Research in International Business and Finance 59 (2022) 101540

Contents lists available at ScienceDirect

Research in International Business and Finance

journal homepage: www.elsevier.com/locate/ribaf

Forecasting volatility of Bitcoin

Lykke Øverland Bergsli [a], Andrea Falk Lind [a], Peter Molnár [a,b,c,d,*], Michał Polasik [c]

[a] Norwegian University of Science and Technology, Trondheim, Norway
[b] University of Stavanger, Stavanger, Norway
[c] Nicolaus Copernicus University in Torun, Torun, Poland
[d] Prague University of Economics and Business, Prague, Czech Republic

ARTICLE INFO

Keywords:
Bitcoin
Volatility forecasting
Realized variance
HAR
GARCH

ABSTRACT

Since Bitcoin price is highly volatile, forecasting its volatility is crucial for many applications, such as risk management or hedging. We study which model is the most suitable for forecasting Bitcoin volatility. We consider several GARCH and two heterogeneous autoregressive (HAR) models and compare them. Since we utilize realized variance estimated from high frequency data as a proxy for true volatility, we can draw sharper conclusions than studies which use only daily data. We find that EGARCH and APARCH perform best among the GARCH models. HAR models based on realized variance perform better than GARCH models based on daily data. Superiority of HAR models over GARCH models is strongest for short-term volatility forecasts.

1. Introduction

Cryptocurrency is a digital payment medium. Starting out as a niche products used in small online communities, cryptocurrencies have now become mainstream, attracting the attention of both financial experts and the general population. With billions of dollars in market capitalization, this market has also become a new arena for speculators.

This paper studies the most popular cryptocurrency, Bitcoin, which is currently traded on more than 500 exchanges. Since Bitcoin is the first cryptocurrency, established in 2008, it provides the longest historical data to study. However, results are likely to apply to other cryptocurrencies as well, due to high correlation between the cryptocurrencies.

Compared to traditional fiat currencies, the Bitcoin price is highly volatile (Blau, 2017; Chu et al., 2017). For risk management purposes, it is therefore important to model and forecast the volatility of Bitcoin. Bitcoin also differs from traditional currencies in that it is not issued by a government, and therefore not attached to a specific country's economy. Kristoufek (2013) argues that the expectations of macroeconomic development, which influence traditional currencies, do not influence Bitcoin prices, since Bitcoin has no underlying economy. This is supported by Bouoiyour and Selmi (2015b), who found that the price of Bitcoin exhibits unpleasant speculative behaviour. The occurrence of bubbles has been confirmed for Bitcoin and other cryptocurrencies (Corbet et al., 2018; Enoksen et al., 2020).

Bitcoin has been considered predominantly a speculative financial asset, rather than a medium of exchange for goods and services, like traditional currencies. This is supported by Dyhrberg (2016), who found that Bitcoin can be classified in between gold and the American dollar. The Bitcoin market is highly speculative, and Bitcoin can therefore be considered mainly an asset and not a currency (Baek and Elbeck, 2015). Baumöhl (2019) study the comovement between cryptocurrencies and traditional currencies. It is hard to place Bitcoin in a certain asset class. Burniske and White (2017) argue that Bitcoin is a distinct asset class, due to its unique characteristics, when compared to other asset classes. This is supported by Bouri et al. (2017), who found that the behaviour of Bitcoin

* Corresponding author.
E-mail address: peto.molnar@gmail.com (P. Molnár).

https://doi.org/10.1016/j.ribaf.2021.101540
Received 19 September 2020; Received in revised form 23 August 2021; Accepted 12 September 2021
Available online 16 September 2021
0275-5319/

makes it suitable for diversification. Whether we regard Bitcoin as a currency or an asset, it is crucial to understand its volatility.

Volatility tells us the extent to which the price of the asset changes over time. Understanding Bitcoin's volatility is a first and crucial step in understanding its risk characteristics, and it is also an important aspect of risk management, market making, portfolio optimization and selection, derivative pricing and hedging, and several other activities. There exists a vast amount of volatility models with different properties. This paper focuses on generalized autoregressive conditional heteroskedasticity (GARCH) models estimated for daily returns data and heterogeneous autoregressive (HAR) models estimated using realized variance calculated from high-frequency data.

Many of the studies about volatility forecasting of Bitcoin consider a few GARCH-type models with a limited number of distributions of the error term (see the literature review in the following section). Also, studies often evaluate in-sample goodness-of-fit, and not how the models perform out-of-sample. We therefore study volatility forecasting for Bitcoin. Our goal is to compare the most commonly used volatility models, GARCH models and HAR models.[1] To evaluate the models, we utilize realized variance as a proxy for the true variance. Among the GARCH family of GARCH models, EGARCH and APARCH perform best. Papers most closely related to our work are Trucíos (2019) and Köchling et al. (2020). Both these papers did a great job in evaluation of various GARCH models. However, we consider also HAR models, and find that the HAR models outperform even the best performing GARCH models.

The rest of this paper is organized as follows: Section 2 provides an overview of the volatility research for cryptocurrencies. Section 3 presents the data. Section 4 is an overview of the volatility models utilized in this study and methods for comparing them. Section 5 presents the results, and Section 6 concludes.

2. Literature review

The original focus in cryptocurrency research has primarily been on technical and legal aspects of cryptocurrencies. However, as the popularity of cryptocurrencies has increased, the financial debate has also intensified, particularly concerning the risk characteristics of cryptocurrencies. Chu et al. (2017) found that cryptocurrencies are highly volatile compared to traditional currencies. Naimy and Hayek (2018) study the volatility of the Bitcoin/USD exchange rate, focusing mainly on generalized autoregressive conditional heteroscedasticity (GARCH), the exponentially weighted moving average (EWMA), and the exponential generalized autoregressive conditional heteroscedasticity (EGARCH) models. Of these three models, they found that EGARCH(1,1) outperforms the others, both in- and out-of-sample. Chu et al. (2017) were the first to provide a paper of GARCH modelling on the seven most popular cryptocurrencies. They fitted 12 different types of GARCH models and used information criteria to evaluate the models. They conclude that IGARCH and GJR-GARCH models provide the best fit in-sample for most of the cryptocurrencies. Katsiampa (2017) explored several conditional heteroskedasticity models with regard to goodness-of-fit to Bitcoin prices, finding that AR-CGARCH gives the best fit. This can indicate that it is important to include both a short-run and a long-run component of the conditional variance. Bouoiyour and Selmi (2015a) found that the volatility of Bitcoin is more influenced by negative shocks than positive ones, giving evidence of leverage effects. Trucíos et al. (2020) used Robust GARCH and GAS models to evaluate Value-at-risk and Expected shortfall of a cryptocurrency portfolio.

Letra (2016) fitted a GARCH(1,1) model on daily data and web content from Google Trends, Wikipedia, and Twitter tweets and conclude that Bitcoin returns are highly driven by popularity, and web content has some degree of predictive power. Charles and Darné (2018) built on Katsiampa's study by detecting jumps in Bitcoin returns. They used a semi-parametric test for jumps and estimated GARCH-type models on filtered data. They conclude that returns of Bitcoin are characterized by the presence of jumps, and that an AR-GARCH estimated on filtered returns gives the best result. The selection of the best model was based on in-sample goodness-of-fit. Catania et al. (2018) argue for using sophisticated volatility models, including leverage effect and time-varying skewness to improve volatility forecasts. Stavroyiannis and Babalos (2017) used data from 2013 to 2016 to estimate a FIAPARCH(1,d,1) model under a skewed student's *t* distribution. They then considered how Bitcoin can be used as a natural hedge, where they conclude that investors cannot benefit from using Bitcoin as a hedge or diversifier with respect to the US market. Balcilar et al. (2017) studied how Bitcoin trading volume can be used to predict volatility using a quantiles-based approach. They conclude that trading volume can predict returns but not volatility, when the market is in median mode. Most studies have been performed on Bitcoin, but Burnie (2018) found that the correlation between cryptocurrencies is strong, especially when one is a fork of the other. A fork is produced when changes to the programming of a currency create incompatibilities between the original version and the new one (Kirk, 2014). We therefore argue that it is reasonable to assume that most findings are also applicable for other cryptocurrencies than Bitcoin.

Some researchers have utilized realized variance (RV) and models based on this in the study of Bitcoin. As realized variance is a precise measure, it has previously been applied to commodities like oil (Haugom et al., 2014), gold and silver (Lyócsa and Molnár, 2016), and electricity (Birkelund et al., 2015), as well as stock markets (Christoffersen et al., 2010), and major exchange rates (Andersen et al., 2001). Li (2017) did a study on the volatility forecasting performance of GARCH, HEAVY, HAR, and logarithmic HAR models on S&P 500 data, using the MSE and QLIKE loss functions. She found that there was no universal winner of the models. Under the MSE loss function, the HEAVY model produced the best forecasts up to three days in the future, and the logarithmic HAR model did best after that, up to a 60 day horizon. Under the QLIKE loss function, the HEAVY model did best the first two days, the HAR model from three to five days, and the GARCH model from six to 60 days. Liu and Maheu (2008) used logarithmic HAR models to identify and date structural breaks in reduced form time-series of RV. They found strong evidence of a structural break in S&P 500 in 1997, and that the evidence was robust when compared to different models, including a GARCH model for the conditional variance of logarithmic

[1] We do not consider stochastic volatility models, as these are more difficult to estimate (Ruiz, 1994; Lord et al., 2010), and therefore, less used by practitioners. Broto and Ruiz (2004) provide a great overview of stochastic volatility models. Recently, stochastic volatility models have been applied to cryptocurrencies (Shi et al., 2020).

Table 1
Summary statistics of Bitcoin log-returns from 01-Jan-14 to 19-Sep-18.

Statistic	Value
Mean	0.00125
Median	0.00157
Max	0.238
Min	− 0.281
Std. Dev.	0.040
Skewness	− 0.335
Kurtosis	8.80
Jarque–Bera	2444.3
p-value	0.00
ADF test statistic	− 11.612
p-value	$<$0.01

realized variance. Dimpfl and Jank (2016) studied the relationship between the volatility of the Dow Jones and retail investors' attention to the stock market. Using the fact that search queries Granger-cause volatility, they found that volatility forecasts could be improved by using a logarithmic HAR model in combination with Google search queries.

Examples of the application of realized variance and the HAR model do exist in Bitcoin research. Baur and Dimpfl (2018) did an in-depth analysis of Bitcoin realized volatility and showed that the volatility of Bitcoin is extreme compared to major fiat currencies. Furthermore, their results suggest that the majority of Bitcoin trading is noise trading. Aalborg et al. (2018) used the HAR-RV model in combination with Google trends, transaction volume, traded volume, unique Bitcoin addresses, returns, and the VIX index to explain and predict the changes in Bitcoin's daily and weekly realized volatility. They found that the past daily, weekly, and monthly realized variance are highly significant and that their model has high explanatory power for the realized variance of Bitcoin. Pichl and Kaizoji (2017) used a HAR-RV model with jumps, as well as the day-to-day distribution of log returns, to model the volatility of Bitcoin. They found that the HAR-RV-J model manages to capture the dynamics of the realized volatility of Bitcoin reasonably well. Kurka (2016) used HAR and GARCH models to examine the patterns and drivers of volatility in Bitcoin. Evidence of the leverage effect was found, as well as high persistence of volatility shocks. Yu (2019) also used the HAR model and found that it is improved when it incorporates leverage effect, and that jumps and economic policy uncertainty are not important for Bitcoin volatility forecasting.

In recent years, some research has been done on volatility modelling of Bitcoin; however, few studies have compared the volatility models both in- and out-of-sample. Investigation of the optimal in-sample window size is also lacking in the literature. While others have previously calculated the realized volatility of Bitcoin using a five or 10 min grid, we use the average of several grids as our proxy, mitigating the variance-bias trade-off. Moreover, most research only considers either GARCH or HAR models, and does not compare the two group of models. The paper most closely related to ours is Caporale and Zekokh (2019). This paper is a very detailed comparison of GARCH models, not only for Bitcoin, but also for Ethereum, Ripple and Litecoin, with primary focus on Value at Risk and Expected Shortfall. However, we study not only GARCH models, but also HAR models and compare them.

Since volatility of Bitcoin is a very important topic, several papers have investigated the impact of external shocks on Bitcoin volatility. These studies find that Bitcoin volatility is increased when geopolitical risks are high (Aysan et al., 2019), economic policy uncertainty is high (Wang et al., 2020), on days of cryptocurrency-related hacking attacks (Lyócsa and Molnár, 2016), and also Mondays and Thursdays (Ma and Tanizaki, 2019). Comovement among cryptocurrencies has been studied too (Beneki et al., 2019; Katsiampa, 2019).

3. Data

According to bitcoin.com there are about 2000 cryptocurrencies and over 500 exchanges currently recorded (Sedgwick, 2018). In this paper we solely study Bitcoin cryptocurrency. This is due to Bitcoin being the largest cryptocurrency in terms of both volume and market capitalization, as well as being the oldest one, meaning more historical data is readily available (CoinMarketCap, 2018). There is also a strong correlation between cryptocurrencies in general (Burnie, 2018), so results are likely to apply to other cryptocurrencies.

Bitcoin prices have been collected from the Bitstamp exchange, as this is one of the oldest exchanges, founded in 2011, and licensed in the EU (Shin, 2016). The choice of using data from one exchange, rather than several, is due the limited price differences across the larger Bitcoin exchanges (Brandvold et al., 2015; Kroeger and Sarkar, 2017). The decision was also driven by data quality, as it was possible to retrieve intraday BTC/USD data from 13-Sep-11 to 19-Sep-18 through the Bitstamp API. Since the exchange is open around the clock, the daily closing price is the USD price of Bitcoin at the last registered trade of each day. These prices were extracted and transformed into log-returns. Data from 1-Jan-14 to 19-Sep-18, 1720 observations in total, has been used throughout this paper, as we consider this time frame representative of the market in terms of both price and volumes. After 2014, we also observe that data is frequent enough to estimate our realized variance precisely for all time grids.

Table 1 presents the summary statistics of the daily log-returns of Bitcoin from the Bitstamp exchange. We can see the data is left skewed and has excess kurtosis, which indicates non-normality. This is confirmed by the Jarque–Bera test. The augmented Dickey–Fuller test statistic rejects the null hypothesis of a unit root, meaning that returns can be viewed as stationary. This can also be seen in Fig. 1, where the log returns have been plotted against time. From this plot we also see evidence of volatility clustering.

Fig. 2 shows the histogram of the log-returns. The distribution exhibits fat tails. This can also be seen from the kurtosis in the

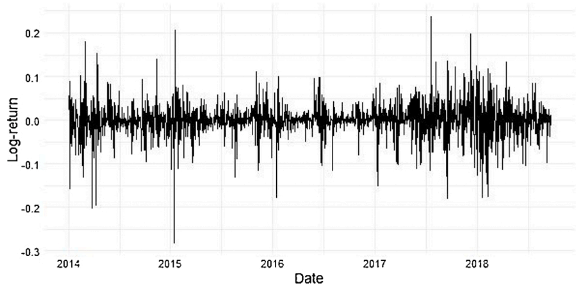


Fig. 1. Plot of daily log-returns from 01-Jan-14 to 19-Sep-18.

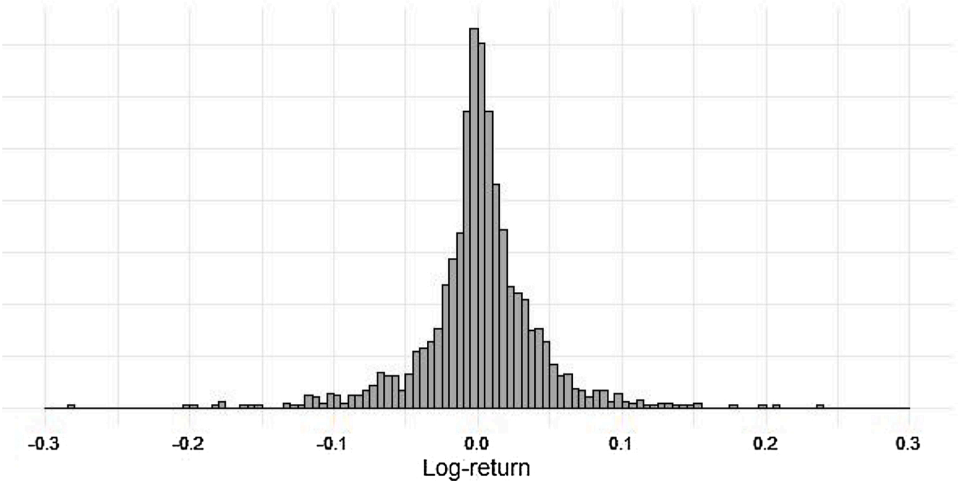


Fig. 2. Histogram of daily log-returns from 01-Jan-14 to 19-Sep-18.

summary statistics. This means that our data is prone to extreme outcomes, which is in accordance with the large price movements of Bitcoin. It is also evident from the summary statistics that the distribution is left skewed, with a mean below the median and a negative skewness. From this we can conclude that the data set is slightly more prone to extreme negative returns than positive ones.

As Bitcoin and other cryptocurrencies are traded 24 hours a day, every day, high frequency data can be used to estimate realized variance (RV) as a variance proxy. An estimate of the daily variance can be constructed using m evenly spaced intraday squared returns (Christoffersen, 2011), with returns defined as:

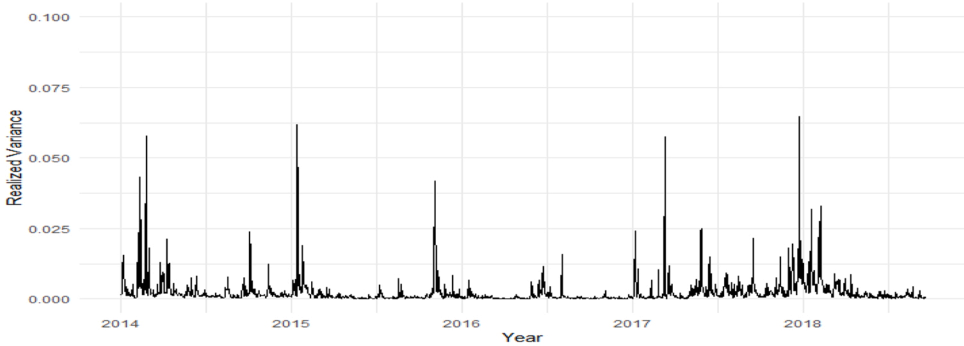
$$R_{t+j/m} = ln\left(S_{t+j/m}\right) - ln\left(S_{t+(j-1)/m}\right) \tag{1}$$

where j defines the jth x-minute return. With m observations within a day, the realized variance can be estimated with the following formula:

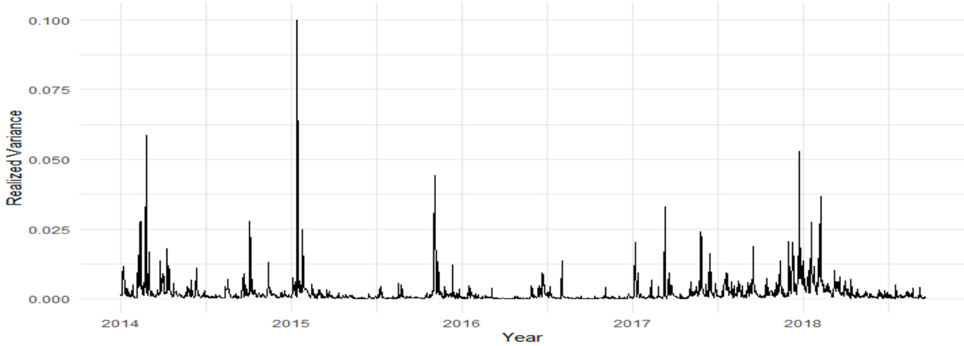
$$\mathrm{RV}_{t+1}^{m} = \sum_{j=1}^{m} R_{t+j/m}^{2} \tag{2}$$

This equation gives the variance for 24 h. We have chosen to calculate the realized variance using a grid of 10, 15, and 30 min, and then taking the average of the estimated realized variances for each of these periods. For this, we use Eq. (2) with $j = 10$, 15, and 30. An equally spaced grid is constructed by taking the last observation for each of these periods. The final realized variance used in our analysis is the average of these three proxies. The variance proxy should satisfy the four following stylized facts of realized variance (Christoffersen, 2011):

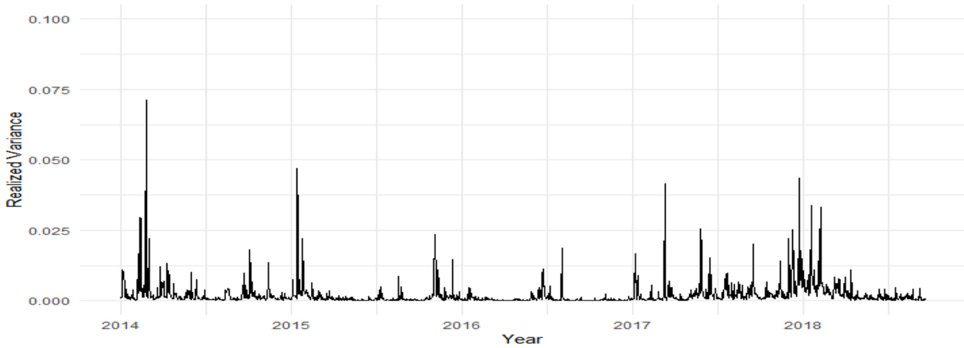
1. The estimated realized variance is a better indicator for true variance than using the squared daily returns.
2. The log of the realized variance will be approximately normally distributed.
3. For many lags, the realized variance will have large positive autocorrelation.
4. Taking the daily return and dividing it by the realized variance estimated, will produce independent and identically distributed standard normally distributed variables.



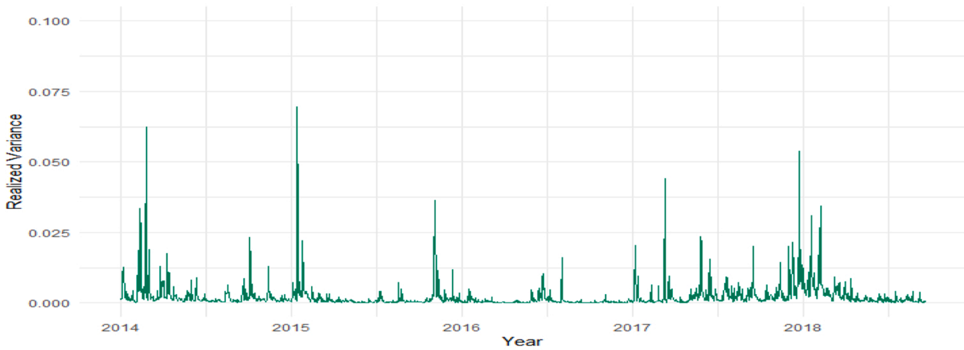
(a) 10 minute grid



(b) 15 minute grid



(c) 30 minute grid



(d) Average for 10, 15 and 30 minute grids

Fig. 3. Realized variance for 10, 15, and 30 min grids and the average realized variance for the grids, for the period 01-Jan-14 to 19-Sep-18.

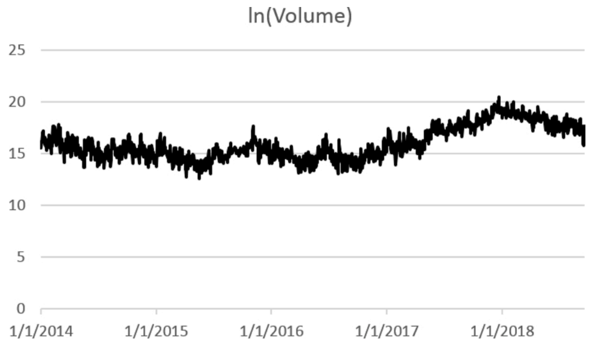


Fig. 4. Logarithm of Bitstamp trading volume measured in USD.

Fig. 3a–c shows the plots of realized variance calculated for each of the mentioned grids. Fig. 3d visualizes the average of these, our variance proxy going forward. From the figures we are able to see that the 10-minute interval grid contains more spikes in the variance compared to when we calculate the realized variance using larger intervals. When using a larger interval, we are less likely to get a biased estimate (Christoffersen, 2011). However, this involves using fewer observations, and estimates will be noisier. By taking an average of different grids, we mitigate part of this variance-bias trade-off.

There are two important issues related to estimation of realized variance that should be mentioned. First, there are more advanced realized volatility measures designed to deal with jumps, see for example Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2012). These estimators allow to decompose total variance into continuous and jump component, as was done previously also for Bitcoin, see Lyócsa et al. (2020). However, most of the models studied in our paper are low-frequency GARCH models based on daily data, and these models forecast the total variance. We therefore purposefully do not remove jump component from the realized variance to keep our realized variance measure comparable with variance estimated from GARCH models. Second, if trading volume is very low, then prices are not necessarily informative. Therefore, we do not include early days of Bitcoin, when trading volume was low, in our sample. Our sample starts in 2014, and already at the first day of our sample, trading volume was over 5 million USD. Time evolution of trading volume is plotted in Fig. 4.

4. Overview of volatility models

This section presents the GARCH and HAR models used in this paper.

4.1. GARCH models

Generalized autoregressive conditional heteroskedacity (GARCH) models illustrate the volatility as a stationary process. Let X_t denote the observed log-returns of the Bitcoin prices, and μ the conditional mean. With a volatility process denoted by σ_t^2, GARCH models can be given by:

$$X_t = \mu_t + \sigma_t Z_t \tag{3}$$

The GARCH-models differ in the specification of the volatility process σ_t, but all these models use an AR(1) process for the conditional mean. The innovation Z_t is distributed according to one of the following distributions: normal distribution (norm), skewed normal distribution (snorm), generalized error distribution (ged), skewed generalized error distribution (sged), Student's t distribution (std), skewed Student's t distribution (sstd), normal inverse Gaussian distribution (nig), generalized hyperbolic distribution (ghyp), or Johnson's S_U-distribution (jsu). In the case of Markov Switching GARCH model with two standard GARCH(1,1) regimes, the two regimes have the same type of the distribution of the innovation, either norm, snorm, ged, sged, std, or sstd.

The GARCH(p,q) process is specified as (Bollerslev, 1986):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{4}$$

Here, p are the number of lags of the conditional variance and q the number of lags of the squared errors. We also need that:

$$\alpha_0 > 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \ldots, q$$
$$\beta_i \geq 0, \quad 2, \ldots, p$$

The Glosten–Jagannathan–Runkle GARCH (GJR-GARCH) model is another form of GARCH-model. It was developed based on the idea that the standard GARCH might not be good enough at capturing the characteristics of monthly excess return (Glosten et al., 1993). It is an asymmetric version of the GARCH(p,q) model which takes leverage effects into account. The conditional variance in the GJR-GARCH(p,q) model is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}(\alpha_i + \gamma_i I_{t-i})Z_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 \tag{5}$$

where I_{t-i} is an indicator function that takes the values:

$$I_{t-i} = \begin{cases} 1 & \text{if } Z_{i-t} < 0, \quad \text{for } i = 1,2,\dots,q \\ 0 & \text{if } Z_{t-i} \geq 0, \quad \text{for } i = 1,2,\dots,q \end{cases}$$

We also need that:

$$\begin{array}{lll} \alpha_0 > 0, & \alpha_i \geq 0, & i = 1,2,\dots,q \\ \beta_i \geq 0, & & i = 1,2,\dots,p \\ \alpha_i + \gamma_i \geq 0, & & i = 1,2,\dots,p \end{array}$$

γ_i is an asymmetry parameter. If it is positive, it indicates that volatility increases relatively more following negative shocks than what it does following positive shocks.

The integrated GARCH (IGARCH) model has a persistent variance property. The equation for the conditional variance is the same as for standard GARCH(p,q), Eq. (4)., but the IGARCH(p,q) is characterized by having:

$$\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j = 1 \tag{6}$$

In this model, the unconditional variance of the model does not exist.

The exponential GARCH (EGARCH) model is another asymmetric GARCH model accounting for leverage effects proposed by Nelson (1991). The conditional variance in the EGARCH(p,q) model is defined by:

$$ln\left(\sigma_t^2\right) = \alpha_0 + \sum_{j=1}^{p}\beta_j ln\left(\sigma_{t-j}^2\right) + \sum_{i=1}^{q}\left[\gamma_i\left(\frac{Z_{t-i}}{\sqrt{\sigma_{t-i}^2}}\right) + \alpha_i\left(\frac{|Z_{t-i}|}{\sqrt{\sigma_{t-i}^2}} - E\left(\frac{|Z_{t-i}|}{\sqrt{\sigma_{t-i}^2}}\right)\right)\right] \tag{7}$$

For $\gamma_i < 0$ negative shocks will have a higher influence than positive shocks of the same magnitude on future volatility, accounting for the leverage effect. If γ_i equals zero, the model indicates perfect symmetry. By using the logarithm, it is not necessary to impose restrictions on the parameters requiring them to be positive.

The asymmetric power ARCH (APARCH) model – which accounts for long memory, power effects and leverage terms – was proposed by Ding et al. (1993). The APARCH(p,q) model is defined as:

$$\sigma_t^\delta = \omega + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^\delta + \sum_{i=1}^{q}\alpha_i(|Z_{t-i}| - \gamma_i Z_{t-i})^\delta \tag{8}$$

The introduction of the power term, $\delta > 0$ as a parameter to be estimated is the main difference from other GARCH-type models where it is usually assumed to be 2. δ plays the role of a Box-Cox transformation of the conditional volatility σ_t, whereas γ_i reflects the leverage effect. The standard GARCH model is a special case of the APARCH model where $\delta = 2$ and $\gamma_i = 0 \ (i = 1,\dots,q)$.

Markov switching GARCH (MSGARCH) models allow the unconditional mean and variance to change from one GARCH regime to another (Bauwens et al., 2010). Mikosch and Stărică (2004) and Hillebrand (2005) show that in the presence of structural changes in the unconditional variance, the estimation of regular GARCH models can create bias upwards in the persistence parameter. In MSGARCH, the parameters are allowed to change from one regime to another. The model allows the unconditional variance to change based on different regimes. Denoting the log-return at time t by y_t, we assume that this variable has a zero mean and is not serially correlated, so:

$$\mathrm{E}[y_t] = 0 \quad \text{and} \quad \mathrm{E}[y_t y_{t-k}] = 0, \quad \text{for} \quad k \neq 0$$

Let $\mathscr{I}_{t-1}$ denote the information we have up to time $t-1$: $\mathscr{I}_{t-1} \equiv \{y_{t-i},\, i > 0\}$. $\mathscr{D}\left(0, \sigma_{k,t}^2, \xi_k\right)$ is a zero-mean continuous distribution with a time-varying variance $\sigma_{k,t}^2$. ξ_k is a vector of additional shape parameters for the distribution. A Markov-switching GARCH specification can then be written as:

$$y_t|(s_t = k, \mathscr{I}_{t-1}) \sim \ \mathscr{D}\left(0, \sigma_{k,t}^2, \xi_k\right) \tag{9}$$

The MSGARCH models can be characterized by an integer-valued stochastic variable s_t. This variable is defined on the discrete space $\{1,\dots,K\}$. It is assumed that this evolves as stated by an unobserved periodic homogeneous Markov chain. The transition matrix has K rows and K columns:

$$\mathrm{P} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,K} \\ \vdots & \ddots & \vdots \\ p_{K,1} & \cdots & p_{K,K} \end{bmatrix}$$

Here $p_{i,j}$ denotes the probability of going from state $s_{t-1} = i$ to state $s_t = j$. We therefore require that the following holds: $\sum_{j=1}^{K} p_{i,j} = 1$, $\forall i \in \{1, \ldots, K\}$. Haas et al. (2004) assume that the conditional variances $\sigma_{k,t}^2$, $k = 1, \ldots, K$ each follow a separate GARCH-process giving K separate processes evolving in parallel. They use the past observations, past variance and a vector of regime dependent parameters, $\boldsymbol{\theta}_k$, to find $\sigma_{k,t}^2$ conditional on regime s_t:

$$\sigma_{k,t}^2 = f\left(y_{t-1}, \sigma_{k,t-1}^2, \boldsymbol{\theta}_k\right) \tag{10}$$

f is a function that takes into account all previous observations and ensures positiveness.

Specification of GARCH models. GARCH(p, q), EGARCH(p, q), IGARCH(p, q), GJR-GARCH(p, q), and APARCH(p, q) models were fitted as described in this section for $p = 1$ and $q = 1$.

4.2. Heterogeneous autoregressive models

The heterogeneous autoregressive (HAR) model is able to capture the long-memory of realized variance (Corsi, 2009), which can be estimated as a simple regression. The regression uses daily, weekly, and monthly realized variance. Since Bitcoin is traded 24 h a day, independent of exchanges' opening hours, we use seven day weeks and 30 day months. Define the simple moving averages:

$$\begin{aligned} RV_{D,t} &= RV_t \\ RV_{W,t} &= RV_{t-6,t} = \frac{RV_{t-6} + RV_{t-5} + \cdots + RV_t}{7} \\ RV_{M,t} &= RV_{t-29,t} = \frac{RV_{t-29} + RV_{t-28} + \cdots + RV_t}{30} \end{aligned} \tag{11}$$

Forecast is constructed by running the following regression:

$$RV_{t+1} = \phi_0 + \phi_D RV_{D,t} + \phi_W RV_{W,t} + \phi_M RV_{M,t} + \varepsilon_{t+1} \tag{12}$$

This equation defines the HAR model, specified by Corsi (2009). Corsi and Renò (2012) changed this model by a log-transformation:

$$ln(RV_{t+1}) = \phi_0 + \phi_D ln\left(RV_{D,t}\right) + \phi_W ln\left(RV_{W,t}\right) + \phi_M ln\left(RV_{M,t}\right) + \varepsilon_{t+1} \tag{13}$$

The advantages of using the logarithmic specification (13) is that it assures positive variance forecasts, and it fits the assumptions of the ordinary least squares (OLS) better.

HAR and logarithmic HAR models for the variance have been estimated using Eqs. (12) and (13), respectively. In the regular HAR, the realized variance is regressed on the previous day's variance, previous weekly variance, and previous monthly variance. For logarithmic HAR, we do the same. However, we now regress the logarithm of the realized variance on the logarithm of the previous day's variance, the average of the previous seven day's logarithmic variances, and the average of the previous 30 day's logarithmic variances.

5. Methods for model comparison

This section describes how models are compared using the information criteria AIC and BIC and the loss functions MAE, MSE, MAPE, and QLIKE. In addition, we describe the model confidence set approach used to establish whether the models have statistically different predictive abilities.

5.1. Information criteria

Information criteria are a common way to compare models in-sample. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are estimates of the relative quality of statistical models. We let k denote the number of parameters in the model, $\widehat{L}$ the maximum value of the likelihood function, and n the number of observations. Then the Akaike information criterion, developed by Akaike (1974), is defined as:

$$\text{AIC} = 2k - 2ln(\widehat{L}) \tag{14}$$

and the Bayesian information criterion, developed by Schwarz et al. (1978), is defined as:

$$\text{BIC} = ln(n)k - 2ln(\widehat{L}) \tag{15}$$

The model with the lowest value of the information criterion is preferred.

5.2. Loss functions

Loss functions are used to evaluate forecasts. The general rule is that the model producing the lowest loss function value is the best at predicting volatility when comparing different models. In our case, it tells us how a model performs out-of-sample compared to realized variance defined in Section 3. We let T denote the number of observations out-of-sample, $\widehat{\sigma}_t^2$ be the variance forecast for day t, and σ_t^2 be the "true" variance for day t.

Mean square error (MSE) is one of the most used loss functions. It is calculated by taking the squared difference between the variance proxy and the variance forecast for each of the out-of-sample observations, and then the average is taken:

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\sigma}_t^2 - \sigma_t^2\right)^2 \tag{16}$$

The literature has shown that a few extreme outcomes can have a large impact on the forecast evaluation, see Andersen et al. (1999) and Poon and Granger (2003). Mean absolute error (MAE) loss function, which mitigates this problem, is calculated as follows:

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|\widehat{\sigma}_t^2 - \sigma_t^2\right| \tag{17}$$

A third loss function is the mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\sigma_t^2 - \widehat{\sigma}_t^2}{\sigma_t^2}\right| \tag{18}$$

A forth loss function we use is QLIKE, which is very suitable for evaluation of volatility forecasts (Patton, 2011):

$$\mathrm{QLIKE} = \frac{1}{T}\sum_{t=1}^{T}\left(ln\left(\widehat{\sigma}_t^2\right) + \frac{\sigma_t^2}{\widehat{\sigma}_t^2}\right) \tag{19}$$

5.3. Model confidence set

The model confidence set (MCS) procedure, developed by Hansen et al. (2011), is an approach for choosing the best model or set of models. The models are evaluated using a user-specified loss function.

$$d_{i\cdot,t} = (m-1)\sum_{j\in M} d_{\mathrm{ij},t}, \quad i = 1,\dots,m \tag{20}$$

where M is a set of m models. Then, the null and alternative hypotheses of equal predictive ability of a model set M can be formulated as:

$$\begin{array}{lll} H_{0,M}: & \mathbb{E}\left[d_{\mathrm{ij}}\right] = 0, & \text{for all } i,j = 1,\dots,m \\ H_{1,M}: & \mathbb{E}\left[d_{\mathrm{ij}}\right] \neq 0, & \text{for some } i,j = 1,\dots,m \end{array} \tag{21}$$

Hansen et al. (2011) show that these hypotheses can be tested using two test statistics:

$$t_{\mathrm{ij}} = \frac{\overline{d}_{\mathrm{ij}}}{\sqrt{\widehat{\sigma}^2\left(\overline{d}_{\mathrm{ij}}\right)}} \quad \text{and} \quad t_{i\cdot} = \frac{\overline{d}_{i\cdot}}{\sqrt{\widehat{\sigma}^2\left(\overline{d}_{i,\cdot}\right)}}, \quad \text{for } i,j \in M \tag{22}$$

Here $\overline{d}_{i\cdot}$ gives the relative sample loss between model i and j, given by $\overline{d}_{\mathrm{ij}} = \frac{1}{m}\sum_{t=1}^{m} d_{\mathrm{ij},t}$. Then $\overline{d}_{i,\cdot} = \frac{1}{m-1}\sum_{j\in M}\overline{d}_{\mathrm{ij}}$ gives the simple loss of model i relative to the average of losses for all models in M. $\widehat{\sigma}^2\left(\overline{d}_{\mathrm{ij}}\right)$ and $\widehat{\sigma}^2\left(\overline{d}_{i,\cdot}\right)$ are bootstrapped estimates of $\sigma^2\left(\overline{d}_{\mathrm{ij}}\right)$ and $\sigma^2\left(\overline{d}_{i,\cdot}\right)$. The test statistics are then constructed as:

$$\mathrm{T}_{R,M} = \max_{i,j\in M}\left|t_{\mathrm{ij}}\right| \quad \text{and} \quad \mathrm{T}_{\max,M} = \max_{i\in M} t_{i\cdot} \tag{23}$$

The relevant distribution under the null hypothesis has to be estimated using bootstrapping since it is asymptotic and nonstandard. If the test statistics are larger than the bootstrapped estimates, the null hypothesis is rejected, and the models compared have different predictive abilities.

6. Results

This section presents the results. First, the in-sample fits of the GARCH and HAR models are evaluated. Next, results from the out-of-sample model comparison are presented.

Table 2
Information criteria values for the 10 best performing GARCH models in-sample.

(a) AIC values	
Model	AIC
MS-SGARCH-sstd	− 7072.0
EGARCH-nig	− 7066.6
EGARCH-ghyp	− 7065.0
APARCH-nig	− 7062.5
EGARCH-jsu	− 7061.7
MS-SGARCH-sged	− 7061.2
APARCH-ghyp	− 7060.9
APARCH-jsu	− 7057.5
EGARCH-ged	− 7056.0
IGARCH-nig	− 7055.1
(b) BIC values	
Model	BIC
EGARCH-nig	− 7023.0
IGARCH-nig	− 7022.4
IGARCH-ged	− 7020.8
EGARCH-jsu	− 7018.1
EGARCH-ged	− 7017.9
IGARCH-ghyp	− 7016.8
EGARCH-ghyp	− 7016.0
IGARCH-sged	− 7014.8
SGARCH-nig	− 7014.5
APARCH-nig	− 7013.5

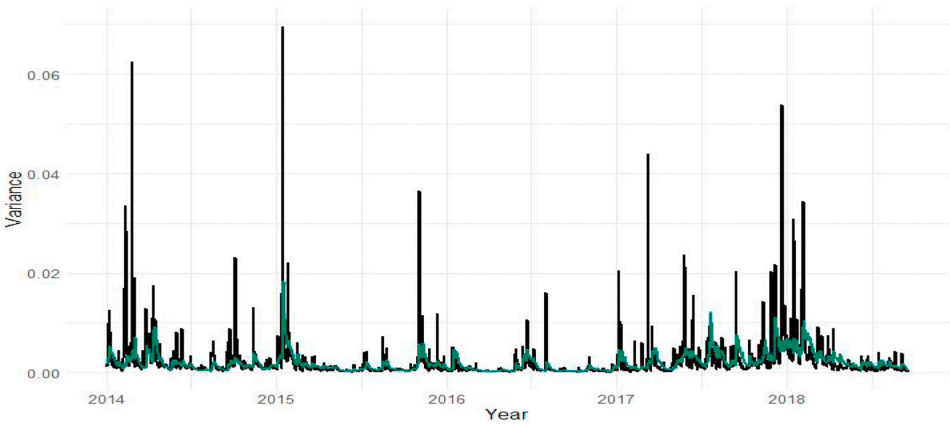


Fig. 5. Fitted values of EGARCH nig (green) against realized variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.1. In-sample comparison

The in-sample comparison tells us how well a model fits the entire data set. The GARCH models were compared relative to each other using the information criteria AIC and BIC. This method is not applicable for comparing GARCH models with HAR models, since the models are estimated on different data. Therefore, the HAR models are compared separately. Lastly, the loss function MSE is used to illustrate the difference in performance between GARCH and HAR models.

6.1.1. GARCH models

All GARCH-type models were fitted on the total data sample consisting of 1720 observations. The values of AIC and BIC for the top 10 models are given in Table 2. We can see that the models selected by AIC and BIC are quite different. The best model according to the AIC is the Markov Switching GARCH-sstd, while the best model according to the BIC is the EGARCH-nig. The AIC tends to favour more

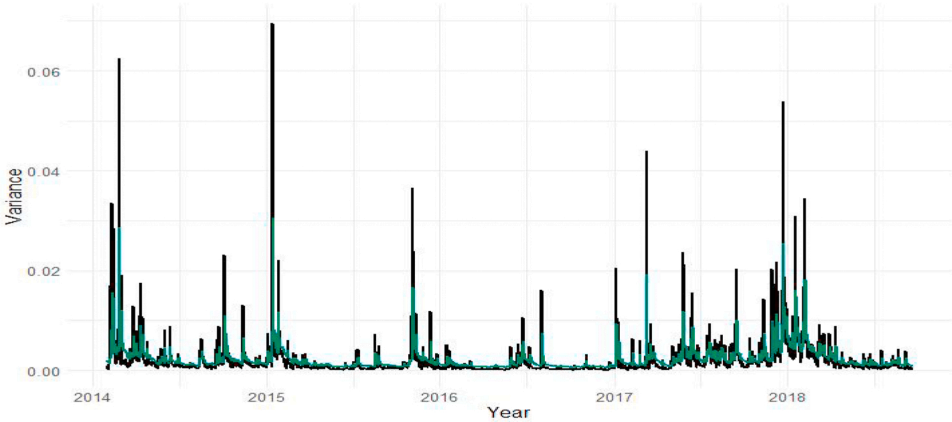


Fig. 6. Fitted values of regular HAR (green) against realized variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

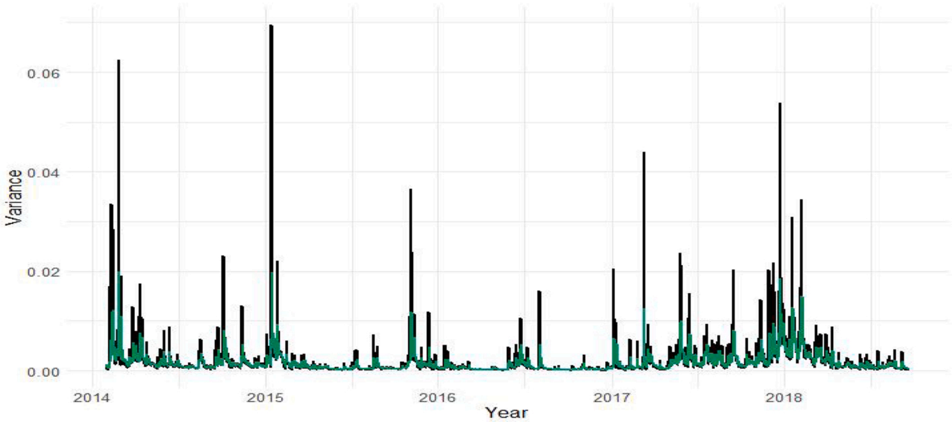


Fig. 7. Fitted values of logarithmic HAR (green) against realized variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Parameters and R^2 values for regular and logarithmic HAR models.

	Regular HAR	Logarithmic HAR
ϕ_0	4.740×10^{-4}**	− 0.7990**
ϕ_D	0.3894**	0.5400**
ϕ_W	0.1358*	0.2336**
ϕ_M	0.2585**	0.1245**
R^2	0.2665	0.6393
Adjusted R^2	0.2652	0.6386

Significance codes: $p = 0$ '**', $p \leq 0.001$ '*'.

complex models that might be overfitting. We see that using the AIC, we get several Markov Switching GARCH and APARCH models in the top 10. These models have more parameters than simpler GARCH models. The BIC penalizes the number of parameters more heavily than the AIC, and therefore favours the simpler models. Fig. 5 shows the conditional volatility of the EGARCH-nig model plotted against the volatility proxy. We can see that the model is not able to properly capture the large jumps in volatility. This is as expected, as GARCH models have a tendency to smooth out variance.

6.1.2. HAR models

Another group of volatility models we estimated are HAR models. We estimated the regular HAR as specified by Eq. (12) and logarithmic

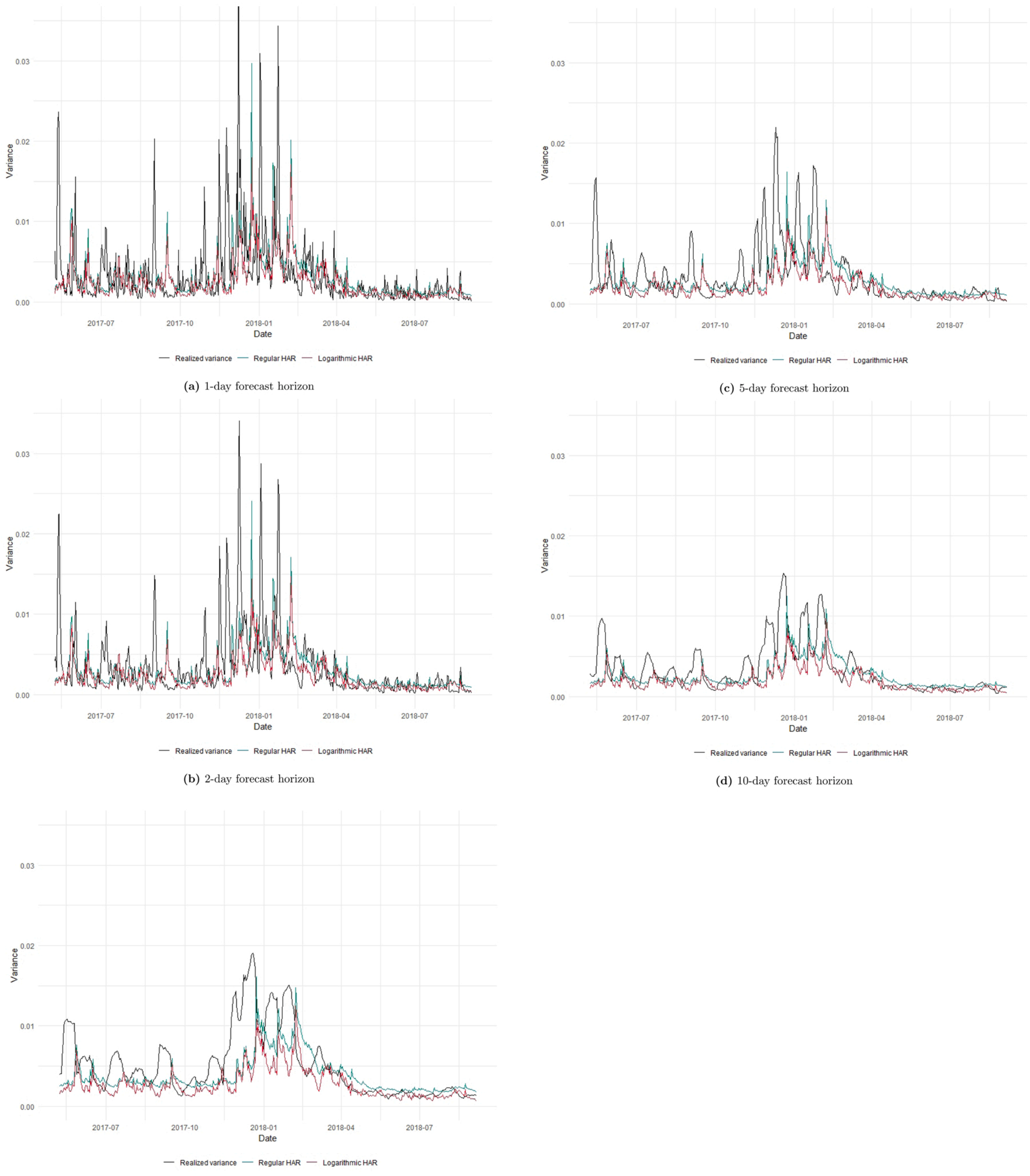
Table 4
In-sample MSE values for regular HAR, logarithmic HAR, and EGARCH-nig.

Model	MSE
Regular HAR	1.451×10^{-5}
Log HAR	1.480×10^{-5}
EGARCH-nig	1.599×10^{-5}

Table 5
Summary of the model confidence sets for the MAE, MSE, MAPE, and QLIKE loss functions for cumulative forecasts on horizons of 1, 2, 5, 10, and 15 days, using a significance level of $\alpha = 0.05$.

Forecast horizon	MAE	MSE	MAPE	QLIKE
1 day	Log HAR	Regular HAR Log HAR APARCH-sstd	Log HAR	Regular HAR
2 days	Log HAR	Regular HAR APARCH-sstd	Log HAR	Regular HAR APARCH-sged
5 days	Regular HAR Log HAR	Regular HAR APARCH-sstd	Log HAR	Regular HAR EGARCH-std EGARCH-sstd EGARCH-jsu APARCH-std APARCH-jsu
10 days	Regular HAR Log HAR	Regular HAR MSGARCH-sstd APARCH-sstd	Log HAR	Regular HAR EGARCH-std EGARCH-sstd EGARCH-jsu MSGARCH-std MSGARCH-sstd APARCH-sged APARCH-std APARCH-sstd APARCH-jsu
15 days	Regular HAR Log HAR	Regular HAR GARCH-norm EGARCH-std EGARCH-sstd EGARCH-nig EGARCH-ghyp EGARCH-jsu IGARCH-norm IGARCH-snorm GJR-GARCH-norm GJR-GARCH-ged GJR-GARCH-sged GJR-GARCH-std GJR-GARCH-sstd GJR-GARCH-nig GJR-GARCH-ghyp GJR-GARCH-jsu MSGARCH-sstd APARCH-ged APARCH-std APARCH-sstd APARCH-jsu	Log HAR	Regular HAR EGARCH-std EGARCH-sstd EGARCH-jsu APARCH-ged APARCH-sstd APARCH-jsu

HAR as specified by Eq. (13). Figs. 6 and 7 plot the regular HAR and the logarithmic HAR against the variance proxy for the entire period, respectively. From the figures we can see that the estimated variances produced by the HAR models fit realized variance more precisely than the EGARCH-nig model in Fig. 5. When comparing the two HAR models, we see that the regular HAR model fits the proxy better in times of high variance, whereas the logarithmic HAR model fits better in times of low variance. The parameter values from the two HAR models are presented in Table 3. All parameters are significant at the 0.001 significance level. For both regular and logarithmic HAR, the previous day's variance is the most important predictor of the current day's variance. Even though logarithmic HAR gives a much larger R^2 value than the regular HAR, the values are not comparable since the logarithmic HAR model is estimated on transformed data.



(a) 1-day forecast horizon

(c) 5-day forecast horizon

(b) 2-day forecast horizon

(d) 10-day forecast horizon

(e) 15-day forecast horizon

Fig. 8. The daily average cumulative realized variance against the predicted daily average cumulative variance for regular HAR (green) and logarithmic HAR (red) for time horizons of 1, 2, 5, 10, and 15 days. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.1.3. Best in-sample fit based on MSE

In order to illustrate the difference between the GARCH and HAR models, we also compare the best GARCH model (EGARCH-nig), and the HAR models in terms of in-sample MSE. The analysis is done using the loss function mean square error, presented in Section 5.2. The loss is calculated using the fitted values from the previously mentioned three models and the realized variance. Results are presented in Table 4. The regular HAR model has the lowest loss function value, and therefore fits the data best.

Table 6
Summary of the model confidence sets for the MAE, MSE, and MAPE loss functions for forecasting horizons of 1, 2, 5, 10, and 15 days for predicting logarithmic variance, using a significance level of $\alpha = 0.05$.

Forecast horizon	MAE	MSE	MAPE
1 day	Log HAR	Regular HAR	Log HAR
		Log HAR	
2 days	Regular HAR	Regular HAR	Regular HAR
	Log HAR	Log HAR	Log HAR
5 days	Regular HAR	Regular HAR	Regular HAR
	Log HAR		
10 days	Regular HAR	Regular HAR	Regular HAR
15 days	Regular HAR	Regular HAR	Regular HAR

6.2. *Comparison of forecasting performance*

The comparison of predictive accuracy of the different models has been determined through four different analyses. We have looked at the forecasting performance of all models on cumulative variance at different horizons. A subset of well-performing models was selected for the analysis of predictive accuracy of logarithmic variance, daily variance, and cumulative variance using varying in-sample window size. The comparison was done using loss functions and the MCS approach presented in Section 5.3.

6.2.1. *Predicting cumulative variance*

We used the K-day cumulative realize variance when evaluating the predicted variance for the K-day period, where K is one, two, five, 10, and 15. When predicting the cumulative variance, all models were fitted on a rolling window of 1220 observations starting on 01-Jan-2014, leaving out 500 observations. The fitted models were then used to predict the variance for periods of one, two, five, 10, and 15 days. For each period, we use the cumulative predictions for that period's variance. To evaluate the forecasts, we compared the models against our variance proxy, the realized variance.

We evaluated the models by calculating the loss functions MSE, MAE, MAPE, and QLIKE, as described in Section 5.2, for each model. The model with the lowest value is identified as the best model. Tables A.1–A.5 in the appendix present the loss function values for the 10 models with the lowest loss function values for each of the four loss functions. From the tables, we observe that both HAR models perform well compared to the other models, especially for shorter horizons. For one-day forecast, logarithmic HAR performs best when considering MAE and MAPE, while regular HAR has the lowest values when considering MSE and QLIKE. The same results apply for the two day-period. For a five day-period, regular HAR has the lowest value for MAE, MSE, and QLIKE, while logarithmic HAR is best when looking at MAPE. For the 10- and 15-day periods, we observe the same results, except where the 15-day period EGARCH-std has the lowest value for QLIKE.

To determine if any of the models are statistically better than the others, we used the model confidence set (MCS) approach, as described in Section 5.3. For each prediction horizon, we used the model with the smallest loss function value for each of the loss functions as a benchmark model and compared it to each of the other models. This test gives us the set of models with forecasting abilities not statistically different from the best model. These models will make up the sets of superior forecasting models. We make the evaluation using a significance level of $\alpha = 0.05$. The test is summarized in Table 5.

Table 5 combined with Tables A.1, A.2, A.3, A.4 and A.5 indicate that the HAR models perform well at predicting the variance of Bitcoin. We especially observe that for one day logarithmic HAR is part of the superior sets of models when using the three loss functions MAE, MSE, and MAPE, while regular HAR is part of the sets when considering MSE and QLIKE. For the two-day period, logarithmic HAR has superior predictive ability when looking at MAE and MAPE, while regular HAR is in the superior sets for MSE and QLIKE. For periods of five, 10, and 15 days, regular HAR is part of all superior sets, except when considering MAPE. Logarithmic HAR performs well in general, especially when considering the loss functions MAE and MAPE but less so when using MSE and QLIKE. These last two loss functions are considered robust, as opposed to MAE and MAPE. It is worth noting that APARCH-sstd is part of the superior sets for MSE for all horizons. Since regular HAR is in the superior sets for all time horizons when looking at both of the robust loss functions, and it is a much less complex model than APARCH, we conclude that regular HAR is the best choice of model for all horizons of the considered models. However, we note that when the horizon increases, several of the GARCH-type models are included in the superior sets of models, especially for QLIKE and MSE. We observe that APARCH-sstd and EGARCH with distributions std, sstd, and jsu will be the best alternative to the HAR models of the GARCH family of models. It is worth noting that EGARCH is less complex than the APARCH model and would therefore probably be the preferred choice in many practical applications.

Fig. 8a–e plots the daily average cumulative realized variance against daily average cumulative variance predicted by HAR and logarithmic HAR for each of the time horizons. We observe again, as in the in-sample-comparison, that the regular HAR has a tendency to follow the realized variance well in times of high volatility, while the opposite seems to be the case for the logarithmic HAR. Especially, for longer time horizons the logarithmic HAR has a tendency to under-predict the variance more than the regular HAR.

6.2.2. *Predicting logarithmic variance*

We also tested ability of some (better performing) models to predict the logarithmic cumulative realized variance. We did not perform this analysis on all the models, but chose one distribution for each of the GARCH-type models. For each horizon, we based our selection on which distribution provided the lowest MSE value for each GARCH model for the specified time horizon, using the analysis conducted in Section 6.2.1. See Table A.11 for which GARCH models this includes for each horizon. The MCS approach was used to evaluate the models,

Table 7
Summary of the model confidence sets for the MAE, MSE, MAPE, and QLIKE loss functions when forecasting the daily volatility for the 1st, 2nd, 5th, 10th, and 15th day ahead, using a significance level of $\alpha = 0.05$.

Forecast horizon	MAE	MSE	MAPE	QLIKE
1 day	Log HAR	Regular HAR Log HAR	Log HAR	Regular HAR
2 days	Log HAR	Regular HAR Log HAR EGARCH-nig	Log HAR	Regular HAR EGARCH-nig
5 days	Regular HAR Log HAR	Regular HAR	Log HAR	Regular HAR APARCH-ged
10 days	Regular HAR Log HAR EGARCH-jsu GJR-GARCH-sged	Regular HAR Log HAR EGARCH-jsu GJR-GARCH-sged GARCH-ghyp IGARCH-ghyp MSGARCH-sged APARCH-ged	Log HAR	Regular HAR GARCH-ghyp EGARCH-jsu GJR-GARCH-sged IGARCH-ghyp
15 days	Regular HAR Log HAR GARCH-ghyp EGARCH-jsu GJR-GARCH-ghyp	Regular HAR IGARCH-ghyp GARCH-ghyp EGARCH-jsu GJR-GARCH-ghyp MSGARCH-sstd APARCH-ged	Log HAR	Regular HAR GJR-GARCH-ghyp MSGARCH-sstd

Table 8
Summary of the model confidence sets for the MSE and QLIKE loss functions when forecasting the cumulative forecast for one, two, five, 10, and 15 days ahead, using a significance level of $\alpha = 0.05$. Models are estimated using 750 observations; loss functions are calculated using 970 out-sample observations.

Forecast horizon	MSE	QLIKE
1 day	Regular HAR Log HAR	Regular HAR Log HAR EGARCH-jsu
2 days	Regular HAR Log HAR	Regular HAR Log HAR EGARCH-jsu
5 days	Regular HAR Log HAR	Regular HAR EGARCH-std EGARCH-jsu
10 days	Regular HAR Log HAR EGARCH-sstd EGARCH-jsu	Regular HAR EGARCH-std EGARCH-sstd EGARCH-jsu
15 days	Regular HAR Log HAR EGARCH-sstd EGARCH-jsu	EGARCH-std EGARCH-sstd EGARCH-jsu

again at a significance level of 5%. Table 6 shows the results. QLIKE requires taking the logarithm of the prediction, but since the logarithmic variance predictions are negative, it cannot be calculated. Therefore, only the loss functions MAE, MSE, and MAPE are used. We observe that regular HAR outperforms the logarithmic version. Logarithmic HAR is only included in the superior sets of models for shorter time horizons, while regular HAR is included in all superior sets, except for MAE and MAPE when considering one day. We also notice that none of the GARCH models are included in any of the superior sets. This supports our finding that the HAR models have the ability to capture the dynamics of the realized variance time series, and this is especially the case for regular HAR, even when predicting logarithmic variance.

6.2.3. Predicting daily variance at different horizons

As another experiment, we wanted to look at the predictive accuracy of different models on daily variance, as opposed to cumulative variance, at different horizons. We estimated all models, as in Section 6.2.1, and predicted the variance for the 1st, 2nd, 3rd, 10th, and 15th day ahead. We then calculated the MAE, MSE, MAPE, and QLIKE loss functions for each model using the daily realized variance. These results are presented in Tables A.6–A.10 in the appendix. Again, we evaluated the models using the MCS approach with a significance level of 5%. For this test, we used the same subset of models as in Section 6.2.2, see Table A.11. We found that logarithmic HAR was included in the sets of superior models for all horizons when considering MAE and MAPE. For the first, second, and 10th day

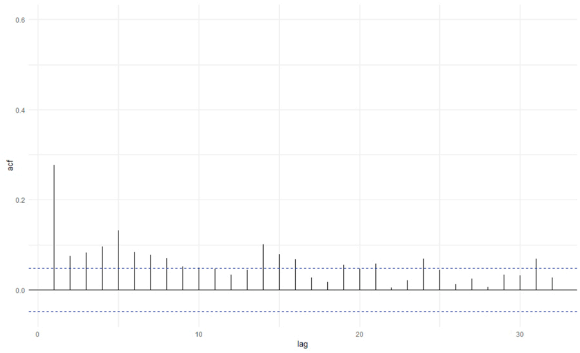


Fig. 9. Autocorrelation function for squared log-returns time series from lag 1 to 35, confidence level 5%.

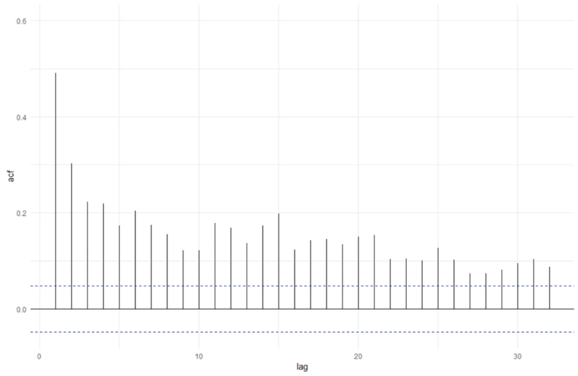


Fig. 10. Autocorrelation function for realized variance from lag 1 to 35, confidence level 5%.

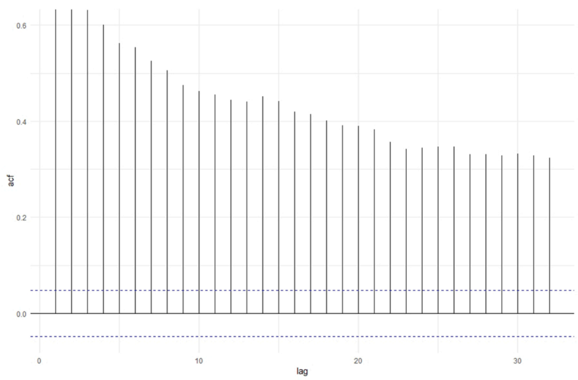


Fig. 11. Autocorrelation function for logarithmic realized variance from lag 1 to 35, confidence level 5%.

prediction, is it included also in the superior sets of models for MSE. Regular HAR is included in all superior sets when considering MSE and QLIKE, as well as MAE for the fifth, 10th, and 15th day prediction. Again, HAR performs well, especially for shorter horizons. When considering horizons of 10 and 15 days ahead, several of the models are included in the superior sets. This is due to increasing uncertainty, making us less able to distinguish the models. Complete results of the MCS approach are summarized in Table 7.

6.2.4. *Varying the in-sample window size*

In all the above analyses, we estimated the models with an in-sample window size of 1220 observations. This is a relatively long history, and the potential risk is that past observations that worsen the fit of our models are included. We investigated this by estimating some of the models using in-sample window sizes of 500, 750, and 1000 observations. We compared both of the models estimated with the same window, and the models estimated with different in-sample windows. From the summary of the MCS approach, Table 5, we chose to do the analysis on the regular HAR, logarithmic HAR, EGARCH-std, EGARCH-sstd, and EGARCH-jsu models. This choice was based on the observation that these models are frequently represented in the sets of superior models. Based on this argument, we could also have chosen to use APARCH-sstd, however, due to its complexity, we have chosen not to. As MSE and QLIKE are robust loss functions, we chose to compare the models using these functions. Again, we use the cumulative forecast for one, two, five, 10, and 15 days.

Tables A.12–A.15 give the MSE and QLIKE loss function values when evaluating the models estimated on different in-sample window sizes, using 500 out-of-sample observations. They show that regular HAR with 1000 in-sample observations gives the lowest MSE value for all time horizons, while regular HAR with 1220 in-sample observations gives the lowest QLIKE value, except for the 15 day horizon, where EGARCH-std with 1220 observations gives the lowest value. We used the MCS approach to evaluate each model's predictive ability when changing the window size. Again, we use a significance level of 5%. In general, all the regular HAR models have the same predictive ability over all horizons. The same is the case for logarithmic HAR. When considering the EGARCH models, we found that the EGARCH models with 1220 in-sample observations are typically superior for shorter horizons, while for horizons of five to 15 days, EGARCH models with 750 and 1000 observations have the same predictive abilities. HAR is a simple model that requires fewer observations to be estimated precisely, while GARCH-type models are more complex and require a larger sample of observations for estimation. Altogether, the length of the estimation window has a very small impact on our conclusions.

We also did an experiment where we compared the models using a larger set of out-of-sample observations to estimate the loss functions. We chose the models estimated on 750 observations, leaving 970 observations out-of-sample. Table A.16 provides the loss function values for the models. We see that regular HAR has the lowest MSE value for all time horizons, except for one day predictions where logarithmic HAR has the lowest MSE value. EGARCH-jsu has lowest QLIKE value for all horizons. Again, we apply the MCS procedure with a significance level of 5%. The results are summarized in Table 8. From the table, we see that when we estimate the models using fewer observations and evaluate them using more out-of-sample observations, the conclusions are consistent with the original findings which used 1220 in-sample observations. Regular and logarithmic HAR appear superior for shorter time horizons, especially when considering MSE. However, we also observe that the EGARCH models are still part of most of the superior sets of models. From this, we determine that our conclusions regarding models estimated on 1220 observations do not change when using smaller in-sample sizes and more out-of-sample observations.

6.2.5. Autocorrelation in squared returns and realized variance

Throughout the previous sections, we have concluded that regular HAR is the preferred model for the considered time horizons. We also observed that logarithmic HAR performed well for the shorter horizon. In order to illustrate why models based on realized variance perform better than models based on squared returns, we plot the autocorrelations (ACF) for the time series of squared returns (Fig. 9), realized variance (Fig. 10) and logarithmic realized variance (Fig. 11).

Fig. 9 plots the autocorrelation of the squared log-returns. From this figure, it appears that this time series is highly autocorrelated for several lags. This shows that the autocorrelation of the squared returns are highly persistent. Fig. 10 and 11 plot the autocorrelation function for the realized variance and the logarithmic realized variance, respectively. These figures indicate strong autocorrelation in both time series. We observe that both regular realized variance and the logarithmic version have a trend where the ACF decreases slowly. The ACF decreases slower for the logarithmic realized variance than for the regular one. The long memory of the process is even more visible when compared to the squared log-returns.

7. Conclusion

This paper analyses which models are most suitable for forecasting volatility of Bitcoin. We consider several GARCH models (GARCH, EGARCH, GJR-GARCH, IGARCH, MSGARCH, and APARCH) models with various error distributions and two HAR models (regular and logarithmic HAR model). The models are compared in terms of both in-sample fit and out-of-sample forecasting performance but emphasis is on the forecasting performance. We study forecasting of volatility one, two, five, 10, and 15 days ahead. As a proxy for true variance, we utilize realized variance calculated from high-frequency data.

We find that from the class of GARCH models, EGARCH and APARCH are the best performing models. HAR models perform better than any of the considered GARCH models. The reason for superiority of HAR models over GARCH models is that HAR models are estimated from realized variance, and as we show, realized variance exhibits much higher degree of autocorrelation than squared returns.

The difference between performance of HAR models and GARCH models is largest particularly for short-term forecasting horizons. However, HAR models perform well for all horizons. These conclusions are robust with respect to the size of the estimation window, as well as choice of loss function. We arrive at similar conclusions when we forecast logarithmic variance, and also when we forecast volatility for a particular day in the future, e.g. volatility for the 10th day in the future, instead of volatility for the 10-day period.

Altogether, we recommend the HAR model based on high-frequency data whenever precise volatility forecasts of Bitcoin are required. In the case that the user seeks a model estimated on daily data, we recommend an EGARCH model as a good alternative.

Authors' contribution

Lykke Øverland Bergsli and Andrea Falk Lind: conceptualization, methodology, data curation, formal analysis, writing – original draft. Peter Molnár: supervision, conceptualization, methodology, writing – original draft, writing – review & editing. Michał Polasik: writing – review & editing.

Acknowledgement

M. Polasik acknowledges the support by the National Science Centre, Poland under Grant No. 2017/26/E/HS4/00858 and P. Molnár acknowledges the Institutional support of the Prague University of Economics and Business no. IP100040.

Appendix A

Table A.1
Loss function values for 1 day cumulative forecast.

(a) MAE loss values

Model	MAE
Logarithmic HAR	1.711×10^{-3}
Regular HAR	1.831×10^{-3}
EGARCH-nom	1.973×10^{-3}
MSGARCH-sged	1.974×10^{-3}
MSGARCH-snorm	1.981×10^{-3}
EGARCH-snorm	1.981×10^{-3}
MSGARCH-norm	1.996×10^{-3}
GARCH-ghyp	2.002×10^{-3}
GARCH-ged	2.002×10^{-3}
GARCH-sged	2.003×10^{-3}

(b) MSE loss values

Model	MSE
Regular HAR	1.563×10^{-5}
Logarithmic HAR	1.616×10^{-5}
IGARCH-sged	1.859×10^{-5}
IGARCH-ghyp	1.859×10^{-5}
IGARCH-ged	1.859×10^{-5}
GARCH-sged	1.860×10^{-5}
GARCH-ghyp	1.861×10^{-5}
GARCH-ged	1.861×10^{-5}
IGARCH-nig	1.862×10^{-5}
GARCH-nig	1.864×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.541
MSGARCH-norm	0.645
MSGARCH-snorm	0.649
GARCH-nig	0.764
GARCH-std	0.764
GARCH-sstd	0.764
GARCH-jsu	0.765
Regular HAR	0.765
GARCH-ghyp	0.766
GARCH-ged	0.766

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 4.972
EGARCH-jsu	− 4.921
EGARCH-nig	− 4.915
APARCH-ghyp	− 4.914
EGARCH-ghyp	− 4.913
APARCH-std	− 4.907
EGARCH-sged	− 4.903
EGARCH-ged	− 4.903
APARCH-std	− 4.900
IGARCH-ghyp	− 4.890

Table A.2
Loss function values for 2 days cumulative forecast.

(a) MAE loss values

Model	MAE
Logarithmic HAR	3.322×10^{-3}
Regular HAR	3.500×10^{-3}
EGARCH-norm	3.862×10^{-3}
MSGARCH-sged	3.875×10^{-3}
EGARCH-snorm	3.880×10^{-3}
MSGARCH-snorm	3.933×10^{-3}
GARCH-ghyp	3.934×10^{-3}
GARCH-sged	3.936×10^{-3}
GARCH-ged	3.937×10^{-3}
GARCH-nig	3.937×10^{-3}

(b) MSE loss values

Model	MSE
Regular HAR	4.880×10^{-5}
Logarithmic HAR	5.234×10^{-5}
IGARCH-ghyp	5.955×10^{-5}
IGARCH-sged	5.956×10^{-5}
IGARCH-ged	5.959×10^{-5}
GARCH-ghyp	5.961×10^{-5}
GARCH-sged	5.961×10^{-5}
GARCH-ged	5.964×10^{-5}
IGARCH-nig	5.966×10^{-5}
GARCH-nig	5.972×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.469
MSGARCH-norm	0.565
MSGARCH-snorm	0.567
GARCH-nig	0.665
GARCH-std	0.666
GARCH-jsu	0.666
GARCH-sstd	0.666
GARCH-ghyp	0.667
MSGARCH-sged	0.667
GARCH-ged	0.667

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 4.238
EGARCH-jsu	− 4.192
EGARCH-nig	− 4.182
EGARCH-ghyp	− 4.180
APARCH-sstd	− 4.176
APARCH-ghyp	− 4.174
APARCH-std	− 4.174
EGARCH-std	− 4.170
EGARCH-sstd	− 4.169
EGARCH-sged	− 4.163

Table A.3
Loss function values for 5 days cumulative forecast.

(a) MAE loss values

Model	MAE
Regular HAR	8.381×10^{-3}
Logarithmic HAR	8.416×10^{-3}
EGARCH-norm	9.424×10^{-3}
EGARCH-snorm	9.434×10^{-3}
MSGARCH-sged	9.520×10^{-3}
EGARCH-sged	9.537×10^{-3}
EGARCH-ged	9.541×10^{-3}
GARCH-ghyp	9.788×10^{-3}
GARCH-nig	9.794×10^{-3}
GARCH-sged	9.797×10^{-3}

(b) MSE loss values

Model	MSE
Regular HAR	2.358×10^{-4}
Logarithmic HAR	2.615×10^{-4}
EGARCH-nig	2.828×10^{-4}
EGARCH-ghyp	2.835×10^{-4}
EGARCH-sged	2.858×10^{-4}
EGARCH-ged	2.861×10^{-4}
EGARCH-jsu	2.863×10^{-4}
MSGARCH-sged	2.867×10^{-4}
IGARCH-ghyp	2.879×10^{-4}
GARCH-ghyp	2.881×10^{-4}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.427
MSGARCH-norm	0.517
MSGARCH-snorm	0.518
Regular HAR	0.566
EGARCH-snorm	0.595
EGARCH-norm	0.603
EGARCH-ged	0.606
EGARCH-sged	0.607
MSGARCH-sged	0.613
GARCH-nig	0.622

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 3.244
EGARCH-jsu	− 3.207
EGARCH-std	− 3.207
EGARCH-sstd	− 3.206
APARCH-sstd	− 3.198
APARCH-std	− 3.196
EGARCH-nig	− 3.186
APARCH-ghyp	− 3.185
EGARCH-ghyp	− 3.182
MSGARCH-std	− 3.171

Table A.4
Loss function values for 10 days cumulative forecast.

(a) MAE loss values

Model	MAE
Regular HAR	1.628×10^{-2}
Logarithmic HAR	1.718×10^{-2}
EGARCH-sged	1.842×10^{-2}
EGARCH-ged	1.842×10^{-2}
EGARCH-snorm	1.848×10^{-2}
EGARCH-norm	1.856×10^{-2}
EGARCH-ghyp	1.870×10^{-2}
EGARCH-nig	1.872×10^{-2}
MSGARCH-sged	1.879×10^{-2}
EGARCH-jsu	1.926×10^{-2}

(b) MSE loss values

Model	MSE
Regular HAR	7.093×10^{-4}
EGARCH-jsu	8.463×10^{-4}
EGARCH-nig	8.518×10^{-4}
EGARCH-ghyp	8.568×10^{-4}
Logarithmic HAR	8.570×10^{-4}
MSGARCH-sged	8.879×10^{-4}
EGARCH-sged	8.905×10^{-4}
EGARCH-ged	8.918×10^{-4}
EGARCH-norm	9.000×10^{-4}
EGARCH-snorm	9.037×10^{-4}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.424
Regular HAR	0.493
MSGARCH-norm	0.493
MSGARCH-snorm	0.495
EGARCH-ged	0.533
EGARCH-sged	0.534
EGARCH-snorm	0.561
EGARCH-norm	0.573
MSGARCH-sged	0.577
MSGARCH-ged	0.578

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 2.494
EGARCH-std	− 2.487
EGARCH-sstd	− 2.487
APARCH-sstd	− 2.80
APARCH-std	− 2.468
EGARCH-jsu	− 2.464
APARCH-ghyp	− 2.488
MSGARCH-std	− 2.446
MSGARCH-sstd	− 2.438
EGARCH-nig	− 2.431

Table A.5
Loss function values for 15 days cumulative forecast.

(a) MAE loss values

Model	MAE
Regular HAR	2.367×10^{-2}
Logarithmic HAR	2.587×10^{-2}
EGARCH-jsu	2.642×10^{-2}
EGARCH-nig	2.646×10^{-2}
EGARCH-ghyp	2.656×10^{-2}
EGARCH-sged	2.702×10^{-2}
EGARCH-ged	2.703×10^{-2}
APARCH-ged	2.711×10^{-2}
GJR-GARCH-ghyp	2.712×10^{-2}
GJR-GARCH-sged	2.714×10^{-2}

(b) MSE loss values

Model	MSE
Regular HAR	1.313×10^{-3}
EGARCH-jsu	1.447×10^{-3}
EGARCH-nig	1.517×10^{-3}
EGARCH-ghyp	1.537×10^{-3}
MSGARCH-sstd	1.607×10^{-3}
GARCH-ghyp	1.611×10^{-3}
IGARCH-ghyp	1.613×10^{-3}
APARCH-ged	1.613×10^{-3}
GARCH-sged	1.613×10^{-3}
GARCH-ged	1.614×10^{-3}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.403
Regular HAR	0.447
MSGARCH-norm	0.465
MSGARCH-snorm	0.472
EGARCH-sged	0.480
EGARCH-ged	0.481
EGARCH-ghyp	0.515
EGARCH-nig	0.520
EGARCH-snorm	0.525
EGARCH-norm	0.536

(d) QLIKE loss values

Model	QLIKE
EGARCH-std	− 2.077
EGARCH-sstd	− 2.077
APARCH-sstd	− 2.071
Regular HAR	− 2.060
APARCH-std	− 2.049
EGARCH-jsu	− 2.037
APARCH-jsu	− 2.036
MSGARCH-sstd	− 2.023
MSGARCH-std	− 2.017
IGARCH-sstd	− 2.004

Table A.6
Loss function values for 1st day ahead daily forecast.

(a) MAE loss values

Model	MAE
Logarithmic HAR	1.711×10^{-3}
Regular HAR	1.831×10^{-3}
EGARCH-nom	1.973×10^{-3}
MSGARCH-sged	1.974×10^{-3}
MSGARCH-snorm	1.981×10^{-3}
EGARCH-snorm	1.981×10^{-3}
MSGARCH-norm	1.996×10^{-3}
GARCH-ghyp	2.002×10^{-3}
GARCH-ged	2.002×10^{-3}
GARCH-sged	2.003×10^{-3}

(b) MSE loss values

Model	MSE
Regular HAR	1.563×10^{-5}
Logarithmic HAR	1.616×10^{-5}
IGARCH-sged	1.859×10^{-5}
IGARCH-ghyp	1.859×10^{-5}
IGARCH-ged	1.859×10^{-5}
GARCH-sged	1.860×10^{-5}
GARCH-ghyp	1.861×10^{-5}
GARCH-ged	1.861×10^{-5}
IGARCH-nig	1.862×10^{-5}
GARCH-nig	1.864×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.541
MSGARCH-norm	0.645
MSGARCH-snorm	0.649
GARCH-nig	0.764
GARCH-std	0.764
GARCH-sstd	0.764
GARCH-jsu	0.765
Regular HAR	0.765
GARCH-ghyp	0.766
GARCH-ged	0.766

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 4.972
EGARCH-jsu	− 4.921
EGARCH-nig	− 4.915
APARCH-ghyp	− 4.914
EGARCH-ghyp	− 4.913
APARCH-std	− 4.907
EGARCH-sged	− 4.903
EGARCH-ged	− 4.903
APARCH-sstd	− 4.900
IGARCH-ghyp	− 4.980

Table A.7
Loss function values for 2nd day ahead daily forecast.

(a) MAE loss values

Model	MAE
Logarithmic HAR	1.936×10^{-3}
Regular HAR	2.019×10^{-3}
MSGARCH-snorm	2.127×10^{-3}
EGARCH-snorm	2.128×10^{-3}
MSGARCH-ged	2.129×10^{-3}
EGARCH-norm	2.131×10^{-3}
MSGARCH-norm	2.156×10^{-3}
MSGARCH-sstd	2.166×10^{-3}
EGARCH-sged	2.191×10^{-3}
EGARCH-ged	2.191×10^{-3}

(b) MSE loss values

Model	MSE
Regular HAR	1.871×10^{-5}
Logarithmic HAR	1.988×10^{-5}
EGARCH-nig	2.112×10^{-5}
EGARCH-ghyp	2.115×10^{-5}
EGARCH-sged	2.116×10^{-5}
EGARCH-ged	2.118×10^{-5}
IGARCH-ghyp	2.125×10^{-5}
IGARCH-sged	2.126×10^{-5}
GARCH-ghyp	2.126×10^{-5}
IGARCH-ged	2.126×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.615
MSGARCH-snorm	0.724
MSGARCH-norm	0.724
MSGARCH-ged	0.862
GJR-GARCH-nig	0.866
GJR-GARCH-jsu	0.868
GJR-GARCH-ghyp	0.868
GJR-GARCH-sstd	0.868
GJR-GARCH-std	0.870
GJR-GARCH-sged	0.871

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 4.893
EGARCH-jsu	− 4.855
APARCH-sstd	− 4.844
APARCH-std	− 4.843
EGARCH-std	− 4.843
EGARCH-sstd	− 4.842
EGARCH-nig	− 4.840
EGARCH-ghyp	− 4.837
APARCH-ghyp	− 4.829
EGARCH-sged	− 4.812

Table A.8
Loss function values for 5th day ahead daily forecast.

(a) MAE loss values

Model	MAE
Logarithmic HAR	2.145×10^{-3}
Regular HAR	2.210×10^{-3}
MSGARCH-norm	2.286×10^{-3}
MSGARCH-snorm	2.298×10^{-3}
EGARCH-snorm	2.329×10^{-3}
EGARCH-sged	2.333×10^{-3}
EGARCH-ged	2.335×10^{-3}
EGARCH-norm	2.344×10^{-3}
MSGARCH-sged	2.349×10^{-3}
MSGARCH-ged	2.362×10^{-3}

(b) MSE loss values

Model	MSE
Regular HAR	2.154×10^{-5}
MSGARCH-sstd	2.165×10^{-5}
MSGARCH-std	2.260×10^{-5}
Logarithmic HAR	2.301×10^{-5}
MSGARCH-sged	2.308×10^{-5}
EGARCH-nig	2.317×10^{-5}
EGARCH-ghyp	2.320×10^{-5}
EGARCH-jsu	2.323×10^{-5}
EGARCH-norm	2.339×10^{-5}
EGARCH-sged	2.341×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.695
MSGARCH-norm	0.810
MSGARCH-snorm	0.824
EGARCH-sged	0.973
EGARCH-ged	0.974
EGARCH-snorm	1.035
MSGARCH-ged	1.035
Regular HAR	1.036
GJR-GARCH-nig	1.052
GJR-GARCH-jsu	1.052

(d) QLIKE loss values

Model	QLIKE
Regular HAR	− 4.789
EGARCH-std	− 4.779
EGARCH-sstd	− 4.778
MSGARCH-sstd	− 4.774
APARCH-sstd	− 4.773
APARCH-std	− 4.762
EGARCH-jsu	− 4.754
MSGARCH-std	− 4.748
APARCH-ghyp	− 4.735
EGARCH-nig	− 4.720

Table A.9
Loss function values for 10th day ahead daily forecast.

(a) MAE loss values

Model	MAE
Logarithmic HAR	2.250×10^{-3}
Regular HAR	2.273×10^{-3}
EGARCH-sged	2.340×10^{-3}
EGARCH-ged	2.342×10^{-3}
EGARCH-ghyp	2.365×10^{-3}
EGARCH-nig	2.369×10^{-3}
EGARCH-snorm	2.372×10^{-3}
EGARCH-norm	2.388×10^{-3}
MSGARCH-norm	2.388×10^{-3}
MSGARCH-snorm	2.402×10^{-3}

(b) MSE loss values

Model	MSE
MSGARCH-sstd	2.234×10^{-5}
Regular HAR	2.236×10^{-5}
EGARCH-sju	2.287×10^{-5}
EGARCH-nig	2.331×10^{-5}
EGARCH-ghyp	2.341×10^{-5}
GJR-GARCH-ghyp	2.361×10^{-5}
GJR-GARCH-nig	2.362×10^{-5}
GJR-GARCH-sged	2.365×10^{-5}
GJR-GARCH-jsu	2.365×10^{-5}
GJR-GARCH-ged	2.367×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.735
MSGARCH-norm	0.835
MSGARCH-snorm	0.840
EGARCH-sged	0.955
EGARCH-ged	0.957
EGARCH-ghyp	1.059
MSGARCH-ged	1.061
EGARCH-nig	1.073
Regular HAR	1.086
MSGARCH-sged	1.125

(d) QLIKE loss values

Model	QLIKE
APARCH-std	− 4.794
EGARCH-std	− 4.788
EGARCH-sstd	− 4.788
APARCH-sstd	− 4.783
Regular HAR	− 4.756
APARCH-ghyp	− 4.745
EGARCH-jsu	− 4.728
GJR-GARCH-std	− 4.724
GJR-GARCH-sstd	− 4.724
GJR-GARCH-jsu	− 4.724

Table A.10
Loss function values for 15th day ahead daily forecast.

(a) MAE loss values

Model	MAE
Regular HAR	2.263×10^{-3}
Logarithmic HAR	2.288×10^{-3}
EGARCH-nig	2.310×10^{-3}
EGARCH-ghyp	2.313×10^{-3}
EGARCH-jsu	2.327×10^{-3}
EGARCH-sged	2.330×10^{-3}
EGARCH-ged	2.331×10^{-3}
EGARCH-snorm	2.379×10^{-3}
EGARCH-norm	2.394×10^{-3}
MSGARCH-sged	2.402×10^{-3}

(b) MSE loss values

Model	MSE
EGARCH-sstd	2.095×10^{-5}
EGARCH-std	2.105×10^{-5}
GJR-GARCH-ghyp	2.123×10^{-5}
GJR-GARCH-nig	2.123×10^{-5}
GJR-GARCH-jsu	2.125×10^{-5}
GJR-GARCH-sged	2.127×10^{-5}
GJR-GARCH-ged	2.128×10^{-5}
GJR-GARCH-std	2.129×10^{-5}
GJR-GARCH-sstd	2.129×10^{-5}
GARCH-ghyp	2.158×10^{-5}

(c) MAPE loss values

Model	MAPE
Logarithmic HAR	0.702
MSGARCH-norm	0.814
MSGARCH-snorm	0.817
EGARCH-sged	0.873
EGARCH-ged	0.875
EGARCH-ghyp	0.967
EGARCH-nig	0.980
MSGARCH-ged	1.054
Regular HAR	1.088
EGARCH-jsu	1.109

(d) QLIKE loss values

Model	QLIKE
APARCH-sstd	− 4.789
EGARCH-std	− 4.788
EGARCH-sstd	− 4.788
APARCH-std	− 4.769
GJR-GARCH-sstd	− 4.753
GJR-GARCH-std	− 4.752
GJR-GARCH-jsu	− 4.751
APARCH-ghyp	− 4.750
GJR-GARCH-nig	− 4.745
GJR-GARCH-ghyp	− 4.745

Table A.11
Superior distribution for each GARCH-type model considering the MSE loss function for forecasting cumulative variance on time horizons of 1, 2, 5, 10, and 15 days.

Forecast horizon	Model	Distribution
1 day	GARCH	sged
	EGARCH	norm
	IGARCH	sged
	GJR-GARCH	norm
	APARCH	ged
	MSGARCH	sstd
2 days	GARCH	ghyp
	EGARCH	nig
	IGARCH	ghyp
	GJR-GARCH	nig
	APARCH	ged
	MSGARCH	sged
5 days	GARCH	ghyp
	EGARCH	nig
	IGARCH	ghyp
	GJR-GARCH	norm
	APARCH	ged
	MSGARCH	sged
10 days	GARCH	ghyp
	EGARCH	jsu
	IGARCH	ghyp
	GJR-GARCH	sged
	APARCH	ged
	MSGARCH	sged
15 days	GARCH	ghyp
	EGARCH	nig
	IGARCH	ghyp
	GJR-GARCH	ghyp
	APARCH	ged
	MSGARCH	sstd

Table A.12
Loss function values when using 500 observations for model estimation and 500 out-of-sample observations.

	1 day		2 days		5 days		10 days		15 days	
Model	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Regular HAR	1.599×10^{-5}	−4.957	5.051×10^{-5}	−4.215	2.407×10^{-4}	3.219	7.164×10^{-4}	−2.467	1.317×10^{-3}	−2.026
Logarithmic HAR	1.608×10^{-5}	−4.899	5.173×10^{-5}	−4.155	2.537×10^{-4}	−3.137	8.083×10^{-4}	−2.334	1.557×10^{-3}	−1.855
EGARCH-std	18.47×10^{-5}	−4.753	73.82×10^{-5})	−4.036	46.49×10^{-4}	−3.077	187.7×10^{-4}	−2.347	42.11×10^{-3}	−1.930
EGARCH-sstd	106.0×10^{-5}	−4.649	429.6×10^{-5}	−3.931	278.5×10^{-4}	−2.975	1174×10^{-4}	−2.246	277.3×10^{-3}	−1.827
EGARCH-jsu	2.228×10^{-5}	−4.908	7.457×10^{-5}	−4.183	3.746×10^{-4}	−3.207	11.81×10^{-4}	−2.465	2.041×10^{-3}	−2.044

Table A.13
Loss function values when using 750 observations for model estimation and 500 out-of-sample observations.

	1 day		2 days		5 days		10 days		15 days	
Model	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Regular HAR	1.573×10^{-5}	−4.965	4.940×10^{-5}	−4.227	2.369×10^{-4}	−3.230	7.066×10^{-4}	−2.475	1.296×10^{-3}	−2.033
Logarithmic HAR	1.617×10^{-5}	−4.887	5.224×10^{-5}	−4.137	2.587×10^{-4}	−3.108	8.409×10^{-4}	−2.294	1.647×10^{-3}	−1.806
EGARCH-std	3.808×10^{-5}	−4.776	13.48×10^{-5}	−4.061	7.418×10^{-4}	−3.103	25.86×10^{-4}	−2.385	5.126×10^{-3}	−1.975
EGARCH-sstd	6.150×10^{-5}	−4.734	23.30×10^{-5}	−4.021	13.97×10^{-4}	−3.063	56.09×10^{-4}	−2.340	12.90×10^{-3}	−1.92
EGARCH-jsu	2.088×10^{-5}	−4.895	6.619×10^{-5}	−4.170	3.148×10^{-4}	−3.198	9.107×10^{-4}	−2.467	1.475×10^{-3}	−2.049

Table A.14
Loss function values when using 1000 observations for model estimation and 500 out-of-sample observations.

	1 day		2 days		5 days		10 days		15 days	
Model	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Regular HAR	1.559×10^{-5}	– 4.970	4.867×10^{-5}	– 4.235	2.352×10^{-4}	– 3.239	7.050×10^{-4}	– 2.487	1.293×10^{-3}	– 2.049
Logarithmic HAR	1.622×10^{-5}	– 4.885	5.257×10^{-5}	– 4.135	2.625×10^{-4}	– 3.107	8.619×10^{-4}	– 2.296	1.715×10^{-3}	– 1.812
EGARCH-std	2.820×10^{-5}	– 4.839	9.656×10^{-5}	– 4.124	4.954×10^{-4}	– 3.173	15.58×10^{-4}	– 2.460	2.755×10^{-3}	– 2.047
EGARCH-sstd	2.890×10^{-5}	– 4.835	9.656×10^{-5}	– 4.120	5.152×10^{-4}	– 3.169	16.42×10^{-4}	– 2.456	2.938×10^{-3}	– 2.043
EGARCH-jsu	2.046×10^{-5}	– 4.899	6.506×10^{-5}	– 4.175	3.022×10^{-4}	– 3.203	8.752×10^{-4}	– 2.472	1.469×10^{-3}	– 2.048

Table A.15
Loss function values when using 1220 observations for model estimation and 500 out-of-sample observations.

	1 day		2 days		5 days		10 days		15 days	
Model	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Regular HAR	1.563×10^{-5}	– 4.972	4.880×10^{-5}	– 4.238	2.358×10^{-4}	– 3.244	7.093×10^{-4}	– 2.494	1.313×10^{-3}	– 2.060
Logarithmic HAR	1.616×10^{-5}	– 4.887	5.234×10^{-5}	– 4.139	2.615×10^{-4}	– 3.117	8.570×10^{-4}	– 2.316	1.702×10^{-3}	– 1.843
EGARCH-std	2.284×10^{-5}	– 4.887	7.590×10^{-5}	– 4.170	3.749×10^{-4}	– 3.207	11.09×10^{-4}	– 2.487	1.767×10^{-3}	– 2.077
EGARCH-sstd	2.300×10^{-5}	– 4.886	7.661×10^{-5}	– 4.169	3.796×10^{-4}	– 3.206	11.25×10^{-4}	– 2.487	1.788×10^{-3}	– 2.077
EGARCH-jsu	1.924×10^{-5}	– 4.921	6.110×10^{-5}	– 4.192	2.863×10^{-4}	– 3.207	8.463×10^{-4}	– 2.464	1.447×10^{-3}	– 2.037

Table A.16
Loss function values when using 750 observation for model estimation and 970 out-of-sample observations.

	1 day		2 days		5 days		10 days		15 days	
Model	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
Regular HAR	1.180×10^{-5}	– 5.532	3.447×10^{-5}	– 4.778	1.545×10^{-4}	– 3.751	4.502×10^{-4}	– 2.994	8.286×10^{-4}	– 2.554
Logarithmic HAR	1.150×10^{-5}	– 5.515	3.466×10^{-5}	– 4.753	1.616×10^{-4}	– 3.664	5.117×10^{-4}	– 2.856	9.971×10^{-4}	– 2.373
EGARCH-std	2.302×10^{-5}	– 5.453	7.825×10^{-5}	– 4.728	4.153×10^{-4}	– 3.740	14.21×10^{-4}	– 3.017	28.01×10^{-4}	– 2.587
EGARCH-sstd	3.508×10^{-5}	– 5.433	12.83×10^{-5}	– 4.710	7.520×10^{-4}	– 3.719	29.65×10^{-4}	– 2.995	67.57×10^{-4}	– 2.563
EGARCH-jsu	1.386×10^{-5}	– 5.540	4.203×10^{-5}	– 4.802	1.883×10^{-4}	– 3.788	5.380×10^{-4}	– 3.054	8.913×10^{-4}	– 2.614

References

Aalborg, H.A., Molnár, P., de Vries, J.E., 2018. What can explain the price, volatility and trading volume of bitcoin? Finance Res. Lett. https://doi.org/10.1016/j.frl.2018.08.010, 1544-6123.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Am. Stat. Assoc. 96 (453), 42–55.

Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: sample frequency vis-a-vis forecast horizon. J. Emp. Finance 6 (5), 457–477.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometr. 169 (1), 75–93.

Aysan, A.F., Demir, E., Gozgor, G., Lau, C.K.M., 2019. Effects of the geopolitical risks on bitcoin returns and volatility. Res. Int. Business Finance 47, 511–518.

Baek, C., Elbeck, M., 2015. Bitcoins as an investment or speculative vehicle? A first look. Appl. Econ. Lett. 22 (1), 30–34.

Balcilar, M., Bouri, E., Gupta, R., Roubaud, D., 2017. Can volume predict Bitcoin returns and volatility? A quantiles-based approach. Econ. Modell. 64, 74–81.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econometr. 2 (1), 1–37.

Baumöhl, E., 2019. Are cryptocurrencies connected to forex? a quantile cross-spectral approach. Finance Res. Lett. 29, 363–372.

Baur, D.G., Dimpfl, T., 2018. Excess Volatility as an Impediment for a Digital Currency.

Bauwens, L., Preminger, A., Rombouts, J.V., 2010. Theory and inference for a Markov switching GARCH model. Econometr. J. 13 (2), 218–244.

Beneki, C., Koulis, A., Kyriazis, N.A., Papadamou, S., 2019. Investigating volatility transmission and hedging properties between bitcoin and ethereum. Res. Int. Business Finance 48, 219–227.
Birkelund, O.H., Haugom, E., Molnár, P., Opdal, M., Westgaard, S., 2015. A comparison of implied and realized volatility in the Nordic power forward market. Energy Econ. 48, 288–294.
Blau, B.M., 2017. Price dynamics and speculative trading in bitcoin. Res. Int. Business Finance 41, 493–499.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometr. 31 (3), 307–327.
Bouoiyour, J., Selmi, R., 2015a. Bitcoin Price: Is It Really That New Round of Volatility Can Be On Way? (MPRA Paper No. 65580). University Library of Munich, Germany.
Bouoiyour, J., Selmi, R., 2015b. What does Bitcoin look like? Ann. Econ. Finance 16, 2.
Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I., 2017. On the hedge and safe haven properties of Bitcoin: is it really more than a diversifier? Finance Res. Lett. 20, 192–198.
Brandvold, M., Molnár, P., Vagstad, K., Valstad, O.C.A., 2015. Price discovery on bitcoin exchanges. J. Int. Financ. Markets Inst. Money 36, 18–35.
Broto, C., Ruiz, E., 2004. Estimation methods for stochastic volatility models: a survey. J. Econ. Surv. 18 (5), 613–649.
Burnie, A., 2018. Exploring the Interconnectedness of Cryptocurrencies using Correlation Networks. arXiv:1806.06632.
Burniske, C., White, A., 2017. Bitcoin: Ringing the Bell for A New Asset Class, Ark Invest. https://research.ark-invest.com/hubfs/1_Download_Files_ARK-Invest/White_Papers/Bitcoin-Ringing-The-Bell-For-A-New-Asset-Class.pdf.
Caporale, G.M., Zekokh, T., 2019. Modelling volatility of cryptocurrencies using Markov-switching Garch models. Res. Int. Business Finance 48, 143–155.
Catania, L., Grassi, S., Ravazzolo, F., 2018. Predicting the Volatility of Cryptocurrency Time–Series (Working Papers No. No 3/2018). Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.
Charles, A., Darné, O., 2018. Volatility Estimation for Bitcoin: Replication and Extension. Manuscript.
Christoffersen, P., 2011. Elements of Financial Risk Management, 2nd ed. Academic Press, pp. 94–98.
Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Volatility dynamics for the S&P500: evidence from realized volatility, daily returns, and option prices. Rev. Financ. Stud. 23 (8), 3141–3189.
Chu, J., Chan, S., Nadarajah, S., Osterrieder, J., 2017. GARCH modelling of cryptocurrencies. J. Risk Financ. Manag. 10 (4), 17.
CoinMarketCap, 2018. Retrieved 18.10.2018. https://coinmarketcap.com/all/views/all/.
Corbet, S., Lucey, B., Yarovaya, L., 2018. Datestamping the bitcoin and ethereum bubbles. Finance Res. Lett. 26, 81–88.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econometr. 7 (2), 174–196.
Corsi, F., Renò, R., 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. J. Business Econ. Stat. 30 (3), 368–380.
Dimpfl, T., Jank, S., 2016. Can internet search queries help to predict stock market volatility? Eur. Financ. Manag. 22 (2), 171–192.
Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. J. Emp. Finance 1 (1), 83–106.
Dyhrberg, A.H., 2016. Bitcoin, gold and the dollar – a GARCH volatility analysis. Finance Res. Lett. 16, 85–92.
Enoksen, F., Landsnes, C.J., Lučivjanská, K., Molnár, P., 2020. Understanding risk of bubbles in cryptocurrencies. J. Econ. Behav. Organ. 176, 129–144.
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. J. Finance 48 (5), 1779–1801.
Haas, M., Mittnik, S., Paolella, M.S., 2004. A new approach to Markov-switching GARCH models. J. Financ. Econometr. 2 (4), 493–530.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. J. Bank. Finance 47, 1–14.
Hillebrand, E., 2005. Neglecting parameter changes in GARCH models. J. Econometr. 129 (1-2), 121–138.
Katsiampa, P., 2017. Volatility estimation for Bitcoin: a comparison of GARCH models. Econ. Lett. 158, 3–6.
Katsiampa, P., 2019. An empirical investigation of volatility dynamics in the cryptocurrency market. Res. Int. Business Finance 50, 322–335.
Kirk, D., 2014. Retrieved 09.10.2018. https://www.tech-recipes.com/rx/48517/cryptocurrency-what-is-a-fork/.
Köchling, G., Schmidtke, P., Posch, P.N., 2020. Volatility forecasting accuracy for bitcoin. Econ. Lett. 191, 108836.
Kristoufek, L., 2013. Bitcoin meets Google trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. Sci. Rep. 3, 3415.
Kroeger, A., Sarkar, A., 2017. Law of One Bitcoin Price? Federal Reserve Bank of Philadelphia.
Kurka, J., 2016. BT Does Bitcoin Have Potential To Co-Function with Fiat Money? (Unpublished Master's Thesis).
Letra, I.J.S., 2016. BT What Drives Cryptocurrency Value? A Volatility and Predictability Analysis (Unpublished Doctoral Dissertation).
Li, X., 2017. BT Evaluating Volatility Forecasting Performance at Different Horizons (Unpublished master's thesis).
Liu, C., Maheu, J.M., 2008. Are there structural breaks in realized volatility? J. Financ. Econometr. 6 (3), 326–360.
Lord, R., Koekkoek, R., Dijk, D.V., 2010. A comparison of biased simulation schemes for stochastic volatility models. Quant. Finance 10 (2), 177–194.
Lyócsa, Š., Molnár, P., 2016. Volatility forecasting of strategically linked commodity ETFs: gold-silver. Quant. Finance 16 (12), 1809–1822.
Lyócsa, Š., Molnár, P., Plíhal, T., Širaňová, M., 2020. Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. J. Econ. Dyn. Control 119, 103980.
Ma, D., Tanizaki, H., 2019. The day-of-the-week effect on bitcoin return and volatility. Res. Int. Business Finance 49, 127–136.
Mikosch, T., Stărică, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Rev. Econ. Stat. 86 (1), 378–390.
Naimy, V.Y., Hayek, M.R., 2018. Modelling and predicting the Bitcoin volatility using GARCH models. Int. J. Math. Modell. Numer. Optim. 8 (3), 197–215.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica: J. Econometr. Soc. 347–370.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econometr. 160 (1), 246–256.
Pichl, L., Kaizoji, T., 2017. Volatility Analysis of Bitcoin Price Time Series.
Poon, S.-H., Granger, C.W., 2003. Forecasting volatility in financial markets: A review. J. Econ. Lit. 41 (2), 478–539.
Ruiz, E., 1994. Quasi-maximum likelihood estimation of stochastic volatility models. J. Econometr. 63 (1), 289–306.
Schwarz, G., et al., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.
Sedgwick, K., 2018. Retrieved 18.10.2018. https://news.bitcoin.com/the-number-of-cryptocurrency-exchanges-has-exploded/.
Shi, Y., Tiwari, A.K., Gozgor, G., Lu, Z., 2020. Correlations among cryptocurrencies: evidence from multivariate factor stochastic volatility model. Res. Int. Business Finance 101231.
Shin, L., 2016. Retrieved 22.10.2018. https://www.forbes.com/sites/laurashin/2016/04/25/7886/#55113c596056.
Stavroyiannis, S., Babalos, V., 2017. Dynamic Properties of the Bitcoin and the US Market.
Trucíos, C., 2019. Forecasting bitcoin risk measures: a robust approach. Int. J. Forecast. 35 (3), 836–847.
Trucíos, C., Tiwari, A.K., Alqahtani, F., 2020. Value-at-risk and expected shortfall in cryptocurrencies' portfolio: a vine copula-based approach. Appl. Econ. 52 (24), 2580–2593.
Wang, P., Li, X., Shen, D., Zhang, W., 2020. How does economic policy uncertainty affect the bitcoin market? Res. Int. Business Finance 101234.
Yu, M., 2019. Forecasting bitcoin volatility: the role of leverage effect and uncertainty. Physica A: Stat. Mech. Appl. 533, 120707.