCH(1, 1), GJR-GARCH(1, 1) and APARCH(1, 1) VaR models. According to the *UC* test results, the VaR forecasts based on the GARCH-EVT model produce a rather accurate out-of-sample proportion of violations the highest number of times. However, GARCH-EVT model rejected the null of correct coverage five times at 5% risk level for the 1000 window. At the 1% risk level, these differences are mostly significant; we obtain *p*-values close to 1 with the EVT-based model. Overall, the one-day ahead backtesting results demonstrate the superiority of the GARCH-EVT models over the GARCH-type models.

Table 6. Out-of-sample VaR Violations and *p*-values of the Unconditional Coverage (*UC*) test on the VaR forecasts.

Model	S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
500 window						5% *level of significance*						
GARCH	6.2%	5%	7.2%	6.8%	6%	5.8%	5.6%	5.8%	6.2%	6.4%	6%	4.6%
UC: *p*-value	(0.235)	(1.000)	**	(0.079)	(0.319)	(0.423)	(0.546)	(0.423)	(0.235)	(0.168)	(0.319)	(0.678)
EGARCH	5.6%	4.6%	5.8%	5.8%	6.8%	5.4%	4.2%	5.2%	6.0%	6.4%	4.4%	4.4%
UC: *p*-value	(0.546)	(0.678)	(0.423)	(0.423)	(0.079)	(0.685)	(0.399)	(0.838)	(0.319)	(0.168)	(0.530)	(0.530)
GJR-GARCH	5.8%	5.2%	5.8%	5.0%	5.2%	4.6%	4.4%	5.0%	6.0%	6.0%	4.8%	4.6%

Continued

UC: *p*-value	(0.423)	(0.838)	(0.423)	(1.000)	(0.838)	(0.678)	(0.530)	(1.000)	(0.319)	(0.319)	(0.836)	(0.678)
APARCH	6.2%	4.5%	5.6%	5.8%	6.2%	5.2%	4.0%	4.8%	5.6%	6.2%	4.6%	4.4%
UC: *p*-value	(0.235)	(0.678)	(0.546)	(0.423)	(0.235)	(0.838)	(0.288)	(0.836)	(0.546)	(0.235)	(0.673)	(0.530)
GARCH-EVT	3.8%	4.2%	5.0%	5.4%	5.0%	4.8%	3.6%	5.0%	5.0%	5.8%	5.0%	4.0%
UC: *p*-value	(0.199)	(0.399)	(1.000)	(0.685)	(1.000)	(0.836)	(0.131)	(1.000)	(1.000)	(0.423)	(1.000)	(0.288)
						1% *level of significance*						
GARCH	1.8%	1.8%	2.2%	1.8%	1.4%	1.8%	1.8%	1%	1.8%	1.6%	1.4%	1.4%
UC: *p*-value	(0.106)	(0.106)	**	(0.106)	(0.397)	(0.106)	(0.106)	(1.000)	(0.106)	(0.215)	(0.397)	(0.397)
EGARCH	1.8%	1.8%	2.2%	2.2%	1.6%	2.6%	1.6%	1.2%	1.6%	1.4%	1.4%	1.6%
UC: *p*-value	(0.106)	(0.106)	**	**	(0.215)	**	(0.215)	(0.663)	(0.215)	(0.397)	(0.397)	(0.215)
GJR-GARCH	1.6%	2.0%	2.2%	2.2%	1.4%	2.2%	1.6%	1.0%	1.2%	1.2%	1.2%	1.4%
UC: *p*-value	(0.215)	**	**	**	(0.397)	**	(0.215)	(1.000)	(0.663)	(0.663)	(0.663)	(0.397)
APARCH	1.4%	1.8%	2.2%	2.0%	1.6%	2.4%	1.4%	0.8%	1.4%	1.4%	1.4%	1.6%
UC: *p*-value	(0.897)	(0.106)	**	**	(0.215)	**	(0.397)	(0.641)	(0.397)	(0.397)	(0.397)	(0.215)
GARCH-EVT	1.0%	1.8%	1.8%	1.8%	1.2%	1.6%	1.6%	0.8%	1.6%	1.2%	1.2%	1.2%
UC: *p*-value	(1.000)	(0.106)	(0.106)	(0.106)	(0.663)	(0.215)	(0.215)	(0.641)	(0.215)	(0.663)	(0.663)	(0.663)
1000 window						5% *level of significance*						
GARCH	4.8% (3)	4.4%	6.3%	5.3%	5.3%	5.3%	5.0%	4.4%	5.6%	5.6%	4.8%	4.1%
UC: *p*-value	(0.770)	(0.375)	(0.069)	(0.666)	(0.666)	(0.666)	(1.000)	(0.375)	(0.393)	(0.393)	(0.770)	(0.178)
EGARCH	5.1% (2)	4.3%	4.3%	4.3%	4.9%	3.9%	4.5%	4.6%	5.6%	5.9%	4.5%	4.0%
LR_{UC} *p*-value	(0.885)	(0.299)	(0.299)	(0.299)	(0.884)	(0.097)	(0.461)	(0.557)	(0.393)	(0.204)	(0.461)	(0.133)
GJR-GARCH	4.5% (4)	4.4%	4.9%	3.9%	4.6%	3.7%	4.2%	3.8%	5.3%	5.5%	3.9%	4.1%
UC: *p*-value	(0.461)	(0.375)	(0.884)	(0.097)	(0.557)	**	(0.233)	(0.070)	(0.666)	(0.475)	(0.097)	(0.178)
APARCH	4.9% (1)	3.9%	4.5%	4.4%	4.8%	4.1%	4.3%	4.0%	5.4%	5.7%	4.3%	4.0%
UC: *p*-value	(0.884)	(0.097)	(0.461)	(0.375)	(0.770)	(0.178)	(0.299)	(0.133)	(0.566)	(0.320)	(0.299)	(0.133)
GARCH-EVT	3.3% (5)	3.5%	3.7%	4.1%	4.2%	3.4%	4.0%	3.6%	4.6%	4.6%	3.9%	3.8%
UC: *p*-value	**	**	**	(0.178)	(0.233)	**	(0.133)	**	(0.557)	(0.557)	(0.097)	(0.069)
						1% *level of significance*						
GARCH	1.5% (4)	1.4% (5)	1.1% (2)	1.0% (1)	0.9% (1)	1.2%	1.5%	0.9%	1.6%	1.3%	0.9%	1.4%
UC: *p*-value	(0.139)	(0.231)	(0.754)	(1.000)	(0.746)	(0.538)	(0.139)	(0.746)	(0.097)	(0.362)	(0.746)	(0.231)
EGARCH	1.5% (4)	1.3% (4)	1.3% (5)	1.3% (5)	0.8% (3)	1.4%	1.2%	0.8%	1.5%	1.1%	1.1%	1.5%
UC: *p*-value	(0.139)	(0.362)	(0.362)	(0.362)	(0.510)	(0.231)	(0.538)	(0.510)	(0.139)	(0.754)	(0.754)	(0.139)
GJR-GARCH	1.2% (2)	1.2% (1)	1.1% (3)	1.2% (5)	0.8%	1.1%	1.2%	0.9%	1.2%	1.2%	0.8%	1.4%
UC: *p*-value	(0.538)	(0.538)	(0.754)	(0.538)	(0.510)	(0.754)	(0.538)	(0.746)	(0.538)	(0.538)	(0.510)	(0.231)
APARCH	1.3% (3)	1.2% (1)	1.1% (3) (2)	1.1% (3)	0.9% (1)	1.3%	1.3%	0.8%	1.4%	1.1%	1.0%	1.5%
UC: *p*-value	(0.362)	(0.538)	(0.754)	(0.754)	(0.746)	(0.362)	(0.362)	(0.510)	(0.231)	(0.754)	(1.000)	(0.139)
GARCH-EVT	1.0% (1)	1.2% (1)	0.9% (1)	0.9% (2)	0.8% (3)	0.8%	1.3%	0.8%	1.3%	0.9%	0.9%	1.3%
UC: *p*-value	(1.000)	(0.538)	(0.746)	(0.746)	(0.510)	(0.510)	(0.362)	(0.510)	(0.362)	(0.746)	(0.746)	(0.362)

The table reports the violations ratios and *p*-values of the unconditional coverage (*UC*) test for the one-day ahead 5% and 1% VaR. ** stands for the null hypothesis of the Kupiec's test being rejected under the significance level 5%. In other words, the violations rates are either underestimated or over estimated.

Table 7. Conditional Coverage (*CC*) tests of Christoffersen (1998) results of backtesting.

Model	S&P	FTSE	DAX	CAC	SMI	EURO	S&P_TSX	N225	KOSPI	HSI	BSESN	SSE
500 window						5% *level of significance*						
GARCH	5.501	8.000	5.380	14.657	5.835	6.032	0.487	3.343	3.425	2.348	3.334	0.117
CC: *p*-value	(0.064)	**	(0.068)	**	(0.054)	**	(0.784)	(0.188)	(0.180)	(0.309)	(0.189)	(0.915)
EGARCH	0.487	3.097	1.714	1.714	4.393	1.734	1.874	0.154	1.759	2.348	0.395	0.395
CC: *p*-value	(0.784)	(0.213)	(0.424)	(0.424)	(0.111)	(0.420)	(0.392)	(0.926)	(0.415)	(0.309)	(0.821)	(0.821)
GJR-GARCH	3.343	4.264	1.714	5.009	1.903	1.012	0.395	2.638	1.759	1.759	0.066	0.177
CC: *p*-value	(0.188)	(0.119)	(0.424)	(0.082)	(0.386)	(0.603)	(0.821)	(0.267)	(0.415)	(0.415)	(0.967)	(0.915)
APARCH	3.425	3.097	1.673	3.734	3.588	4.264	2.543	2.469	0.487	2.007	0.177	0.395
CC: *p*-value	(0.180)	(0.213)	(0.433)	(0.155)	(0.166)	(0.119)	(0.280)	(0.291)	(0.784)	(0.367)	(0.915)	(0.821)
GARCH-EVT	1.749	4.523	0.497	7.098	4.739	5.339	4.295	0.060	0.426	1.605	2.016	1.176
CC: *p*-value	(0.417)	(0.104)	(0.780)	**	(0.094)	(0.069)	(0.117)	(0.970)	(0.808)	(0.488)	(0.365)	(0.555)
						1% *level of significance*						
GARCH	2.943	2.943	5.869	2.943	0.918	2.943	4.739	0.101	2.943	4.104	0.918	3.805
CC: *p*-value	(0.230)	(0.230)	(0.053)	(0.230)	(0.632)	(0.230)	(0.094)	(0.951)	(0.230)	(0.128)	(0.632)	(0.149)
EGARCH	2.943	2.943	5.869	5.915	1.799	9.669	4.104	0.336	1.799	0.918	0.918	4.104
CC: *p*-value	(0.230)	(0.230)	(0.053)	(0.052)	(0.407)	**	(0.128)	(0.845)	(0.407)	(0.632)	(0.632)	(0.128)
GJR-GARCH	1.799	4.323	5.869	5.915	0.918	5.915	4.104	0.101	0.336	0.336	0.336	3.805
CC: *p*-value	(0.407)	(0.115)	(0.053)	(0.052)	(0.632)	(0.052)	(0.128)	(0.951)	(0.845)	(0.845)	(0.845)	(0.149)
APARCH	0.918	2.943	5.869	4.323	1.799	7.702	3.805	0.282	0.918	0.918	0.918	4.104
CC: *p*-value	(0.632)	(0.230)	(0.053)	(0.115)	(0.407)	**	(0.149)	(0.869)	(0.632)	(0.632)	(0.632)	(0.128)
GARCH-EVT	0.101	2.943	2.906	2.943	0.336	1.798	4.104	0.282	1.798	0.336	0.336	3.901
CC: *p*-value	(0.951)	(0.229)	(0.234)	(0.229)	(0.845)	(0.407)	(0.128)	(0.869)	(0.407)	(0.845)	(0.845)	(0.142)
1000 window						5% *level of significance*						
GARCH	7.265	7.271	5.615	13.737	5.504	3.471	0.855	4.771	4.994	0.738	1.233	1.874
CC: *p*-value	**	**	(0.060)	**	(0.064)	(0.176)	(0.652)	(0.092)	(0.082)	(0.691)	(0.540)	(0.392)
EGARCH	0.747	1.102	1.102	1.807	2.579	4.068	0.996	0.348	1.878	1.699	0.544	2.354
CC: *p*-value	(0.688)	(0.576)	(0.576)	(0.405)	(0.275)	(0.131)	(0.608)	(0.840)	(0.391)	(0.428)	(0.762)	(0.308)
GJR-GARCH	2.224	4.771	2.579	6.006	0.751	4.210	1.454	3.434	1.861	0.831	2.896	1.874
CC: *p*-value	(0.329)	(0.092)	(0.275)	(0.050)	(0.687)	(0.122)	(0.483)	(0.180)	(0.394)	(0.660)	(0.235)	(0.392)
APARCH	4.391	4.068	2.325	1.397	2.898	4.518	1.093	2.502	0.781	1.171	1.093	2.354
CC: *p*-value	(0.111)	(0.131)	(0.313)	(0.497)	(0.235)	(0.104)	(0.579)	(0.286)	(0.677)	(0.557)	(0.579)	(0.308)
GARCH-EVT	10.532	7.396	5.591	6.843	6.085	14.207	3.328	4.952	1.929	0.707	5.861	3.501
CC: *p*-value	**	**	(0.061)	**	**	**	(0.189)	(0.084)	(0.381)	(0.702)	(0.053)	(0.174)
						1% *level of significance*						
GARCH(1, 1)	3.704	1.807	0.320	0.202	0.268	0.672	3.704	0.250	3.597	2.833	0.268	3.183
CC: *p*-value	(0.157)	(0.405)	(0.852)	(0.904)	(0.875)	(0.715)	(0.157)	(0.883)	(0.166)	(0.243)	(0.875)	(0.204)

Continued

EGARCH	7.823	1.173	1.147	1.173	0.563	1.835	2.669	0.547	3.704	0.343	0.343	3.704
CC: *p*-value	**	(0.556)	(0.564)	(0.556)	(0.755)	(0.399)	(0.263)	(0.761)	(0.157)	(0.842)	(0.842)	(0.157)
GJR-GARCH	2.669	0.672	0.320	0.672	0.563	0.343	2.669	0.250	0.672	0.672	0.563	3.183
CC: *p*-value	(0.263)	(0.715)	(0.852)	(0.715)	(0.755)	(0.842)	(0.263)	(0.883)	(0.715)	(0.715)	(0.755)	(0.204)
APARCH	2.833	0.672	0.320	0.343	0.268	1.173	2.833	0.547	3.183	0.343	0.202	3.704
CC: *p*-value	(0.243)	(0.715)	(0.852)	(0.842)	(0.875)	(0.556)	(0.243)	(0.761)	(0.204)	(0.842)	(0.904)	(0.157)
GARCH-EVT	2.973	0.672	0.249	0.268	0.563	0.563	2.833	0.547	1.173	0.268	0.268	2.833
CC: *p*-value	(0.226)	(0.715)	(0.883)	(0.875)	(0.755)	(0.755)	(0.243)	(0.761)	(0.556)	(0.875)	(0.875)	(0.243)

The table reports the *p*-values of the conditional coverage (*CC*) test of Christoffersen (1998) for the one-day ahead 5% and 1% VaR. ** stands for the null hypothesis of the Christoffersen's test being rejected under the significance level 5%.

Table 7 presents conditional coverage (*CC*) test statistic values and *p*-values in parentheses for the 500 and 1000 windows at two levels of significance of 95% and 99% which are considered to reflect extreme market conditions. For the conditional coverage test, likewise, a good model should accept the null hypothesis, that is, correctly identifying the number of violations and being independent. The results of the *CC* test checking both the correct coverage and the lack of dependence of order one in VaR violations seem to support the GARCH-EVT model. The conditional EVT model yield the highest success rate for both tests with higher *p*-values that are statistically significant in most of the cases demonstrating the supremacy of the model over the other competing models. The poorest fit corresponds to the 5% risk level because, the null of proper specification had to be rejected for S&P 500, FTSE 100, CAC, SMI and EURO. As in the case of the *UC* test, the *CC* test results demonstrate that the GARCH VaR model rendered the worst fit with the null rejected in three cases at 5% level of significance for both the 500 and 1000 windows.

In general, we observe that conditional EVT-based models give the best one day-ahead VaR forecasts according to the *UC* and *CC* backtesting results. Moreover, an EGARCH(1, 1) specification leads to a substantial reduction in the rejection frequencies. A heavy-tailed conditional distribution is of fundamental importance for both the GARCH-EVT and GARCH specifications, and delivers excellent results at both risk levels. Thus, we conclude that it is feasible to discriminate between the estimation methods based on an analysis of the VaR forecast accuracy.

5. Conclusion

In recent times, VaR has become the most common risk measure used by financial institutions to assess market risk of financial assets. Since VaR models often focus on the behavior of asset returns in the left tail, it is important that the models are calibrated such that they do not underestimate or overestimate the proportion of outliers, as this will have significant effects on the allocation of economic capital for investments. Stock market indices are normally characte-

rized by high volatility and extreme price shocks unlike financial assets such as currencies exchange rates and securities market prices. The GARCH-EVT approach allows us to model the tails of the time-varying conditional return distribution. The conditional extreme value theory has been proved to be one of the most successful in estimating market risk. The implementation of this method in the framework of the POT model requires choosing a threshold return for fitting the generalized Pareto distribution. Threshold choice involves balancing bias and variance. The GARCH-EVT model performs relatively well in estimating the risk for all stock indices. Empirical backtesting results demonstrate that the conditional EVT and the E-GARCH skewed Student's *t* models are the most appropriate techniques in measuring and forecasting risk since they outperform the competing conventional methods and are ranked as the top two models in most cases. Backtesting procedures indicate that regardless of the choice of the tail, approximately the same accuracy of VaR prediction is provided. The GARCH-EVT model provides a significant improvement in forecasting value-at-risk over the widely used conventional GARCH models. This study may be extended by considering more robust models such as the MSGARCH-EVT-Copula model that can also capture the structural breaks and dependence structure between stock markets. The measurement of market risk can also be implemented using expected shortfall which is a coherent risk measure. Given that the financial markets are complex, dynamic and dependent on other markets, selection of diversified investment portfolio is another important area for further research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bień-Barkowska, K. (2020) Looking at Extremes without Going to Extremes: A New Self-Exciting Probability Model for Extreme Losses in Financial Markets. *Entropy*, **22**, 789. https://doi.org/10.3390/e22070789

[2] Abad, P., Benito, S. and Lpez, C. (2014) A Comprehensive Review of Value at Risk Methodologies. *The Spanish Review of Financial Economics*, **12**, 15-32. https://doi.org/10.1016/j.srfe.2013.06.001

[3] Scarrott, C. and MacDonald, A. (2012) A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. *REVSTAT-Statistical Journal*, **10**, 33-60.

[4] McNeil, A.J. and Frey, R. (2000) Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance*, **7**, 271-300. https://doi.org/10.1016/S0927-5398(00)00012-8

[5] Fernandez, V. (2003) Extreme Value Theory and Value at Risk. *Revista de Análisis Económico*, **18**, 86-102.

[6] Gencay, R. and Selcuk, F. (2004) Extreme Value Theory and Value-at-Risk: Relative Performance in Emerging Markets. *International Journal of Forecasting*, **20**, 287-303. https://doi.org/10.1016/j.ijforecast.2003.09.005

[7] Chan, K.F. and Gray, P. (2006) Using Extreme Value Theory to Measure Val-

ue-at-Risk for Daily Electricity Spot Prices. *International Journal of Forecasting*, **22**, 283-300. https://doi.org/10.1016/j.ijforecast.2005.10.002

[8] Ghorbel, A. and Trabelsi, A. (2008) Predictive Performance of Conditional Extreme Value Theory in Value-at-Risk Estimation. *International Journal of Monetary Economics and Finance*, **1**, 121-148. https://doi.org/10.1504/IJMEF.2008.019218

[9] Marimoutou, V., Raggad, B. and Trabelsi, A. (2009) Extreme Value Theory and Value at Risk: Application to Oil Market. *Energy Economics*, **31**, 519-530. https://doi.org/10.1016/j.eneco.2009.02.005

[10] Allen, D.E., Singh, A.K. and Powell, R.J. (2011) Extreme Market Risk—An Extreme Value Theory Approach. The Econometric Society.

[11] Soltane, H.B., Karaa, A. and Bellalah, M. (2012) Conditional VaR Using GARCH-EVT Approach: Forecasting Volatility in Tunisian Financial Market. *Journal of Computations & Modelling*, **2**, 95-115.

[12] Singh, A.K., Allen, D.E. and Robert, P.J. (2013) Extreme Market Risk and Extreme Value Theory. *Mathematics and Computers in Simulation*, **94**, 310-328. https://doi.org/10.1016/j.matcom.2012.05.010

[13] Karmakar, M. and Shukla, G.K. (2015) Managing Extreme Risk in Some Major Stock Markets: An Extreme Value Approach. *International Review of Economics & Finance*, **35**, 1-25. https://doi.org/10.1016/j.iref.2014.09.001

[14] Totic, S. and Bo-ovic, M. (2016) Tail Risk in Emerging Markets of South Eastern Europe. *Applied Economics*, **48**, 1785-1798. https://doi.org/10.1080/00036846.2015.1109037

[15] Zhang, Z. and Zhang, H.-K. (2016) The Dynamics of Precious Metal Markets VaR: A GARCH-EVT Approach. *Journal of Commodity Markets*, **4**, 14-27. https://doi.org/10.1016/j.jcomm.2016.10.001

[16] Omari, C.O., Mwita, P.N. and Waititu, A.G. (2017) Using Conditional Extreme Value Theory to Estimate Value-at-Risk for Daily Currency Exchange Rates. *Journal of Mathematical Finance*, **7**, 846-870. https://doi.org/10.4236/jmf.2017.74045

[17] Stoyanov, S., Loh, L. and Fabozzi, F.J. (2017) How Fat Are the Tails of Equity Market Indices. *International Journal of Finance & Economics*, **22**, 181-200. https://doi.org/10.1002/ijfe.1577

[18] Tabasi, H., Yousefi, V., Tamošaitienė, J. and Ghasemi, F. (2019) Estimating Conditional Value at Risk in the Tehran Stock Exchange Based on the Extreme Value Theory Using GARCH Models. *Administrative Sciences*, **9**, 40. https://doi.org/10.3390/admsci9020040

[19] Jadhav, D. and Ramanathan, T. (2009) Parametric and Non-Parametric Estimation of Value-at-Risk. *The Journal of Risk Model Validation*, **3**, 51. https://doi.org/10.21314/JRMV.2009.034

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

[21] Omari, C.O., Mwita, P.N. and Gichuhi, A.W. (2018) Currency Portfolio Risk Measurement with Generalized Autoregressive Conditional Heteroscedastic-Extreme Value Theory-Copula Model. *Journal of Mathematical Finance*, **8**, 457-477. https://doi.org/10.4236/jmf.2018.82029

[22] Cifter, A. (2011) Value-at-Risk Estimation with Wavelet-Based Extreme Value Theory: Evidence from Emerging Markets. *Physica A*: *Statistical Mechanics and Its Applica-*

tions, **390**, 2356-2367. https://doi.org/10.1016/j.physa.2011.02.033

[23] Huang, C.K., North, D. and Zewotir, T. (2017) Exchangeability, Extreme Returns and Value-at-Risk Forecasts. *Physica A*: *Statistical Mechanics and Its Applications*, **477**, 204-216. https://doi.org/10.1016/j.physa.2017.02.080

[24] Rocco, M. (2014) Extreme Value Theory in Finance: A Survey. *Journal of Economic Surveys*, **28**, 82-108. https://doi.org/10.1111/j.1467-6419.2012.00744.x

[25] Echaust, K. and Just, M. (2020) Value at Risk Estimation Using the GARCH-EVT Approach with Optimal Tail Selection. *Mathematics*, **8**, 114. https://doi.org/10.3390/math8010114

[26] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

[27] Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica Econometrica*, **59**, 347-370. https://doi.org/10.2307/2938260

[28] Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993) On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, **48**, 1779-1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

[29] Ding, Z., Granger, C.W. and Engle, R.F. (1993) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, **1**, 83-106. https://doi.org/10.1016/0927-5398(93)90006-D

[30] Taylor, S. (1986) Modeling Financial Time Series. John Wiley & Sons, Hoboken.

[31] Engle, R.F. and Lee, G. (1999) A Permanent and Transitory Component Model of Stock Return Volatility. Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger. Oxford University Press, New York.

[32] Hansen, B.E. (1994) Autoregressive Conditional Density Estimation. *International Economic Review*, **35**, 705-730. https://doi.org/10.2307/2527081

[33] Zhu, D. and Galbraith, J.W. (2010) A Generalized Asymmetric Student-t Distribution with Application to Financial Econometrics. *Journal of Econometrics*, **157**, 297-305. https://doi.org/10.1016/j.jeconom.2010.01.013

[34] Fernández, C. and Steel, M.F. (1998) On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, **93**, 359-371. https://doi.org/10.1080/01621459.1998.10474117

[35] Balkema, A.A. and de Haan, L. (1974) Residual Life Time at Great Age. *The Annals of Probability*, **2**, 792-804. https://doi.org/10.1214/aop/1176996548

[36] Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, **3**, 119-131. https://doi.org/10.1214/aos/1176343003

[37] Embrechts, P., Resnick, S.I. and Samorodnitsky, G. (1999) Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal*, **3**, 30-41. https://doi.org/10.1080/10920277.1999.10595797

[38] Hosking, J.R. and Wallis, J.R. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, **29**, 339-349. https://doi.org/10.1080/00401706.1987.10488243

[39] Kupiec, P.H. (1995) Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, **3**, 73-84. https://doi.org/10.3905/jod.1995.407942

[40] Christoffersen, P.F. (1998) Evaluating Interval Forecasts. *International Economic*

Review, **39**, 841-862. https://doi.org/10.2307/2527341

[41] Ardia, D., Bluteau, K., Boudt, K. and Catania, L. (2018) Forecasting Risk with Markov-Switching GARCH Models: A Large-Scale Performance Study. *International Journal of Forecasting*, **34**, 733-747. https://doi.org/10.1016/j.ijforecast.2018.05.004