The Predictive Performance of Extreme Value Analysis Based-Models in Forecasting the Volatility of Cryptocurrencies

Cyprian Omari*, Anthony Ngunyi

Department of Statistics and Actuarial Science, Dedan Kimathi University of Technology, Nyeri, Kenya
Email: *cyomari@dkut.ac.ke, antony.ngunyi@dkut.ac.ke

Abstract

This paper implements the analysis of volatility behaviour of the eight major cryptocurrencies (Bitcoin, Ethereum, Ripple, Litecoin, Monero, Stellar, Dash and Tether) for the period starting from October 13th 2015 to November 18th 2019. The GARCH-type models with heavy-tailed distributions are fitted to filter the conditional volatility exhibited by cryptocurrencies. Extreme value analysis based on the peak over threshold approach is then used to model the extreme tail behaviour of the cryptocurrencies. The predictive performance of the GARCH-EVT model in forecasting Value-at-Risk is evaluated at both 5% and 1% levels of significance. The backtesting results demonstrate the superiority of the GARCH-EVT model in both out-of-sample forecasts and goodness-of-fit properties to cryptocurrency returns and forecasting Value-at-Risk. Overall, the empirical results of this study recommend the heavy-tailed GARCH-EVT based model for modelling and forecasting the volatility of cryptocurrencies.

Keywords

Backtesting, Bitcoin, Cryptocurrency, Extreme Value Theory, GARCH-EVT, Value-at-Risk

1. Introduction

Cryptocurrencies have attracted a lot of attention since Bitcoin was first proposed by Nakamoto [1]. They are highly volatile and show extreme tail movements as compared to traditional financial markets and fiat currencies. This provides a new investment asset category to investors, practitioners, and policymakers in financial markets and portfolio management. Bitcoin is one of the

most traded and still, the largest cryptocurrency, representing about 62.24% of the total estimated cryptocurrencies capitalisation as of March 2021 (https://coinmarketcap.com) [2]. As of March 28, 2021, the cryptocurrencies market capitalization was valued at about US $1517b. Remarkable growth has also been witnessed in other important digital currencies like Ethereum, Ripple, and Litecoin which are among the top ten cryptocurrencies by market capitalization. Despite being largely unregulated by government institutions, cryptocurrency prices and exchanges exhibit most stylized facts from established exchanges [3]. Nevertheless, these cryptocurrencies are characterized by periods of high volatility, large shocks and extreme price jumps.

Accurate forecasts of volatility and hence Value-at-Risk is important to investors, practitioners, and policymakers for making informed decisions and portfolio risk management. It is also important to utilize a model capable of capturing the stylized characteristics and volatility dynamics of cryptocurrencies by combining conventional and novel techniques [4]. The Generalized Autoregressive Conditional Heteroscedastic (GARCH) model and its variants are famous volatility models for modelling traditional financial time series as well as for cryptocurrencies. The popularity of GARCH-type models for describing the dynamics of cryptocurrencies volatility is due to their deterministic dependence of the conditional variance on past observations.

Several studies have employed variants of GARCH-type models for several cryptocurrencies to select the best volatility model or a superior set of models. Fakhfekh and Jeribi [5] applied various GARCH-type models with different error distributions to sixteen of the most popular cryptocurrencies and found that the TGARCH model with double exponential distribution provided the best fit. Ngunyi *et al.* [6] applied several GARCH-type models with different error distributions to eight of the most popular cryptocurrencies and found that the asymmetric GARCH models with long memory property and heavy-tailed innovations provided the best fit for all cryptocurrencies. Chu *et al.* [7] using GARCH models with different error distributions concluded that the IGARCH (1, 1) model estimates the Bitcoin volatility better than the competing models. Therefore, the selection of the appropriate distribution of cryptocurrencies returns is also a major challenge in cryptocurrencies risk management.

Alternatively, extreme value theory could be useful to better understand the characteristics of the extreme tail distribution of cryptocurrencies. However, only a few attempts have been made so far to examine extreme price movements of different cryptocurrencies. In the recent past, a limited number of studies have investigated the tail behaviour of cryptocurrencies using extreme value theory. Borri [8] modelled the conditional tail-risk in four major cryptocurrencies and the results showed that these cryptocurrencies are highly exposed to tail-risk within the crypto market contexts. Gangwal and Longin [9] presented an extreme value analysis of the returns of Bitcoin and showed that the returns followed a Frèchet distribution; Begušić *et al.* [10] also provided evidence that ex-

treme prices of Bitcoin are considerably more frequent, implying that Bitcoin exhibits heavier tails than stock returns. Zhang *et al.* [11] utilized extreme value analysis to investigate the tail risk behaviour of the high-frequency (hourly) log-returns of the four most popular cryptocurrencies estimating value at risk and expected shortfall with varying thresholds. The empirical results found that Ripple was the riskiest cryptocurrency exhibiting the largest potential gain or loss for both positive and negative (hourly) log-returns at every percentile and threshold while Bitcoin was the least risky cryptocurrency.

In a Value-at-Risk context, Gkillas and Katsiampa [12] apply extreme value theory to estimate Value at Risk and Expected shortfall as measures of tail risk for five cryptocurrencies. Likitratcharoen *et al.* [13] predicted the Value at Risk (VaR) of Bitcoin, Ethereum and Ripple using historical and Gaussian parametric, VaR. Their backtesting results show that the historical VaR model is suitable for measuring cryptocurrency risk over delta normal VaR only for a high confidence level of critical values.

The objective of this study is twofold. First, a comprehensive in-sample volatility modelling is implemented utilizing a variety of GARCH-type models to account for volatility clustering and leverage effects present in cryptocurrency returns. The probability distributions assumed for the standardized innovations include the Skewed Student-*t*, skewed Generalized error (GED), generalized hyperbolic (GHYP), Johnson's SU distributions. Second, we apply the GARCH-EVT model that combines the conditional heteroscedastic model and extreme value theory to examine the tail behaviour of eight major cryptocurrencies. The GARCH models and GARCH-EVT model are then used to estimate the out-of-sample 1-day-ahead Value at Risk (VaR) forecasts. The forecasting performance is evaluated using unconditional and conditional coverage tests to backtest the accuracy of VaR forecasts. The accuracy of forecast estimates is evaluated to determine which technique most accurately models extreme market risk on the eight cryptocurrencies.

The research contributes to the literature in two ways. First, it fits GARCH-type models using heavy-tailed innovations distributions to account for volatility clustering, asymmetry and leverage effects present in cryptocurrency returns. Second, it provides more accurate results based on a hybrid model combining conditional heteroscedastic model and extreme value analysis, namely the generalized Pareto distribution (GPD). The GPD is the only non-degenerate distribution that approximates asymptotically the limiting distribution of exceedances. We, therefore, consider only the relevant information of extremes providing more accurate risk estimates. The remaining part of the paper is organised as follows: Section 2 describes the methodology; GARCH modelling with selected innovations distribution, extreme value theory, value-at-risk estimation and backtesting procedures. Section 3 presents data description, empirical results and a discussion of the backtesting results. Finally, Section 4 concludes the study.

2. Methodology

2.1. GARCH Modelling

The generalized autoregressive conditional heteroscedastic (GARCH) model (Engle, [14]; Bolleslev, [15]) constitutes a benchmark in financial econometrics that is commonly used to estimate and forecast volatility of financial returns.

Let r_t denote the daily log returns of the corresponding cryptocurrencies data series at time t for $t = 1, \cdots, n$, computed as the logarithm of prices at the end of day t divided by the price at the end of the preceding day $t - 1$, $r_t = \ln(p_t / p_{t-1})$. The GARCH model can be specified as:

$$r_t = \mu_t + \sigma_t z_t \tag{1}$$

where μ_t denotes the conditional mean and σ_t denotes the volatility process, (σ_t^2 being the conditional variance). z_t the innovations, are independent and follow a distribution with zero mean and unit variance. For brevity, all selected GARCH models are restricted to a maximum order of one ($p = q = 1$). The parsimonious GARCH (1, 1) models tend to be more flexible, efficient and significant than higher order models in the out-of-sample analysis [16].

In this study, several GARCH-type specifications are considered namely the Standard GARCH (SGARCH), IGARCH (1, 1), EGARCH (1, 1), GJR-GARCH (1, 1), Asymmetric Power ARCH (APARCH) (1, 1), Threshold GARCH (TGARCH) (1, 1) and Component GARCH (CGARCH) (1, 1), to model the time-varying volatility of the selected cryptocurrencies. All of the GARCH-type models selected follow the specification in Equation (1); however, they differ in the conditional variance specification.

The conditional variance for the standard GARCH (SGARCH) (1, 1) process is given by:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{2}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$ to ensure a uniquely stationary process and positive conditional variance. The GARCH (1, 1) model captures volatility clustering in the data through the persistence parameter $\alpha + \beta$. However, if the persistence parameter $\alpha + \beta$ equals 1, the GARCH model converges to the Integrated GARCH model, where the long term volatility bears an infinite process.

The Integrated GARCH (IGARCH) model is a special version of SGARCH (1, 1) model where, the persistence parameter ($\alpha + \beta$) is equal to 1 and typically allows a unit root under the GARCH process. Thus, the conditional variance in the IGARCH (1, 1) is expressed in Equation (3), given that β is set equal to $(1 - \alpha)$ with parameter restrictions $\omega > 0$, $\alpha \geq 0$ and $1 - \alpha \geq 0$:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + (1 - \alpha) \sigma_{t-1}^2. \tag{3}$$

In both the SGARCH and IGARCH models, the impact of positive and negative news on the conditional variance is assumed to be symmetrical. These models restrict all coefficients to be greater than zero and thus cannot explain the

negative correlation between return and volatility. Some long-memory GARCH-type models are also introduced to forecast cryptocurrencies price volatility by capturing some stylized facts such as asymmetry and fat tails in the cryptocurrency price return innovations and to provide better VaR's computations.

The exponential GARCH (EGARCH) model by Nelson [17], incorporates the asymmetric impact of positive and negative shocks on volatility whereby the latter is believed to produce greater levels of volatility, despite having the same magnitude. This model is specified in logarithmic form, which suggests that parameters are unrestricted, and are thereby allowed to take negative values while ensuring a positive conditional variance. In addition, the conditional variance is written as a function of past standardized innovations, instead of past innovations. The volatility dynamics of an EGARCH (1, 1) can be expressed as:

$$\log_e \sigma_t^2 = \omega + \alpha_1 z_{t-1} + \gamma_1 \left(|z_{t-1}| - E|z_{t-1}| \right) + \beta_1 \log_e \left(\sigma_{t-1}^2 \right) \tag{4}$$

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

The Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model by Glosten et al. [18] is similar to EGARCH (1, 1) in incorporating the asymmetric impact of positive and negative shocks. The conditional variance responds asymmetrically via the use of an indicator function I. The volatility equation of a GJR-GARCH (1, 1) model is given as:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 I_{t-1} \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{5}$$

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. [19] allows for both leverage and the Taylor effect, named after Taylor [20] who observed that the sample autocorrelation of absolute returns were usually larger than that of squared returns.

The APARCH (1, 1) model can be expressed as:

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

where $\delta \in Re^+$, 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 [21] 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

$$\sigma_t^2 = q_t + \alpha_1 \left(z_{t-1}^2 - q_{t-1} \right) + \beta_1 \left(\sigma_{t-1}^2 - q_{t-1} \right) \tag{7}$$

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

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

The Nonlinear GARCH (NGARCH) model of Higgins *et al.* [22] is given by

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1} + \beta_1 \sigma_{t-1}^2 \qquad (8)$$

The Nonlinear Asymmetric GARCH (NAGARCH) model of Engle and Ng [23] is a model with the specification:

$$\sigma_t^2 = \omega + \alpha\left(\varepsilon_{t-1} - \theta\sigma_{t-1}\right)^2 + \beta\sigma_{t-1}^2 \qquad (9)$$

where $\alpha \geq 0, \beta \geq 0, \omega > 0$ and $\alpha\left(1 + \theta^2\right) + \beta < 1$, which ensures the non-negativity and stationarity of the variance process.

For stock returns, the parameter θ is usually estimated to be positive; in this case, it reflects a phenomenon referred to as the "leverage effect", signifying that negative returns increase future volatility by a larger amount than positive returns of the same magnitude.

For each GARCH-type model, the innovation process z_t is allowed to follow one of the following four skewed and heavy-tailed distributions: the Skewed Student-*t*, skewed Generalized error (GED), generalized hyperbolic (GHYP), Johnson's SU distributions since the cryptocurrencies returns have heavier tails than the normal distribution.

The skewed Student-t (SST) distribution by Azzalini and Capitanio [24], has a density given by

$$f\left(x;\delta,\nu,\mu,\beta\right) = \frac{1}{\delta}t_\nu\left(\frac{x-\mu}{\delta}\right)2T_{\nu+1}\left(\beta\left(\frac{x-\mu}{\delta}\right)\sqrt{\frac{\nu+1}{\left(\frac{x-\mu}{\delta}\right)^2 + \nu}}\right) \qquad (10)$$

where t_ν is the density of standard Student t distribution with ν degrees of freedom and $T_{\nu+1}$ is the distribution function of the standard Student t distribution with $\nu+1$ degrees of freedom.

The skewed generalized error distribution (SGED) by Theodossiou [25] is given by

$$f_{\text{SGED}}\left(x;\mu,\sigma,k,\lambda\right) = \frac{C}{\sigma}\exp\left(-\frac{\left|x-\mu+\delta\sigma\right|^k}{\left[1-\text{sign}\left(x-\mu+\delta\sigma\right)\lambda\right]^k \theta^k \sigma^k}\right) \qquad (11)$$

where

$$C = \frac{k}{2\theta}\Gamma\left(\frac{1}{k}\right)^{-1},$$

$$\theta = \Gamma\left(\frac{1}{k}\right)^{\frac{1}{2}}\Gamma\left(\frac{3}{k}\right)^{-\frac{1}{2}}S\left(\lambda\right)^{-1},$$

$$\delta = 2\lambda AS\left(\lambda\right)^{-1},$$

$$S(\lambda) = \sqrt{1 + 3\lambda^2 - 4A^2\lambda^2},$$

and

$$A = \Gamma\left(\frac{2}{k}\right)\Gamma\left(\frac{1}{k}\right)^{-\frac{1}{2}}\Gamma\left(\frac{3}{k}\right)^{-\frac{1}{2}},$$

μ and σ are the mean and standard deviation parameters respectively, λ is a skewness parameter, sign is the sign function, and $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1}e^z dz$ is the gamma function. The scaling parameter k and λ satisfy the following constraints $k > 0$ and $-1 < \lambda < 1$. The parameter k controls the height and tails of the density function and the skewness parameter λ controls the rate of descent of the density around the mode of the random variable x, where $\mathrm{mode}(x) = \mu - \delta\sigma$.

The generalized hyperbolic (GH) distribution by Barndorff-Nielsen [26] is given by

$$f(x; \lambda, \alpha, \beta, \mu, \delta) = \frac{\left(\delta\sqrt{\alpha^2 - \beta^2}\right)^\lambda (\delta\alpha)^{1/2-\lambda}}{\sqrt{2\pi}\delta K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}\left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{\lambda/2-1/4}$$
$$\times \exp(\beta(x-\mu))K_{\lambda-1/2}\left(\alpha\delta\sqrt{1 + \frac{(x-\mu)^2}{\delta^2}}\right) \tag{12}$$

where K_λ is the modified third-order Bessel function. The density is defined under the following parameter restrictions.

$$\delta \geq 0 \quad \text{and} \quad |\beta| < \alpha \quad \text{if} \quad \lambda > 0$$
$$\delta > 0 \quad \text{and} \quad |\beta| < \alpha \quad \text{if} \quad \lambda = 0$$
$$\delta > 0 \quad \text{and} \quad |\beta| \leq \alpha \quad \text{if} \quad \lambda < 0$$

The class of generalized hyperbolic distribution variants can be obtained by changing the values of the parameter λ; hence, λ is called the class-defining parameter.

The Johnson system of distributions consists of families of distributions that, through specified transformations, can be reduced to the standard normal random variable. A random variable X from the Johnson translation system is represented as a transformation of the normal distribution given by

$$X = \xi + \lambda r^{-1}\left(\frac{Z - \gamma}{\delta}\right)$$

where Z is a standard normal random variable, γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter and $r(\cdot)$ denotes one of the following normalizing transformations:

$$r(y) = \begin{cases} y & \text{for the } S_N \text{ (normal) family,} \\ \log(y) & \text{for the } S_L \text{ (lognormal) family,} \\ \log(y/(1-y)) & \text{for the } S_B \text{ (bounded) family,} \\ \log\left(y + \sqrt{y^2 + 1}\right) & \text{for the } S_U \text{ (unbounded) family} \end{cases}$$

where $X > \xi$ and $\lambda = 1$ for the S_L family; ξ feasible combination of the skewness and kurtosis values. The cryptocurrency returns considered in this study have skewness and kurtosis values that correspond to Johnson's S_U-distribution. Thus, we only consider the S_U family of the Johnson translation system. The reparameterized Johnson SU distribution, as discussed in Rigby and Stasinopoulos [27], is a four-parameter distribution denoted by JSU (μ, σ, ν, τ), with mean μ and standard deviation σ for all values of the skew and shape parameters ν and τ respectively.

The parameters of all GARCH-type models are estimated using Maximum Likelihood, since it is generally consistent and efficient, and provides asymptotic standard errors that are valid under non-normality. The most appropriate GARCH-type model is the one that minimizes the Kullback-Leibler distance between the model and the observed values. The selection is based on information criteria namely; the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

2.2. Extreme Value Theory and the Peaks-over-Threshold Model

In this section, we describe how to obtain the quantile z_q by applying EVT techniques to the distribution of GARCH-models filtered innovations. The Peak-over-threshold (POT) modelling approach is illustrated as follows. First, we fix a sufficiently high threshold u and assume that excess residuals over this threshold follow a generalized Pareto distribution (GPD) with tail index ξ.

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\dfrac{y}{\beta}\right) & \text{if } \xi = 0, \end{cases} \tag{13}$$

where $\beta > 0$ is scale parameter and the support is $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ when $\xi < 0$. ξ is the shape parameter, which governs the tail behaviour of $G_{\xi,\beta}(y)$. Consider a general distribution function F and the corresponding excess distribution above the threshold u defined by:

$$F_u(y) = \Pr(X - u \leq | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y \geq 0 \tag{14}$$

For $0 \leq y$, Balkema and De Haan [28] and Pickands [29] showed that for a large class of distributions F it is possible to find a positive measurable function $\sigma(u)$ such that

$$\lim_{u \to x_F} \sup_{0 \leq y \leq x_F - u} \left| F_u(y) - G_{\xi, \beta(u)}(y) \right| = 0 \tag{15}$$

The GPD is generalized in the sense that it subsumes several other specific distributions under its parametrization. When $\xi > 0$, the distribution function $G_{\xi,\beta}$ is the parameterized version of a heavy-tailed ordinary Pareto distribution; when $\xi = 0$ we have a light-tailed exponential distribution and when $\xi < 0$ we have a short-tailed Pareto type II distribution.

The tail of the underlying distribution is assumed to begin at the threshold u, with N the random variables of exceeding observations. For a random sample of size n the proportion of extremes is then N/n. Assuming that the N_u excesses over the threshold are independently and identically distributed (i.i.d) with exact GPD distribution, the parameters ξ and β are estimated by maximum likelihood. Smith [30] showed that maximum likelihood estimates $\hat{\xi}$ and $\hat{\beta}$ of the GPD parameters ξ and β are consistent and asymptotically normal as $N_u \to \infty$ provided $\xi > -1/2$. Even under the weaker assumption that the excesses are i.i.d from $F_u(y)$ which is only approximately GPD he also obtained unbiased and asymptotically normal results for ξ and β provided a sufficient rate of convergence.

By setting $x = u + y$, the following equality holds for points $x > u$ in the tail of F obtained from Equation (14):

$$1 - F(x) = \left(1 - F(u)\right)\left(1 - F_u(x - u)\right) \tag{16}$$

The first term, $\left(1 - F(u)\right)$, can be estimated non-parametrically using the random proportion of the data on the tail N/n and we can also estimate the term $1 - F_u(x - u)$, by approximating the excess distribution, $F_u(y)$ with a GPD fitted by maximum likelihood, to get the tail estimator:

$$\hat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\left(\frac{x - \hat{u}}{\hat{\beta}}\right)\right)^{-\frac{1}{\hat{\xi}}}, \tag{17}$$

For x number of observations in the tail is fixed to be $N = k$, this gives us a random threshold at the $(k+1)$th order statistic. The GPD with parameter ξ and β is fitted to the data $Z_{(1)} - Z_{(k+1)}, \cdots, Z_{(k)} - Z_{(k+1)}$, the excess amounts over the threshold for all residuals exceeding the threshold. The tail estimator for $F_Z(z)$ is then given by

$$\hat{F}_Z(z) = 1 - \frac{k}{n}\left(1 + \hat{\xi}\left(\frac{z - z_{(k+1)}}{\hat{\beta}}\right)\right)^{-1/\hat{\xi}}, \tag{18}$$

For $q > 1 - k/n$, we can invert Equation (18) to get

$$\hat{z}_q = z_{k+1} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1 - q}{k/n}\right)^{-\hat{\xi}} - 1\right) \tag{19}$$

which is the q-th quantile of the data distribution.

2.3. Measure of Value-at-Risk

Value at Risk (VaR) is a measure of risk that determines the losses that may happen in extreme events for a given confidence level. The main parameters of VaR are the significance level (confidence level $1 - \alpha$) and the risk horizon (h), which is the period of time in terms of trading days.

Consider $(X_t, t \in Z)$ a strictly stationary time series representing daily observations of the negative log-return of a financial asset price. The dynamics of X_t is assumed to be given by:

446

$$X_t = \mu_t + \sigma_t Z_t \qquad (20)$$

where the innovations Z_t follow a strict white noise process, independent and identically distributed, with zero mean, unit variance and marginal distribution function $F_Z(z)$. We assume that μ_t and σ_t are both measurable with respect to F_{t-1} the information about the return process available up to time $t-1$.

Let $F_X(x)$ denote the marginal distribution of (X_t) and, for a horizon $h \in N$, let $F_{X_{t+1}+\cdots+X_{t+h}|F_t}(x)$ denote the predictive distribution of the return over the next h days, given information on returns up to and including day t. For $0 < q, 1$, the q-th unconditional quantile for the marginal distribution is denoted by:

$$x_q = \inf\left\{x \in \mathbb{R} : P(X > x) \le 1 - q\right\} = \inf\left\{x \in \mathbb{R} : F_X(x) \ge q\right\}, \qquad (21)$$

and a conditional quantile is a quantile of the predictive distribution for the return over the next h days denoted by

$$x_q^t(h) = \inf\left\{x \in \mathbb{R} : F_{X_{t+1}+\cdots+X_{t+h}|F_t}(x) \ge q\right\} \qquad (22)$$

We are principally interested in estimating unconditional and conditional quantiles in the tails of negative log-returns for the 1-step predictive distribution. Since

$$F_{X+t|F_t}(x) = P\left\{\mu_{t+1} + \sigma_{t+1}Z_{t+1} \le x \mid F_t\right\} = F_Z\left(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\right)$$

The quantile is denoted by x_q^t and simplify to

$$x_q^t = \mu_{t+1} + \sigma_{t+1}z_q \qquad (23)$$

where z_q is the upper q-th quantile of the marginal distribution of Z_t which by assumption does not depend on t. Mathematically, VaR is the q-th quantile of the underlying distribution of returns.

To estimate risk measure, VaR for the cryptocurrency market, our main interest is on extreme value theory-based models: we consider only the conditional GPD approach and conventional GARCH models.

The Peak over Threshold: Conditional GPD Approach

Different approaches have been proposed in the literature to estimate risk measures. The unconditional GPD has the advantages that it focuses directly on the tail of the distribution. However, it doesn't recognize the fact that returns are not i.i.d. The econometric models of volatility such as the GARCH-process under different innovation's distributions yield VaR estimates which reflects the current volatility dynamics. The weakness of this GARCH modelling approach is that it focuses on modelling the whole conditional return distribution as time-varying, and not only the tail distribution that is of interest. This approach may sometimes fail to accurately estimate risk measures like VaR.

In order to overcome the drawbacks of each of the above methods, McNeil

and Frey [31] proposed to combine ideas from these two approaches. By first filtering, the returns with a GARCH model is that we get essentially i.i.d. series on which it is straightforward to apply the EVT technique. The advantage of this GARCH?EVT combination lies in its ability to capture conditional heteroscedasticity in the time series through the GARCH framework, while simultaneously, modelling the extreme tails behaviour through the EVT method. The conditional GPD produces a VaR, which reflects the current volatility background. The combined approach denoted conditional GPD, may be presented in the following three steps:

Step 1: Fit a GARCH-type model to the return data by quasi-maximum likelihood. Estimate μ_{t+1} and σ_{t+1} from the fitted model and extract the standardized residuals z_t.

Step 2: Consider the standardized residuals computed in Step 1 to be realizations of a white noise process, and estimate the tails of the innovations using extreme value theory. Next, compute the quantiles of the innovations.

Step 3: Construct VaR from parameters estimated in steps 1 and 2.

Assuming that the volatility dynamics of log-returns can be represented by Equation (2). Given the 1-step forecasts μ_{t+1}, σ_{t+1} and the estimate quantile of standardized residuals series, $\text{VaR}_{t+1}(Z)$, using the Equation (19) the VaR for the return series can be estimated as:

$$\widehat{\text{VaR}}_{t+1}^{q} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{z}_q \tag{24}$$

2.4. Statistical Backtesting of Model-Based VaR Forecasts

To back-test the accuracy for the estimated VaRs, we computed the empirical failure rates. By definition, the failure rate is the number of times returns (in absolute values) exceed the forecasted VaR. If the model is correctly specified, the failure rate should be equal to the specified VaR's level. In this study, the backtesting VaR is based on the Kupiec's [32] and Christoffersen [33] for unconditional and conditional coverage tests.

For purposes of implementing VaR forecast tests, the first step is to define the "hit sequence" of VaR violations:

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t+1} < -\text{VaR}_{t+1}^{\alpha} \\ 0 & \text{if } r_{t+1} \geq -\text{VaR}_{t+1}^{\alpha} \end{cases} \tag{25}$$

where $\text{VaR}_{t+1}^{\alpha}$ is the VaR prediction at time $t+1$ for risk quantile level α. Under the null hypothesis of correct specification the hit sequence should be an independent Bernoulli distributed variable.

$$H_0 : I_{t+1} \sim Bernoulli(\alpha), \tag{26}$$

$$f(I_{t+1}, p) = (1-p)^{1-I_{t+1}} p^{I_{t+1}}. \tag{27}$$

The accuracy and reliability of VaR methodology are tested by evaluating the out-of-sample performance of the estimated VaR forecasts. The backtesting procedure consists of comparing the out-of-sample VaR estimates with actual

realized loss in the next period. For a VaR forecast model to be accurate in its predictions, then the average hit sequence or hit ratio or the failure rate over the full sample should be equal α for the $(1-\alpha)\%$ quantile VaR (*i.e.*, for 95% VaR). 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 $(1-\alpha)\%$, the model overestimates risk.

The unconditional coverage (UC) test uses the fraction/ratio of observed violations for a particular risk model π and compares it with p. For this purpose the likelihood Bernoulli function is required and is given by:

$$L(\pi) = \pi(1-\pi)^{1-I_{t+1}} \pi^{I_{t+1}} = (1-\pi)^{T_0} \pi^{T_1} \tag{28}$$

where T_0, T_1 are the number of 0 s and 1 s in the sample $(T = T_0 + T_1)$. The maximum likelihood estimator is $\hat{\pi} = T_1/T$. The null hypothesis can be tested by means of the following likelihood ratio test:

$$LR_{uc} = -2\ln\left(\frac{L(\alpha)}{L(\hat{\pi})}\right) \sim \chi_1^2 \tag{29}$$

Under the null hypothesis that the VaR model is correct LR_{uc} is asymptotically chi-square distributed with one degree of freedom. However, this test focuses only on the number of exceptions.

In practice, situations arise when the VaR model passes the unconditional coverage test but all violations are clustered. To reject a VaR model with clustered violations, a test of independence of the hit sequence is required. Suppose the hit sequence is assumed to exhibit time dependence and follows a first-order Markov sequence with the following transition probability matrix:

$$\pi_1 = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix} \tag{30}$$

where $\pi_{ij} = \Pr(I_t = j \mid I_{t+1} = i)$, π_{01} is the probability of getting a violation tomorrow given no violation today, π_{11} is the probability of getting a violation tomorrow given today is also a violation. Then the corresponding likelihood function is given as:

$$L(\pi_1) = (1-\pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1-\pi_{11})^{T_{10}} \pi_{11}^{T_{11}}, \tag{31}$$

where T_{ij} is the number of observations with a j following i. If the hit sequence is independent over time, the probability of a violation tomorrow does not depend on today having a violation or not. Hence, the null hypothesis in the independence test is $H_0 : \pi_{01} = \pi_{11} = \pi$. The transition probability matrix will take the form:

$$\hat{\pi} = \begin{pmatrix} 1-\hat{\pi} & \hat{\pi} \\ 1-\hat{\pi} & \hat{\pi} \end{pmatrix} \tag{32}$$

Then, independence can be tested using a likelihood ratio test statistics defined as follows:

$$LR_{ind} = -2\ln\left(\frac{L(\hat{\pi})}{L(\hat{\pi}_1)}\right) \sim \chi_1^2. \tag{33}$$

Ultimately, VaR users are interested in being able to test simultaneously whether the hit sequence is independent and the average number of violations is correct. The conditional coverage (CC) test jointly examines whether the percentage of exceptions is statistically equal to the one expected and the serial independence of the exception indicator. A sequence of VaR forecasts at-risk level α has the correct conditional coverage if $\{I_t(\alpha); t = 1, \cdots, T\}$ is an independent and identically distributed sequence of Bernoulli random variables with parameter α. In this test, the null hypothesis takes the form: $H_0 : \pi_{01} = \pi_{11} = \alpha$. To test this hypothesis a joint test of independence of the hit sequence and the unconditional coverage of the VaR forecasts is required. Thus, under the null hypothesis of the expected proportion of exceptions equals α and the failure process is independent, the appropriate likelihood ratio test statistic is of the form:

$$LR_{cc} = -2\ln\left(\frac{L(\alpha)}{L(\hat{\pi}_1)}\right) \sim \chi_2^2 \tag{34}$$

Under the null hypothesis the likelihood ratio statistic, LR_{cc}, is asymptotically Chi-square distributed, with two degree of freedom. Note also that $LR_{cc} = LR_{uc} + LR_{ind}$.

3. Data Description and Empirical Results

3.1. Data Description

In this study, the data set consists of daily closing prices (in US dollars) of the eight largest cryptocurrencies in terms of market capitalization traded from August 8, 2015, to September 16, 2020 (1859 observations). The data are publicly available online at https://coinmarketcap.com/coins/. We only considered cryptocurrencies with no missing values, which resulted in eight cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), Monero (XMR), Stella (XLM), Dash (DASH) and Tether (USTD). **Figure 1** presents time series plots for the cryptocurrency daily trading prices for the given period of August 8, 2015, to September 16, 2020. The sample period covers both relatively volatile and stable periods, with phases of price fluctuations and occasional extreme price jumps. All the cryptocurrencies display visible patterns of volatility clustering dynamics over time.

In order to expressively visualize some features for each cryptocurrency data, daily returns were computed by $r_t = \ln(p_t/p_{t-1})$ with p_t denoting daily closing price in time (t). The data adjustment procedure is applied to obtain stationary time-series for the returns of the cryptocurrencies considering heteroscedasticity. **Figure 2** presents the dynamic evolution of log return series for all cryptocurrencies and illustrates the stylized feature of leptokurtosis that arises from a pattern of time-varying volatility clustering in the cryptocurrencies where periods of high (low) volatility are followed by periods of high (low) volatility.
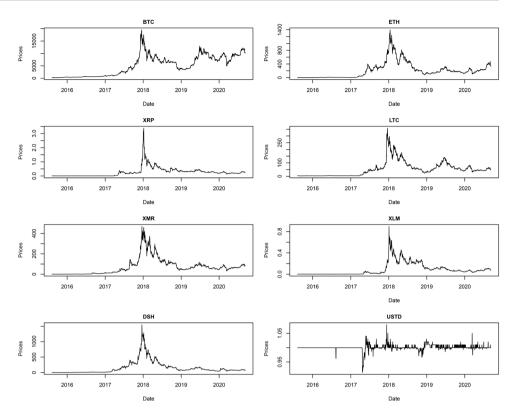
Figure 1. Daily prices of the eight major cryptocurrencies for the period starting from August 8, 2015 to September 16, 2020.
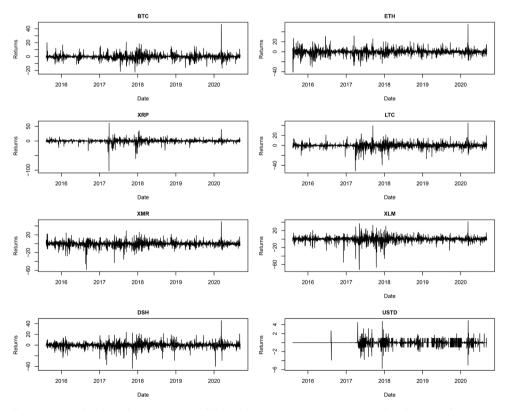

Figure 2. Daily logarithmic returns of the eight major cryptocurrencies for the period starting from August 8, 2015 to September 16, 2020.

Table 1 reports summary statistics of cryptocurrencies and statistical test results. All cryptocurrencies record a negative mean which is close to zero except for Tether while standard deviation values are all slightly above zero value. Except for Bitcoin, all the other cryptocurrencies are significantly negatively skewed. Additionally, all series have excess-kurtosis implying fat-tails and non-normally distributed. Concerning the normal distribution, the Jarque-Bera test suggests that all the cryptocurrencies are not distributed normally. To test stationarity in cryptocurrencies return series, the Augmented Dickey and Fuller (ADF) test is used. The results of the ADF test accepted the null hypotheses, meaning that all the series were non-stationary at all levels. The significant autoregressive conditional heteroscedasticity (ARCH) confirmed the presence of ARCH effects in all the cryptocurrencies studied. The Ljung-Box Q statistics on lag (20) of squared returns confirmed the significant ARCH effects.

Table 1. Descriptive statistics and statistical test results for eight cryptocurrencies for the period starting from August 8, 2015 to September 8, 2020.

Stat.	BTC	ETH	XRP	LTC	XMR	XLM	DASH	USTD
Obs.	1858	1858	1858	1858	1858	1858	1858	1858
Min	−0.2251	−0.4102	−1.0274	−0.5103	−0.5846	−0.7231	−0.4377	−0.0572
Max	0.4647	0.5507	0.6163	0.4490	0.4942	0.4099	0.4593	0.0492
Mean	−0.0197	−0.0033	−0.0018	−0.0014	−0.0026	−0.0019	−0.0017	0.0000
Std.Dev	0.0398	0.0628	0.0662	0.0550	0.0639	0.0735	0.0577	0.0062s
Skew	0.9545	−0.0725	−2.9354	−0.7616	−0.6839	−1.9885	−0.6287	−0.2846
Kurt.	13.8551	7.9767	45.7677	13.0484	9.7859	18.6155	8.8248	16.3052
JB	15,183.00	4942.10	165,208.00	13,396.00	7579.80	28,121.00	6169.00	20,660.00
ADF	−11.474*	−10.896*	−10.604*	−11.732*	−11.148*	−10.552*	−10.689*	−14.547*
$Q(5)$	3.9460	7.7223	35.873	5.8803	17.31	17.831	10.465	248.71
p-value	(0.5572)	(0.1722)	(0.0000)	(0.3180)	(0.0039)	(0.0032)	(0.0631)	(0.0000)
$Q^2(5)$	37.428*	169.730*	236.510*	74.070*	88.279*	478.840*	69.495*	344.870*
$Q(10)$	16.262	9.542	55.307	22.717	39.446	29.572	18.686	253.800
p-value	(0.0924)	(0.4816)	(0.0000)	(0.0118)	(0.0000)	(0.0010)	(0.0444)	(0.0000)
$Q^2(10)$	63.62*	203.93*	269.56*	114.10*	190.46*	488.06*	112.89*	396.68*
LM(5)	31.424*	121.37*	174.76*	61.27*	74.34*	325.03*	54.04*	327.29*
LM(10)	49.982*	132.00*	194.42*	87.308*	141.63*	329.20*	81.378*	344.88*

Note: Std Dev (Standard deviation), J.B. (Jarque-Bera), ARCH (autoregressive conditional heteroscedasticity), ADF (augmented Dickey and Fuller), the value of J.B., ADF, ARCH(5), ARCH(10), and $Q(5)$, $Q(10)$, $Q^2(5)$, $Q^2(10)$ Ljung are statistically significant for * at 1%.

3.2. Parameter Estimates of GARCH-Type Models

In this section, results from the estimated GARCH-type models are presented. The sampled period is divided into two sub-sample periods: the in-sample period extending from October 13th 2015 till December 3rd 2018, and the out-of-sample period covering the period from December 4th 2018 till November 18th 2019. In-sample returns are used to estimate the parameters of the selected models, subject to the assumptions and constraints of each model. Accordingly, the calculated in-sample parameters are applied to forecast the volatilities for both the in-sample and out-of-sample periods. First, we estimate GARCH, EGARCH, GJRGARCH, APARCH, CSGARCH, NGARCH and NAGARCH models concerning long memory test results to account for the long memory properties of our cryptocurrency returns.

Table 2 presents BIC values of the fitted GARCH-type specifications: GARCH, EGARCH, GJRGARCH, APARCH, CSGARCH, NGARCH and NAGARCH under different error distributions. The skewed generalized error distribution has minimum BIC values for Bitcoin, Ethereum, Ripple and Litecoin. Skewed-Student's-t distribution, which accounts for both asymmetry and heavy tails, is selected as the most suitable distribution for modelling this data set. Thus, the results deduce that the use of fat-tailed distribution to describe innovations distribution is justified.

Table 3 (*Panel A*) reports the estimation results of the NGARCH model with selected innovations distribution. The mean parameters are not significantly different from zero for all eight cryptocurrency price returns indicating that the GARCH components are covariance stationery. The GARCH (1, 1)-type model results reveal that the lagged conditional volatility for each cryptocurrency is statistically significant. In addition, the shock squared term in the variance equation is statistically significant, which means the lagged volatility and current news immediately reflect in the price of the cryptocurrencies. It is observed that under different distributional assumptions, the parameters vary, implying that the distributional assumption does have a certain effect on the estimation process. The skewness parameter, having a very low p-value, is quite significant. Moreover, the shape parameters for both the Student's-t and skewed-t distributions are significantly high, confirming the presence of heavy tails in the series. The results further show that the p-values of the GARCH parameters are very low except for LTC and ETHM, indicating that these parameters are also highly significant.

For the goodness-of-fit test (*Panel B*), the diagnostic results reveal that the NGARCH specifications filter the serial autocorrelation, conditional volatility dynamics and leverage effects present in cryptocurrencies return series. The Box-Pierce and ARCH-LM tests do not reject the null hypothesis of a correct model specification and show the power of the NGARCH model to take into account the major stylized facts of time series prices behaviour. However, the NGARCH model fails to capture extreme events normally experienced in the cryptocurrency markets. The standardized residuals of the NGARCH model are closer approximately independently and identically distributed (i.i.d) which is a

standard requirement for extreme value theory to be applied. Therefore, we can apply successfully EVT methods to i.i.d residual series. Obviously, in what follow we choose the NGARCH-EVT approach to compute the one-day-ahead VaR for all cryptocurrencies. The forecast performance of this model should be evaluated for the out-of-sample period and using more accurate performance criteria.

Table 2. The Bayesian Information Criterion (BIC) for GARCH model selection.

Dist.	BTC	ETH	XRP	LTC	XMR	XLM	DASH	USTD
Skewed Student t distribution								
GARCH	−4.1179	−3.1004	−3.5211	−3.5104	−2.9913	−2.9936	−3.2543	−9.3414
EGARCH	−4.1395	−3.1103	−3.5448	−3.5620	−3.0071	−3.0091	−3.2616	−4.0378
GJRGARCH	−4.0966	−3.0975	−3.5201	−3.5127	−2.9896	−2.9944	−3.2503	−9.3405
APARCH	−4.1434	−3.1085	−3.3154	−3.5816	−3.0135	−3.0144	−3.2577	**−9.6565**
CSGARCH	−4.0905	−3.0927	−3.5126	−3.5039	−2.9842	−2.9898	−3.2498	−9.1919
NGARCH	**−4.1608**	**−3.1125**	**−3.5578**	**−3.5851**	**−3.0162**	**−3.0171**	**−3.2617**	−9.3486
NAGARCH	−4.0962	−3.0964	−3.5184	−3.5087	−2.9896	−2.9906	−3.2502	9.3369
skewed Generalized error distribution								
GARCH	−4.1345	−3.1201	−3.5694	−3.5374	−2.9931	−3.0198	−3.2564	−6.1006
EGARCH	−4.1585	−3.1286	−3.5666	−3.5702	−3.0046	−3.0152	−3.2617	−6.1105
GJRGARCH	−4.1325	−3.1173	−3.5513	−3.5367	−2.9899	−3.0038	−3.2524	−5.7956
APARCH	−4.1708	−3.1254	−3.5739	−3.5885	−3.0091	−3.0190	−3.2577	−6.0873
CSGARCH	−4.1266	−3.1123	−3.5424	−3.5312	−2.9859	−2.9980	−3.2488	**−7.1108**
NGARCH	**−4.1770**	**−3.1293**	**−3.5758**	**−3.5925**	**−3.0124**	**−3.0211**	**−3.2617**	−6.0965
NAGARCH	−4.1312	−3.1160	−3.5494	−3.5344	−2.9903	−2.993	−3.2525	−6.0838
Generalised Hyperbolic Distribution (GHYP)								
GARCH	−4.1515	−3.1165	−3.5493	−3.5344	−2.9934	−3.0011	−3.2568	−9.0120
EGARCH	−4.1555	−3.1252	−3.5644	−3.5700	−3.0067	−3.0149	−3.2629	−8.7340
GJRGARCH	−4.1287	−3.1139	−3.5491	−3.5339	−2.9909	−3.0030	−3.2528	**−9.0356**
APARCH	−4.1610	−3.1223	−3.5719	−3.5885	−3.0122	−3.0191	−3.2591	−8.9999
CSGARCH	−4.1228	−3.1088	−3.5401	−3.5283	−2.9864	−2.9975	−3.2512	−8.9256
NGARCH	**−4.1745**	**−3.1262**	**−3.5739**	**−3.5921**	**−3.0143**	**−3.0212**	**−3.2631**	−9.0081
NAGARCH	−4.1274	−3.1125	−3.5470	−3.5315	−2.9913	−2.9986	−3.2528	−9.0071
Johnson's SU-distribution								
GARCH	−4.1152	−3.1078	−3.5368	−3.5237	−2.9944	−2.9989	−3.2579	−9.6638
EGARCH	−4.1458	−3.1168	−3.5553	−3.5690	−3.0094	−3.0136	−3.2648	−8.4328
GJRGARCH	−4.1119	−3.1048	−3.5358	−3.5257	−2.9926	−3.0001	−3.2539	−9.6581
APARCH	−4.1524	−3.1148	−3.5646	−3.5881	−3.0156	−3.0186	−3.2610	9.6621
CSGARCH	−4.1060	−3.1002	3.5284	−3.5170	−2.9875	−2.9954	−3.2505	−9.5075
NGARCH	**−4.1696**	**−3.1188**	**−3.5677**	--	**−3.0180**	**−3.0211**	**−3.2650**	**−9.6682**
NAGARCH	−4.1114	−3.1037	−3.5340	−3.5219	−2.9927	−2.9960	−3.2538	−9.6584

Table 3. Estimation results of NGARCH (1, 1) models with selected innovations distribution.

Par.	BTC	ETH	XRP	LTC	XMR	XLM	DASH	USTD
Distribution	sged	sged	sged	sged	jsu	ghyp	jsu	jsu
μ	−0.0008	−0.0013	0.0019	−0.0000	−0.0013	−0.0028	−0.0001	0.0000
	()	()	()	()	()	()	()	()
ω	0.0109	0.0079	0.015	0.0139	0.0206	0.0164	0.0043	0.0000
α_1	0.1461	0.2094	0.2773	0.1655	0.1672	0.2568	0.1865	0.0500
β_1	0.8509	0.7834	0.7241	0.8309	0.8057	0.7232	0.8149	0.9001
λ	0.3554	0.7598	0.5420	0.3518	0.3852	0.5929	0.8867	1.996
skew	1.0389	0.9319	0.9475	0.9950	−0.299	−0.0268	−0.0924	0.0064
shape	0.8817	0.9672	0.8614	0.8855	1.3893	0.5999	1.2326	0.9991
ghlambda						0.3947		
Diagnostic checks								
Log likelihood	3906.733	2933.499	3348.302	3363.79	2830.046	2836.835	3059.544	9008.062
AIC	−4.1978	−3.1502	−3.5758	−3.6133	−3.0388	−3.0450	−3.2858	−9.6890
BIC	−4.1770	−3.1293	−3.5758	−3.5925	−3.0180	−3.0212	−3.2650	−9.6682
LQ(5)	10.67	9.730	16.84	14.45	20.63	18.79	12.47	55.74
p-value	(0.0059)	(0.0108)	(0.0000)	(0.0006)	(0.0000)	(0.0000)	(0.0021)	(0.0000)
$LQ^2(5)$	2.9197	2.497	21.360	4.690	3.1178	6.666	8.106	0.0005
p-value	(0.4219)	(0.5064)	(0.0000)	(0.1796)	(0.3859)	(0.0625)	(0.0277)	(1.0000)
LM(5)	2.6130	0.9699	1.0839	1.6695	3.4978	3.5470	1.1665	0.0077
p-value	(0.3511)	(0.7419)	(0.7082)	(0.5489)	(0.2253)	(0.2197)	(0.6842)	(0.9997)

3.3. Parameter Estimates of the GARCH-EVT Model

In Extreme value theory (EVT) modelling, Peak over threshold (POT) approach is normally used to estimate the parameters of the generalized Pareto distribution (GPD). The POT method generally depends on the selection of the threshold. In this study, an optimal threshold value is set at 90% quantile of the total observations to estimate the GPD parameters for both left and right tails. Table 4 presents parameter estimates of the fitted GPD with their corresponding standard errors enclosed in brackets for both the left and right tails of the cryptocurrencies standardized residuals. The shape parameter (ξ) is positive and significantly different from zero for all cryptocurrencies indicating heavy-tailed distributions and a finite variance. This also implies that the tail distribution of cryptocurrencies belongs to Frechet class which is heavy-tailed. However, the shape parameter is negative except for Ethereum on the left tail. The scale parameters are also positive and significant for all cryptocurrencies both for the left and right tails.

3.4. Forecasting Performance Analysis

To evaluate the out-of-sample performance of the VaR forecast models, we used

Table 4. Parameter estimates of the Generalized Pareto model for selected u for the daily log returns of the eight crypto-currencies.

Par.	BTC	ETH	XRP	LTC	XMR	XLM	DASH	USTD
				Right Tail (*positive*)				
Obs.	1858	1858	1858	1858	1858	1858	1858	1858
u	1.0760	1.0462	1.0560	1.0288	1.1421	1.0879	1.0558	1.0421
n.exceed	186	186	186	186	186	186	186	186
Shape (ξ)	0.1222	0.0444	0.0256	0.1918	0.0811	0.0102	0.0619	0.1346
(s.e)	(0.0849)	(0.0824)	(0.0733)	(0.0903)	(0.0791)	(0.0688)	(0.0761)	(0.0703)
Scale (β)	0.6974	0.6748	0.6733	0.6525	0.6089	0.6584	0.6181	1.4611
(s.e)	(0.0779)	(0.0744)	(0.0698)	(0.0753)	(0.0656)	(0.0662)	(0.0652)	(0.1477)
				Left Tail (*negative*)				
u	1.1107	1.1381	1.1143	1.1076	1.1447	1.1880	1.0980	0.9902
n.exceed	186	186	186	186	186	186	186	186
Shape (ξ)	0.0842	−0.0899	0.0918	0.0165	0.0659	0.0236	0.0464	0.1676
(s.e)	(0.0860)	(0.0823)	(0.0831)	(0.0774)	(0.0746)	(0.0842)	(0.0805)	(0.0518)
Scale (β)	0.6663	0.8812	0.7926	0.7645	0.6119	0.7125	0.7355	1.2986
(s.e)	(0.0752)	(0.0969)	(0.0877)	(0.0815)	(0.0639)	(0.0795)	(0.0800)	(0.1155)

a rolling windows scheme with a window size of 1358 days and 500 days are reserved for the out-of-sample forecast period. The evaluation is based on the one-step-ahead forecast that is produced from a series of rolling sample size with an estimation window of 1358 observations kept constant and simply rolled forward after every 25 days. The advantage of a rolling window procedure is two-fold: to assess the stability of the model over time and the accuracy of the forecasting. Stability amounts to examining whether the coefficients are time-invariant. The one-day ahead VaR is calculated at 95% and 99% confidence levels. Both levels of confidence are used for out-of-sample backtesting of VaR, following Basel II Backtesting Requirements, which stipulates that backtesting of VaR needs to be done on confidence levels other than 99%. Backtesting is used to evaluate the relative performance of conventional GARCH models and the GARCH-EVT approach to forecast value at risk. Kupiec's unconditional coverage and Christoffersen's conditional coverage tests are used at two different levels of significance of 95% and 99% which are considered to reflect extreme market conditions.

Table 5 presents VaR forecast violation percentages and *p*-values in parentheses of unconditional coverage tests for GARCH (1, 1), EGARCH (1, 1), APARCH (1, 1), NGARCH (1, 1) and GARCH (1, 1)-EVT models with skewed-*t* distribution for eight cryptocurrencies returns. The exceedances involve counting the number of actual realized returns that exceed the VaR forecast and com-

paring this number with the expected number of exceedances. The closer the observed number of exceedances is to the hypothetically expected number, the more preferable the model is for estimating accurate forecasts. More exceedances indicate that the model underestimates Value at Risk and fewer exceedances indicate that the model overestimates Value at Risk. The expected exceedances are 25 for the 95% confidence level and 5 for 99% confidence level. The null hypothesis of Kupiec's unconditional coverage test assumes that the probability of occurrence of variations equals the expected level of significance. Under the null hypothesis, a good model should be the one that does not reject the null hypothesis. Hence, the test with a p-value greater than 0.05 for the unconditional coverage test indicates that the number of violations is statistically equal to the expected. These backtesting results demonstrate that GARCH-EVT clearly outperforms GARCH benchmark VaR predictors.

Table 5. VaR forecast violations of the cryptocurrencies in terms of actual and expected exceedances and Unconditional Coverage (UC) results.

α	5%		1%	
Tails:	Left Tail	Right Tail	Left Tail	Right Tail
BTC				
GARCH	3.4%	5.0%	1.2%	0.4%
UC: p-value	(0.082)	(1.000)	(0.663)	(0.125)
EGARCH	4.2%	5.0%	1.2%	0.4%
UC: p-value	(0.399)	(1.000)	(0.663)	(0.125)
APARCH	4.2%	4.6%	1.2%	0.2%
UC: p-value	(0.399)	(0.678)	(0.663)	**
NGARCH	4.2%	4.8%	1.2%	0.2%
UC: p-value	(0.399)	(0.836)	(0.663)	**
GARCH-EVT	2.8%	4.8%	1.0%	0.4%
UC: p-value	**	(0.836)	(1.000)	(0.125)
ETH				
GARCH	4.8%	7.0%	0.8%	1.2%
UC: p-value	(0.836)	(0.052)	(0.641)	(0.663)
EGARCH	4.8%	6.8%	1.0%	1.0%
UC: p-value	(0.386)	(0.079)	(1.000)	(1.000)
APARCH	5.0%	6.6%	1.0%	0.8%
UC: p-value	(1.000)	(0.117)	(1.000)	(0.641)
NGARCH	4.8%	6.8%	0.8%	1.0%
UC: p-value	(0.836)	(0.079)	(0.641)	(1.000)
GARCH-EVT	5.0%	6.8%	0.6%	1.2%
UC: p-value	(1.000)	(0.079)	(0.331)	(0.663)

Continued

	XRP			
GARCH	3.8%	3.6%	0.8%	0.8%
UC: *p*-value	(0.199)	(0.131)	(0.641)	(0.641)
EGARCH	3.8%	4.2%	0.4%	0.8%
UC: *p*-value	(0.199)	(0.399)	(0.125)	(0.641)
APARCH	4.4%	4.0%	0.4%	0.8%
UC: *p*-value	(0.530)	(0.288)	(0.125)	(0.641)
NGARCH	4.4%	4.0%	0.4%	0.8%
UC: *p*-value	(0.530)	(0.288)	(0.125)	(0.641)
GARCH-EVT	4.8%	3.4%	1.0%	0.8%
UC: *p*-value	(0.836)	(0.082)	(1.000)	(0.641)
	LTC			
GARCH	5.0%	3.4%	1.2%	0.8%
UC: *p*-value	(1.000)	(0.082)	(0.663)	(0.641)
EGARCH	4.4%	3.6%	2.2%	0.6%
UC: *p*-value	(0.530)	(0.131)	**	(0.331)
APARCH	5.0%	4.2%	2.0%	0.8%
UC: *p*-value	(1.000)	(0.399)	**	(0.641)
NGARCH	5.0%	3.4%	1.8%	0.6%
UC: *p*-value	(1.000)	(0.082)	(0.106)	(0.331)
GARCH-EVT	4.8%	2.4%	1.4%	0.4%
UC: *p*-value	(0.836)	**	(0.397)	(0.125)
	XMR			
GARCH	5.2%	5.4%	0.8%	1.6%
UC: *p* -value	(0.838)	(0.685)	(0.641)	(0.215)
EGARCH	4.8%	4.6%	0.6%	1.6%
UC: *p*-value	(0.836)	(0.678)	(0.331)	(0.215)
APARCH	4.4%	5.0%	0.6%	1.4%
UC: *p*-value	(0.530)	(1.000)	(0.331)	(0.397)
NGARCH	4.2%	5.4%	0.6%	1.6%
UC: *p*-value	(0.399)	(0.685)	(0.331)	(0.215)
GARCH-EVT	3.4%	4.8%	0.8%	1.8%
UC: *p*-value	(0.082)	(0.836)	(0.641)	(0.106)
	XLM			
GARCH	4.2%	4.0%	0.8%	1.2%
UC: *p* -value	(0.399)	(0.288)	(0.641)	(0.663)
EGARCH	4.4%	4.2%	0.8%	0.8%
UC: *p*-value	(0.530)	(0.399)	(0.641)	(0.641)
APARCH	4.6%	4.8%	1.0%	1.0%

Continued

UC: p-value	(0.678)	(0.836)	(1.000)	(1.000)
NGARCH	5.0%	4.6%	1.0%	1.0%
UC: p-value	(1.000)	(0.678)	(1.000)	(1.000)
GARCH-EVT	5.8%	3.6%	1.4%	0.8%
UC: p-value	(0.423)	(0.131)	(0.397)	(0.641)
DASH				
GARCH	3.4%	5.2%	--	0.8%
UC: p-value	(0.082)	(0.838)	--	(0.641)
EGARCH	3.5%	5.0%	--	0.8%
UC: p-value	(0.082)	(1.000)	--	(0.641)
APARCH	3.6%	5.0%	--	1.0%
UC: p-value	(0.131)	(1.000)	--	(1.000)
NGARCH	3.4%	5.0%	--	0.8%
UC: p-value	(0.082)	(1.000)	--	(0.641)
GARCH-EVT	4.0%	4.4%	--	0.8%
UC: p-value	(0.288)	(0.530)	--	(0.641)
USTD				
GARCH	3.6%	1.6%	2.6%	1.2%
UC: p-value	(0.131)	**	**	(0.663)
EGARCH	2.2%	2.0%	1.4%	1.0%
UC: p-value	**	**	(0.397)	(1.000)
APARCH	1.8%	1.4%	1.6%	1.4%
UC: p-value	**	**	(0.215)	(0.397)
NGARCH	1.8%	1.4%	1.8%	1.4%
UC: p-value	**	**	(0.106)	(0.397)
GARCH-EVT	0.2%	0.4%	--	0.4%
UC: p-value	**	**	--	(0.125)

Table 6 also presents test statistic and p-values in parentheses of conditional coverage tests for GARCH (1, 1), EGARCH (1, 1), APARCH (1, 1), NGARCH (1, 1) and GARCH (1, 1)-EVT models with skewed-t distribution. 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 null hypothesis of the conditional coverage test indicates that the probability of occurrence of the violations equals the expected significance level and the violation is independently distributed through time. The empirical results suggest that the combined GARCH-EVT model performs best in estimating out-of-sample VaR forecasts in the specified backtesting period and this makes it relatively better in forecasting VaR. The superior performance is attributed to the combined approachability to appropriately capture the statistical features of the data.

Table 6. Conditional Coverage (CC) results of backtesting.

α	5%		1%	
Tails:	Left Tail	Right Tail	Left Tail	Right Tail
BTC				
GARCH	3.294	0.060	0.336	2.369
CC: p-value	(0.193)	(0.970)	(0.845)	(0.306)
EGARCH	1.874	0.060	0.336	2.369
CC: p-value	(0.392)	(0.097)	(0.845)	(0.306)
APARCH	1.874	2.396	0.336	4.817
CC: p-value	(0.392)	(0.302)	(0.845)	(0.090)
NGARCH	1.874	2.469	0.336	4.817
CC: p-value	(0.392)	(0.291)	(0.845)	(0.090)
GARCH-EVT	6.728	2.469	0.101	2.369
CC: p-value	**	(0.306)	(0.951)	(0.291)
ETH				
GARCH	0.615	3.897	0.282	0.336
CC: p-value	(0.735)	(0.142)	(0.869)	(0.845)
EGARCH	0.615	3.296	4.480	0.101
CC: p-value	(0.735)	(0.192)	(0.106)	(0.951)
APARCH	0.426	2.778	4.480	0.282
CC: p-value	(0.808)	(0.249)	(0.106)	(0.869)
NGARCH	0.615	3.296	5.679	0.101
CC: p-value	(0.735)	(0.192)	(0.058)	(0.951)
GARCH-EVT	0.426	3.133	0.979	0.336
CC: p-value	(0.808)	(0.209)	(0.613)	(0.845)
XRP				
GARCH	3.347	3.624	0.282	0.282
CC: p-value	(0.188)	(0.163)	(0.869)	(0.869)
EGARCH	1.750	0.727	2.369	0.282
CC: p-value	(0.417)	(0.695)	(0.306)	(0.869)
APARCH	1.334	1.176	2.369	0.282
CC: p-value	(0.513)	(0.555)	(0.306)	(0.869)
NGARCH	1.334	1.176	2.369	0.282
CC: p-value	(0.513)	(0.555)	(0.306)	(0.869)
GARCH-EVT	0.066	4.221	0.101	0.282
CC: p-value	(0.967)	(0.121)	(0.951)	(0.869)

Continued

	LTC			
GARCH	0.060	3.294	0.336	0.282
CC: *p*-value	(0.970)	(0.193)	(0.845)	(0.869)
EGARCH	2.424	2.453	5.915	0.979
CC: *p*-value	(0.298)	(0.293)	(0.052)	(0.613)
APARCH	2.638	0.727	4.323	0.282
CC: *p*-value	(0.267)	(0.695)	(0.115)	(0.869)
NGARCH	0.060	3.294	2.943	0.979
CC: *p*-value	(0.970)	(0.193)	(0.230)	(0.613)
GARCH-EVT	2.469	9.329	0.918	2.369
CC: *p*-value	(0.291)	**	(0.632)	(0.306)
	XMR			
GARCH	0.344	3.255	0.282	1.799
CC: *p*-value	(0.842)	(0.196)	(0.869)	(0.407)
EGARCH	0.066	2.396	0.979	1.799
CC: *p*-value	(0.967)	(0.302)	(0.613)	(0.407)
APARCH	0.395	0.426	0.979	0.918
CC: *p*-value	(0.821)	(0.808)	(0.613)	(0.632)
NGARCH	2.557	0.366	0.979	1.799
CC: *p*-value	(0.279)	(0.833)	(0.613)	(0.407)
GARCH-EVT	4.220	2.469	0.282	2.943
CC: *p*-value	(0.121)	(0.291)	(0.869)	(0.229)
	XLM			
GARCH	2.557	2.797	0.282	0.336
CC: *p*-value	(0.279)	(0.247)	(0.869)	(0.845)
EGARCH	0.395	0.727	0.282	0.282
CC: *p*-value	(0.821)	(0.695)	(0.869)	(0.869)
APARCH	0.177	0.066	0.101	0.101
CC: *p*-value	(0.915)	(0.967)	(0.951)	(0.951)
NGARCH	0.426	0.177	0.101	0.101
CC: *p*-value	(0.808)	(0.915)	(0.951)	(0.951)
GARCH-EVT	4.223	3.624	0.918	0.282
CC: *p*-value	(0.121)	(0.163)	(0.632)	(0.869)

Continued

	DASH			
GARCH	4.221	2.901	--	0.282
CC: p-value	(0.121)	(0.234)	--	(0.869)
EGARCH	4.221	0.060	--	0.282
CC: p-value	(0.121)	(0.970)	--	(0.869)
APARCH	3.624	0.060	--	0.101
CC: p-value	(0.163)	(0.970)	--	(0.951)
NGARCH	4.221	0.060	--	0.282
CC: p-value	(0.121)	(0.970)	--	(0.869)
GARCH-EVT	2.797	2.424	--	0.282
CC: p-value	(0.247)	(0.297)	--	(0.869)
	USTD			
GARCH	22.024	16.631	23.788	0.336
CC: p-value	**	**	**	(0.845)
EGARCH	21.493	12.552	3.805	0.101
CC: p-value	**	**	(0.149)	(0.951)
APARCH	21.344	19.051	9.760	0.918
CC: p-value	**	**	**	(0.632)
NGARCH	21.344	19.051	9.813	0.918
CC: p-value	**	**	**	(0.632)
GARCH-EVT	42.759	37.001	--	2.369
CC: p-value	**	**	--	(0.306)

4. Conclusion

Cryptocurrencies unlike conventional financial assets such as currencies exchange rates and stock prices are characterized by high volatility and extreme price movements. This paper employed GARCH-type models and extreme value theory to model the volatility and tail behaviour of the cryptocurrencies returns. Modelling the tail behaviour of the returns of cryptocurrencies is of utmost importance for both investors and policy-makers. The GARCH-EVT approach is implemented in modelling the tail distribution of cryptocurrencies return series and forecasting out-of-sample value at risk. The back-testing results demonstrate the superiority of the heavy-tailed GARCH-EVT models in forecasting out-of-sample value at risk. Overall, the model provides a significant improvement in forecasting value-at-risk over the widely used conventional GARCH models. This study can be extended by considering intra-day cryptocurrencies data and more

robust models such as the GARCH-EVT-Copula model that can also capture the dependence structure of between cryptocurrencies.

Conflicts of Interest

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

References

[1] Nakamoto, S. (2008) Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin, 4. https://bitcoin.org/bitcoin.pdf

[2] coinmarketcap.com. (2021) Overview of Available Cryptocurrencies. https://coinmarketcap.com

[3] Schnaubelt, M., Rende, J. and Krauss, C. (2019) Testing Stylized Facts of Bitcoin Limit Order Books. *Journal of Risk and Financial Management*, **12**, Article No. 25. https://doi.org/10.3390/jrfm12010025

[4] Peng, Y., Albuquerque, P.H.M., de Sá, J.M.C., Padula, A.J.A. and Montenegro, M.R. (2018) The Best of Two Worlds: Forecasting High-Frequency Volatility for Cryptocurrencies and Traditional Currencies with Support Vector Regression. *Expert Systems with Applications*, **97**, 177-192. https://doi.org/10.1016/j.eswa.2017.12.004

[5] Fakhfekh, M. and Jeribi, A. (2020) Volatility Dynamics of Cryptocurrencies Returns: Evidence from Asymmetric and Long Memory GARCH Models. *Research in International Business and Finance*, **51**, Article ID: 101075. https://doi.org/10.1016/j.ribaf.2019.101075

[6] Ngunyi, A., Mundia, S. and Omari, C. (2019) Modelling Volatility Dynamics of Cryptocurrencies Using GARCH Models. *Journal of Mathematical Finance*, **9**, 591-615. https://doi.org/10.4236/jmf.2019.94030

[7] Chu, J., Chan, S., Nadarajah, S. and Osterrieder, J. (2017) GARCH Modelling of Cryptocurrencies. *Journal of Risk and Financial Management*, **10**, Article No. 17. https://doi.org/10.3390/jrfm10040017

[8] Borri, N. (2019) Conditional Tail-Risk in Cryptocurrency Markets. *Journal of Empirical Finance*, **50**, 1-19. https://doi.org/10.1016/j.jempfin.2018.11.002

[9] Gangwal, S. and Longin, F. (2018) Extreme Movements in Bitcoin Prices: A Study Based on Extreme Value Theory. *ESSEC Working Paper Ser*, **8**, 1-17.

[10] Begušić, S., Kostanjčar, Z., Stanley, H.E. and Podobnik, B. (2018) Scaling Properties of Extreme Price Fluctuations in Bitcoin Markets. *Physica A*: *Statistical Mechanics and its Applications*, **510**, 400-406. https://doi.org/10.1016/j.physa.2018.06.131

[11] Zhang, Y., Chan, S. and Nadarajah, S. (2019) Extreme Value Analysis of High-Frequency Cryptocurrencies. *High Frequency*, **2**, 61-69. https://doi.org/10.1002/hf2.10032

[12] Gkillas, K. and Katsiampa, P. (2018) An Application of Extreme Value Theory to Cryptocurrencies. *Economics Letters*, **164**, 109-111. https://doi.org/10.1016/j.econlet.2018.01.020

[13] Likitratcharoen, D., Ranong, T.N., Chuengsuksomboon, R., Sritanee, N. and Pansriwong, A. (2018) Value at Risk Performance in Cryptocurrencies. *The Journal of Risk Management and Insurance*, **22**, 11-28.

[14] Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50**, 987-1007.

https://doi.org/10.2307/1912773

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

[16] Hansen, P.R. and Lunde, A. (2005) A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH (1, 1)? *Journal of Applied Econometrics*, **20**, 873-889. https://doi.org/10.1002/jae.800

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

[18] 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

[19] 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

[20] Taylor, S. (1986) Modelling Financial Time Series. John Wiley & Sons, Great Britain.

[21] Engle, R. F. and Lee, G. (1999) A Long-Run and Short-Run Component Model of Stock Return Volatility. In: *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, 475-497.

[22] Higgins, M.L. and Bera, A.K. (1992) A Class of Nonlinear ARCH Models. *International Economic Review*, **33**, 137-158. https://doi.org/10.2307/2526988

[23] Engle, R. F. and Ng, V. K. (1993) Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, **48**, 1749-1778. https://doi.org/10.1111/j.1540-6261.1993.tb05127.x

[24] Azzalini, A. and Capitanio, A. (2003) Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew *T*-Distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 367-389. https://doi.org/10.1111/1467-9868.00391

[25] Theodossiou, P. (2015) Skewed Generalized Error Distribution of Financial Assets and Option Pricing. *Multinational Finance Journal*, **19**, 223-266. https://doi.org/10.17578/19-4-1

[26] Barndorff-Nielsen, O. (1978) Hyperbolic Distributions and Distributions on Hyperbolae. *Scandinavian Journal of Statistics*, **5**, 151-157.

[27] Rigby, R.A. and Stasinopoulos, D.M. (2005) Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507-554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

[28] 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

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

[30] Smith, R. L. (1987) Approximations in Extreme Value Theory. North Carolina University at Chapel Hill Center for Stochastic Processes.

[31] 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

[32] 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

[33] Christoffersen, P.F. (1998) Evaluating Interval Forecasts. *International Economic Review*, **39**, 841-862. https://doi.org/10.2307/2527341