Modelling Volatility of Cryptocurrencies Using Markov-Switching Garch Models

Guglielmo Maria Caporale, Timur Zekokh

Modelling Volatility of Cryptocurrencies Using Markov-Switching Garch Models

Abstract

This paper aims to select the best model or set of models for modelling volatility of the four most popular cryptocurrencies, i.e. Bitcoin, Ethereum, Ripple and Litecoin. More than 1,000 GARCH models are fitted to the log returns of the exchange rates of each of these cryptocurrencies to estimate a one-step ahead prediction of Value-at-Risk (VaR) and Expected Shortfall (ES) on a rolling window basis. The best model or superior set of models is then chosen by backtesting VaR and ES as well as using a Model Confidence Set (MCS) procedure for their loss functions. The results imply that using standard GARCH models may yield incorrect VaR and ES predictions, and hence result in ineffective risk-management, portfolio optimisation, pricing of derivative securities etc. These could be improved by using instead the model specifications allowing for asymmetries and regime switching suggested by our analysis, from which both investors and regulators can benefit.

Guglielmo Maria Caporale*
Department of Economics and Finance
Brunel University
United Kingdom – London, UB8 3PH
Guglielmo-Maria.Caporale@brunel.ac.uk

Timur Zekokh
Higher School of Economics
Moscow / Russia

*corresponding author

July 2018

1. Introduction

Modelling volatility is crucial for risk management. Following the global financial crisis of 2008, the Basel III international regulatory framework for banks has imposed more stringent capital requirements, and enhanced risk management systems have been developed. Since then the international financial system has had to face a new challenge, namely the introduction of decentralised cryptocurrencies, the first being Bitcoin, which was created in 2009 (Nakamoto, 2009). Unlike traditional currencies, cryptocurrencies are based on cryptographic proof, which provides many advantages over traditional payment methods (such as credit cards) including high liquidity, lower transaction costs, and anonymity (these features are discussed by Fantazzini et al., 2016).

Interest in Bitcoin and other cryptocurrencies has risen considerably in recent years. Their market capitalisation increased from approximately 18 billion US dollars at the beginning of 2017 to nearly 600 billion at the end of that year, [1] and high returns have attracted new investors. In addition, two big exchanges, i.e. the Chicago Mercantile Exchange (CME) and the Chicago Board Options Exchange (CBOE), started to trade futures on Bitcoin.[2] As a result of these developments, central banks have been facing the question of whether or not cryptocurrencies should be regulated, given the numerous technical and legal issues involved.

Further, cryptocurrencies are highly volatile and consequently it is important to estimate appropriate risk metrics, which can be used for calculating capital requirements, margins, hedging and pricing derivatives etc. It is well known that standard GARCH models can produce biased results if the series display structural breaks (Bauwens et al. (2010, 2014)); these are likely to occur in the case of cryptocurrencies, and therefore a suitable modelling approach should be used. Ardia et al. (2017) suggest estimating in such cases Markov-Switching GARCH models, whose parameters are allowed to change over time according to a discrete latent variable.

The aim of this paper is to find the best model or set of models for modelling volatility of the four most popular cryptocurrencies, i.e. Bitcoin, Ethereum, Ripple and Litecoin. More than 1,000 GARCH models are fitted to the log returns of the exchange rates of each of these cryptocurrencies to estimate a one-step ahead prediction of Value-at-Risk (VaR) and Expected Shortfall (ES) on a rolling window basis. The best model or superior set of models is then chosen by backtesting VaR and ES as well as using a Model Confidence Set (MCS) procedure for their loss functions.

---

[1] Source: https://coinmarketcap.com/
[2] Source: Thomson Reuters

The paper is organised as follows. Section 2 briefly discusses the relevant literature; Section 3 provides a description of the data; Section 4 outlines the methodology; Section 5 presents the empirical results; finally, Section 6 offers some concluding remarks.

2. Literature Review

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are the most commonly used in the literature for modelling volatility and estimating Value-at-Risk (VaR) and Expected Shortfall (ES). The original Autoregressive Conditional Heteroskedasticity (ARCH) specification was introduced by Engle (1982) and then extended by Bollerslev (1986), who put forward the GARCH framework. Additional specifications were then developed: the exponential GARCH (EGARCH) model of Nelson (1991), the threshold GARCH (TGARCH) model of Zakoian (1994), the Student's t-GARCH model of Bollerslev (1987), the GJRGARCH model of Glosten et al. (1993) and many others (see Bollerslev et al. (1992), Bollerslev et al. (1994) and Engle (2004)).

Recent studies have shown that structural breaks result in biased estimates of GARCH models and poor volatility forecasts (Bauwens et al. (2010, 2014)). To overcome this problem Markov-switching GARCH models (MSGARCH) have been proposed, whose parameters can change over time according to a discrete latent (i.e., unobservable) variable (Ardia et al., 2017); to make computations easier a normal (or mixture) distribution) is typically assumed. This framework has been used in recent papers to analyse various type of assets: commodity prices (Alizadeh et al. (2008)), stock returns (Henry (2009)), exchange rate returns (Wilfling (2009), Bohl et al. (2011)), etc. Ardia et al. (2017) estimated a large set of MSGARCH specifications allowing for various functional forms as well as distributions of the errors. They compared two estimation techniques, namely Maximum Likelihood (ML) and Markov chain Monte Carlo (MCMC) procedures, and found that MSGARCH outperforms single regime models in the case of stock prices, but not of stock indices and currencies.

GARCH models have also been used for modeling the volatility of cryptocurrencies. Glaser et al. (2014) estimated a standard GARCH (1, 1). Gronwald (2014) reported that an autoregressive jump-intensity GARCH model fits the Bitcoin data better than a standard GARCH. Dyhrberg (2016) estimated an asymmetric GARCH for Bitcoin arguing that it can be used for hedging.

Bouoiyour and Selmi (2016) compared different specifications including EGARCH, Asymmetric Power ARCH (APARCH), weighted GARCH and component GARCH with multiple thresholds by using in-sample criteria such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQC); they concluded that, despite a noticeable decrease in its volatility, Bitcoin cannot yet

be considered a mature currency. Katsiampa (2017) found that the AR-CGARCH model gives the best fit for Bitcoin, which means that accounting for both the short- and long-term components of the conditional variance is important. Chu et al. (2017) fitted 12 GARCH specifications with different distributions for the innovation process to seven cryptocurrencies; the Integrated GARCH (1, 1) (IGARCH) model with normal innovations was found to yield the smallest values of AIC, the corrected AIC (AICc), BIC, HQC and the Consistent AIC (CAIC) in the case of Bitcoin, Dash, Litecoin, Maidsafecoin and Monero; instead the GJRGARCH (1, 1) and GARCH (1, 1) model with normal innovations produced the smallest values for Dogecoin and Ripple respectively.

3. Data Description

The series analysed are the daily closing prices of Bitcoin, Ethereum, Ripple and Litecoin. The Bitcoin data were taken from the Coindesk Price Index and cover the period from 18 July 2010 to 30 April of 2018; for the other three series the data source is CoinMarketCap.[3] The end date is the same for all series, whilst the start date differs: it is 7 August 2015 for Ethereum, 4 August 2013 for Ripple, and 28 April 2013 for Litecoin. Prices were transformed into log returns by taking first differences of their logarithm (see Figure 1).
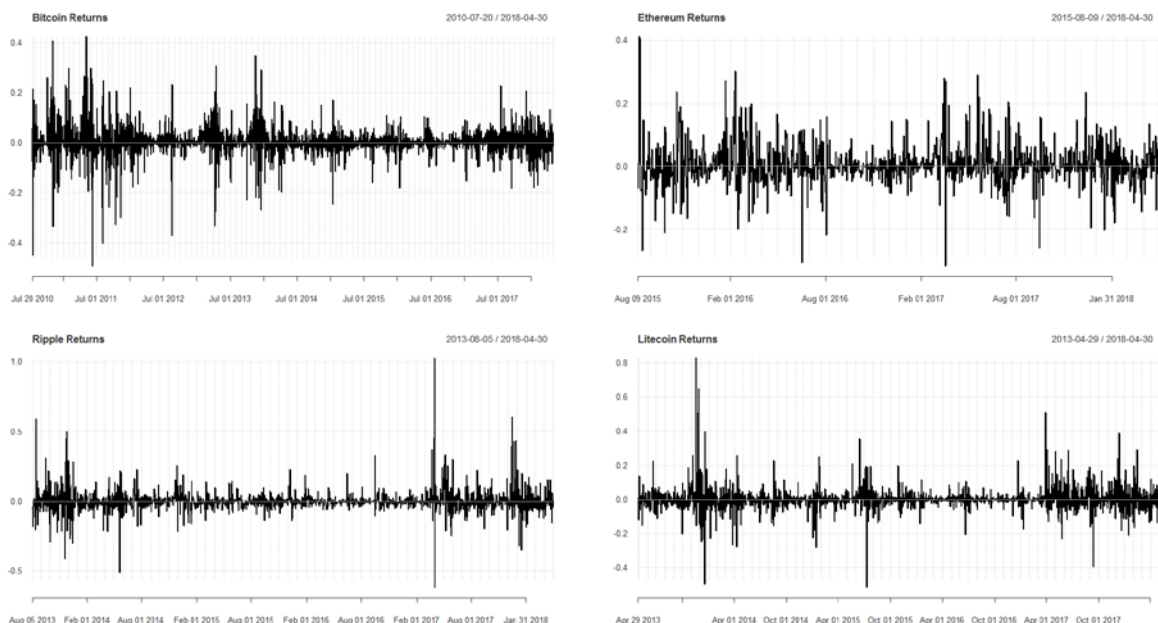

Figure 1. Log returns

Summary statistics for the log returns are shown in Table 1.

Table 1 Descriptive statistics of log returns

Statistics	Bitcoin	Ethereum	Ripple	Litecoin
Number of obs.	2 842	996	1 730	1 828
Minimum	-0.491	-0.315	-0.616	-0.514
Q1	-0.013	-0.027	-0.023	-0.019
Median	0.002	0	-0.003	0
Mean	0.004	0.007	0.003	0.002
Q3	0.023	0.036	0.022	0.018
Maximum	0.425	0.412	1.027	0.829
Std Dev.	0.058	0.072	0.080	0.069
Skewness	-0.326	0.515	1.998	1.777
Kurtosis	11.811	4.347	26.629	24.961

As can be seen, they are negatively skewed in the case of Bitcoin and positively skewed in all other cases. All four cryptocurrencies exhibit leptokurtosis. The histograms of the log returns series are shown in Figure 2.
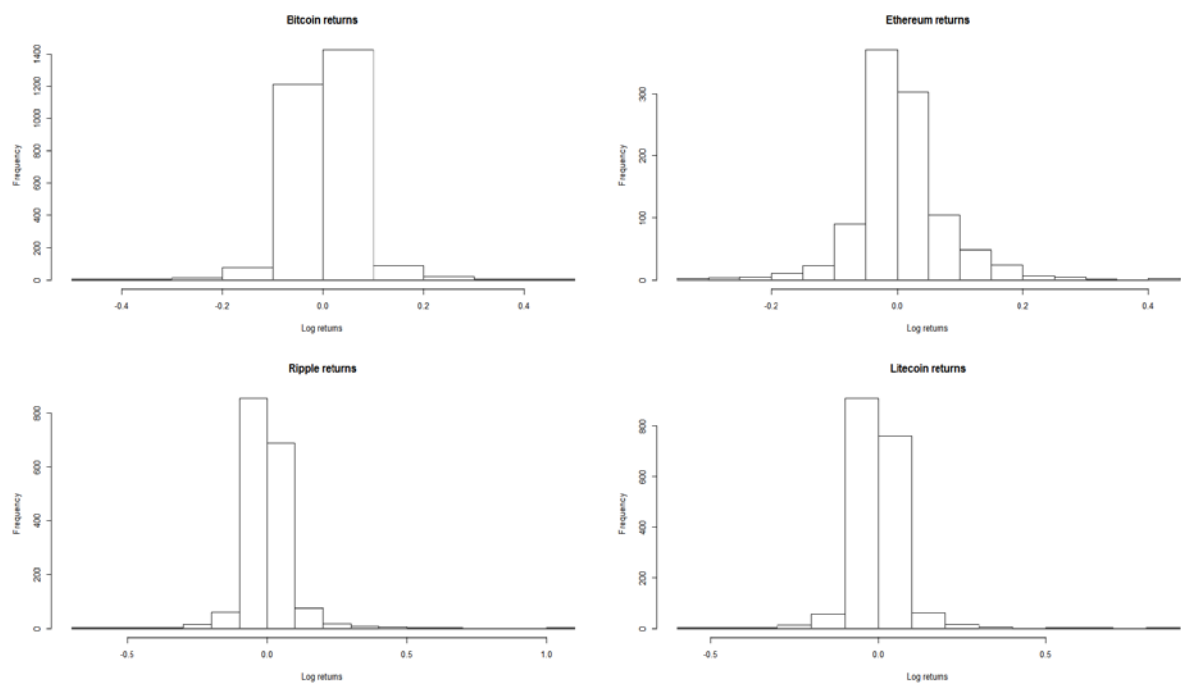

Figure 2 Histograms of log returns

4. Methodology

4.1 GARCH models

Let $y_t \in R$ be the percentage log-returns of the financial asset (exchange rate) of interest at time t. The general Markov-Switching GARCH specification has the following form:

$$y_t|(S_t = k, I_{t-1}) \sim F\big(0, h_{k,t}, \theta_k\big),$$

where $F\left(0, h_{k,t}, \theta_k\right)$ is a continuous distribution with zero mean, time-varying variance $h_{k,t}$, and additional shape parameters contained in the vector θ_k. Following Ardia et al. (2017), the integer-valued stochastic variable S_t, defined on the discrete space {1, …, K}, is assumed to evolve according to an unobserved first-order ergodic homogeneous Markov chain with transition probability matrix $\boldsymbol{P} = \{p_{i,j}\}^k_{i,j=1}$, with $p_{i,j} = P[S_t = j|S_{t-1} = i]$, with I_{t-1} being the information set available at time t-1.

As in Haas et al. (2004), the conditional variance of y_t is assumed to follow a GARCH process. This is not restricted to be the standard GARCH model:

$$h_{k,t} = \omega\left(y_{t-1}, h_{k,t-1}, \theta_k\right),$$

where $\omega(\cdot)$ – is a I_{t-1} – measurable function which defines the filter for the conditional variance and also ensures its positiveness. By contrast, a variety of GARCH specifications are considered, in particular:

SGARCH (Bollerslev (1986))

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

EGARCH (Nelson (1991))

$$\ln(\sigma_t^2) = w + \sum_{i=1}^{p} \left\{\alpha_i * \left(\frac{|\varepsilon_{t-i}|}{\sigma_{t-i}} - E\left[\frac{|\varepsilon_{t-i}|}{\sigma_{t-i}}\right]\right) + \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right\} + \sum_{j}^{q} \ln(\sigma_{t-j}^2).$$

GJRGARCH (Glosten et al. (1993))

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \left[\alpha_i + \gamma_i I_{(\varepsilon_{t-i}>0)}\right]\varepsilon_{t-i}^2 + \sum_{1}^{q} \beta_j \sigma_{t-j}^2.$$

TGARCH (Zakoian (1994))

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i[(1 - \gamma_i)\varepsilon_{t-i}^+ - (1 + \gamma_i)\varepsilon^-] + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

As for distribution mixture models, suppose that:

$$y_t \sim DM(p_1, \ldots, p_k; \mu_1, \ldots, \mu_k; h_1, \ldots, h_k),$$

where DM is a mixture of densities with the following form:

$$\vartheta(y) = \sum_{1}^{k} p_i f_i(y), \sum_{1}^{k} p_i = 1, f_i(y) = f(y; \mu_i; h_i),$$

where $[p_1, \ldots, p_k]$ is the positive mixing law and f denotes the density function.

As an example of a distribution mixture model, consider the normal mixture models: according to Alexander and Lazar (2006), these can be seen as Markov switching GARCH

models in a restricted form, where the transition probabilities are independent of the past state. It is assumed that K variances follow some mixture of distributions. For example, the normal mixture standard GARCH (1, 1) will be defined as follows:

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2, for\ i = 1, \dots, K.$$

According to the Cifter and Ozun (2007) the overall conditional variance will then be:

$$\sigma_t^2 = \sum_{i=1}^{K} p_i \sigma_{it}^2 + \sum_{i=1}^{K} p_i \mu_i^2.$$

4.2 Value-at-Risk Backtesting

Value-at-Risk at level α is defined as the maximum loss one could expect to incur with probability α over a specific period. Mathematically:

$$VaR_\alpha(X) = \inf\{x | F_X(x) \geq \alpha\} = F_X^-(x),$$

where $VaR_\alpha(X)$ is VaR at level α of a random variable X, $F_X(x)$ is the cumulative distribution function of X. Basically, VaR is the α-quantile of the underlying distribution.

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

$$I_{t+1} = 1, if\ r_{t+1} < -VaR_{t+1}^\alpha,$$
$$I_{t+1} = 0, if\ r_{t+1} > -VaR_{t+1}^\alpha,$$

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

$$H_0: I_{t+1} \sim Bernoulli(\alpha),$$
$$f(I_{t+1}, p) = (1 - p)^{1 - I_{t+1}} p^{I_{t+1}}.$$

The unconditional coverage (UC) test of Kupiec (1995) uses the fraction of observed violations for a particular risk model – π and compares it with p. For this purpose the likelihood Bernoulli function is needed:

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

where T_0, T_1 are the number of 0s and 1s in the sample ($T = T_0 + T_1$). The maximum likelihood estimator is:

$$\hat{\pi} = \frac{T_1}{T_0 + T_1},$$

The hypothesis of interest can be tested by means of the following likelihood ratio test:

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

Under the null hypothesis that the model is correct LR_{uc} is asymptotically χ_1^2.

However, this test focuses only on the number of exceptions. A situation can arise when the model passes the unconditional coverage test but all violations are concentrated. In

order to reject a VaR with clustered violations a test of independence of the hit sequence is needed. Suppose that this exhibits time dependence and follows a first-order Markov sequence with the following transition probability matrix:

$$\Pi_1 = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix},$$

where π_{01} is the probability of getting a violation tomorrow given no violation today.

Then the corresponding likelihood function is the following:

$$L(\Pi_1) = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}},$$

where T_{ij} is the number of observations with a j following i.

If the hit sequence is independent over time, then $\pi_{01} = \pi_{11} = \pi$. The transition probability matrix will have the following form:

$$\widehat{\Pi} = \begin{pmatrix} 1 - \hat{\pi} & \hat{\pi} \\ 1 - \hat{\pi} & \hat{\pi} \end{pmatrix}.$$

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

$$LR_{ind} = -2 \ln \left[\frac{L(\hat{\pi})}{L(\widehat{\Pi_1})} \right] \sim \chi_1^2.$$

It is important for VaR users to be able to test simultaneously whether the hit sequence is independent and the average number of violations is correct; the conditional coverage (CC) test developed by Christoffersen (1998) makes this possible. A sequence of VaR forecasts at risk level α has the correct conditional coverage if $\{I_t^a; t = 1, \dots, H\}$ is an independent and identically distributed sequence of Bernoulli random variables with parameter α. To test this hypothesis a joint test of independence of the series and the unconditional coverage of the VaR forecasts the likelihood ratio defined as follows is carried out:

$$LR_{cc} = -2 \ln \left[\frac{L(\alpha)}{L(\widehat{\Pi_1})} \right] \sim \chi_2^2.$$

This is equivalent to testing the null $\pi_{01} = \pi_{11} = \alpha$. Note also that $LR_{cc} = LR_{uc} + LR_{ind}$.

An additional test is the dynamic quantile (DQ) one introduced by Engle and Manganelli (2004). It is based on a linear regression of the hit variable on a set of explanatory variables including a constant, the lagged values of the hit variable and any useful function of past information. Let us denote $Hit_t(\alpha) = I_t(\alpha) - \alpha$. Under the correct model specification, the following moment conditions are satisfied: $E[Hit_t(\alpha)] = 0, E[Hit_t(\alpha)|I_{t-1}] = 0, E[Hit_t(\alpha)Hit_{t`}(\alpha)] = 0$ for $t \neq t`$. The linear regression is the following:

$$Hit_t(\alpha) = \delta + \sum_{k=1}^{K} \beta_k Hit_{t-1}(\alpha) + \sum_{k=1}^{K} \gamma_k g[Hit_{t-k}(\alpha), Hit_{t-k-1}(\alpha), \dots] + \varepsilon_t,$$

where $g(.)$ is a function of past information. The null hypothesis of conditional efficiency is equivalent to testing whether the coefficients are jointly equal to zero:

$$H_0: \delta = \beta_1 = \cdots = \beta_k = \gamma_1 = \cdots = \gamma_k = 0, \forall\, k = 1, \dots, K.$$

However, these tests do not provide any insights into the magnitude of exceedance and therefore do not enable the researcher to make model comparisons.

A final test is due to González-Rivera et al. (2004) and McAleer and Da Veiga (2008). It uses the asymmetric linear losses incurred by VaR forecasts. The quantile loss (QL) is given by the following function:

$$QL_t^a = \big(\alpha - \mathrm{I_t}(\alpha)\big)(y_t - VaR_t^a).$$

4.3 Expected Shortfall Backtesting

VaR is an elicitable risk measure, where a statistics $\phi(Y)$ of a random variable Y is said to be elicitable if it minimises the expected value of a scoring function S (Acerbi and Szekely (2014)):

$$\phi = arg\min_x E[S(x,Y)].$$

It has been shown that VaR has several shortcomings, and that in particular is not able to capture tail risks beyond the α-quantile (Artzner et al. (1999), Danielsson et al. (2001), Basel Committee (2013)). For these reasons, ES was introduced (Artzner et al. (1997, 1999)). To put it simply, ES is the expected loss given that the loss L exceeds VaR ($ES = E[L|L > VaR]$). Mathematically:

$$ES_\gamma(X) = \frac{1}{\gamma}\int_0^\gamma VaR_\alpha(X)\, d\alpha,$$

where $ES_\gamma(X)$ is ES of a random variable X at level γ.

The transition from VaR to ES as the main market risk metric (Basel Committee, 2016) makes it necessary to have a reliable backtesting procedure of ES.

McNeil and Frey (2000) suggested an exceedance residual (ER) test that involves fitting a (volatility) model to the return data, estimating predictions of the mean and volatility, and then calculating the implied model residuals. Next, extreme value theory (EVT) can be applied, while treating the residuals as a white noise process, to model the tail of the marginal distribution. Lastly, the EVT model can be used to estimate the upper quantile of the marginal distribution. One can then test whether the expected value of ER is zero against the one-sided alternative that it is greater than zero or, equivalently, that ES is systematically underestimated.

The most recent test is the regression-based one of Bayer and Dimitriadis (2018), known as the ESR test. These authors use a joint regression framework for the quantile and

the ES, and present two-sided and one-sided versions of the test. Suppose that $\widehat{e}_t$ is the ES forecast and Y_t is the log return. Then one can regress returns on the ES forecast as follows:

$$Y_t = \alpha + \beta * \hat{e}_t + u_t^e,$$

where $ES_\gamma(u_t^e|\mathcal{F}_{t-1}) = 0$. One can then test the joint hypothesis that α is equal to 0 and β is equal to 1, $H_0: (\alpha, \beta) = (0, 1)$, against the alternative that $H_1: (\alpha, \beta) \neq (0, 1)$.

As the functional ES is not elicitable, these authors estimate semi-parametrically the following system:

$$Y_t = \alpha_0 + \beta_0 * \widehat{e}_t + u_t^q,$$
$$Y_t = \alpha_1 + \beta_1 * \widehat{e}_t + u_t^e,$$

where $VaR_\gamma(u_t^q|\mathcal{F}_{t-1}) = 0$ and $ES_\gamma(u_t^e|\mathcal{F}_{t-1}) = 0$.

A Wald statistic is then carried out using the parameters (α, β):

$$T_{ESR} = ((\alpha, \beta)' - (0, 1)')' \widehat{\Sigma_{ES}^{-1}} ((\alpha, \beta)' - (0, 1)')',$$

where $\widehat{\Sigma_{ES}}$ is an estimator for the (asymptotic) covariance matrix of the M-estimator of the parameters (α, β). Hence, the test statistics follows asymptotically a χ_2^2 distribution.

The test described above has been named the bivariate ESR test by Bayer and Dimitriadis (2018). Since in their opinion the possibility of underestimating risk is the main issue for regulators, they also suggest another regression-based backtesting procedure for the ES:

$$Y_t - \widehat{e}_t = \alpha + u_t^e,$$

where $ES_\gamma(u_t^e|\mathcal{F}_{t-1}) = 0$ and the null hypothesis is that α is zero. These are *t*-tests based on the asymptotic covariance.

4.4 Model Confidence Set

The backtesting procedures described above cannot help to select the best GARCH specification (even QL is not sufficiently informative). For this purpose we shall use instead the model confidence set (MCS) procedure introduced by Hansen et al. (2011). This consists of a sequence of tests that is used to construct a set of "superior" models (SSM) for which the null hypothesis of equal predictive ability (EPA) is not rejected.

Specifically, consider a set, M^0, that contains a finite number of models ($i = 1, ..., m_0$); these are evaluated in terms of a specified loss function for the i-th model at time t, $L_{i,t}$, where $t = 1, ..., n$.

Let the performance variable be:

$$d_{ij,t} = L_{i,t} - L_{j,t} \ \forall \ i, j \ \in M^0.$$

Then define the relative sample loss statistics

$$\bar{d}_{ij} = \frac{1}{n} \sum_{t=1}^{n} d_{ij,t},$$

and

$$\bar{d}_{i.} = \frac{1}{m} \sum_{j \in M} \bar{d}_{ij}.$$

Finally assume that

$$\mu_{ij} = E\big(d_{ij}\big) \text{ and } \mu_{i.} = E(d_{i.})$$

is finite and does not depend on $t \; \forall \; i, j \in M^0$.

Therefore, the set of superior models is defined as

$$M^* = \big\{i \in M^0 : \mu_{ij} \leq 0 \; \forall \; j \in M^0\big\}.$$

In order to eliminate inferior elements of the set M^0 the following EPA hypotheses are tested

$$H_{0,M} : \mu_{ij} = 0 \; \forall \; i, j \in M,$$

$$H_{A,M} : \mu_{ij} \neq 0 \; for \; some \; i, j \in M$$

or

$$H_{0,M} : \mu_{i.} = 0 \; \forall \; i, j \in M,$$

$$H_{A,M} : \mu_{i.} \neq 0 \; for \; some \; i, j \in M$$

where $M \subset M^0$.

From these statistics the following t-statistics can be constructed

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \text{ and } t_{i.} = \frac{\bar{d}_{i.}}{\sqrt{\widehat{var}(\bar{d}_{i.})}} \text{ for } i, j \in M.$$

The two EPA hypotheses map into the two test statistics

$$T_{R,M} = \max_{i,j \in M} |t_{ij}| \text{ and } T_{max,M} = \max_{i \in M} t_i.$$

As mentioned before, the MCS procedure is a sequential testing procedure that removes the worst model that removes the worst model at each step, until the hypothesis of EPA is not rejected for all models in the SSM. The choice of the worst model to be eliminated is made using an elimination rule that is coherent with the statistic tests defined above, namely:

$$e_{max,M} = arg \max_{i \in M} \frac{\bar{d}_{i.}}{\widehat{var}(\bar{d}_{i.})}, e_{R,M} = arg \max_{i} \left\{ \sup_{j \in M} \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \right\}.$$

To summarise, the MCS procedure consist of the following steps:

1. set $M = M^0$,

2. Test for the EPA hypothesis; if EPA is not rejected terminate the algorithm and set $M = M^*$, otherwise use the elimination rule and find the worst model,

3. Remove the worst model, and go to step 2.

The MCS procedure requires a loss function. For the VaR the QL will be used. However, the choice of a loss function is not straightforward. Various papers have attempted to develop a consistent scoring function. Here we shall use the same loss function for ES backtesting as in Bayer and Dimitriadis (2018), which belongs to a class of functions originally introduced by Fissler and Ziegel (2016) in the context of forecast evaluation. They derived a consistent scoring function for (VaR_α, ES_α) in the following way.

Let $\alpha \in (0, 1)$. Also let $\mathcal{F}$ be a class of distribution functions on $\mathcal{R}$ with finite first moments and unique α-quantiles, and $A_0 = \{x \in \mathcal{R}^2 : x_1 \geq x_2\}$. A scoring function S is then defined as follows: $A_0 \times \mathcal{R} \to \mathcal{R}$ of the form

$$S(x_1, x_2, y) = \left(1_{\{y \leq x_1\}} - \alpha\right)G_1(x_1) - 1_{\{y \leq x_1\}}G_1(y) + G_2(x_2)\left(x_2 - x_1 + \frac{1}{\alpha}1_{\{y \leq x_1\}}(x_1 - y)\right)$$

$$- \mathcal{G}_2(x_2) + a(y),$$

where $G_1, G_2, \mathcal{G}_2, a : \mathcal{R} \to \mathcal{R}, \mathcal{G}_2' = G_2, a$ is $\mathcal{F}$-integrable and $1_{(-\infty; x_1]}G_1$ is $\mathcal{F}$-integrable $\forall x_1 \in \mathcal{R}$, is $\mathcal{F}$-consistent for (VaR_α, ES_α) if G_1 is increasing and $\mathcal{G}_2$ is increasing and convex. If $\mathcal{G}_2$ is strictly increasing and strictly convex, then S is strictly $\mathcal{F}$-consistent for (VaR_α, ES_α). Bayer and Dimitriadis (2018) used the scoring function defined above as follows:

$$S(VaR_\alpha^t, ES_\alpha^t, y_t) = \frac{1}{-ES_\alpha^t}\left(ES_\alpha^t - VaR_\alpha^t + \frac{(VaR_\alpha^t - y_t)1_{\{y \leq x_1\}}}{\alpha}\right) + \log(-ES_\alpha^t),$$

which is the 0-homogeneous loss function (see Dimitriadis and Bayer (2017), Taylor (2017), Patton et al. (2017), Barendse (2017) and Nolde and Ziegel (2017)).

5. Empirical Results

As mentioned above, we estimate the following GARCH specifications: SGARCH, EGARCH, GJRGARCH and TGARCH. Normal (norm), skewed normal (snorm), Student's t (std), skewed Student's t (sstd), generalized error (GED) and skewed generalized error distributions (sged) are used. Finally, the mixture parameter is included; if this takes the value of one (TRUE) then the model becomes the distribution mixture GARCH, if it is instead set equal to zero (FALSE) it yields the MSGARCH specification. In total, 1176 GARCH models were estimated for each cryptocurrency.

The results were obtained from 1-step ahead VaR and ES predictions. A moving window with refitting at every step was used for the estimates and the predictions. The window size is 70% of the total number of observations. The models that did not fail the backtesting procedures were then used in the MCS procedure in order to obtain the best model or set of models. The p-values in the MCS procedure was set equal to 30%.

5.1 Value-at-Risk Results

Table 2 reports p-values for the backtesting procedures and the MCS procedure for the superior set of models (SSM) for Bitcoin. As can be seen, 24 models out of 1176 satisfy the VaR and ES backtesting procedures and were selected by MCS as models with equal predictive power with respect to QL. Note that none of the single regime models was selected for the SSM.

Standard GARCH models prevail in the first and second regimes. Interestingly, the normal distribution prevails in the first regime, but whilst the Student's t distribution is appropriate for 70% of the models in the second regime. Mixture models represent approximately 60% of those selected for the SSM. It is also noteworthy that in the first regime a variety of specifications appear to be appropriate, whilst in the second standard GARCH and GJRGRACH account for more than 90% of the specifications. The same pattern emerges for the distribution functions: in the second regime the Student's t and normal distribution are chosen in 87.5% of the cases. Interestingly, no models with a skewed GED distribution are selected for the SSM, and none with skewed distributions in the second regime. Specifications that account for leverage effects at least in one of the regimes represent more than 75% of those selected for the SSM.

Table 2 P-values of Bitcoin backtesting procedures and MCS procedure for VaR loss function

	UC	CC	DQ	ER	ESR	MCS
tGARCH-sGARCH-sstd-std-TRUE	63%	84%	99%	68%	87%	100%
tGARCH-gjrGARCH-sstd-std-TRUE	61%	83%	100%	81%	80%	100%
eGARCH-sGARCH-snorm-std-TRUE	86%	89%	99%	71%	97%	100%
sGARCH-sGARCH-norm-std-TRUE	87%	91%	100%	66%	94%	100%
sGARCH-sGARCH-norm-std-FALSE	85%	91%	100%	58%	96%	100%
sGARCH-gjrGARCH-norm-std-FALSE	85%	91%	100%	59%	96%	100%
gjrGARCH-eGARCH-ged-std-TRUE	59%	81%	100%	66%	71%	100%
gjrGARCH-gjrGARCH-norm-std-FALSE	85%	91%	100%	67%	89%	99%
eGARCH-sGARCH-sstd-std-TRUE	60%	82%	100%	68%	56%	98%
sGARCH-gjrGARCH-snorm-std-FALSE	87%	90%	99%	67%	95%	97%
gjrGARCH-sGARCH-norm-ged-FALSE	87%	90%	100%	81%	68%	97%
eGARCH-sGARCH-std-std-TRUE	60%	82%	100%	54%	62%	96%
sGARCH-sGARCH-snorm-std-FALSE	87%	90%	100%	67%	99%	95%
gjrGARCH-gjrGARCH-std-std-TRUE	85%	91%	97%	57%	73%	94%
sGARCH-sGARCH-snorm-std-TRUE	87%	91%	100%	65%	94%	94%
eGARCH-gjrGARCH-norm-norm-TRUE	87%	90%	99%	84%	85%	94%
sGARCH-gjrGARCH-snorm-norm-TRUE	87%	90%	99%	72%	91%	78%
sGARCH-tGARCH-std-std-TRUE	89%	92%	100%	77%	81%	74%
sGARCH-sGARCH-norm-ged-FALSE	59%	81%	100%	65%	81%	74%
tGARCH-sGARCH-norm-ged-TRUE	59%	81%	100%	54%	91%	71%
tGARCH-sGARCH-std-norm-FALSE	59%	81%	100%	72%	65%	67%
gjrGARCH-sGARCH-norm-std-FALSE	85%	91%	100%	59%	97%	67%
sGARCH-sGARCH-std-norm-FALSE	59%	81%	100%	58%	84%	66%
gjrGARCH-sGARCH-std-std-TRUE	60%	82%	92%	59%	92%	66%

Table 3 shows the results for Ethereum. Only one model was selected. It is a mixture of GJRGARCH with Student's t distribution in the first regime and TGARCH with skewed Student's t distribution in the second regime.

Table 3 P-values of Ethereum backtesting procedures and MCS procedure for VaR loss function

	UC	CC	DQ	ER	ESR	MCS
gjrGARCH-tGARCH-std-sstd-TRUE	96%	97%	99%	100%	65%	100%

Table 4 reports the results for Ripple. Only eight models satisfied all backtesting procedures and were selected by the MCS procedure. For the first regime models with standard GARCH or TGARCH were chosen by this procedure. For the second regime 50% of the models selected for the SSM have a TGARCH specification and 25% of them has either a standard GARCH or GJRGARCH specification. In the first regime normal and skewed normal distributions prevail, while in the second the Student-t and skewed Student-t distributions are found to be appropriate for all models. The same percentage of mixture models and Markov-switching models are selected for the SSM. Specifications accounting for leverage effects at least in one of the regimes represent more than 75% of those in the SSM.

Table 4 P-values of Ripple backtesting procedures and MCS procedure for VaR loss function

	UC	CC	DQ	ER	ESR	MCS
sGARCH-sGARCH-snorm-sstd-TRUE	30%	57%	100%	54%	66%	100%
tGARCH-gjrGARCH-sstd-sstd-FALSE	11%	28%	87%	89%	61%	100%
sGARCH-tGARCH-sged-sstd-FALSE	11%	28%	96%	100%	84%	100%
sGARCH-tGARCH-snorm-std-TRUE	12%	29%	95%	6%	58%	100%
tGARCH-gjrGARCH-norm-std-FALSE	30%	57%	96%	99%	10%	99%
tGARCH-tGARCH-norm-sstd-FALSE	44%	68%	94%	100%	10%	98%
sGARCH-tGARCH-norm-std-TRUE	11%	28%	84%	100%	99%	69%
sGARCH-sGARCH-snorm-std-TRUE	94%	95%	100%	79%	85%	64%

Table 5 shows that MCS selects the highest percentage of models satisfying the backtesting procedures in the case of Litecoin; in total, 33 models out of 1176 were chosen. In the first regime TGARCH models represent 40% of those selected, standard GARCH ones 30% and EGARCH and GJRGARCH the rest; the most common distribution is the normal, followed by the skewed normal and Student's t. In the second regime standard GARCH models exceed 50%, and the Student's t and skewed Student's t distributions are also selected in more than 50% of the cases for the SSM. Mixture models represent more than 80% of those in the SSM. Specifications allowing for leverage at least in one regime account for approximately 76% of those in the SSM.

Table 5 P-values of Litecoin backtesting procedures and MCS procedure for VaR loss function

	UC	CC	DQ	ER	ESR	MCS
gjrGARCH-tGARCH-norm-sstd-TRUE	51%	78%	62%	100%	85%	100%
tGARCH-sGARCH-snorm-sstd-TRUE	52%	79%	100%	56%	51%	100%
sGARCH-sGARCH-norm-sstd-FALSE	83%	91%	97%	60%	83%	100%
sGARCH-sGARCH-norm-std-FALSE	83%	91%	93%	70%	82%	100%
gjrGARCH-gjrGARCH-norm-sged-TRUE	50%	78%	99%	53%	59%	100%
gjrGARCH-tGARCH-sstd-std-TRUE	51%	78%	100%	63%	67%	100%
tGARCH-sGARCH-sged-sstd-TRUE	51%	78%	100%	50%	71%	100%
sGARCH-sGARCH-std-sstd-TRUE	83%	93%	100%	58%	90%	99%
tGARCH-gjrGARCH-std-sstd-TRUE	84%	94%	100%	57%	51%	99%
sGARCH-sGARCH-norm-sged-TRUE	83%	91%	100%	54%	76%	98%
sGARCH-sGARCH-snorm-std-TRUE	53%	75%	99%	63%	99%	98%
sGARCH-sGARCH-sstd-std-TRUE	83%	93%	100%	66%	80%	98%
gjrGARCH-sGARCH-sstd-std-TRUE	83%	93%	100%	68%	54%	97%
tGARCH-tGARCH-snorm-ged-TRUE	53%	80%	100%	86%	79%	97%
sGARCH-sGARCH-std-std-TRUE	51%	78%	100%	57%	74%	97%
tGARCH-tGARCH-norm-ged-TRUE	50%	78%	100%	58%	65%	95%
tGARCH-gjrGARCH-ged-sstd-TRUE	86%	94%	100%	86%	65%	93%
tGARCH-sGARCH-ged-sstd-TRUE	51%	78%	100%	66%	79%	93%
gjrGARCH-gjrGARCH-std-std-TRUE	50%	78%	100%	52%	54%	93%
tGARCH-sGARCH-sstd-std-TRUE	84%	94%	100%	88%	56%	84%
gjrGARCH-sGARCH-std-sged-TRUE	50%	78%	99%	54%	66%	81%
eGARCH-sGARCH-ged-sged-TRUE	50%	78%	100%	54%	67%	78%
gjrGARCH-eGARCH-norm-std-FALSE	50%	78%	100%	57%	53%	76%
eGARCH-tGARCH-snorm-ged-TRUE	50%	78%	100%	57%	60%	74%
sGARCH-tGARCH-snorm-ged-TRUE	50%	78%	100%	52%	54%	72%
eGARCH-gjrGARCH-snorm-sged-TRUE	50%	78%	100%	52%	53%	69%
tGARCH-sGARCH-sged-std-TRUE	51%	78%	99%	56%	57%	68%
sGARCH-sGARCH-snorm-sged-TRUE	83%	93%	100%	51%	85%	64%
tGARCH-sGARCH-sstd-sstd-TRUE	82%	91%	100%	86%	51%	63%
tGARCH-eGARCH-sstd-snorm-TRUE	85%	95%	100%	69%	53%	53%
tGARCH-sGARCH-norm-ged-FALSE	51%	78%	99%	99%	79%	52%
sGARCH-eGARCH-std-sged-FALSE	83%	91%	99%	64%	81%	38%
tGARCH-sGARCH-std-ged-FALSE	51%	78%	100%	94%	82%	34%

5.2 Expected Shortfall Results

Table 6 reports p-values for all backtesting procedures and the MCS procedure to choose the best set of models for Bitcoin. As can be seen 25 models out of 1176 satisfy the VaR and ES backtesting procedures and were selected by MCS as models with equal predictive power with respect to the joint loss function. Note that none of the single regime models was selected for the SSM.

The results from the MCS procedure with a joint loss function are mostly the same as those for QL, since SSM is constructed from the same set of models that did not fail the VaR and ES backtesting procedures. Standard GARCH models prevail in the first and second regimes. Interestingly, the normal distribution prevails in the first regime, but in the second the Student's t distribution is selected in 70% of the cases. Mixture models represent 60% of those in the SSM. In the first regime there is more variety of specifications, whilst in the second standard GARCH and GJRGRACH represent 90% of the chosen models. Similarly, the Student's t and normal distribution prevail in the second regime being selected in 88% of

the cases. No models with a skewed GED distribution are included in the SSM, and none with a skewed distribution in the second regime. Specifications accounting for leverage effects at least in one of the regimes represent 80% of those in the SSM.

Table 6 P-values of Bitcoin backtesting procedures and MCS procedure for joint loss function

	UC	CC	DQ	ER	ESR	MCS
eGARCH-sGARCH-snorm-std-TRUE	86%	89%	99%	71%	97%	100%
eGARCH-sGARCH-sstd-std-TRUE	60%	82%	100%	68%	56%	100%
gjrGARCH-gjrGARCH-std-std-TRUE	85%	91%	97%	57%	73%	100%
tGARCH-sGARCH-sstd-std-TRUE	63%	84%	99%	68%	87%	100%
sGARCH-sGARCH-norm-std-FALSE	85%	91%	100%	58%	96%	100%
sGARCH-sGARCH-norm-std-TRUE	87%	91%	100%	66%	94%	100%
gjrGARCH-sGARCH-norm-ged-FALSE	87%	90%	100%	81%	68%	100%
gjrGARCH-gjrGARCH-norm-std-FALSE	85%	91%	100%	67%	89%	100%
tGARCH-gjrGARCH-sstd-std-TRUE	61%	83%	100%	81%	80%	100%
sGARCH-gjrGARCH-norm-std-FALSE	85%	91%	100%	59%	96%	100%
eGARCH-sGARCH-std-std-TRUE	60%	82%	100%	54%	62%	100%
sGARCH-sGARCH-norm-ged-FALSE	59%	81%	100%	65%	81%	100%
sGARCH-sGARCH-std-norm-FALSE	59%	81%	100%	58%	84%	100%
gjrGARCH-eGARCH-ged-std-TRUE	59%	81%	100%	66%	71%	99%
tGARCH-sGARCH-std-norm-FALSE	59%	81%	100%	72%	65%	99%
tGARCH-sGARCH-norm-ged-TRUE	59%	81%	100%	54%	91%	99%
sGARCH-gjrGARCH-snorm-std-FALSE	87%	90%	99%	67%	95%	99%
gjrGARCH-sGARCH-std-std-TRUE	60%	82%	92%	59%	92%	99%
sGARCH-sGARCH-snorm-std-FALSE	87%	90%	100%	67%	99%	97%
sGARCH-tGARCH-std-std-TRUE	89%	92%	100%	77%	81%	97%
eGARCH-gjrGARCH-norm-norm-TRUE	87%	90%	99%	84%	85%	93%
sGARCH-gjrGARCH-snorm-norm-TRUE	87%	90%	99%	72%	91%	79%
tGARCH-gjrGARCH-norm-std-TRUE	86%	91%	100%	67%	89%	48%
sGARCH-gjrGARCH-norm-norm-TRUE	87%	90%	99%	61%	88%	43%
gjrGARCH-sGARCH-norm-std-FALSE	85%	91%	100%	59%	97%	39%

Table 7 shows the results for Ethereum, which are essentially the same as for QL.

Table 7 P-values of Ethereum backtesting procedures and MCS procedure for joint loss function

	UC	CC	DQ	ER	ESR	MCS
gjrGARCH-tGARCH-std-sstd-TRUE	96%	97%	99%	100%	65%	100%

Table 8 reports the results for Ripple. Only five models did not fail the backtesting procedures and were selected by the MCS procedure. In the first regime standard GARCH or TGARCH models were chosen by MCS procedure; in the second a TGARCH specification is selected for the SSM in 60% of the cases, and GARCH or GJRGARCH ones in 20% of them. Normal and skewed normal distributions prevail in the first regime, whilst Student's t and skewed Student's t distributions are chosen in all cases in the second. Markov-switching and mixture models represent respectively 60% and 40% of those included in the SSM. Specifications with leverage effects at least in one of the regimes represent 80% of the total in the SSM.

Table 8 P-values of Ripple backtesting procedures and MCS procedure for joint loss function

	UC	CC	DQ	ER	ESR	MCS
sGARCH-sGARCH-snorm-sstd-TRUE	30%	57%	100%	54%	66%	100%
sGARCH-tGARCH-sged-sstd-FALSE	11%	28%	96%	100%	84%	100%

sGARCH-tGARCH-snorm-std-TRUE	12%	29%	95%	6%	58%	91%
tGARCH-tGARCH-norm-sstd-FALSE	44%	68%	94%	100%	10%	79%
tGARCH-gjrGARCH-sstd-sstd-FALSE	11%	28%	87%	89%	61%	45%

Table 9 show the results of the MCS procedure for Litecoin with a joint loss function. In total 47 models were chosen, the highest percentage of models satisfying the VaR and ES backtesting procedures. Standard GARCH is chosen in one third of the cases in the first regime, and in approximately half of them in the second. The Student's t distribution and its skewed version prevail in both regime, whilst the percentages for the normal distribution and its skewed version are 40% and 4% respectively in the two regimes. The corresponding percentages for the GED distribution are 44% and 18% respectively. Mixture models are selected in 80% of the cases for the SSM. Specifications with leverage effects are also chosen in 80% of the cases at least in one regime.

Table 9 P-values of Litecoin backtesting procedures and MCS procedure for joint loss function

	UC	CC	DQ	ER	ESR	MCS
sGARCH-sGARCH-norm-std-FALSE	83%	91%	93%	70%	82%	100%
sGARCH-sGARCH-norm-sstd-FALSE	83%	91%	97%	60%	83%	100%
gjrGARCH-tGARCH-norm-sstd-TRUE	51%	78%	62%	100%	85%	100%
tGARCH-sGARCH-snorm-sstd-TRUE	52%	79%	100%	56%	51%	100%
tGARCH-sGARCH-sged-sstd-TRUE	51%	78%	100%	50%	71%	100%
gjrGARCH-gjrGARCH-norm-sged-TRUE	50%	78%	99%	53%	59%	100%
sGARCH-sGARCH-std-sstd-TRUE	83%	93%	100%	58%	90%	100%
sGARCH-sGARCH-snorm-std-TRUE	53%	75%	99%	63%	99%	100%
tGARCH-sGARCH-ged-sstd-TRUE	51%	78%	100%	66%	79%	99%
sGARCH-sGARCH-sstd-std-TRUE	83%	93%	100%	66%	80%	97%
gjrGARCH-sGARCH-std-sged-TRUE	50%	78%	99%	54%	66%	97%
tGARCH-gjrGARCH-std-sstd-TRUE	84%	94%	100%	57%	51%	96%
sGARCH-sGARCH-norm-sged-TRUE	83%	91%	100%	54%	76%	95%
sGARCH-sGARCH-std-std-TRUE	51%	78%	100%	57%	74%	95%
gjrGARCH-tGARCH-sstd-std-TRUE	51%	78%	100%	63%	67%	94%
tGARCH-tGARCH-norm-ged-TRUE	50%	78%	100%	58%	65%	94%
gjrGARCH-sGARCH-sstd-std-TRUE	83%	93%	100%	68%	54%	92%
tGARCH-sGARCH-sstd-sstd-TRUE	82%	91%	100%	86%	51%	91%
eGARCH-tGARCH-snorm-ged-TRUE	50%	78%	100%	57%	60%	89%
tGARCH-sGARCH-std-ged-FALSE	51%	78%	100%	94%	82%	88%
eGARCH-sGARCH-snorm-sged-TRUE	50%	78%	100%	50%	53%	88%
sGARCH-eGARCH-std-sged-FALSE	83%	91%	99%	64%	81%	87%
tGARCH-tGARCH-snorm-ged-TRUE	53%	80%	100%	86%	79%	86%
tGARCH-gjrGARCH-ged-sstd-TRUE	86%	94%	100%	86%	65%	85%
gjrGARCH-eGARCH-norm-std-FALSE	50%	78%	100%	57%	53%	85%
tGARCH-sGARCH-sstd-std-TRUE	84%	94%	100%	88%	56%	85%
sGARCH-sGARCH-sged-sstd-FALSE	50%	78%	100%	55%	60%	82%
gjrGARCH-eGARCH-sstd-ged-TRUE	50%	78%	99%	53%	67%	79%
eGARCH-gjrGARCH-snorm-sged-TRUE	50%	78%	100%	52%	53%	79%
sGARCH-tGARCH-snorm-ged-TRUE	50%	78%	100%	52%	54%	71%
eGARCH-sGARCH-ged-sged-TRUE	50%	78%	100%	54%	67%	70%
sGARCH-tGARCH-norm-ged-TRUE	50%	78%	100%	52%	60%	69%
gjrGARCH-gjrGARCH-std-std-TRUE	50%	78%	100%	52%	54%	68%
sGARCH-eGARCH-std-snorm-TRUE	50%	78%	97%	51%	59%	64%
sGARCH-eGARCH-ged-sged-TRUE	50%	78%	100%	53%	63%	63%
sGARCH-tGARCH-sged-sstd-FALSE	50%	78%	100%	63%	68%	61%
eGARCH-sGARCH-snorm-sged-FALSE	50%	78%	100%	52%	58%	59%
tGARCH-eGARCH-sstd-snorm-TRUE	85%	95%	100%	69%	53%	59%
tGARCH-sGARCH-norm-ged-FALSE	51%	78%	99%	99%	79%	56%
gjrGARCH-eGARCH-std-std-TRUE	50%	78%	96%	61%	72%	55%
eGARCH-tGARCH-sstd-sstd-TRUE	53%	80%	100%	71%	69%	52%
sGARCH-sGARCH-snorm-sged-TRUE	83%	93%	100%	51%	85%	50%
tGARCH-sGARCH-sged-std-TRUE	51%	78%	99%	56%	57%	43%
sGARCH-eGARCH-std-sged-TRUE	50%	78%	98%	55%	68%	42%
sGARCH-gjrGARCH-sstd-ged-TRUE	50%	78%	100%	58%	71%	40%
gjrGARCH-tGARCH-snorm-ged-FALSE	51%	78%	99%	96%	64%	38%
eGARCH-sGARCH-std-std-TRUE	50%	78%	100%	62%	59%	35%

6. Conclusions

This paper has used VaR and ES backtesting as well as the MCS procedure to select the best model or superior set of GARCH volatility models for four of the main cryptocurrencies, namely Bitcoin, Ethereum, Ripple and Litecoin. Two-regime GARCH models are found to produce better VaR and ES predictions than single-regime models. In particular, models that

allow for asymmetry prevail in the superior set of models on the basis of both VaR and joint loss functions. Mixture GARCH models prevail for almost all cryptocurrencies with both a quantile loss function and a joint loss function.

On the whole, our findings are consistent with those reported in the existing literature showing that cryptocurrencies exhibit extreme volatility and leverage effects. They also indicate that using standard GARCH models may yield incorrect VaR and ES predictions, and hence result in ineffective risk-management, portfolio optimisation, pricing of derivative securities etc. These could be improved by using instead the model specifications suggested by our analysis, from which both investors and regulators (such as the US Securities and Exchange Commission (SEC), which is planning to regulate the cryptocurrency exchanges) can benefit.

Future work will use intraday data to address the issue of the large number of the observations required by some of the tests carried out in the present paper, and also estimate multivariate GARCH models, e.g. to examine the linkages between Altcoin and Bitcoin.

References

Acerbi., C. and Szekely, B. (2014). Backtesting expected shortfall. Risk Magazine, 27, 76–81

Alexander, C. and Lazar, E. (2006). Normal mixture GARCH (1, 1): Applications to exchange rate modelling. Journal of Applied Econometrics, 21(3), 307-336.

Alizadeh, A.H., Nomikos, N.K., Pouliasis, P.K. (2008). A Markov regime switching approach for hedging energy commodities. Journal of Banking and Finance 32, 1970-1983.

Ardia, D., Bluteau, K., Boudt, K., Catania, L. (2017). Forecasting risk with Markov-switching GARCH models: A large-scale performance study. Available at SSRN: https://ssrn.com/abstract=2918413

Artzner, P. F. (1997). Thinking coherently. Risk, 10(11), 68-71.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. Mathematical finance, 9(3), 203-228.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. Mathematical finance, 9(3), 203-228.

Barendse, S. (2017). Interquantile Expectation Regression. Tinbergen Institute Discussion Paper 2017-034/III. Available at https://ssrn.com/abstract=2937665

Basel Committee (2013). Fundamental review of the trading book: A revised market risk framework. Technical report, Bank for International Settlements. Available at http://www.bis.org/publ/bcbs265.pdf.

Basel Committee (2016). Minimum capital requirements for Market Risk. Technical report, Bank for International Settlements. Available at http://www.bis.org/bcbs/publ/d352.pdf.

Bauwens, L., Backer, B.D., Dufays, A. (2014). A Bayesian method of change-point estimation with recurrent regimes: Application to GARCH models. Journal of Empirical Finance 29, 207-229. doi:10.1016/j.jempfin.2014.06.008.

Bauwens, L., Preminger, A., Rombouts, J.V.K. (2010). Theory and inference for a Markov switching GARCH model. Econometrics Journal 13, 218-244. doi:10.1111/j.1368-423X.2009.00307.x

Bayer, S. and Dimitriadis, T. (2018). Regression Based Expected Shortfall Backtesting. arXiv preprint arXiv:1801.04112.

Bohl, M.T., Salm, C.A., Wilfling, B. (2011). Do individual index futures investors destabilize the underlying spot market? Journal of Futures Markets 31, 81-101.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327. doi:10.1016/0304-4076(86)90063-1

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542-547. doi:10.2307/1925546

Bollerslev, T., Chou, R.Y., Kroner, K. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5-59. doi:10.1016/0304-4076(92)90064-X.

Bollerslev, T., Engle, R.F., Nelson, D.B. (1994). Handbook of Econometrics. North-Holland. chapter 49. pp. 2959-3038

Bouoiyour, J., Selmi, R. (2016). Bitcoin: A beginning of a new phase? Econ. Bull.36 (3), 1430–1440. http://www.accessecon.com/Pubs/EB/2016/Volume36/EB-16-V36-I3-P142.pdf

Christoffersen, P.F. (1998). Evaluating interval forecasts. International Economic Review 39, 841-862.

Chu, J., Chan, S., Nadarajah, S., and Osterrieder, J. (2017). GARCH Modelling of Cryptocurrencies. Journal of Risk and Financial Management, 10(4), 17.

Cifter, A., Ozun, A. (2007). The Predictive Performance of Asymmetric Normal Mixture GARCH in Risk Management: Evidence from Turkey. MPRA Paper 2489.

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O., and Shin, H. S. (2001). An Academic Response to Basel II. Financial Markets Group Special Papers, available at https://EconPapers.repec.org/RePEc:fmg:fmgsps:sp130.

Dimitriadis, T. and Bayer, S. (2017). A Joint Quantile and Expected Shortfall Regression Framework. arXiv:1704.02213 [math.ST].

Dyhrberg, A.H. (2016a). Bitcoin, gold and the dollar - A GARCH volatility analysis. Financ. Res. Lett. 16, 85–92. https://doi-org.lcproxy.shu.ac.uk/10.1016/j.frl.2015.10.008.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1008. doi:10.2307/1912773

Engle, R.F. (2004). Risk and volatility: Econometric models and financial practice. American Economic Review 94, 405-420. doi:10.1257/0002828041464597

Fantazzini, D., Nigmatullin, E., Sukhanovskaya, V., & Ivliev, S. (2016). Everything you always wanted to know about bitcoin modelling but were afraid to ask. Available at SSRN: https://ssrn.com/abstract=2794622

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. Annals of Statistics, 44(4):1680–1707.

Glaser, F., Haferhorn, M., Weber, M.C., Zimmarmann, K., Siering, M.b (2014). Bitcoin – Asset or currency? Revealing users' hidden intentions. ECIS 2014 Tel Aviv.

Glosten, L.R., Jagannathan, R., Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801. doi:10.1111/j.1540-6261.1993.tb05128.x

González-Rivera, G., Lee, T.H., Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, Value-at-Risk, and predictive likelihood. International Journal of Forecasting 20, 629-645.

Gronwald, M. (2014). The economics of Bitcoins - Market characteristics and price jumps (No. 5121). CESifo Working Paper, (5121). Available at SSRN: https://ssrn.com/abstract=2548999.

Haas, M., Mittnik, S., Paolella, M.S. (2004). A new approach to Markov-switching GARCH models. Journal of Financial Econometrics 2, 493-530. doi:10.1093/jjfinec/nbh020

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. Econometrica, 79(2), 453-497.

Henry, O.T. (2009). Regime switching in the relationship between equity returns and short-term interest rates in the UK. Journal of Banking and Finance 33, 405-414.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. Economics Letters 158: 3–6.

Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3.

Manganelli, S., Engle, R.F. (2004). CAViaR: Conditional autoregressive Value at Risk by regression quantiles. Journal of Business & Economic Statistics 22, 367-381.

McAleer, M., Da Veiga, B. (2008). Single-index and portfolio models for forecasting Value-at-Risk thresholds. Journal of Forecasting 27, 217-235.

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(3–4):271–300.

Nakamoto, S. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System. https://Bitcoin.org/Bitcoin.pdf.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347-370. doi:10.2307/2938260

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. The Annals of Applied Statistics, 11(4):1833–1874

Patton, A. J., Ziegel, J. F., and Chen, R. (2017). Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk). arXiv:1707.05108 [q-fin.EC].

Taylor, J. W. (2017). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. Journal of Business & Economic Statistics, 0(0):1–13.

Wilfling, B. (2009). Volatility regime-switching in European exchange rates prior to monetary unification. Journal of International Money and Finance 28, 240-270.

Zakoian, J.M. (1994). Threshold heteroskedastic models. Journal of Economic Dynamics & Control 18, 931-955. doi:10.1016/0165-1889(94)90039-6