CQE Center for Quantitative Economics

Forecasting Volatility in Cryptocurrency Markets

Mawuli Segnon [†] and Stelios Bekiros [#]

79/2019

[†] Department of Economics, University of Münster, Germany

[#] Department of Economics, European University Institute, Florence, Italy

wissen.leben
WWU Münster

Forecasting Volatility in Cryptocurrency Markets

Mawuli Segnon
Department of Economics, Institute for Econometric and Economic Statistics, University of Münster, Germany

and

Stelios Bekiros*
Department of Economics, European University Institute, Florence, Italy

Abstract: In this paper, we revisit the stylized facts of cryptocurrency markets and propose various approaches for modeling the dynamics governing the mean and variance processes. We first provide the statistical properties of our proposed models and study in detail their forecasting performance and adequacy by means of point and density forecasts. We adopt two loss functions and the model confidence set (MSC) test to evaluate the predictive ability of the models and the likelihood ratio test to assess their adequacy. Our results confirm that cryptocurrency markets are characterized by *regime shifting*, *long memory* and *multifractality*. We find that the Markov switching multifractal (MSM) and FIGARCH models outperform other GARCH-type models in forecasting bitcoin returns volatility. Furthermore, combined forecasts improve upon forecasts from individual models.

Keywords Bitcoin, Multifractal processes, GARCH processes, Model confidence set, Likelihood ratio test

JEL classification C22, C53, C58

*Corresponding author: Stelios Bekiros, Department of Economics, European University Institute, Florence, Italy. E-mail: Stelios.Bekiros@eui.eu.

1 Introduction

Since the seminal paper of Nakamoto (2008) Bitcoin has witnessed a rapid development and attracted attention from both the investment and research communities. Some of the research questions are the center of the debate are: What type of asset is bitcoin? What is its fundamental value? What are the statistical properties of this new market? And are the extent volatility models appropriate for modeling and forecasting volatility in cryptocurrency markets? However, the focus of this paper is more on questions related to the forecasting performance and the adequacy of the volatility models used in the literature.

For analysts and other market participants, the forecasting of volatility becomes important task because such forecasts represent an important ingredient in risk assessment and allocation, and derivatives pricing theory. A growing number of studies has investigated in considerable detail the statistical properties of bitcoin returns. Pichl and Kaizoji (2017) found that crypto-currency markets are even more volatile than foreign exchange markets. Volatility clusterings have been observed by Chu et al. (2017), Bouri et al. (2017), Katsiampa (2017), Bariviera (2017), Bau et al. (2018), Stavroyiannis (2018) and Catania and Grassi (2017), among others. Osterrieder and Lorenz (2017) and Begušić et al. (2018) have studied the unconditional distribution of bitcoin returns and found that it has more probability mass in the tails than that of foreign exchange and stock market returns. Regime-switching behaviors are detected by Bariviera et al. (2017), and Balcombe and Fraser (2017) and Thies and Molnár (2018) have identified structural breaks in the volatility process of bitcoin via a Bayesian framework. Recently, Lahmiri et al. (2018) and Lahmiri and Bekiros (2018) have pointed out that bitcoin markets are characterized by long memory and multifractality.

There is a large body of evidence that crypto-currency markets share the most important stylized facts of foreign exchange, stock and commodity markets. These empirical observations have motivated the use of the GARCH-type models (cf. Chu et al., 2017; Bouri et al., 2017; Katsiampa, 2017; Bariviera, 2017; Bau et al., 2018; Stavroyiannis,

2018; Catania and Grassi, 2017), the Markov switching GARCH (cf. Ardia et al., 2018), and the stochastic volatility model (cf. Phillip et al., 2018).

Our objective in this paper is to revisit some stylized facts of bitcoin markets and propose new econometrics models that may produce accurate volatility forecasts. In contrast to previous studies that for simplicity consider a constant conditional mean, in this paper we assume that the mean process follows an autoregressive fractionally integrated moving average (ARFIMA), and we model the variance process using GARCH-type processes, Markov switching GARCH (MS-GARCH) and Markov switching multifractal (MSM) processes. The choice of MSM processes as a candidate for modeling and forecasting bitcoin returns volatility is based on the findings of previous studies that have found the MSM processes more appropriate and robust for forecasting volatility in foreign exchange, stock and commodity markets (cf. Calvet and Fisher, 2004; Lux, 2008; Lux and Morales-Arias, 2010; Lux et al., 2016). We evaluate the performance of our models in the form of point and density forecasts. We consider two loss functions, namely square and absolute errors, and apply the model confident set (MCS) and the likelihood ratio tests of Berkowitz (2001) to evaluate the forecasting performance and adequacy of our proposed models. Our empirical results indicate that the FIGARCH and the MSM outperform other specifications. We also find that forecast combinations yield a gain in forecast accuracy. However, the results of the likelihood test suggest that none of the models are well specified.

The rest of the paper is organized as follows. Section 2 presents the data analysis and some stylized facts. In Section 3 we describe our econometric models and Section 4 provides some statistical properties of the models. The empirical study is presented in Section 5 and Section 6 concludes

2 Data analysis

Our data set consists of daily spot exchange rates (p_t) in units of US dollars and is from the Bloomberg terminal. The price observations range from 01/01/2013 to 28/11/2018

and fluctuate between 13.28 USD on January 2, 2013 and 18,571.57 USD on December 18, 2017. We compute the percentage continuously compounded returns as

$$r_t = 100 * [\ln(p_{t-1}) - \ln(p_t)], \tag{1}$$

where p_t denotes the price of bitcoin in USD at a time t.

Figure 1 illustrates the time evolution of bitcoin prices, and their log-returns and squared returns. The descriptive statistics of daily, weekly and monthly log-returns are reported in Table 1. The daily returns exhibit high variability, negative skewness and excess kurtosis. These deviations from the Normal distribution, as in Figure 4, are confirmed by the Jarque-Bera test that rejects the null hypothesis of normality. To gain insight into how the unconditional distribution of bitcoin returns under time aggregation changes, we also report the descriptive statistics of weekly and monthly bitcoin returns in Table 1. Figure 4 illustrates the unconditional distributions of bitcoin returns at different time frequencies. We observe that under time aggregation, the unconditional distributions of bitcoin returns do not converge to Normal, as shown in Figure 4. The unconditional distributions of weekly or monthly returns are still characterized by asymmetries and excess kurtosis. We applied the Augmented-Dicker-Fuller (ADF) unit-root test of Dickey and Fuller (1979), which suggests stationarity of the log-returns. We also applied the Phillip-Perron unit root test, which rejects the null of unit root. To confirm the verdict of the ADF test, we also have adopted the KPSS test for stationarity, and it cannot reject the null hypothesis.

Figure 3 shows the autocorrelation functions for log-returns, absolute and squared log-returns. Absolute and squared log-returns are highly correlated, and this observation conforms with the Ljung-Box statistic, $Q(8)$. To measure the degree of persistence we compute the Hurst exponents via Detrended Fluctuation Analysis (DFA), the results of which are reported in Table 1. The values for the bitcoin log-returns are slightly above 0.5, as in Table 1. For absolute and squared returns, the tendency is clear that the Hurst index values are significantly above 0.5, indicating the presence of high persistence in

the volatility process. To gain insight into how the unconditional distribution of returns behaves in the tails, we compute the so-called Hill estimator for the tail index. The results for daily, weekly and monthly returns are less than 3 and suggest that the tails of the unconditional distribution may behave differently from that of stock, foreign exchange and commodity markets (cf. Begušić et al., 2018).

We applied the modified iterated cumulative squares (ICSS) algorithm, as developed by Sansó et al. (2004), to detect the structural breaks that may occur in the unconditional variance. We refer the reader to Sansó et al. (2004) for more detail on the framework and technical issues related to the test. We detected three break points, and the dates at which the break points occurred are: 15/04/2014, 24/11/2017 and 22/02/2018. Furthermore, We applied a multifractal detrented fluctuation analysis (MFDFA) recently developed by Kantelhardt et al. (2002), which is based on the local root-mean square (RMS). The MFDFA permits the estimation of the multifractal spectrum of power law exponents from the bitcoin price changes and compares its characteristics to those of monofractal time series. We first compute the q-order Hurst exponent as slopes of regression lines for each q-order RMS, and the results are depicted in Figure 5. We observe that the slopes of the regression lines vary with q-order, in other words, are q-order dependent. The Hq's values decrease with increasing segment sample size, suggesting that the small segments are able to distinguish between the local periods with high and low volatility. We observe, for bitcoin price changes, a large arc that depicts the multifractal spectrum. However, we note that the reliability of tools used here to detect multifractality is subject to debate in the literature (cf. Barndorff-Nielsen and Prause, 2001; Lux, 2004). For example, Barndorff-Nielsen and Prause (2001) show that one can obtain the appearance of scaling due to the presence of fat tails in the absence of true scaling.

In general, we observe that bitcoins share the typical salient features of financial assets such as *volatility clustering*, *fat tails*, *asymmetry*, *long memory* and *multifractality*. However, we observe that the Ljung-Box statistic, $Q(8)$ indicates the existence of autocorrelation in the log-returns, and the value of Hurst exponent for log-returns is signif-

icantly greater than 0.5, indicating the presence of high persistence in bitcoin returns. These observations suggest that the bitcoin market is not as efficient as stock or foreign exchange markets, which display a complete lack of predictability (cf. Lahmiri et al., 2018).

3 Theoretical framework

We assume that returns $\{r_t\}$ in crypto-currencies markets follow an autoregressive fractionally integrated moving average process (cf. Granger and Joyeux, 1980; Hosking, 1981), given by the following equation

$$\Phi(\mathrm{L})\,(1-\mathrm{L})^{d}\,(r_t-\mu)=\Theta\,(\mathrm{L})\,\epsilon_t. \tag{2}$$

The lag polynomials in *eq.* 2 are defined as: $\Phi(\mathrm{L}) = 1-\phi_1 L-\cdots-\phi_p L^p$ and $\Theta(\mathrm{L}) = 1-\theta_1 L-\cdots-\theta_q L^q$ where the p and q are autoregressive and moving average orders, respectively. $d \in (-1/2, 1/2)$, L is the lag operator and $(1-L)^d$ is the fractional differencing operator that is given by

$$(1-\mathrm{L})^{d}=\sum_{k=0}^{\infty}\frac{\Gamma(k-d)\mathrm{L}^{k}}{\Gamma(-d)\Gamma(k+1)}, \tag{3}$$

with $\Gamma(\cdot)$ being the gamma function.

In general, the innovation process, ϵ_t, in Eq. (2) can be formalized as follows:

$$\epsilon_t = u_t\sigma_t, \tag{4}$$

where u_t is a sequence of independent identically distributed normal random variables with zero mean and unit variance.

In this study, we proposed various traditional and modern models for capturing the time-varying dynamics of σ_t:

1. *Markov-switching multifractal (MSM) model*:

In this framework, volatility is modeled as product of k random volatility components, $M_t^{(1)}, M_t^{(2)}, \ldots, M_t^{(k)}$, and a scaling factor $\bar{\sigma}$

$$\sigma_t^2 = \bar{\sigma}^2 \prod_{j=1}^{k} M_t^{(j)}. \tag{5}$$

The dynamics governing the random volatility components (also called multipliers) determines the unique framework that characterizes the multifractal models. At date t, each multiplier $M_t^{(j)}$ is drawn from the base distribution M (to be specified) with positive support. Depending on its rank within the hierarchy of multipliers, $M_t^{(j)}$ changes from one period to the next, with probability γ_j, and remains unchanged with probability $1 - \gamma_j$, providing a spectrum of low and high frequencies of multiplier renewal.

We adopt both a parametric and non-parametric specification for the transition probabilities. Our objective is to investigate whether the non-parametric one can provide a satisfactory fit to this new market, and thus reduce the number of estimated parameters in the model:

a) The parametric specification was proposed by Calvet and Fisher (2001) and ensures the convergence of the discrete-time MSM model to the Poisson multifractal process in the continuous-time limit. The k transition probabilities are given by

$$\gamma_j = 1 - (1 - \gamma_k)^{(b^{j-k})}, \tag{6}$$

where $\gamma_k \in (0, 1)$ and $b > 1$, $j = 1, \ldots, k$.

b) The non-parametric specification was proposed by Lux (2008) and is given by

$$\gamma_j = 2^{j-k}, \qquad j = 1, \ldots, k. \tag{7}$$

The transition matrix related to the jth multiplier is given by

$$P_j = \begin{pmatrix} 1 - 0.5\gamma_j & 0.5\gamma_j \\ 0.5\gamma_j & 1 - 0.5\gamma_j \end{pmatrix}.$$

To finalize the specification of the MSM model we draw each multiplier, $M_t^{(j)}$ (in the event of a change) from a binomial distribution with support $\{m_0, 2 - m_0\}$, $1 < m_0 < 2$, and (binomial) probability 0.5, implying the unconditional expectation $\mathbb{E}(M_t^j) = 1$. If we assume stochastic independence among the multipliers, the transition matrix of the vector $M_t \equiv (M_t^{(1)}, \ldots, M_t^{(k)})'$ becomes the $2^k \times 2^k$ matrix $P = P_1 \otimes P_2 \otimes \cdots \otimes P_k$, where $\otimes$ denotes the Kronecker product. Using the binomial base distribution[2] for the single multipliers implies the finite support $\Gamma \equiv \{m_0, 2 - m_0\}^k$ for M_t and allows implementing of the maximum likelihood approach.

Remark. *Note that a higher k increases the number of regimes (which is 2^k), and generates proximity to long memory over a longer number of lags, but comes at an additional computational cost in our maximum likelihood approach. In contrast to the traditional Markov switching models, in which the number of parameters to be estimated doubles with an additional regime, the number of parameters in MSM model remains constant with an increasing number of regimes. We note that the MSM processes exhibit only* apparent *long memory with an asymptotic hyperbolic decay of the autocorrelation of absolute powers over a finite horizon and does not obey the traditional definition of long memory. This means that the asymptotic power-law behavior of autocorrelation functions at the limit or with a divergence of the spectral density (cf. Beran, 1994).*

2. *Markov-switching GARCH model*

[2] Liu et al. (2007) find that assuming other base distributions, such as lognormal and gamma, makes little difference in empirical applications

Following Haas et al. (2004), we define the Markov switching GARCH (MS-GARCH) model as:

$$\epsilon_t = u_t \sigma_{\delta_t,t}, \tag{8}$$

where $\{\delta_t, t \in \mathbb{Z}\}$ is a Markov chain with finite-state space $S = \{1, 2, \ldots, q\}$ and an irreducible and primitive $q \times q$ transition matrix, P, whose element p_{ij}, are given by

$$P = [p_{ij}] = \left[P(\delta_t = j | \delta_{t-1} = i)\right], \; i, j = 1, \ldots, q. \tag{9}$$

Furthermore, it is assumed that u_t and δ_t are independent and that the $(q \times 1)$ vector $\sigma_t^{(2)} = (\sigma_{1t}^2, \sigma_{2t}^2, \ldots, \sigma_{qt}^2)'$ of regime variances follows the standard GARCH(1,1)

$$\sigma_t^{(2)} = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^{(2)}. \tag{10}$$

3. *Short-memory GARCH-type models*:

   We consider a general class of GARCH$(1, 1)$ models proposed by He and Terasvirta (1999) of the form

$$\sigma_t^\kappa = g(u_{t-1}) + c(u_{t-1}) \sigma_{t-1}^\kappa, \tag{11}$$

   with $Pr(\sigma_t^\kappa > 0) = 1$, $\kappa > 0$, and where $\{u_t\}$ is a sequence of i.i.d. standard normal random variables, and $g(x), c(x)$ are nonnegative functions. This class of GARCH-type models includes, among others, the specifications of Bollerslev (1986) (standard GARCH), Glosten et al. (1993) (GJR-GARCH), Nelson (1991) (EGARCH), and Ding et al. (1993) (APARCH).

4. *Long-memory GARCH-type models*:

   To reproduce the long-term dependence of bitcoin returns volatility as documented

in the high Hurst coefficients of absolute and squared returns, meaning that these volatility measures are characterized by a slowly decaying autocorrelation function rather than an exponentially decaying one (as imposed, for instance, by the baseline GARCH approach) various long-memory GARCH-type models have been developed. These models obey the following ARCH(∞) representation:

$$\sigma_t^{\kappa} = g(\epsilon_t) + c(\epsilon_t). \tag{12}$$

This class of models includes:

(a) HYGARCH(p,d,q) model for $\kappa = 2$, $g_t \equiv \omega/\beta(1)$, $c_t = \Psi(\mathrm{L})\,\epsilon_t^2$ with $\Psi(\mathrm{L}) = 1 - \frac{\Phi(\mathrm{L})}{\beta(\mathrm{L})}\left[1 - \tau\left(1 - (1-\mathrm{L})^d\right)\right]$. The lag polynomials $\Phi(\mathrm{L})$, $\beta(\mathrm{L})$ are given by $\Phi(\mathrm{L}) = 1 - \sum_{i=1}^{\max(p,q)} \phi_i \mathrm{L}^i$ and $\beta(\mathrm{L}) = 1 - \sum_{i=1}^{p} \beta_i \mathrm{L}^i$ with L being the lag operator.

(b) HYGARCH(p,d,q) model reduces to FIGARCH(p,d,q) model for $\tau = 1$.

Remark. *Long-range dependence shows up in the fact that in principle, all available past data should be used in constructing of forecasts of future volatility (while in GARCH, its short-range dependence makes it sufficient to use the filtered realization of the conditional variance, σ_t, at the forecast origin, time t). This feature makes the long memory processes more appropriate for bitcoin returns volatility forecasting.*

4 Statistical properties of the theoretical models

In this section we show that our proposed models for modeling and forecasting of bitcoin market volatility are stationary, ergodic and exhibit high-order moments.

Assumption 1. *The roots of the characteristic polynomials $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle, the parameter $d \in (-0.5, 0.5)$.*

Assumption 2. *The random volatility components* $M_t^1, M_t^2, \ldots, M_t^k$ *with* $E(M_t^j) = 1$, $j = 1, \ldots, k$, *are nonnegative and independent of each other at any time and* $\gamma_j \in (0, 1)$.

Assumption 3. *Let define a matrix* $G = [G_{ji}]$, *with* $G_{ji} = p_{ji}(\beta + \alpha e_i')$ *and* e_i *is the ith* $(q \times 1)$ *unit vector,* $i, j = 1, \ldots, q$ *and assume that the spectral radius of* G, $\rho(G) < 1$.

Proposition 1. *Under Assumption 1 and 2, the ARFIMA*(p, d, q)*-MSM*(k) *model given by 2, 5 and 6 has a unique, second-order stationary solution. It follows that* $\{r_t, \epsilon_t, \sigma_t\}$ *are strictly stationary, ergodic and invertible.*

Proof. Under Assumption 2, the conditions of *Theorem* 1 in (Chapter 1, section 12 Shiryaev, 1995) are satisfied. It follows that the chain underlying the dynamics of multipliers M_t^j is geometrically ergodic. The ergodic distribution is given by $\pi_l = 1/2^k$, $l = 1, \ldots, 2^k$. Under Assumptions 1 and 2, $\{r_t, \epsilon_t, \sigma_t\}$ are strictly stationary, ergodic and invertible. ■

Proposition 2. *Under Assumption 1, the ARFIMA(p,d,q)-GARCH-class model given by 2 and 11 has a unique,* $\alpha\kappa-$*order stationary solution. It follows that* $\{r_t, \epsilon_t, \sigma_t\}$ *are strictly stationary, ergodic and invertible.*

Proof. Under Assumption 1 and the conditions of *Theorem* 2.1 in Ling and McAleer (2002a) with a constant mean process replaced by a stationary univariate ARFIMA process, $\{r_t, \epsilon_t, \sigma_t\}$ are strictly stationary, ergodic and invertible. ■

Proposition 3. *Under Assumptions 1 and 3, the ARFIMA(p,d,q)-MS-GARCH model given by 2 and 8 to 10 has a unique strictly stationary and ergodic solution with the finite second-order moment. It follows that* $\{r_t, \epsilon_t, \sigma_t\}$ *are strictly stationary, ergodic and invertible with the finite second-order moment.*

Proof. Under Assumption 3 the conditions of *Corollary* 1 in Liu (2006) are met. Combined with a stationary univariate ARFIMA process under Assumption 1, it follows that $\{r_t, \epsilon_t, \sigma_t\}$ are strictly stationary, ergodic and invertible with the finite second-order moment. ■

Proposition 4. *Under Proposition 1, it follows that the 2mth moments of* $\{r_t, \epsilon_t, \sigma_t\}$ *are finite, where m is a strictly integer.*

Proof. Under Proposition 1 and the conditions of Theorem 1 in (Chapter 1, section 12 Shiryaev, 1995), the $2mth$ moments of $\{r_t, \epsilon_t, \sigma_t\}$ are finite. With the specification à la Lux (2008) that sets the parameter b in the formula of the transition probabilities derived by Calvet and Fisher (2001) to 2, it is obvious that all the elements of the transition matrix of the chain underlying the multipliers are strictly positive, $\gamma_j \in (0, 1)$, without any restriction on the parameter b. ■

Proposition 5. *Under Proposition 2, it follows that the mκth moments of the* $\{r_t, \epsilon_t, \sigma_t\}$ *exist.*

Proof. Under Proposition 2 and the conditions of Theorem 2.2 in Ling and McAleer (2002a), the $m\kappa th$ moments of the $\{r_t, \epsilon_t, \sigma_t\}$ exist. ■

Remark. *The second moment and autocovariances of the MSM(k) for a binomial distribution of the multipliers are available in Lux (2008). As pointed out in Ling and McAleer (2002a), Proposition 5 cannot easily be extended to higher orders of the class of GARCH processes defined by* eq. 11. *However, for GARCH(p,q) of Bollerslev (1986), Ling (1999) provides a sufficient condition for the existence of the 2mth moment. Ling and McAleer (2002b) extend the results of Ling (1999) to establish the necessary and sufficient higher-order moment conditions for GARCH(p,q) and asymmetric power GARCH(p,q) of Ding et al. (1993).*

Denoting $\rho(h) = cov(r_t, r_{t-h})/var(r_t)$, the autocorrelation function of the process defined by *eq.* 2 and $\rho_q(h) = cov(|\epsilon_t|^q, |\epsilon_{t-h}|^q)/var(|\epsilon_t|^q)$ the autocorrelation of ϵ_t for every moment q and every integer h. Consider two arbitrary numbers κ_1 and κ_2 in the open interval $(0, 1)$. The following set of integers $S_k = \{h : \kappa_1 k \leq \log 2(h) \leq \kappa_2 k\}$ contains a large range of intermediate lags.

Proposition 6. *Under Assumption 1, it follows that* $\rho(n) \sim c|h|^{2d-1}$ *as* $h \to \infty$, *where c is a constant.*

Proof. Under Proposition 2 and Theorem 2.4 in Hosking (1981), $\rho(h)$ is proportional to $|h|^{2d-1}$. ■

Proposition 7. *Under Assumption 2, it follows that* $\ln \rho_q(h) \sim -\psi(q) \ln h$ *as* $k \to \infty$, *where* $\psi(q) = \log_2 \left(\frac{E(M^q)}{[E(M^{q/2})]^2} \right)$.

Proof. Under Proposition 2 and the proof of the Proposition 1 in Calvet and Fisher (2004), $\rho(h)$ is proportional to $\psi(q) \ln h$. ■

Remark. *Covariance stationarity in long-memory GARCH models requires that* $\Psi(1) < 1$. *We refer the reader to Conrad and Haag (2006) and Conrad (2010) for more details on covariance stationarity and non-negativity conditions for this class of models. We note that this condition is not fulfilled by the FIGARCH model. However, using the results in Bougerol and Picard (1992) one can show that the FIGARCH model is strictly stationary.*

5 Empirical study

5.1 Estimation of models

All the models are estimated via the maximum likelihood approach. Note that we estimate the mean and variance processes separately without an asymptotic loss in efficiency (cf. Ling and Li, 1997). The optimal lag in the ARFIMA(p, d, q) is obtained based on the AIC, and the parameters are well estimated, cf. Table 2. The diagnostic tests show that the ARFIMA can properly capture the dynamics underlying the mean processes of bitcoin returns. Furthermore, residuals from the model estimation are uncorrelated and exhibit ARCH effects that justify the use of GARCH-type models.

For the volatility models, we first determine the optimal number of volatility components in the MSM model using the log-likelihood values reported in Table 3, which indicate that $k = 8$ may be the best and stable choice. From now on, we estimate the MSM model with $k = 8$ and adopt the different specifications for the transition probabilities.

We observe only a slight difference between the estimates of the binomial parameter $\hat{m}_0$ and the scaling factor in both model specifications. We have fixed lags in GARCH-type models to $p = q = 1$. The motivation for this choice is that GARCH(1,1) processes are very simple, but effective in modeling the clustering effects (cf. Bollerslev et al., 1994). The estimates of the parameters in the MSM, GARCH-type (short- and long-memory) and MS-GARCH model with two regimes are well estimated and reported in Table 4. To investigate which long-memory GARCH specification is appropriate for the bitcoin returns, we test the null hypothesis that the estimate of τ is equal to one and cannot be rejected at any confidence level. This result indicates that the FIGARCH is the most appropriate long-memory model for bitcoin returns.

5.2 Point forecasts

We first focus on the point forecasts. To produce volatility forecasts from our proposed models in Section 2, we split our data set into appropriate in-sample and out-of-sample periods and adopt a rolling forecasting scheme that ensures a fixed number of observations being used for the estimation over the out-of-sample period. The splitting point is based on the identified break points[3] using the modified ICSS algorithm as developed by Sansó et al. (2004). The in-sample covers the period from 01/01/2013 until 24/11/2017, and the out-of-sample from 25/11/2017 until 28/11/2018. For each model, we compute the volatility forecasts for four different horizons, $h = 1, 5, 10, 20$ trading days.

5.3 Forecasting evaluation criteria

To evaluate the forecast performance of our proposed models, we adopt two well-known accuracy measures, namely the root mean square error and mean absolute deviation that are given by

$$\text{RMSE}_j = \left(\frac{1}{n} \sum_{i=1}^{n} \left(\sigma^2_{T+i} - \hat{\sigma}^2_{T+i,j} \right)^2 \right)^{1/2}, \quad \text{MAE}_j = \frac{1}{n} \sum_{i=1}^{n} |\sigma^2_{T+i} - \hat{\sigma}^2_{T+i,j}|,$$

[3]We refer the reader to Pesaran and Timmermann (2007) for more details on selection of an optimal window in the presence of breaks.

respectively. j denotes a particular model in our portfolio, n is the number of out-sample forecast observations and T the forecast origin.

In addition, we utilize the model confidence set (MCS) test that was recently developed by Hansen et al. (2011) to assess the forecast performance of our proposed models. The basic idea of the MCS approach is to derive from an initial set of competing models, $\mathcal{M}_0$ without a predefined benchmark model, a set of superior models, $\mathcal{M}^*$ at forecasting horizon h, with a given confidence level. Formally, we have

$$\mathcal{M}^* = \{i \in \mathcal{M}_0 | \, \mathbb{E}\left(d_{i,j}^h\right) \leq 0 \; \forall j \in \mathcal{M}_0\},$$

where $d_{i,j}^h = g(e_{t,i,h}) - g(e_{t,j,h})$ is the loss differential between models i and j and $e_{t,i,h} = \hat{\sigma}_{t,i,h}^2 - \sigma_{t,h}^2$ represents the model-specific forecast errors at date $t+h$ and the loss function $g(\cdot)$ either denotes the squared error loss $g(e_{t,i,h}) = e_{t,i,h}^2$ or the absolute error loss $g(e_{t,i,h}) = |e_{t,i,h}|$. Based on the expected loss functions, competing models are ranked and the worst performing model at each step is eliminated. This sequential elimination continues until the null hypothesis of equal loss differentials for all models cannot be rejected:

$$H_0 : \mathbb{E}\left(d_{i,j}^h\right) \leq 0 \; \forall i, j \in \mathcal{M}.$$

The test statistic used under the null is either the range statistic, T_r, or the semi-quadratic statistic, T_{sq}, that are given by

$$T_r = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{i,j}|}{\sqrt{\hat{var}(\bar{d}_{i,j})}}, \quad T_{sq} = \sum_{i \neq j} \frac{(\bar{d}_{i,j})^2}{\sqrt{\hat{var}(\bar{d}_{i,j})}}.$$

We refer the reader to Hansen et al. (2011) for details on the framework of the MCS approach concerning the impact of the (i) forecasting schemes used, (ii) the relationships between models under comparison and (iii) its relationship with existing tests in the literature.

5.4 Forecasting results

For each model specification we compute the root mean square error (RMSE) and mean absolute error at different forecasting horizons $h = 1, 5, 10, 15, 20$. The results are reported in Table 5. Based on the RMSE criterion, the parametric MSM model seems to be the best, followed by the short-memory GARCH models, the FIGARCH and the MS-GARCH model. The non-parametric MSM model does not perform well. The results of the MCS test show that all models perform well at all forecasting horizons, although we note that the parametric MSM model seems to be the only one that cannot be outperformed at any confidence level. The non-parametric specification version of the MSM model does not perform well here. One reason might be that the non-parametric MSM model lacks of sufficient flexibility for bitcoin markets.

According to the MAE criterion, the FIGARCH clearly outperforms its competitors. The superior model set does only contain the FIGARCH model, whereas other models are consistently at all forecasting horizons excluded. However when we consider their ranked positions according to the MCS test, we note that the second best model is the parametric MSM model followed by the GJR model.

5.4.1 Forecast combination

The difficulty of finding a uniformly best model which outperforms other models based on the square loss function at both short and long horizons motivates us also to try weighted average forecast combinations. Granger and Teräsvirta (1999) and Aiolfi and Timmermann (2006) pointed out that it is often preferable to combine forecasts from competitive models in a linear manner, thereby hopefully generating superior predictions. Following this idea, we combine forecasts of GARCH-type and MSM models in order to explore the complementarities of two classes of volatility model. Our combination strategy consists of weighted linear combinations of both volatility models. The new predictor, f^* is given by

$$f^*_{t,h} = (1-\lambda)\, f^M_{t,h} + \lambda f^G_{t,h}, \tag{13}$$

where $f^M_{t,h}$ and $f^G_{t,h}$ denotes the forecasts of MSM and GARCH-type models at different forecast horizons, respectively.

The estimates of optimal weights related to each forecast are obtained via the forecast encompassing test developed by Harvey et al. (1998) for non-nested models, based on least squares regression given by

$$\varepsilon^M_{t,h} = \lambda \Delta\varepsilon^{M,G}_{t,h} + \epsilon_t, \tag{14}$$

where $\varepsilon^M_{t,h}$ and $\varepsilon^G_{t,h}$ denote the forecasting errors from the MSM and GARCH-type models at the forecast horizon h, respectively, and ϵ_t an *i.i.d.* normally distributed error term. $\Delta\varepsilon^{M,G}_{t,h} = \varepsilon^M_{t,h} - \varepsilon^G_{t,h}$.

The RMSE of combined forecasts are reported in Table 6. The RMSE values of the combined models are, in most cases, smaller than those of individual models, indicating that forecast combinations yield improvement in forecasting accuracy. To determine whether the observed difference between RMSE values of combined and individual models is statistically significant, we apply the MCS test. The test results show that all models are contained in the superior model set, suggesting that combined and individual models perform well, see Table 6.

5.5 Density forecast

To evaluate the forecasting performance of our proposed volatility models, we go beyond the point forecasts and adopt the density forecasts that shed substantial light on the accuracy of the shape of return residuals distribution. The methodology for assessing density forecasts is based on the integral transform that goes back to Rosenblatt (1952) and Diebold et al. (1998).

Let us denote by $\{p_t(x_t|\Omega_t)\}_{t=1}^{\infty}$ a sequence of densities identifying the data generat-

ing process governing the residual returns x_t and $\{f_t(x_t|\Omega_t)\}_{t=1}^{\infty}$, the sequence of one-step-ahead density forecasts produced by any volatility model. Following Diebold et al. (1998) we test the null hypothesis

$$H_0 : \{f_t(x_t|\Omega_t)\}_{t=1}^{\infty} = \{p_t(x_t|\Omega_t)\}_{t=1}^{\infty}.$$

Diebold et al. (1998) completed the work of Rosenblatt (1952) and showed that under the null hypothesis, the probability integral transform, $z_t = \int_{-\infty}^{x_t} f_t(y)dy$, is *i.i.d.* uniformly distributed. Based on these results and the theory of random numbers simulation, we implement the Berkowitz likelihood ratio test to assess whether the probability integral transform series, $\{z_t\}_{t=1}^{T}$, are *i.i.d.* $U(0,1)$.

The computation of one-step-ahead density and the integral transform using GARCH-type models, two state-Markov switching GARCH model and the Markov switching multifractal model is straightforward.

1. *GARCH-type models*:

   Under a normality assumption, the one-step conditional probability density function is normal and the integral transform can easily be obtained.

2. *MSM model*:

   It is not obvious how to derive the one-step conditional probability density function. The conditional density, given past information Ω_{t-1} has the following form

$$f(\epsilon_{t+1}|\Omega_t) = \sum_{j=1}^{n} f(\epsilon_{t+1}|M_{t+1} = m_j)P(M_{t+1} = m_j \mid \Omega_t), \tag{15}$$

   where $m_1, \ldots, m_n$ are the $n = 2^k$ variations of the volatility components, i.e. from $m_1 = (m_0, \ldots, m_0)$ to $m_n = (2-m_0, \ldots, 2-m_0)$. Due to the fact that the innovations in (4) are i.i.d. $N(0,1)$, the density of one-step-ahead return ϵ_{t+1} conditional on

volatility state M_{t+1} is Gaussian,

$$f(\epsilon_{t+1}|M_{t+1} = m_j) = \frac{1}{\bar{\sigma} g(m_j)} \phi\left(\frac{\epsilon_{t+1}}{\bar{\sigma} g(m_j)}\right), \tag{16}$$

where $\phi(.)$ is the standard normal density and $g(m_j) = \sqrt{\prod_{i=1}^{k} m_j^{(i)}}$ with $m_j^{(i)}$ being the i-th element of vector m_j.

The integral transform is given by

$$F(\epsilon_{t+1}|\Omega_t) = \int_{-\infty}^{\epsilon_{t+1}} f(x_t|\Omega_t) dx_t. \tag{17}$$

Inserting (15) into (17) we obtain

$$\begin{aligned} F(\epsilon_{t+1}|\Omega_t) &= \int_{-\infty}^{\epsilon_{t+1}} f(x_{t+1}|\Omega_t) dx_{t+1} \\ &= \int_{-\infty}^{\epsilon_{t+1}} \sum_{j=1}^{n} f(x_{t+1}|M_t = m_j) P(M_t = m_j|\Omega_t) dx_{t+1}. \end{aligned} \tag{18}$$

The density $f(\epsilon_t|M_{t-1} = m_j)$ is Lebesgue integrable. Due to linearity, $F(\epsilon_{t+1}|\Omega_t)$ becomes

$$\begin{aligned} F(\epsilon_{t+1}|\Omega_t) &= \sum_{j=1}^{n} P(M_t = m_j|\Omega_t) \int_{-\infty}^{\epsilon_{t+1}} \left[\bar{\sigma} g(m_j)\right]^{-1} \phi\left[x_{t+1}/\bar{\sigma} g(m_j)\right] dx_{t+1} \\ &= \sum_{j=1}^{n} P(M_t = m_j|\Omega_t) \Phi\left(\frac{\epsilon_{t+1}}{\bar{\sigma} g(m_j)}\right), \end{aligned} \tag{19}$$

where $\Phi(.)$ is the standard normal cumulative distribution function.

3. *MS-GARCH model*:

The one-step conditional probability density function is a mixture of 2 regime-dependent distributions:

$$f(\epsilon_{t+1}|\Omega_t) = \sum_{k=1}^{2} \pi_{k,t+1} f(\epsilon_{t+1}|\delta_{t+1} = k, \Omega_t),$$

where $\pi_{k,t+1} = \sum_{i=1}^{2} P_{i,k}\kappa_{i,t}$, with $\kappa_{i,t} = P[\delta_t = i|\omega_t]$, $i = 1, 2$ are the filtered probabilities at time t

5.6 Likelihood ratio test

The likelihood ratio test is a more powerful tool for evaluating density forecasts. Using a simple transformation to normality, Berkowitz (2001) obtained the following proposition:

- If the sequence $z_t = \int_{\infty}^{x_t} f(u)du$ is distributed as an *i.i.d.* $U(0, 1)$, then

$$v_t = \Phi^{-1}\left[\int_{\infty}^{x_t} f(u)du\right] \quad \text{is an } i.i.d. \quad N(0, 1). \tag{20}$$

With the new sequence v, one can test the joint null hypothesis (H_0) of independence and normality against a first-order autoregressive AR(1) with mean and variance different from 0 and 1, respectively.

The likelihood ratio test statistic is given by

$$LR = -2\Big[L(0, 1, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})\Big], \tag{21}$$

where $L(0, 1, 0)$ is the value of the log-likelihood function under H_0 and $L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})$ is the estimated log-likelihood function associated with the AR(1) process. Under the null hypothesis, the test statistic is chi-squared distributed with three degrees of freedom, $\chi^2(3)$.

One of the shortcomings of the likelihood ratio test is that it may happen that we cannot reject the null hypothesis, although the sequence v is not normally distributed. To avoid this, we also applied the Jarque-Bera test as a complement to the Berkowitz's test, (cf. Dowd, 2004).

We use the out-of-sample rolling scheme as described above. We estimate each model, and then forecast the densities and probability integral transforms (z). The forecasting probability integral transforms are one, again transformed via the simulation theory to obtain v series that are used for the likelihood ratio test. The results are presented in Table 7. The null hypothesis is rejected for all volatility models. It seems that no models used in this study are successful in capturing the dynamic structure of bitcoin markets.

6 Conclusion

In this paper we have revisited the stylized facts of cryptocurrency markets and proposed various approaches for modeling and forecasting bitcoin returns volatility. We have briefly discussed the statistical properties of the models used, and evaluated and compared the out-of-sample forecasting ability the models via two loss functions and the model confidence set (MCS) test. Furthermore, we have also evaluated the adequacy of the models, using the likelihood ratio test of Berkowitz (2001). We found that cryptocurrency markets share the most stylized facts of other financial markets such as *volatility clusterings* and *fat tails*. However, we note that the persistence in bitcoin marekts is more pronounced and that the tails have more probability mass than observed in the stock, foreign exchange and commodity markets. Our out-of-sample empirical results show that the parametric MSM and FIGARCH models outperform other GARCH-type models in forecasting bitcoin returns volatility at both short and long horizons. We also found that forecast combinations yield improvements in forecasting accuracy. Density forecasts results indicate that all the models used in this study fail to capturing properly the dynamics of bitcoin returns.

References

Aiolfi, M. and A. Timmermann (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics 135*, 31–53.

Ardia, D., K. Bluteau, and M. Rüede (2018). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters*.

Balcombe, K. and I. Fraser (2017). Do bubbles have an explosive signature in markov switching models? *Economic Modelling 66*, 81–100.

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters 161*, 1–4.

Bariviera, A. F., M. J. Basgall, W. Hasperué, and M. Naiouf (2017). Some stylized facts of the Bitcoin market. *Physica A: Statistical Mechanics and its Applications 484*, 82–90.

Barndorff-Nielsen, O. E. and K. Prause (2001). Apparent scaling. *Finance and Stochastics 5*, 103–113.

Bau, D. G., T. Dimpfl, and K. Kuck (2018). Bitcoin, gold and the dollar - a replication and extension. *Finance Research Letters 25*, 103–110.

Begušić, S., Z. Kostanjčar, H. E. Stanley, and B. Podobnik (2018). Scaling properties of extreme price fluctuations in Bitcoin markets. *Physica A: Statistical Mechanics and its Applications 510*, 400–406.

Beran, J. (1994). *Statistics for Long-memory Processes*. New York: Chapman and Hall.

Berkowitz, J. (2001). Testing density forecasts, with application to risk management. *Journal of Business and Economic Statistics 12*, 465–474.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T., R. F. Engle, and D. Nelson (1994). *Handbook of Econometrics*, Volume 4, Chapter ARCH models, pp. 2961–3038. Elsevier Science BV, Amsterdam.

Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics 52*, 115–127.

Bouri, E., G. Azzi, and A. H. Dyhrberg (2017). On the return-volatility relationship in the Bitcoin market around price crash of 2013. *Economics 11*, 1–17.

Calvet, L. and A. Fisher (2001). Forecasting multifractal volatility. *Journal of Econometrics 105*, 27–58.

Calvet, L. and A. Fisher (2004). Regime-switching and the estimation of multifractal processes. *Journal of Financial Econometrics 2*, 44–83.

Catania, L. and S. Grassi (2017). Modelling crypto-currencies financial time-series. Available at SSRN: https://ssrn.com/abstract=3028486 or http://dx.doi.org/10.2139/ssrn.3028486.

Chu, J., S. Chan, S. Nadarajah, and J. Osterrieder (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management 10*, 1–15.

Conrad, C. (2010). Non-negativity conditions for the hyperbolic GARCH model. *Journal of Econometrics 157*, 441–457.

Conrad, C. and B. R. Haag (2006). Inequality constraints in the fractionally integrated GARCH model. *Journal of Financial Econometrics 4*, 413–449.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–431.

Diebold, F., T. Gunther, and A. Tay (1998). Evaluating density forecasts with application to financial risk management. *International Economic Review 39*, 863–883.

Ding, Z., C. Granger, and R. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*, 83–106.

Dowd, K. (2004). A modified Berkowitz backtest. *Risk 17*, 86–87.

Glosten, L., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and volatility of the nominal excess return on stocks. *Journal of Finance 46*, 1779–1801.

Granger, C. W. and T. Teräsvirta (1999). A simple nonlinear time series model with missleading linear properties. *Economics Letters 62*, 161–165.

Granger, C. W. J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–29.

Haas, M., S. Mittnik, and M. S. Paolella (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics 2*, 493–530.

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

Harvey, D. I., S. J. Leybourne, and P. Newbold (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics 16*, 254–259.

He, C. and T. Terasvirta (1999). Properties of moments of a family of GARCH processes. *Journal of Econometrics 92*, 173–192.

Hosking, J. R. (1981). Fractional differencing. *Biometrika 68*, 165–176.

Kantelhardt, J. W., S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley (2002). Multifractal detrented fluctuation analysis of nonstationary time series. *Physica A 316*, 87–114.

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

Lahmiri, S. and S. Bekiros (2018). Chaos, randomness and multi-fractality in Bitcoin market. *Chaos, Solitions and Fractals 106*, 28–34.

Lahmiri, S., S. Bekiros, and A. Salvi (2018). Long-range memory, distribution variation and randomness of bitcoin voaltility. *Chaos, Solitions and Fractals 107*, 43–48.

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedasticity model. *Journal of Applied Probability 36*, 1–18.

Ling, S. and W. Li (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association 92*, 1184–1194.

Ling, S. and M. McAleer (2002a). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Econometric Theory 18*, 722–729.

Ling, S. and M. McAleer (2002b). Stationary and the existence of moments of a GARCH processes. *Journal of Econometrics 106*, 109–117.

Liu, J. C. (2006). Stationarity of a markov-switching GARCH model. *Journal of Financial Econometrics 4*, 573–593.

Liu, R., T. di Matteo, and T. Lux (2007). True and apparent scaling: The proximity of the Markov-switching multifractal model to long-range dependence. *Physica A 383*, 35–42.

Lux, T. (2004). Detecting multi-fractal properties in asset returns. *International Journal of Modern Physics 15*, 481–491.

Lux, T. (2008). The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility. *Journal of Business and Economic Statistics 26*, 194–210.

Lux, T. and L. Morales-Arias (2010). Forecasting volatility under fractality, regime-switching, long memory and Student-*t* innovations. *Computational Statistics and Data Analysis 54*, 2676–2692.

Lux, T., M. Segnon, and R. Gupta (2016). Forecasting crude oil price volatility and value-at-risk: Evidence from historical and recent data. *Energy Economics 56*, 117–133.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*, 347–370.

Osterrieder, J. and J. Lorenz (2017). A statistical risk assessment of Bitcoin and its extreme tail behaviour. *Annals of Financial Economics 12*.

Pesaran, M. H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics 137*, 134–161.

Phillip, A., J. S. K. Chan, and P. Shelton (2018). A new look at cryptocurrencies. *Economics Letters 163*, 6–9.

Pichl, L. and T. Kaizoji (2017). Volatility analysis of bitcoin time series. *Quantitative Finance and Economics 1*, 474–485.

Rosenblatt, R. F. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–472.

Sansó, A., V. Arragó, and J. L. Carrion (2004). Testing for change in the unconditional variance of financial time series. *Revista de Economiá Financiera 4*, 32–53.

Shiryaev, A. (1995). *Probability (Graduate Texts in Mathematics)* (2n Edition ed.). Springer Verlag.

Stavroyiannis, S. (2018). Value-at-risk and related measures for the Bitcoin. *Journal of Risk Finance 19*, 127–136.

Thies, S. and P. Molnár (2018). Bayesian change point analysis of Bitcoin returns. *Finance Research Letters 27*, 223–227.

Tables and Figures

Table 1: Descriptive statistics of bitcoin returns

	Bitcoin returns		
	Daily	Weekly	Monthly
No of Obs	1539	309	71
Mean	0.374	1.861	8.101
Standard deviation	5.950	13.093	32.874
Skewness	-0.546	0.068	1.923
Kurtosis	24.990	8.740	10.272
Hurst exponent	0.635 [0.394 0.589]	0.684 [0.312 0.663]	0.939 [0.123 0.940]
Tail index	2.864	2.791	1.796
ADF	-28.648 (-2.865)	-10.169 (-2.871)	-5.106 (-2.905)
ADF*	-28.718 (-3.414)	-10.277 (-3.426)	-5.156 (-3.477)
PP	-39.571 (-2.865)	-16.790 (-2.871)	-7.198 (-2.905)
PP*	-39.629 (-3.414)	-16.911 (-3.426)	-7.291 (-3.477)
KP	0.351 (0.463)	0.339 (0.463)	0.225 (0.463)
KP*	0.212 (0.146)	0.206 (0.146)	0.141 (0.146)
$Q(8)$	37.410 (15.507)	15.437 (15.507)	11.652 (15.507)
Jarque-Bera	3.109E+4 (5.953)	424.392 (5.779)	200.180 (5.240)

Note: ADF* and ADF denote the augmented Dickey-Fuller statistics in a regression with (i) intercept and time trend and (ii) intercept only. PP* and PP are the Phillips-Perron adjusted t-statistics of the lagged dependent variable in a regression with (i) intercept and time trend and (ii) intercept only. KP* and KP represent the KPSS test statistics using residuals from regressions with (i) intercept and time trend and (ii) intercept only. The critical values at the 1% level are reported in parentheses. Q(8) denotes the Ljung-Box test for serial correlation at lag 8.

Table 2: Estimates of ARFIMA(2,d,2) using daily bitcoin returns

	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{d}$
Estimates	0.378	-0.079	-0.821	-0.018	-0.757	0.058
	(0.211)	(0.072)	(0.070)	(0.085)	(0.080)	(0.023)
			Diagnostics			
Log-lik			-2729.45			
AIC			5472.9			
BIC			16231.9			
	Residuals		Absolute residuals		Squared residuals	
Hurst exponents	0.583 {0.394 0.589}		0.892		0.909	
Q(8)	7.670 [0.466]		723.701 [0.000]		280.359 [0.000]	
ARCH(1)	164.533 [0.000]					

Note: The entries in parentheses are the standard error of the estimation. Q(8) denotes the Ljung-Box test at lag 8 and ARCH(1) is the Engle's test of heteroscedasticity at lag 1. The numbers in the square brackets are the p-values of both tests. AIC and BIC represent the Akaike and Bayesian information criterion, respectively.

Table 3: Estimation results for the MSM model

k	1	2	3	4	5	6	7	8	9	10
$\hat{m}_0$	1.604	1.664	1.732	1.644	1.679	1.621	1.513	1.499	1.467	1.576
$\hat{\sigma}$	5.873	5.922	5.865	5.376	4.345	3.290	6.328	5.853	5.434	3.028
$\hat{b}$	-	2.004	2.036	3.526	4.018	9.553	1.374	2.120	2.254	3.457
$\hat{\gamma}_k$	0.071	0.999	0.499	0.999	0.941	0.569	0.250	0.927	0.999	0.941
$\log(L)$	-4579.234	-4550.357	-4344.951	-4355.123	-4319.739	-4326.140	-4314.768	-4299.350	-4299.241	-4300.857

Note: k is the number of volatility components (also called multipliers) used in the estimation procedures. $\log(L)$ denotes the logarithm of the likelihood function.

Table 4: Estimation results using bitcoin prices from January 1, 2013 to November 28, 2018

	GARCH	EGARCH	GJR-GARCH	APARCH	MS-GARCH		FIGARCH	HYGARCH	MSM	MSM*
Regimes					1	2				
ω	0.628	0.168	0.580	0.524	1.392	0.040	2.686	2.044		
	(**0.062**)	(**0.012**)	(**0.057**)	(**0.278**)	(**0.542**)	(**0.022**)	(**0.570**)	(**0.758**)		
α	0.176	0.337	0.195	0.163	0.335	0.044				
	(**0.011**)	(**0.016**)	(**0.013**)	(**0.022**)	(**0.067**)	(**0.013**)				
β	0.824	0.958	0.833	0.836	0.838	0.819	0.728	0.714		
	(**0.009**)	(**0.003**)	(**0.008**)	(**0.018**)	(**0.028**)	(**0.034**)	(**0.056**)	(**0.061**)		
γ		0.025	-0.055	0.089						
		(**0.009**)	(**0.014**)	(**0.036**)						
δ				1.893						
				(**0.422**)						
ϕ							0.092	0.109		
							(**0.056**)	(**0.062**)		
d							0.816	0.782		
							(**0.079**)	(**0.085**)		
τ								1.027		
								(**0.024**)		
p_{ii}					0.454	0.619				
					(**0.064**)	(**0.044**)				
m_0									1.509	1.502
									(**0.013**)	(**0.018**)
$\bar{\sigma}$									5.893	5.844
									(**0.504**)	(**0.571**)
b										2.253
										(**0.154**)
γ_k										0.971
										(**0.026**)

Note: The numbers in parentheses are standard errors of the estimations.

Table 5: Point forecast evaluation results

Models	Forecast horizons									
	h=1		h=5		h=10		h=15		h=20	
	RMSE	p_{MCS}	RMSE	p_{MCS}	RMSE	p_{MCS}	RMSE	p_{MCS}	RMSE	p_{MCS}
GARCH	55.541	0.957*	56.216	0.551*	52.269	0.827*	54.724	0.627*	56.403	0.542*
GJR	55.650	0.957*	55.556	0.551*	51.890	0.936*	53.947	0.862*	55.200	0.682*
EGARCH	55.902	0.860*	55.775	0.551*	52.148	0.906*	54.131	0.862*	56.087	0.542*
APARCH	55.699	0.957*	56.033	0.551*	52.930	0.256*	55.734	0.396*	57.050	0.521*
FIGARCH	56.043	0.957*	56.588	0.551*	54.975	0.256*	56.002	0.627*	57.028	0.555*
MSGARCH	56.255	0.574*	75.310	0.059*	78.910	0.052*	80.310	0.077*	81.310	0.108*
MSM^{NP}	57.900	0.060*	57.352	0.217*	53.407	0.256*	54.734	0.627*	56.095	0.555*
MSM^{P}	55.270	1.000*	54.423	1.000*	51.781	1.000*	53.203	1.000*	54.368	1.000*
	MAE	p_{MCS}	MAE	p_{MCS}	MAE	p_{MCS}	MAE	p_{MCS}	MAE	p_{MCS}
GARCH	32.447	0.003	33.703	0.002	32.899	0.001	35.509	0.002	38.088	0.000
GJR	32.040	0.003	32.735	0.002	31.881	0.001	34.196	0.002	36.305	0.000
EGARCH	33.483	0.001	34.519	0.002	34.412	0.001	37.628	0.001	40.699	0.000
APARCH	32.384	0.003	33.542	0.002	33.340	0.001	36.192	0.002	39.134	0.000
FIGARCH	27.286	1.000*	26.624	1.000*	25.544	1.000*	26.214	1.000*	26.488	1.000*
MSGARCH	33.503	0.001	53.353	0.000	53.160	0.001	63.160	0.000	64.160	0.000
MSM^{NP}	36.975	0.000	38.455	0.000	37.541	0.000	38.991	0.001	40.1561	0.000
MSM^{P}	33.896	0.001	33.005	0.002	31.315	0.001	32.141	0.002	32.454	0.000

Note: The entries are RMSE and MAE values and proportion of times each model is in the superior set using square and absolute loss function. The MCS test is performed based on the T_{sq} at a 95% confidence level. MSM^{NP} and MSM^{P} denote the non-parametric and parametric MSM model, respectively. The forecasts in $\mathcal{M}^{*}_{95\%}$ are identified by one asterisk.

Table 6: RMSEs and MCS results for the combined forecasts

Models	Forecast horizons									
	h=1		h=5		h=10		h=15		h=20	
	RMSE	p_{MCS}	RMSE	p_{MCS}	RMSE	p_{MCS}	RMSE	p_{MCS}	RMSE	p_{MCS}
Individual forecasts										
GARCH	55.541	0.502	56.216	0.502	52.269	0.444	54.724	0.543	56.403	0.314
GJR	55.650	0.502	55.556	0.737	51.890	0.444	53.947	0.609	55.200	0.425
EGARCH	55.902	0.417	55.775	0.404	52.148	0.407	54.131	0.503	56.087	0.268
APARCH	55.699	0.502	56.033	0.617	52.930	0.323	55.734	0.368	57.050	0.353
FIGARCH	56.043	0.502	56.588	0.557	54.975	0.323	56.002	0.418	57.028	0.389
MSGARCH	56.255	0.502	75.310	0.118	78.910	0.115	80.310	0.157	81.310	0.116
MSM^{NP}	57.900	0.395	57.352	0.211	53.407	0.260	54.734	0.435	56.095	0.237
MSM^{P}	55.270	0.502	54.423	0.948	51.781	0.444	53.203	0.618	54.368	0.431
Combined forecasts										
MSM^{P} + GARCH	55.134	0.502	54.419	0.948	51.146	0.766	52.702	0.618	53.800	0.431
MSM^{P} + GJR	55.164	0.502	54.380	0.951	51.109	1.000	52.580	1.000	53.560	1.000
MSM^{P} + EGARCH	55.270	0.502	54.408	0.951	51.450	0.444	52.914	0.618	54.159	0.425
MSM^{P} + APARCH	55.186	0.502	54.399	0.951	51.302	0.444	52.817	0.618	53.675	0.431
MSM^{P} + FIGARCH	54.400	1.000	54.256	1.000	51.777	0.444	53.201	0.618	54.368	0.431
MSM^{P} + MS-GARCH	55.268	0.502	54.372	0.951	51.636	0.444	53.068	0.609	54.096	0.425

Note: The entries are RMSE and MAE values and proportion of times each model is in the superior set using square and absolute loss function. The MCS test is performed based on the T_{sq} at a 95% confidence level. MSM^{P} denote the parametric MSM model. All the forecasts are in $\mathcal{M}^{*}_{95\%}$.

Table 7: Density forecast evaluation results

Models	LR statistics	P_{LR}	JB statistics	P_{JB}
GARCH	518.117	0.000	7.464	0.024
GJR	513.918	0.000	7.806	0.020
EGARCH	525.800	0.000	6.685	0.035
APARCH	517.648	0.000	7.538	0.023
FIGARCH	463.755	0.000	11.643	0.003
MS-GARCH	463.370	0.000	18.188	0.000
MSM	473.796	0.000	16.461	0.000

Note: We test the null hypothesis of i.i.d. normal with mean zero and variance unity against an AR(1) process. The Jarque-Bera test is used to complement the LR test of Berkowitz (2001) as suggested by Dowd (2004).

Table 8: Results of encompassing test for non-nested models at different forecasting horizons

	Forecast horizons				
	1	5	10	15	20
Model 1 vs. Model 2					
MSM vs. GARCH					
ENC-T	0.823	0.131	1.529	1.067	1.269
	(**0.206**)	(**0.448**)	(**0.064**)	(**0.144**)	(**0.103**)
$\hat{\lambda}$	0.366	0.043	0.429	0.331	0.316
	[0.335]	[0.237]	[0.174]	[0.154]	[0.140]
MSM vs. GJR					
ENC-T	0.844	0.499	1.819	1.215	1.421
	(**0.200**)	(**0.309**)	(**0.035**)	(**0.113**)	(**0.078**)
$\hat{\lambda}$	0.319	0.159	0.481	0.402	0.412
	[0.329]	[0.258]	[0.190]	[0.167]	[0.152]
MSM vs. EGARCH					
ENC-T	-0.033	-0.279	1.181	0.706	0.582
	(**0.513**)	(**0.610**)	(**0.119**)	(**0.240**)	(**0.281**)
$\hat{\lambda}$	-0.018	-0.115	0.407	0.327	0.246
	[0.432]	[0.318]	[0.231]	[0.201]	[0.180]
MSM vs. APARCH					
ENC-T	0.750	0.343	1.373	1.053	1.121
	(**0.227**)	(**0.366**)	(**0.086**)	(**0.147**)	(**0.132**)
$\hat{\lambda}$	0.288	0.106	0.350	0.265	0.309
	[0.335]	[0.233]	[0.165]	[0.140]	[0.123]
MSM vs. FIGARCH					
ENC-T	1.998	0.918	0.108	0.073	-0.022
	(**0.023**)	(**0.180**)	(**0.457**)	(**0.471**)	(**0.509**)
$\hat{\lambda}$	0.420	0.209	0.032	0.027	-0.010
	[0.151]	[0.172]	[0.175]	[0.190]	[0.205]
MSM vs. MS-GARCH					
ENC-T	0.093	-0.450	0.729	0.527	0.748
	(**0.463**)	(**0.674**)	(**0.233**)	(**0.299**)	(**0.228**)
$\hat{\lambda}$	0.041	-0.047	0.024	0.009	0.005
	[0.325]	[0.070]	[0.021]	[0.008]	[0.003]

Note: we test the null hypothesis that forecasts from model 1 encompass those of model 2 and ENC-T denotes the associated test statistics at different forecasting horizons. The values in parentheses are the p-values of the tests. $\hat{\lambda}$s are the estimates of the slope parameter λ in the forecast encompassing regression *eq.* 14. The values in square brackets are the standard errors of the estimation.

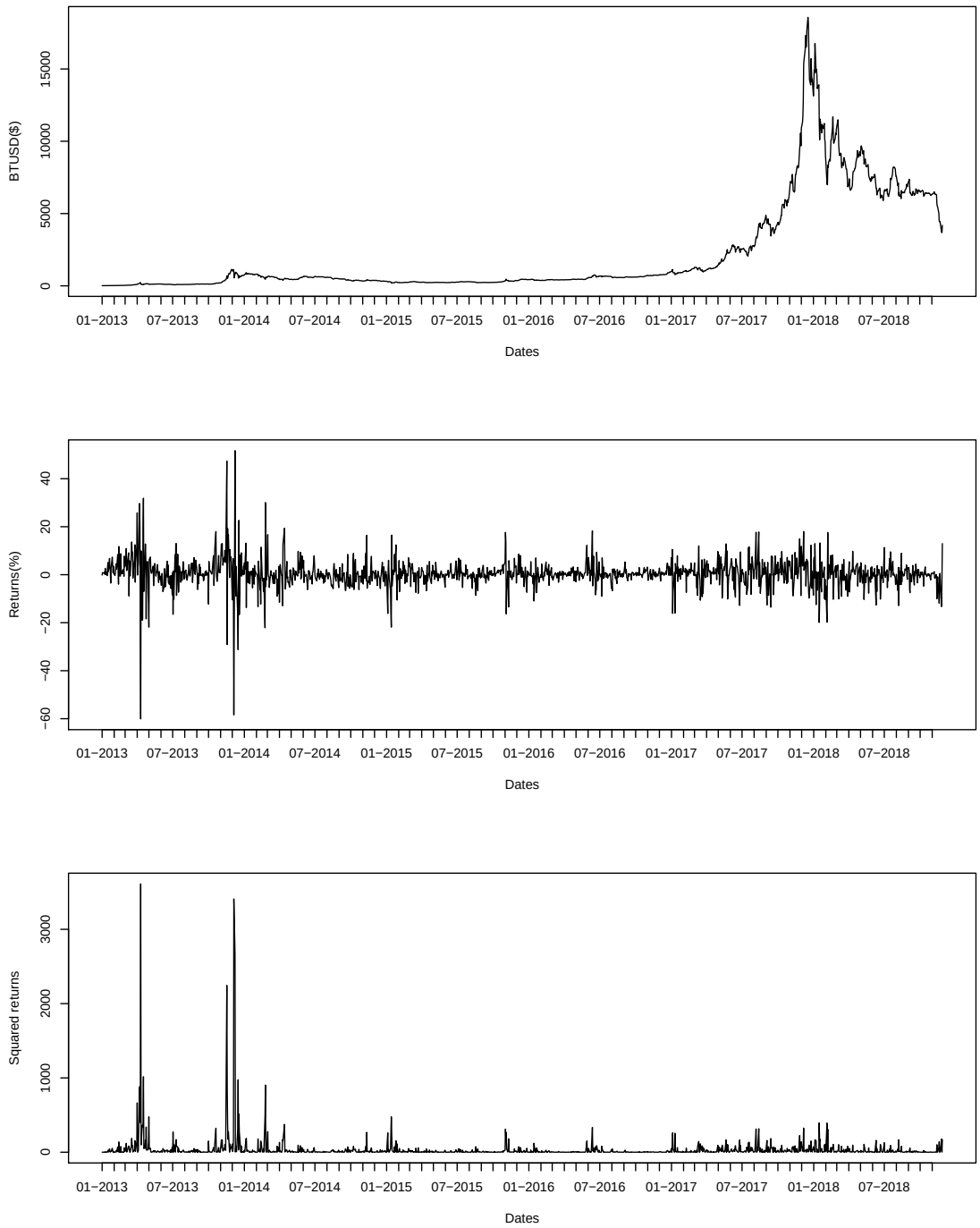

Figure 1: Plot of bitcoin USD exchange rates, the returns and squared returns

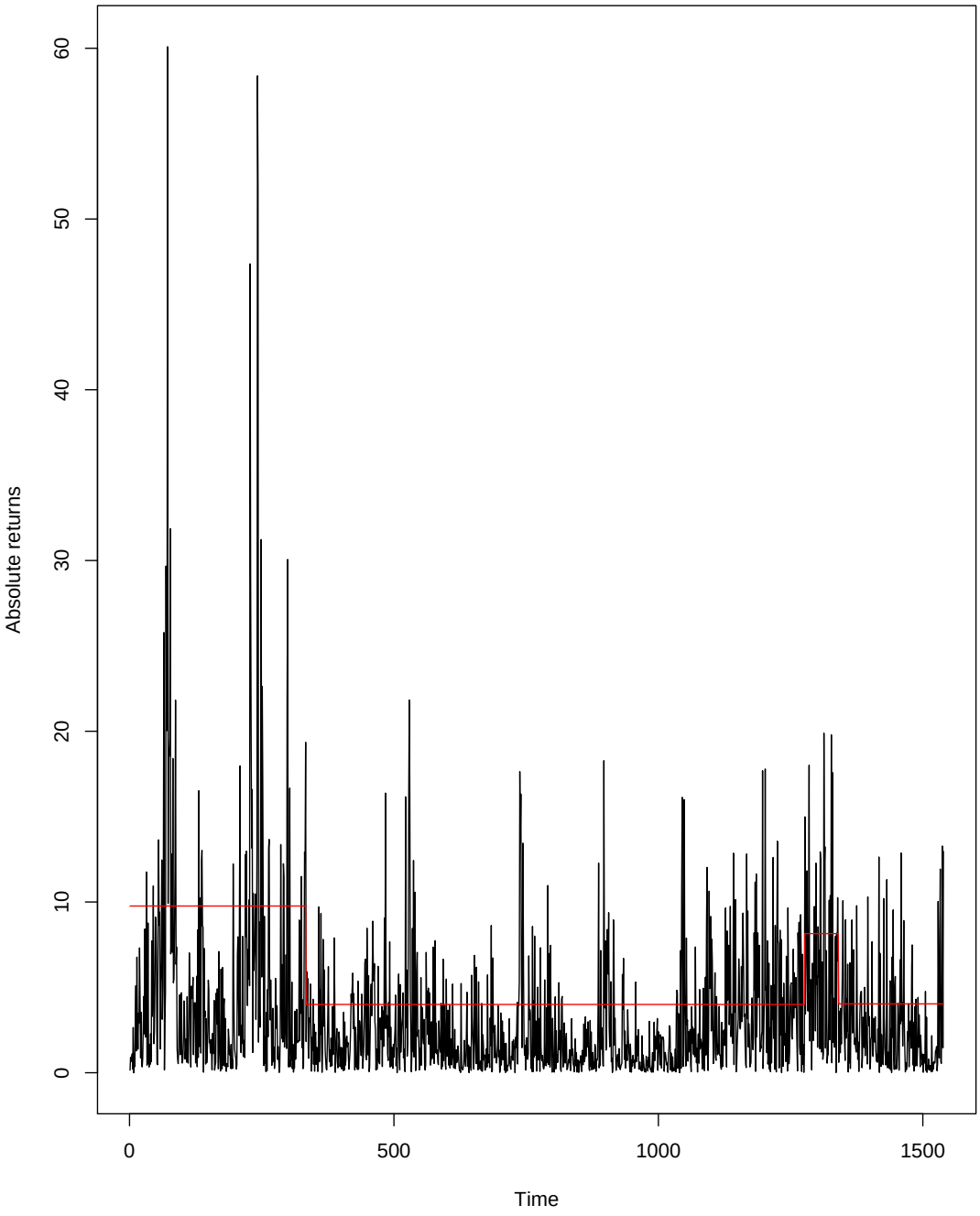

Figure 2: Plot of bitcoin absolute returns and standard deviation for the regimes defined by the structural breaks identified by the modified ICSS algorithm.

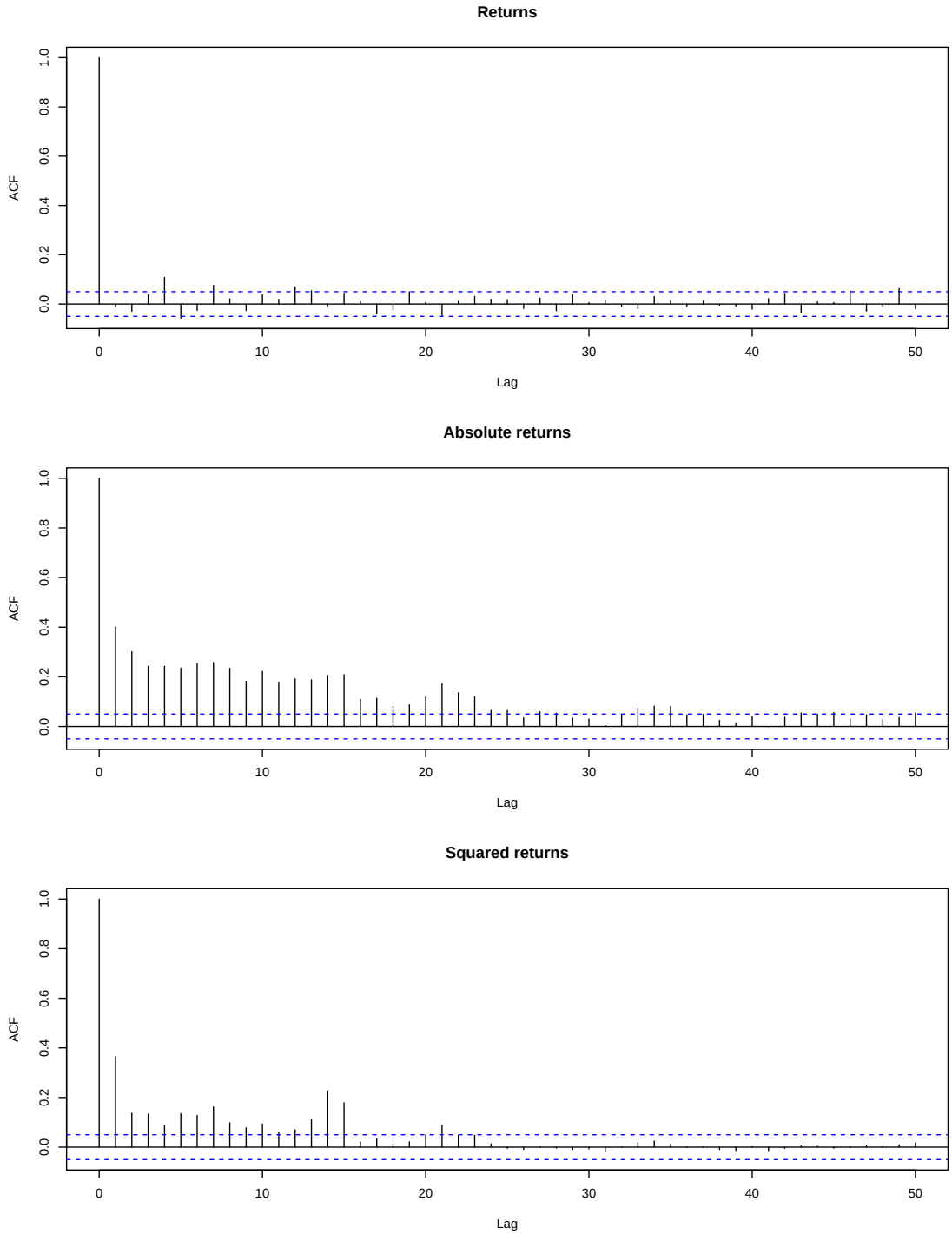

Figure 3: Plot of autocorrelation functions

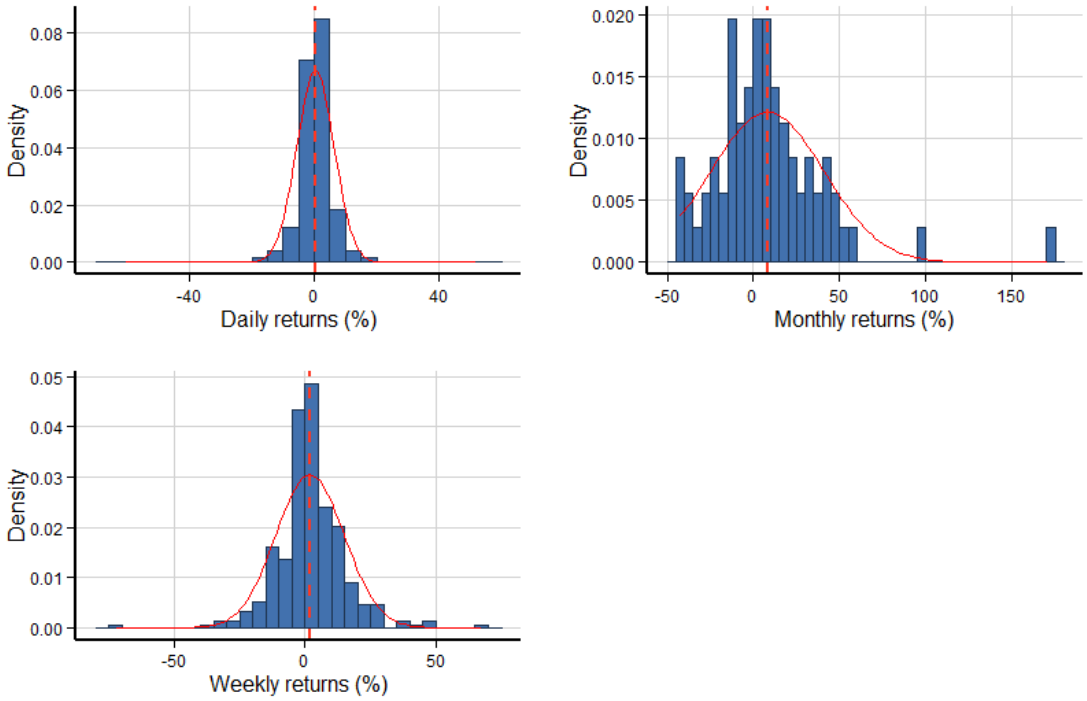

Figure 4: Plot of bitcoin returns distributions

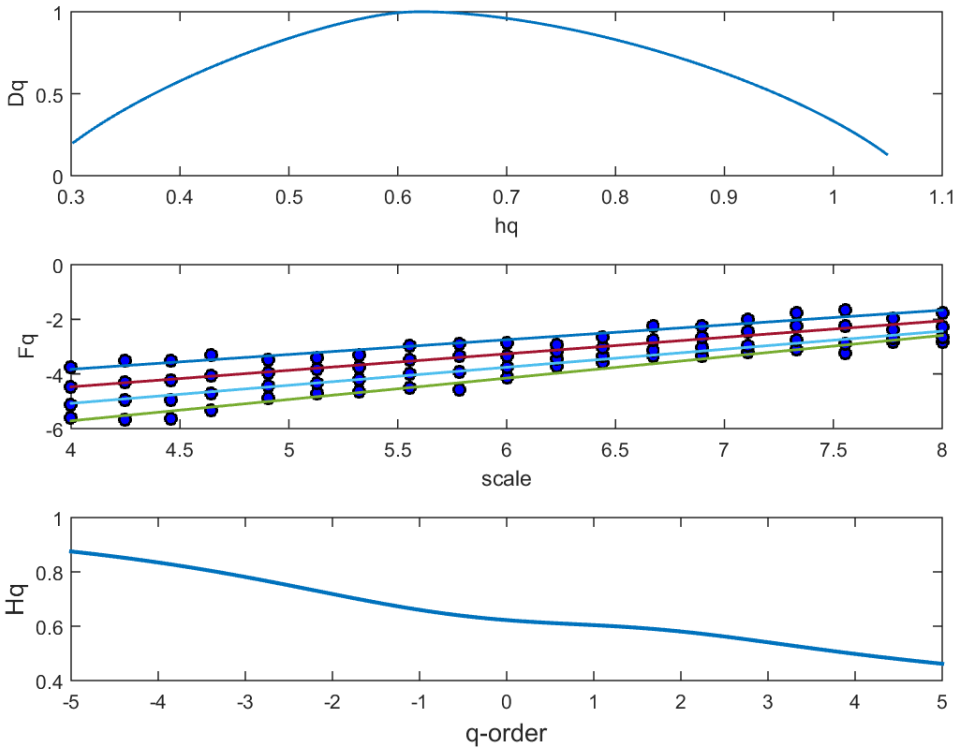

Figure 5: The panel at the top depicts the multifractal spectrum of bitcoin returns, the panel in the middle the scaling functions (Fq) and the panel at the bottom the q-order Hurst exponents (Hq) for bitcoin returns.