DOI: 10.1111/eufm.12254

ORIGINAL ARTICLE

EUROPEAN FINANCIAL MANAGEMENT WILEY

Forecasting the volatility of Bitcoin: The importance of jumps and structural breaks

Dehua Shen[1] | **Andrew Urquhart**[2] | **Pengfei Wang**[1]

[1]College of Management and Economics, Tianjin University, Tianjin, China

[2]ICMA Centre, Henley Business School, University of Reading, Reading, UK

Correspondence
Pengfei Wang, College of Management and Economics, Tianjin University, 92 Weijin Road, Tianjin 300072, China.
Email: pengfeiwang@tju.edu.cn

Abstract

This paper studies the volatility of Bitcoin and determines the importance of jumps and structural breaks in forecasting volatility. We show the importance of the decomposition of realized variance in the in-sample regressions using 18 competing heterogeneous autoregressive (HAR) models. In the out-of-sample setting, we find that the HARQ-F-J model is the superior model, indicating the importance of the temporal variation and squared jump components at different time horizons. We also show that HAR models with structural breaks outperform models without structural breaks across all forecasting horizons. Our results are robust to an alternative jump estimator and estimation method.

KEYWORDS

bitcoin, jumps, realized volatility, structural breaks, volatility forecasting

JEL CLASSIFICATION

C53, G15, G17

1 | INTRODUCTION

Bitcoin has received a great deal of attention since it was first proposed by Nakamoto (2008), and this attention has come from the media, governments, regulators, as well as from investors, who have been attracted to Bitcoin by its huge increase in price during 2017. However, this surge in price has

We wish to thank John Doukas (the Editor) and an anonymous referee for their insightful and constructive suggestions. We are grateful to Chardin Wese Simen for his valuable comments on earlier versions of this paper. This work is supported by the National Natural Science Foundation of China (71701150, 71790594, and U1811462). Any remaining errors are our own.

been accompanied by huge volatility and uncertainty regarding the future price path of this popular cryptocurrency. There is growing evidence that Bitcoin offers substantial diversification to investors when included in portfolios (Kajtazi & Moro, 2019; Platanakis & Urquhart, 2019) and that technical trading rules generate significant returns to investors (Hudson & Urquhart, 2019). Therefore, forecasting the volatility of Bitcoin is of great interest, and this paper provides a comprehensive overview of the forecasting ability of various time-series models derived from the innovative heterogeneous autoregressive (HAR) specification of Corsi (2009). We consider 18 HAR models, and the analysis is conducted in sample, and, more importantly, out of sample for Bitcoin from January 2012 to September 2018. Our results show that the inclusion of jumps is important when forecasting Bitcoin volatility at all forecasting horizons. Specifically, we find that the HARQ-F-J model provides the best out-of-sample forecast of volatility for a 1-day horizon, indicating the importance of the temporal variation and squared jump components at different horizons. For the 1-week and 1-month forecast horizons, we find that a number of models that include jumps are superior to models without jumps. Also, we find that the inclusion of structural breaks in each HAR model improves the forecasting ability of these models when considering forecast horizons of 1 day, 1 week, and 1 month. Therefore, our results indicate the importance of the temporal variation and squared jump components, separated at different horizons, as well as structural breaks, in forecasting Bitcoin volatility through competing HAR models.

Since the availability of high-frequency data has become more common, there is ample evidence of the economic value of forecasting volatility using intraday data. Most studies find that simple autoregressive structures such as HAR models provide much better forecasting ability than generalized autoregressive conditional heteroskedasticity (GARCH)-type models that employ daily data (see, for instance, Andersen & Bollerslev, 1997, 1998; Andersen, Bollerslev, Diebold, & Ebens, 2001; Andersen, Bollerslev, Diebold, & Labys, 2003; Giot & Laurent, 2007; Koopman, Jungbacker, & Hol, 2005).[1] This improvement comes from the fact that GARCH models employ daily data while HAR models are able to capture more information contained in intraday data.

The literature on cryptocurrencies is growing, with many papers reporting the inefficiency of Bitcoin (Khuntia & Pattanayak, 2018; Urquhart, 2016; Nadarajah and Chu, 2017; Tiwari, Jana, Das, & Roubaud, 2018), the hedging and diversification benefits (Borri, 2019; Bouri, Molnár, Azzi, Roubaud, & Hagfors, 2017; Corbet, Meegan, Larkin, Lucey, & Yarovaya, 2018; Urquhart & Zhang, 2018), the existence of bubbles (Cheah & Fry, 2015; Corbet, Lucey, & Yarovaya, 2018), investor attention (Shen, Urquhart, & Wang, 2019; Urquhart, 2018), and the behavior of Bitcoin returns (Corbet & Katsiampa, 2018; Phillip, Chen, & Peiris, 2018; Urquhart, 2017; Katsiampa, 2018).[2] However, there is limited literature examining the volatility dynamics of Bitcoin, with Katsiampa (2017) the first to explore the optimal conditional heteroskedasticity model with regard to goodness of fit to Bitcoin and finding that an autoregressive component GARCH (AR-CGARCH) model is the most appropriate, indicating both the short- and long-run component of the conditional variance. Chaim and Laurini (2018) show that jumps in volatility are permanent in Bitcoin, while jumps in returns are contemporaneous. They also show that large jumps in mean returns are all negative and associated with hacks and forks. Catania, Grassi, and Ravazzolo (2019) compare the abilities of several alternative univariate and multivariate models to predictor cryptocurrencies and show large, statistically significant improvements in the point forecasting of Bitcoin when using combinations of univariate models, while Katsiampa, Corbet, and Lucey (2019) show strong interdependencies between cryptocurrency volatilities and that time-varying conditional correlations of volatility exist between cryptocurrencies. Gronwald (2019) shows

[1]Andersen, Bollerslev, Christoffersen, and Diebold (2006) provides an excellent survey.

[2]See Corbet, Lucey, Urquhart, and Yarovaya (2019) for a recent review of the empirical literature on cryptocurrencies.

that Bitcoin price dynamics is particularly influenced by extreme price movements, more so than in the markets for crude oil or gold, which is possibly a result of the immaturity of the market. Kalyas, Papakyriakou, Sakkas, and Urquhart (2019) show that the Bitcoin crash risk is lower when economic policy uncertainty is high, indicating the hedging ability of Bitcoin against economic policy uncertainty.

All of the previously mentioned studies employ daily data, but there is growing evidence that high-frequency data are useful in predicting future volatility, especially the decomposition between the continuous and the jump component, as well as the separation between negative and positive intraday returns. Jumps have a long history in finance and have traditionally been estimated from daily data (see, for example, Andersen, Benzoni, & Lund, 2002; Eraker, 2004; Eraker, Johannes, & Polson, 2003). However, given the upsurge in the availability of high-frequency data, more and more studies have gained insights from the intraday behavior of volatility. Andersen and Bollerslev (1998) were the first to use intraday data to measure volatility when they proposed realized volatility (RV), and since then high-frequency data applications have developed rapidly with a strong focus on forecasting financial markets. In more recent work, Corsi (2009) proposes the heterogeneous autoregressive model of realized volatility (HAR-RV) and shows that this model is markedly better than the traditional GARCH model and the autoregressive fractionally integrated moving average model of realized volatilty (ARFIMA-RV) at forecasting volatility. Since then, many studies have examined the use of the HAR-RV model as well as modifications to the model, such as the HAR-RV-J, HAR-RV-CJ, and HAR-RSV models. Many studies have shown that these HAR-type models offer better forecasting ability than GARCH, stochastic volatility (SV), and ARFIMA-RV models (see, for example, Andersen, Bollerslev, & Meddahi, 2011; Çelik & Ergin, 2014; Sévi, 2014), indicating the importance of intraday volatility in forecasting future volatility.

This study provides a comprehensive empirical analysis of the forecasting accuracy of various time series models derived from the HAR specification proposed by Corsi (2009). We study 18 competing HAR models that forecast volatility from 1 day to 2 months, where we conduct in-sample and, more importantly, out-of-sample analysis of Bitcoin from January 1, 2012 to September 3, 2018. Specifically, we find that the HARQ-F-J model provides the best out-of-sample forecast of volatility for 1-day, 1-week, and 1-month horizons, indicating the importance of the separating temporal variation and squared jump components at different horizons. Also, we find that the inclusion of structural breaks in each HAR model improves the forecasting ability of these models when considering forecast horizons of 1 day, 1 week, and 1 month. Therefore, our results indicate the importance of the temporal variation and squared jump components, separated at different horizons, as well as structural breaks, in forecasting Bitcoin volatility through competing HAR models. We conduct a number of robustness checks where we employ the novel jump-robust estimator of Andersen, Dobrev, and Schaumburg (2012), as well as using a weighted least squares (WLS) approach rather than ordinary least squares (OLS), and our findings remain consistent. Therefore, we find that jumps are quite common in Bitcoin and that jumps as well as structural breaks are important components when forecasting the volatility of Bitcoin.

Our findings are consistent with the results from Scaillet, Treccani, and Trevisan (2018) in that jumps are frequent events in the Bitcoin market, and therefore in our paper the inclusion of jumps improves the forecasting power of our HAR models. As argued in Scaillet et al. (2018), the presence of “whales” (big-money Bitcoin players who show their hand in the Bitcoin market) can have a large impact on the price if they make such a big order in the market. These large impacts subsequently cause the jumps in the Bitcoin price, and the likelihood of “whales” in the Bitcoin market is much higher than in equity markets. Therefore, Bitcoin (and other cryptocurrencies) are more susceptible to jumps than more mature markets. Also, the Bitcoin market trades 24 hours a day, seven days a week, which enables traders from all over the world to trade at whatever time is suitable for them. Consequently, unlike international stock markets with certain trading times, not all Bitcoin traders are active in the market at the same time, and this can lead to certain traders being able to have a large price impact when liquidity in the market is fairly low. We also find that the inclusion of structural breaks in our HAR models produces better forecasts of Bitcoin volatility. Similar to jumps, structural breaks are inherent in the Bitcoin market due to "whales" and the market structure of Bitcoin. This can be clearly seen in Figure 2, where we report the time-series graph of the price of Bitcoin during our sample period. In a more mature, less volatile market, the inclusion of structural breaks in HAR models may not be as useful, but in the immature Bitcoin market we find that they are very useful. With Bitcoin futures introduced in December 2017, investors are now able to hedge Bitcoin much more easily, which may prevent the huge bubbles experienced in 2015 and 2016 (Shiller, 2017) and reduce the number of jumps and structural breaks in the Bitcoin price.

The remainder of the paper is organized as follows. Section 2 introduces the data employed. Section 3 provides the methodology, where we explain how we measure RV, jumps, as well the 18 different HAR models employed in this study. Section 4 presents the empirical results. Section 5 provides some robustness checks. Finally, Section 6 provides a summary of the findings and concludes.

2 | DATA

We obtain Bitcoin tick data from www.bitcoincharts.com and focus on Bitstamp from January 1, 2012 to September 30, 2018 since it is one of the longest-standing Bitcoin exchanges and thus provides sufficient liquidity.[3] The sample period is chosen due to data availability, as before this date Bitcoin lacked intraday liquidity. Bitcoin trades 24 hours a day, seven days a week, and therefore we have a continuous time series throughout our sample period.[4] Before computing the variances, a sampling frequency needs to be chosen. It is well documented in the literature that microstructure noise (due to bid–ask spreads, non-synchronous trading and price discreteness) may impact on the realized variance estimator at high frequency (see, for example, Andersen & Bollerslev, 1997, 1998). To deal with this issue, we plot the volatility signature in Figure 1, which stabilizes around the 5-minute frequency and therefore we use this as our sampling frequency.[5]

Table 1 presents summary statistics for the different measures of variance (RV, Bi-power variation and Jump) where we annualize by multiplying by the square root of 365. We can see that RV is quite large for Bitcoin, while we find strong evidence of jumps in this cryptocurrency. Figures 2 and 3 present the time-series graph of the daily price and returns of Bitcoin over our sample period and show the dramatic price rise and large volatility associated with Bitcoin.

3 | METHODOLOGY

This section provides an overview of estimating volatility from intraday data, the jump detection tests, and then we introduce our competing models.

[3]Although the volatility of Bitcoin futures may be of interest to forecast, Bitcoin futures only started trading on the Chicago Board Options Exchange on December 10, 2017 and therefore there is a limited sample size. Future work will no doubt examine the volatility forecasting of Bitcoin futures when the market is more mature, and more data are available.

[4]Since Bitcoin trades 24/7, unlike traditional equity markets which trade only five days a week, our 1-week horizons are seven full days, rather than five.

[5]Liu, Patton, and Sheppard (2015) show that it is difficult to beat the standard 5-minute realized variance when forecasting volatility and the use of the 5-minute frequency is consistent with other studies such as Sévi (2014), Wen (2016), and Behrendt and Schmidt (2018).

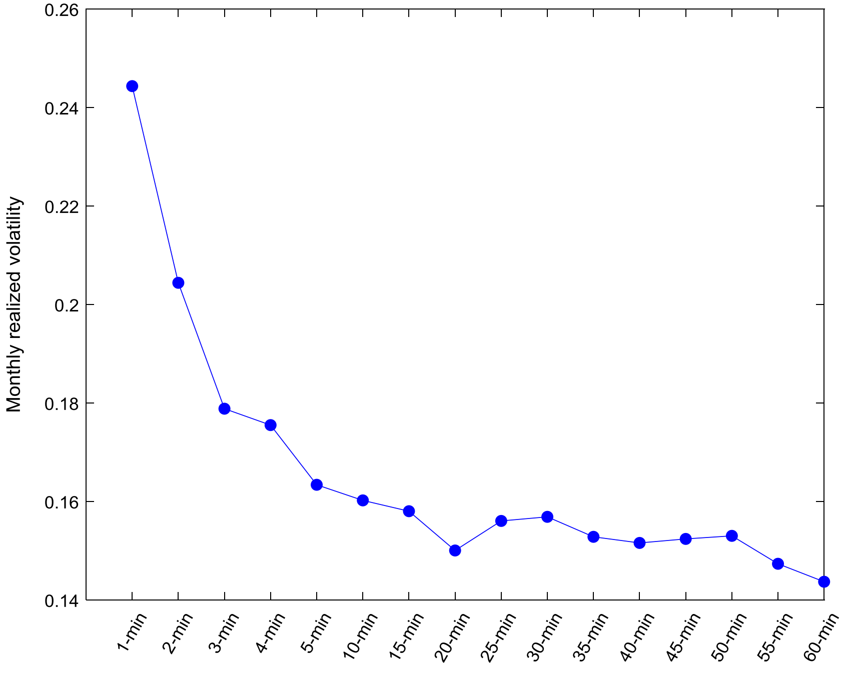

FIGURE 1 The volatility signature plot of Bitcoin
This figure presents the monthly realized volatility signature during our sample. The volatility stabilizes around the 5-min frequency

3.1 | Estimating volatility

In order to calculate the RV from the 5-minute data, we define any given day t, and the RV is computed as the sum of the squared intraday returns $r_{t,j}$ at a given sampling frequency $1/M$:

$$RV_{t,M} = \sum_{j=1}^{M} r_{t,j}^2, \tag{1}$$

where M is the number of intervals in the trading day. Now that we have RV, we need to disentangle jumps and from the continuous component of RV. Barndorff-Nielson and Shephard (2004) propose the bi-power variation (BPV) measure, which is computed as the scaled summation of the product of adjacent absolute returns:

$$BPV_{t,M} = \delta_1 \sum_{j=1}^{M-1} |r_{t,j}||r_{t,j+1}|, \tag{2}$$

where $\delta_1 \equiv 2^{p/2}\Gamma\left(\frac{1/\left(2(p+1)\right)}{\Gamma\left(\frac{1}{2}\right)}\right) = E(|Z|_p)$ denotes the mean of the absolute value of standard normal random variable Z.

TABLE 1 Descriptive statistics of the annualized measures of RV

This table reports the statistics of the different measures of realized variance, where we annualize by multiplying by the square root of 365. RV is realized volatility, BPV is bi-power variation, and J(BPV) and J(MED) are jumps esimated with BPV and median realized variance, respectively. “Std.”, “Max,” and “Min” denote the standard deviation, maximum value, and minimum value, respectively.

	Mean	Std.	Skewness	Kurtosis	Max	Min
RV	0.1051	0.5636	406.5385	10721.5314	18.1669	0.0000
BPV	0.0936	0.5025	391.1610	9747.5904	15.6355	0.0000
J(BPV)	0.0134	0.0802	384.3653	9,900.7244	2.5314	0.0000
J(MED)	0.0115	0.1586	818.6443	37,854.3913	7.4070	0.0000
$\sqrt{\text{RV}}$	1.0164	0.9935	134.0653	1,604.9687	18.6312	0.0745
$\sqrt{\text{BPV}}$	0.9457	0.9381	135.5804	1,615.5471	17.2843	0.0000
$\sqrt{\text{J(BPV)}}$	0.3019	0.4012	118.4355	1,265.7981	6.9542	0.0000
$\sqrt{\text{J(MED)}}$	0.2006	0.4203	225.7501	5,127.5704	11.8948	0.0000

3.2 | Jump detection

We employ the adjusted jump ratio statistic, which has been shown to have power against several empirically realistic calibrated stochastic volatility jump diffusion models and the best empirical properties in Huang and Tauchen (2005), to detect jumps in our study. The test statistic is

$$ZJ_{BPV}(t,M) = \sqrt{M}\frac{\left(RV_{t,M} - BPV_{t,M}\right)RV_{t,M}^{-1}}{\left(\left(\zeta_1^{-4} + 2\zeta_1^{-2} - 5\right)\max\left\{1,\, TQ_{t,M}BPV_{t,M}^{-2}\right\}\right)^{1/2}}, \tag{3}$$

where $TQ_{t,M}$ is the realized tripower quarticity, written as $TQ_{t,M} = M\zeta_{4/3}^3\sum_{j=1}^{M-2}|r_{t,j+1}|^{4/3}|r_{t,j+2}|^{4/3}$, which converges in probability to the integrated quarticity. The ZJ_{BPV} statistic follows a standard normal distribution and allows formal testing for the presence of jumps.

Using the test statistic in Eq. (3) and a significance level of α which we set to 0.999, we extract significant jumps as follows:[6]

$$J_{t,\alpha}(M) = \left[RV_{t,M} - BPV_{t,M}\right] \times I\left[ZJ_{BPV}(t,M) > \Phi_\alpha\right], \tag{4}$$

where $I[\cdot]$ is the indicator function which identifies the significance of the $ZJ_{BPV}(t,M)$ statistic in excess of a given critical value of the Gaussian distribution Φ_α. The continuous path of realized variance can be identified as

$$C_{t,\alpha}(M) = BPV_{t,M} \times I\left[ZJ_{BPV}(t,M) > \Phi_\alpha\right] + RV_{t,M} \times I\left[ZJ_{BPV}(t,M) \leq \Phi_\alpha\right]. \tag{5}$$

[6]We follow Bajgrowicz et al. (2015) and Prokopczuk, Symeonidis, and Wese Simen (2016), and choose this criterion to allay concerns that the test may be driven by false positives.

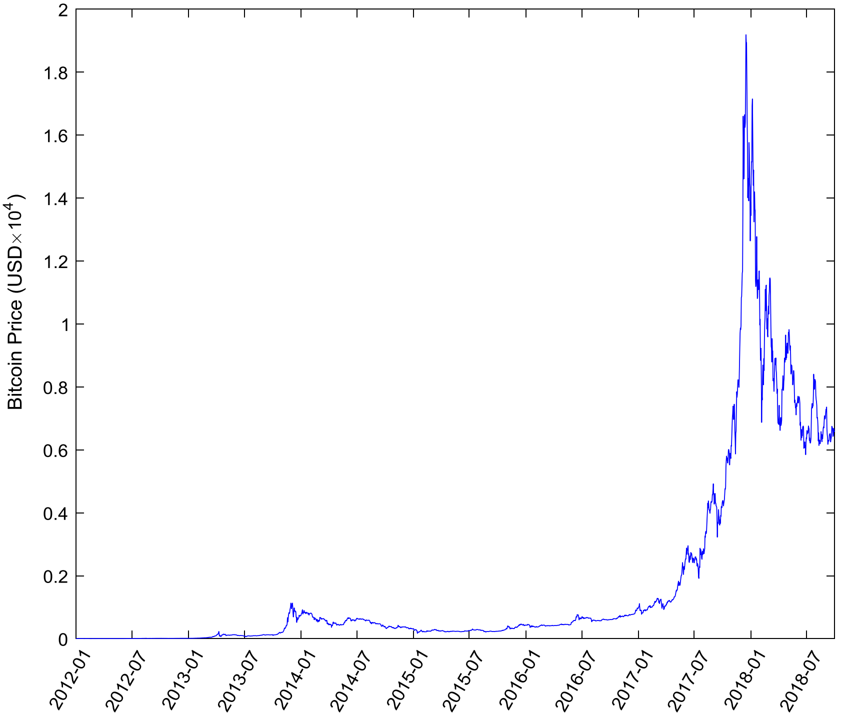

FIGURE 2 Time-series graph of the price of Bitcoin
This figure presents the Bitstamp Bitcoin price (in US dollars) with dramatic boom and bust, during the sample period from January 1, 2012 to September 30, 2018. Data source: www.bitcoincharts.com

Therefore, if a jump is present, the squared jump component equals the difference between RV and the BPV and the continuous component equals the BPV. If there are no jumps, the jump component is naturally zero and the continuous component equals the RV.

3.3 | Semivariances and signed jumps

Realized semivariances have been shown to be important in volatility forecasting (Barndorff-Nielson, Kinnebrock, & Shephard, 2010) since negative and positive returns will have different impacts on the volatility. The daily negative realized semivariances estimator is

$$RSV_{t,M}^{-} = \sum_{j=1}^{M} r_{t,j}^{2} \times I_{|r_{t,j}<0|}, \tag{6}$$

and the positive realized semivariances estimator is

$$RSV_{t,M}^{+} = \sum_{j=1}^{M} r_{t,j}^{2} \times I_{|r_{t,j}>0|}. \tag{7}$$

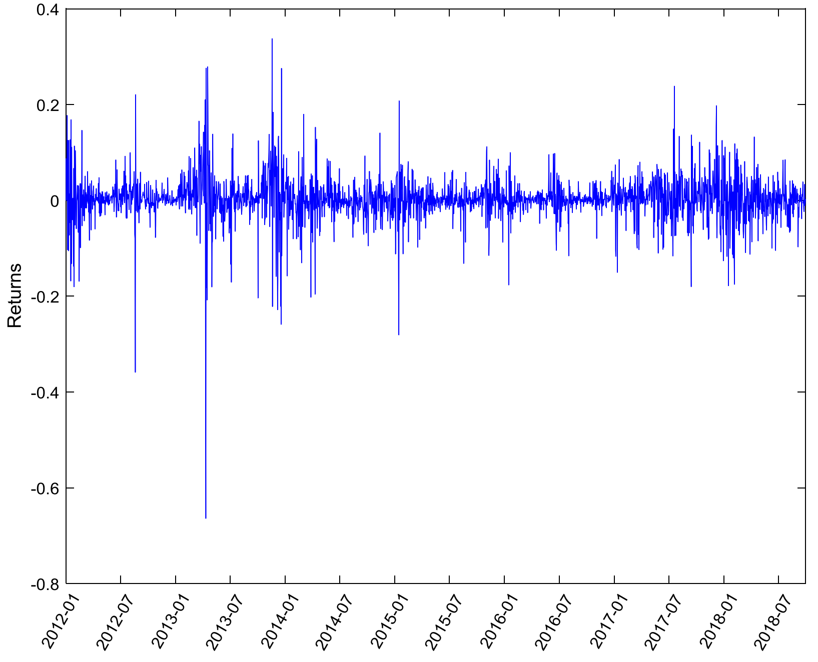

FIGURE 3 **Time-series graph of the returns of Bitcoin**
This figure presents the daily returns of Bitcoin price (in US dollars) on the Bitstamp exchange from January 1, 2012 to September 30, 2018. Data source: www.bitcoincharts.com

Using these estimators, we follow Patton and Sheppard (2011) and define signed jumps as the difference between positive and negative realized semivariances:

$$\Delta J_{t,M} = RSV_{t,M}^{+} - RSV_{t,M}^{-}. \tag{8}$$

3.4 | HAR models

We now introduce the 18 different HAR models we examine in this paper.

3.4.1 | HAR-RV model

Based on his heterogeneous market hypothesis, Corsi (2009) proposes the HAR-RV model given by

$$R\bar{V}_{t+H}^{d} = \beta_0 + \beta_1 RV_t^d + \beta_7 RV_t^w + \beta_{30} RV_t^m + \varepsilon_{t+H}, \tag{9}$$

where $R\bar{V}_{t+H}^{d}$ is the average realized volatility between time t and $t+H$. $R\bar{V}_{t+1}^{d}$ represents the 1-day future realized volatility, while $R\bar{V}_{t+7}^{d}$ represents the 1-week future realized volatility and

$\bar{RV}_{t+30}^{d}$ represents the 1-monith future realized volatility. $RV_t^d = RV_t^{d0} + r_{t,n}^2$ is the daily realized volatility, $RV_t^w = (RV_t^d + RV_{t-1}^d + \cdots + RV_{t-6}^d)/7$ is the weekly realized volatility, while $RV_t^m = (RV_t^d + RV_{t-1}^d + \cdots + RV_{t-29}^d)/30$ is the monthly realized volatility.

3.4.2 | HAR-RV-J model

To determine whether the inclusion of a jump component can help forecast volatility, Andersen, Bollerslev, and Diebold (2007) develop the HAR-RV-J model by adding the daily discontinuous jump variation to the HAR-RV model such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_1 RV_t^d + \beta_7 RV_t^w + \beta_{30} RV_t^m + \beta_{SQJ1} J_t^d + \varepsilon_{t+H}, \tag{10}$$

where J_t^d is the daily discontinuous jump variation that we defined in Eq. (4).

3.4.3 | HAR-CJ model

The HAR-CJ model, proposed by Andersen et al. (2007), separates out the continuous and squared jump components at different time horizons. This model can assess the role of these different volatility components in forecasting volatility such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_{C1} C_t^d + \beta_{C7} C_t^w + \beta_{C30} C_t^m + \beta_{SQJ1} J_t^d + \beta_{SQJ7} J_t^w + \beta_{SQJ7} J_t^m + \varepsilon_{t+H}, \tag{11}$$

where C_t^d is the daily continuous sample path variation as measure, $C_t^w = (C_t^d + C_{t-1}^d + \cdots + C_{t-4}^d)/7$ is the weekly continuous sample path variation, and $C_t^m = (C_t^d + C_{t-1}^d + \cdots + C_{t-29}^d)/30$ is the monthly continuous path variation. The jump component is also decomposed in the following way: $J_t^w = (J_t^d + J_{t-1}^d + \cdots + J_{t-4}^d)/7$ is the weekly jump variation, and $J_t^m = (J_t^d + J_{t-1}^d + \cdots + J_{t-29}^d)/30$ is the monthly jump variation.

3.4.4 | HAR-PS model

This model is the basic specification of Patton and Sheppard (2011) which decomposes the 1-day lagged realized variance into a positive and negative component using realized semivariances such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_1^+ RSV_t^+ + \beta_1^- RSV_t^- + \beta_7 RV_t^w + \beta_{30} RV_t^m + \varepsilon_{t+H}. \tag{12}$$

3.4.5 | HAR-PSL model

This model is similar to the previous HAR-PS model but includes a term for the leverage effect, and checks whether the superior significance of negative realized semivariance does not come from a leverage effect:

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_1^{+} RSV_t^{+} + \beta_1^{-} RSV_t^{-} + \gamma RV_t I_{[r_t<0]} + \beta_7 RV_t^{w} + \beta_{30} RV_t^{m} + \varepsilon_{t+H}. \tag{13}$$

3.4.6 | HAR-RSV model

This model by Patton and Sheppard (2011) decomposes realized variance into a positive and negative component using realized semivariances, since the model assumes that positive and negative realized semivariances can have different predictive abilities for the short, medium and long term. The HAR-RSV model is such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_1^{+} RSV_t^{d+} + \beta_7^{+} RSV_t^{w+} + \beta_{30}^{+} RSV_t^{m+} + \beta_1^{-} RSV_t^{d-} + \beta_7^{-} RSV_t^{w-} + \beta_{30}^{-} RSV_t^{m-} + \varepsilon_{t+H}, \tag{14}$$

where RSV_t^{d+} represents the daily positive realized semivariance, RSV_t^{w+} represents the weekly positive realized semivariance, and RSV_t^{m+} represents the monthly positive realized semivariance, while RSV_t^{d-} represents the daily negative realized semivariance, RSV_t^{w-} represents the weekly negative realized semivariance, and RSV_t^{m-} represents the monthly negative realized semivariance.

3.4.7 | HAR-RSV-J model

Chen and Ghysels (2011) propose the HAR-RSV-J model, which is similar to the HAR-RSV model but includes the lagged daily discontinuous jump variation such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_1^{+} RSV_t^{d+} + \beta_7^{+} RSV_t^{w+} + \beta_{30}^{+} RSV_t^{m+} + \beta_1^{-} RSV_t^{d-} + \beta_7^{-} RSV_t^{w-} + \beta_{30}^{-} RSV_t^{m-} + \beta_{SQJ1} J_t^{d} + \varepsilon_{t+H}. \tag{15}$$

3.4.8 | HARQ-F model

Recently, Bollerslev, Patton, and Quaedvlieg (2016) proposed the HARQ-type model by incorporating realized quarticity (RQ) into the basic models (where $RQ = (M/3)\sum_{j=1}^{M} r_{t,j}^2$):

$$\bar{RV}_{t+H}^{d} = \beta_0 + \left(\beta_1 + \beta_{RQ1}\sqrt{RQ_t^{d}}\right) RV_t^{d} + \left(\beta_7 + \beta_{RQ7}\sqrt{RQ_t^{m}}\right) RV_t^{w} + \left(\beta_{30} + \beta_{RQ30}\sqrt{RQ_t^{m}}\right) RV_t^{m} + \varepsilon_{t+H}. \tag{16}$$

where $RQ_t^{w} = (RQ_t^{d} + RQ_{t-1}^{d} + \dots + RQ_{t-6}^{d})/7$ is the weekly realized quarticity while $RQ_t^{m} = (RQ_t^{d} + RQ_{t-1}^{d} + \dots + Rq_{t-29}^{d})/30$ is the monthly realized quarticity. Since Eq. (16) contains all parameters of measurement of error variance compared to the simplified model described in Subsection 3.4.10, this model is termed the HARQ-F model.

3.4.9 | HARQ-F-J model

The HARQ-F-J model is a new specification where the jump component with a positive sign is considered based on the HARQ-F model proposed by Bollerslev et al. (2016), which has significantly improved forecasting accuracy by incorporating temporal variation. We further add the jump component to the HARQ model:

$$\bar{RV}_{t+H}^{d} = \beta_0 + \left(\beta_1 + \beta_{RQ1}\sqrt{RQ_t^d}\right)RV_t^d + \left(\beta_7 + \beta_{RQ7}\sqrt{RQ_t^m}\right)RV_t^w + \left(\beta_{30} + \beta_{RQ30}\sqrt{RQ_t^m}\right)RV_t^m + \beta_{SQJ1}J_t^d + \varepsilon_{t+H}, \tag{17}$$

where J_t^d is the daily discontinuous jump variation.

3.4.10 | HARQ model

Bollerslev et al. (2016) also observe that there exists substantial estimation bias in daily RV, while the attenuation bias is much less severe in the weekly and monthly ones. Therefore, they simplify Eq. (16) to the function of daily $RQ^{1/2}$:

$$\bar{RV}_{t+H}^{d} = \beta_0 + \left(\beta_1 + \beta_{RQ1}\sqrt{RQ_t^d}\right)RV_t^d + \beta_7 RV_t^w + \beta_{30} RV_t^m + \varepsilon_{t+H}, \tag{18}$$

which is referred to the HARQ model.

3.4.11 | HARQ-J model

As the HAR-Q-F model has its simplified version, the new specification of HARQ-F-J model can be also termed HARQ-J by only focusing on the daily realized quarticity and jump component, which is the second new specification:

$$\bar{RV}_{t+H}^{d} = \beta_0 + \left(\beta_1 + \beta_{RQ1}\sqrt{RQ_t^d}\right)RV_t^d + \beta_7 RV_t^w + \beta_{30} RV_t^m + \beta_{SQJ1}J_t^d + \varepsilon_{t+H}. \tag{19}$$

3.4.12 | HAR-RV-SJ model

This model introduces the notion of the signed jump and is similar to the previous HAR-RV-J model, but now the lagged daily discontinuous jump variation is replaced with the lagged daily signed jump such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_7 RV_t^w + \beta_{30} RV_t^m + \beta_{C1} C_t^d + \beta_{J1}\Delta J_t^d + \varepsilon_{t+H}, \tag{20}$$

where ΔJ_t^d is the daily signed jump.

3.4.13 | HAR-CSJ model

This model, suggested by Sévi (2014), considers signed jumps over a longer interval than 1 day and considers jumps over a short period of time, as well as taking into account the signs of the jumps such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_{C1}C_t^d + \beta_{C7}C_t^w + \beta_{C30}C_t^m + \beta_{J1}\Delta J_t^d + \beta_{J7}\Delta J_t^w + \beta_{J30}\,\Delta J_t^m + \varepsilon_{t+H}. \quad (21)$$

3.4.14 | HAR-RV-SJd model

This model, proposed by Patton and Sheppard (2011), discriminates between positive and negative signed jumps such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_7 RV_t^d + \beta_{30}RV_t^w + \beta_{C1}C_t + \beta_{J1}^{+}\Delta J_t^d I_{[\Delta J_t>0]} + \beta_{J1}^{-}\,\Delta J_t^d I_{[\Delta J_t<0]} + \varepsilon_{t+H}. \quad (22)$$

3.4.15 | HAR-CSJd model

The HAR-CSJd model, proposed by Sévi (2014), separates positive and negative signed jumps at various time horizons such that

$$\begin{aligned} \bar{RV}_{t+H}^{d} = \beta_0 + \beta_{J1}^{+}\Delta J_t I_{\left[\Delta J_t>0\right]} + \beta_{J1}^{-}\Delta J_t^d I_{\left[\Delta J_t<0\right]} + \beta_{C1}C_t^d + \\ \beta_{J7}^{+}\Delta J_t^w I_{\left[\Delta J_t^w>0\right]} + \beta_{J7}^{-}\Delta J_t^w I_{\left[\Delta J_t^w<0\right]} + \beta_{C7}C_t^w + \\ \beta_{J30}^{+}\Delta J_t^m I_{\left[\Delta J_t^m>0\right]} + \beta_{J30}^{-}\Delta J_t^m I_{\left[\Delta J_t^m<0\right]} + \beta_{C30}C_t^m + \varepsilon_{t+H}. \end{aligned} \quad (23)$$

3.4.16 | HAR-J model

Proposed by Andersen et al. (2007), this model is a simple extension of the HAR-RV model in that it replaces the most recent RV with a continuous and jump component such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_{C1}C_t^d + \beta_7 RV_t^w + \beta_{30}RV_t^m + \beta_{J1}J_t^d + \varepsilon_{t+H}. \quad (24)$$

3.4.17 | HAR-RJ model

The previous model can be criticized in that it ignores the sign of the jumps and therefore Tauchen and Zhou (2011) propose the HAR-RJ model such that

$$\bar{RV}_{t+H}^{d} = \beta_0 + \beta_{C1}C_t^d + \beta_7 RV_t^w + \beta_{30}RV_t^m + \beta_{Rj}RJ_t + \varepsilon_{t+H}, \quad (25)$$

where $RJ_t = \text{sign}(r_t)\cdot\sqrt{J_t}$.

TABLE 2 List of HAR models

This table reports the 18 competing heterogeneous autoregressive models that are examined in this paper. We present model name, reference, and equation number in this study.

Model number	Model name	Reference	Equation number in this study
1	HAR-RV	Corsi (2009)	(9)
2	HAR-RV-J	Andersen et al. (2007)	(10)
3	HAR-CJ	Andersen et al. (2007)	(11)
4	HAR-PS	Patton and Sheppard (2011)	(12)
5	HAR-PSL	Patton and Sheppard (2011)	(13)
6	HAR-RSV	Patton and Sheppard (2011)	(14)
7	HAR-RSV-J	Chen and Ghysels (2011)	(15)
8	HARQ-F	Bollerslev et al. (2016)	(16)
9	HARQ-F-J	New specification	(17)
10	HARQ	Bollerslev et al. (2016)	(18)
11	HARQ-J	New specification	(19)
12	HAR-RV-SJ	Patton and Sheppard (2011)	(20)
13	HAR-CSJ	Sévi (2014)	(21)
14	HAR-RV-SJd	Patton and Sheppard (2011)	(22)
15	HAR-CSJd	Sévi (2014)	(23)
16	HAR-J	Prokopczuk et al. (2016)	(24)
17	HAR-RJ	Prokopczuk et al. (2016)	(25)
18	HAR-ARJ	Prokopczuk et al. (2016)	(26)

3.4.18 | HAR-ARJ model

The final model, suggested by Prokopczuk et al (2016), decomposes RJ into positive and negative jumps to determine whether the variation of negative jumps has a more pronounced impact on future volatility than that of positive jumps, such that

$$RV_{t+h} = \beta_0 + \beta_{C1}C_t^d + \beta_{C7}C_t^w + \beta_{C30}C_t^m + \beta_{RJ^+}RJ_t^+ + \beta_{RJ^-}RJ_t^- + \varepsilon_{t+H}. \tag{26}$$

Table 2 provides an overview of the 18 different HAR models explored in this study which includes the traditionally popular HAR models, as well as more advanced and innovative models.

4 | EMPIRICAL RESULTS

This section presents our main results. We begin by comparing the predictive ability of our models in an in-sample setting, and then present a comprehensive and rigorous analysis of their out-of-sample performance.

4.1 | In-sample analysis

We begin by analyzing the in-sample predictive power of our competing models introduced in Section 3.4. We do this by estimating all of the models via OLS regressions and also report the adjusted R^2. We consider different forecasting horizons, notably 1-day, 1-week, and 1-month volatility, where we use Newey–West corrected standard errors.[7]

Table 3 reports our results for models 1–9 for each forecasting horizon, and we see that in the HAR-RV model, all coefficients are statistically significant at the 10% level for all forecasting horizons and the adjusted R^2 for the 1-day, 1-week, and 1-month forecasting horizons is 31.8%, 34.6%, and 13.2%, respectively. The HAR-RV-J model, which encompasses the jump component, shows that the jump component is positive and statistically significant, which is found across all forecasting horizons, which indicates that volatility increases following a jump event. The jump component magnitude decreases monotonically from the 1-day forecasting horizon to the 1-month forecasting horizon. The adjusted R^2 of the HAR-RV-J model is larger than that of the HAR-RV model, indicating that the HAR-RV-J model incorporating jumps predicts Bitcoin volatility more accurately than the standard HAR-RV model. This is consistent with the results of Corsi et al. (2010) who find a positive and significant effect of the jump component when measuring the squared jumps using their threshold indicator. Therefore, we can conclude that jumps significantly increase the impact of lagged volatility that is a highly persistent component.

The results for the HAR-RV-CJ model show that the continuous jump component for 1 day is positive and statistically significant across all forecasting horizons and has a higher adjusted R^2 than that of the HAR-RV-J model at 1-week and 1-month horizons, indicating its superiority. The HAR-PS model of Patton and Sheppard (2011) shows that the negative semivariance is statistically significant; however, the adjusted R^2 is smaller than that of the HAR-RV-CJ model, therefore indicating that the decomposition between positive and negative semivariances does not contribute to improving the fit of the predictive regressions. The HAR-PSL model shows that the leverage component is statistically significant and the high adjusted R^2 shows that this model is the best model for the in-sample 1-day forecasting horizon. For the HAR-RSV model of Patton and Sheppard (2011) and the extension HAR-RSV-J model of Chen and Ghysels (2011), we find that the jump component is statistically significant, as are the RV components decomposed into positive and negative realized semivariances. In the last two rows of Table 3, we report the HARQ-F model of Bollerslev et al. (2016) and the HARQ-F-J model shown in Equation (17), and find that the new components incorporated in this model are all statistically significant at the 1-day forecasting horizon, and most are statistically significant at the 1-week and 1-month forecasting horizons. Regarding the goodness of fit of our models, we find that the model with the highest adjusted $\boldsymbol{R}^2$ over the 1-day horizon is the HAR-PSL, while the HARQ-F-J model has the highest adjusted $\boldsymbol{R}^2$ over the 1-week and 1-month forecasting horizons.

Table 4 presents the findings for models 10–18, where we find that the models incorporating realized quarticity, HARQ and HARQ-J, have larger adjusted R^2 than the standard HAR models. The signed jump component of the HAR-RV-SJ model is negative and statistically significant at the 5% level. The explanatory power of this compared to the HAR-RSV-J model is smaller, indicating that there is no specific in-sample gain in considering signed jumps. We do show, however, that there is an additional increase in explanatory power when signed jumps are considered separately,

[7]We do not include the standard errors to save space, but the full results are available upon request from the corresponding author.

TABLE 3 In-sample results of models 1–9

This table reports in-sample estimation results of models 1–9. The results in Panels A, B, and C show the 1-day, 1-week, and 1-month forecast horizon, respectively. All regressions are estimated using Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) corrected standard errors. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	β_0	β_1	β_7	β_{30}	β_{SQ1}	β_{C1}	β_{SQ7}	β_{C7}	β_{SQ30}	β_{C30}	β_1^-	β_1^+	β_7^-	β_7^+	β_{30}^-	β_{30}^+	γ	β_{RQ1}	β_{RQ7}	β_{RQ30}	adj.R^2
Panel A: 1-day horizon																					
HAR-RV	0.002***	0.509***	0.100***	0.098*																	0.318
HAR-RV-J	0.001**	0.271***	0.125***	0.125**	1.915***																0.329
HAR-RV-CJ	0.001**				2.112***	0.280***	0.093	0.345	0.084	0.507											0.328
HAR-PS	0.001***		0.079**	0.096*							0.964***	0.143									0.320
HAR-PSL	0.002***		0.337***	0.140***							0.657***	−0.142					0.092*				0.337
HAR-RSV	0.002***										1.119***	−0.018	−1.589**	1.522***	−0.681	0.853					0.322
HAR-RSV-J	0.002**				2.193***						1.064***	−0.464***	−1.772***	1.724***	0.323	0.020					0.336
HARQ-F	0.001	0.376***	0.010	0.504***														0.046***	0.108**	−0.519***	0.325
HARQ-F-J	0.001	0.219***	−0.017	0.539***	1.751***													0.024*	0.125**	−0.528***	0.333
Panel B: 1-week horizon																					
HAR-RV	0.002***	0.311***	0.108***	0.1371***																	0.346
HAR-RV-J	0.002***	0.213***	0.118***	0.148***	0.792***																0.351
HAR-RV-CJ	0.002***				0.847***	0.233***	0.121	0.017	−0.080	2.033***											0.354
HAR-PS	0.002***		0.116***	0.138***							0.125	0.460***									0.347
HAR-PSL	0.002***		0.069***	0.129***							0.062	0.402***					0.137***				0.349
HAR-RSV	0.003***										0.090	0.504***	1.470***	−1.084***	−2.722***	2.645***					0.351
HAR-RSV-J	0.003***				0.670***						0.074	0.368***	1.413***	−1.023***	−2.415***	2.389***					0.354
HARQ-F	0.001*	0.206***	0.166***	0.614***														0.037***	0.000	−0.709***	0.367
HARQ-F-J	0.001*	0.156***	0.158***	0.625***	0.563***													0.030***	0.005	−0.712***	0.369
Panel C: 1-month horizon																					
HAR-RV	0.004***	0.091***	0.081***	0.098***																	0.132
HAR-RV-J	0.004***	0.018	0.088***	0.106***	0.588***																0.137
HAR-RV-CJ	0.004***				0.166***	0.070***	0.102***	0.121	0.265***	−0.961***											0.156
HAR-PS	0.004***		0.084***	0.098***							0.030	0.140***									0.131
HAR-PSL	0.004***		−0.020	0.080***							0.048	0.157***					0.099***				0.131
HAR-RSV	0.005***										0.103	0.077	0.886***	−0.659**	−5.204***	4.809***					0.165
HAR-RSV-J	0.005***				0.426***						0.091	−0.009	0.850***	−0.619**	−5.008***	4.646***					0.168
HARQ-F	0.002***	0.086***	0.324***	0.498***														0.003	−0.158***	−0.721***	0.203
HARQ-F-J	0.001***	0.028	0.314***	0.511***	0.644***													−0.005	−0.152***	−0.724***	0.210

TABLE 4 In-sample results of models 10–18

This table reports in-sample estimation results of models 10–18. The results in Panels A, B, and C show the 1-day, 1-week, and 1-month forecast horizon, respectively. All regressions are estimated using Newey and West (1987) HAC corrected standard errors. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	β_0	β_1	β_7	β_{30}	β_{C1}	β_{C7}	β_{C30}	$\Delta J1$	$\Delta J7$	$\Delta J30$	ΔJ_1^-	ΔJ_1^+	ΔJ_7^-	ΔJ_7^+	ΔJ_{30}^-	ΔJ_{30}^+	β_{RJ}	$\beta_{RJ}+$	$\beta_{RJ}-$	β_{SQ1}	β_{RQ1}	adj.R^2
Panel A: 1-day horizon																						
HARQ	0.002***	0.387***	0.154***	0.106**																	0.044***	0.322
HARQ-J	0.001***	0.234***	0.150***	0.126**																1.714***	0.023*	0.329
HAR-RV-SJ	0.002***		0.089**	0.087*	0.601***			−0.288**														0.311
HAR-RV-CSJ	0.002***				0.585***	−0.093	0.108	0.147	1.927***	0.819												0.314
HAR-RV-SJD	0.002***		0.137***	0.065	0.619***						−0.168	−3.005***										0.323
HAR-RV-CSJD	0.002***				0.346***	0.030	0.086				−2.759***	1.007***	3.593***	0.197	−0.278	2.218*						0.328
HAR-J	0.002***		0.104***	0.089*	0.566***																	0.310
HAR-RJ	0.002***		0.191***	0.096*	0.386***												−1.693***					0.341
HAR-ARJ	0.002***		0.188***	0.116**	0.245***													−0.117	−2.675***			0.347
Panel B: 1-week horizon																						
HARQ	0.002***	0.223***	0.147***	0.143***																	0.032***	0.351
HARQ-J	0.002***	0.172***	0.145***	0.150***																0.573***	0.025***	0.353
HAR-RV-SJ	0.002***		0.120***	0.133***	0.321***			0.223**														0.343
HAR-RV-CSJ	0.003***				0.334***	0.159***	−0.023	0.562***	−0.755**	2.449***												0.347
HAR-RV-SJD	0.003***		0.137***	0.125***	0.328***						0.267***	−0.763***										0.347
HAR-RV-CSJD	0.003***				0.210***	0.248***	−0.069				−0.828***	0.943***	−0.333	−1.919***	2.440***	3.445***						0.354
HAR-J	0.002***		0.108***	0.132***	0.348***																	0.342
HAR-RJ	0.003***		0.134***	0.134***	0.294***												−0.505***					0.348
HAR-ARJ	0.002***		0.133***	0.146***	0.207***													0.477*	−1.117***			0.353
Panel C: 1-month horizon																						
HARQ	0.004***	0.105***	0.074***	0.097***																	−0.005	0.132
HARQ-J	0.004***	0.041*	0.073***	0.105***																0.712***	−0.014**	0.139
HAR-RV-SJ	0.004***		0.087***	0.097***	0.090***			0.080														0.129
HAR-RV-CSJ	0.005***				0.100***	0.120***	−0.231*	−0.005	−0.707**	5.098***												0.163
HAR-RV-SJD	0.004***		0.098***	0.092***	0.094***						0.108	−0.556***										0.133
HAR-RV-CSJD	0.005***				0.042*	0.129***	−0.225***				−0.603***	0.241*	−0.758**	−0.432	4.748***	5.317***						0.167
HAR-J	0.004***		0.083***	0.096***	0.100***																	0.129
HAR-RJ	0.004***		0.083***	0.096***	0.099***												−0.005					0.129
HAR-ARJ	0.004***		0.081***	0.107***	0.021													0.872***	−0.551***			0.138

depending on their sign. We find that positive and negative 1-day jumps are significant across all forecasting horizons; however; these models do not offer greater explanatory power than the HAR-RSV-J model reported in Table 3. We also study the HAR-J model of Andersen et al. (2007) and show that this simple model does not add any explanatory power. Finally, we examine two further models, namely the HAR-RJ model of Tauchen and Zhou (2011) and the HAR-ARJ of Prokopczuk et al. (2016). We find that the jump component is negative and statistically significant, indicating that volatility decreases following a jump event. When we decompose the jump component into positive and negative jumps in the HAR-ARJ model, we find that the negative jump component dominates its positive counterpart (consistent with Prokopczuk et al., 2016). The adjusted R^2 is also high for the HAR-ARJ model, as it is the model that has the most explanatory power for the 1-day forecasting horizon, and it is quite high compared to the other models for the 1-week and 1-month forecasting horizons.

Combining Tables 3 and 4 together, we find that the model with the highest adjusted R^2 over the 1-day horizon is HAR-ARJ, while over the 1-week and 1-month horizons it is the HARQ-F-J.

4.2 | Out-of-sample analysis

We now turn our attention to the out-of-sample performance of the competing models since this is the only way to gauge the forecasting performance between different models (Giot & Laurent, 2007). The vast majority of the literature employs the Diebold–Mariano–White (DMW) statistic developed by Diebold and Mariano (1995) and West (1996). This statistic compares the forecast ability of competing models by generating a loss function that is a measure of the difference between the realized value and the forecast in a pseudo-out-of-sample forecasting environment. However, the DMW test is inappropriate when comparing nested models, and therefore we use the Clark and West (2007) statistic when comparing the 18 competing models.[8] The adjusted mean squared prediction error suggested by Clark and West (2007) is

$$\hat{f}_{t+h} = (\hat{e}_{1|t+h})^2 - (\hat{e}_{2|t+h})^2 + (\hat{y}_{1|t+h} - \hat{y}_{2|t+h})^2, \tag{30}$$

where $\hat{e}_{1|t+h}$ is the null model forecast errors for horizon h, $\hat{e}_{2|t+h}$ is the alternative model forecast errors for horizon h, and $\hat{y}_{1|t+h}$ and $\hat{y}_{2|t+h}$ are null and alternative model forecast values. Then the CW statistic for horizon h is calculated by

$$CW = \frac{\sqrt{N}\bar{f}_h}{\sqrt{\mathrm{Var}(\hat{f}_{t+h} - \bar{f}_h)}}, \tag{31}$$

where $\bar{f}_h$ is the sample average of $\hat{f}_{t+h}$, N is the forecast sample number, and Var (•) is the sample variance. Rejection of the null model signifies that the forecast errors from the alternative model are significantly smaller than those from the null. Hence, significant and positive CW statistics indicate that the alternative model is the preferred one.

For the 1-day forecast horizon, Panel A of Table 5 shows that the HAR-RV model significantly outperforms most of other models except for HARQ, HARQ-J, HARQ-F, and

[8]In unreported results, we also use the DMW statistic to compare non-nested models and use the CW statistic to compare nested models, consistent with Corsi and Renò (2012). Our results are qualitatively similar and are available upon request from the corresponding author.

TABLE 5 CW statistics for HAR models for 1-day and 1-week horizon

This table reports the CW statistics for forecasting horizons of 1 day (Panel A) and 1 week (Panel B). A positive result indicates that the model whose name is in the first row outperforms the model whose name is provided in the first column. The statistic is computed using Newey and West (1987) HAC. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HAR-RV-J	HAR-RV-CJ	HAR-PS	HAR-PSL	HAR-RSV	HAR-RSV-J	HARQ	HAR-Q-J	HAR-Q-F	HAR-Q-F-J	HAR-RV-SJ	HAR-RV-CSJ	HAR-RV-SJD	HAR-RV-CSJD	HAR-J	HAR-RJ	HAR-ARJ
Panel A: 1-day horizon																	
HAR-RV	−0.240	−0.017	−0.897	−1.328	−1.057	−0.721	2.086**	1.610*	3.634***	4.529***	−0.707	−0.893	−0.691	−0.825	0.690	−1.613	−1.154
HAR-RV-J		0.884	−0.262	−0.936	−0.059	−0.864	1.867**	2.444***	2.464***	4.432***	1.461*	1.477*	1.364*	−0.722	1.432*	−1.250	−2.120
HAR-RV-CJ			−0.070	−0.750	0.231	−0.781	2.105**	2.618***	2.669***	4.600***	1.728**	1.919**	1.693**	−0.678	1.724**	−0.386	−1.531
HAR-PS				−1.970	0.649	−0.646	1.502*	2.198**	1.904**	2.884***	1.106	0.911	0.988	−0.904	1.057	0.713	0.866
HAR-PSL					3.202***	1.304*	2.246**	3.227***	2.558***	3.688***	3.025***	2.464***	2.816***	−0.442	2.143**	2.427***	2.843***
HAR-RSV						−0.477	2.756***	4.717***	3.233***	5.718***	4.839***	0.859	4.144***	−0.804	2.524***	3.451***	3.731***
HAR-RSV-J							1.547*	1.749**	1.660**	1.904**	1.607*	1.035	1.551*	−0.805	1.416*	1.428*	1.407*
HARQ								−0.185	5.054***	2.120**	−0.335	−0.389	−0.389	−0.809	−0.077	−1.090	−0.805
HARQ-J									1.888**	3.689***	0.477	0.678	0.431	−0.801	1.067	−1.140	−1.278
HARQ-F										−0.597	−0.739	−0.719	−0.689	−0.855	−0.992	−1.397	−1.257
HARQ-F-J											0.901	1.545*	1.002	−0.673	1.687**	−0.268	−0.889
HAR-RV-SJ												0.475	0.162	−0.933	0.995	−0.613	−0.227
HAR-RV-CSJ													3.767***	−0.785	4.237***	3.022***	2.542***
HAR-RV-SJD														−0.710	1.927**	1.254	1.251
HAR-RV-CSJD															1.220	1.232	1.275
HAR-J																−1.439	−0.936
HAR-RJ																	0.007
Panel B: 1-week horizon																	
HAR-RV	1.408*	1.820**	0.995	0.483	1.121	1.624*	1.149	2.794***	17.811***	20.172***	0.856	0.230	0.619	−0.438	−0.229	−2.907	−0.118
HAR-RV-J		0.836	0.567	0.281	0.930	1.322*	0.649	1.292*	12.586***	17.141***	0.642	0.388	0.498	−0.491	0.073	−2.847	−3.841
HAR-RV-CJ			1.289*	1.021	1.745**	2.157**	1.525*	2.448***	13.171***	16.793***	1.192	1.058	1.034	−0.111	1.187	−0.015	0.379
HAR-PS				−1.391	0.533	1.226	0.333	1.515*	21.860***	20.032***	0.319	−0.585	−0.302	−0.133	−0.919	−1.828	−0.253
HAR-PSL					1.690**	2.136**	2.222**	2.931***	21.330***	20.305***	2.674***	0.640	1.967**	0.092	1.456*	−0.262	0.905
HAR-RSV						2.028**	9.100***	9.114***	21.796***	21.027***	7.796***	−3.219	7.205***	−0.826	8.747***	6.474***	6.369***
HAR-RSV-J							4.238***	6.606***	15.534***	19.195***	3.105***	−0.889	2.860***	−1.258	5.474***	6.124***	7.065***
HARQ								2.057**	20.833***	18.788***	1.527*	−0.039	0.881	0.001	0.469	−1.066	0.511
HARQ-J									18.066***	20.090***	0.619	−0.027	0.353	−0.340	−0.412	−2.553	−0.921
HARQ-F										−0.539	−11.785	−3.904	−11.414	−1.961	−12.026	−10.518	−9.381
HARQ-F-J											−5.069	−1.343	−4.981	−1.109	−9.132	−8.368	−8.198
HAR-RV-SJ												−0.706	−1.055	0.418	−0.520	−1.665	0.724
HAR-RV-CSJ													10.589***	0.186	10.039***	8.256***	8.259***
HAR-RV-SJD														0.894	1.360*	0.033	1.958**
HAR-RV-CSJD															2.050**	2.106**	2.428***
HAR-J																−1.940	0.480
HAR-RJ																	2.313**

HARQ-F. As expected from the in-sample analysis, the HARQ-F and HARQ-F-J models are very good as they outperform almost all the other models, and most of the time significantly so. The HARQ and HARQ-J models both do quite well and significantly outperform most other models, but not as well as the HARQ-F and HARQ-F-J models. The worst models are the HAR-RV-CSJD and HAR-PSL, which underperform most of all models. Interestingly, the models with signed jumps or signed variances do not offer any improved performance. Considering the 1-week horizon in Panel B, we find similar results in that the HARQ-F and HARQ-F-J models are preferred, as they also significantly outperform all other models. Therefore, over the 1-day and 1-week forecasting horizons we can conclude that the HARQ-F and HARQ-F-J models offer the best performance; however, the CW statistic for the comparison between HARQ-F and HARQ-F-J is not significant across the 1-day and 1-week horizons. Table 6 reports the out-of-sample results for the 1-month forecasting horizon, where we find that the HARQ-F-J model outperforms all other models, indicating that it gives the best out-of-sample performance. Interestingly, however, the second best model is the newly developed HARQ-F model of Bollerslev et al. (2016), which outperforms all other models (most of the time significantly so) except the HARQ-F-J model.

To summarize, our findings show although modest findings in the in-sample period, the HARQ-F-J model offers the best out-of-sample performance. This points to the well-known in-sample overfitting issue which does not translate well into good out-of-sample forecasts. Therefore, in an out-of-sample setting, we find that temporal variation and squared jump components, separated at different horizons, offer the best forecast of the volatility of Bitcoin.

5 | ROBUSTNESS CHECK

In this section we add robustness to our analysis by considering HAR models with structural breaks as well as an alternative jump estimator by employing the nearest-neighbor estimator of Andersen et al (2012). We add a further robustness check by also employing the WLS estimator in addition to the OLS estimator.

5.1 | Structural breaks

In a recent study, Wen, Gong, and Cai (2016) introduced HAR models with structural breaks and showed that these models can help explain the volatility of crude oil futures. As shown in Figure 2, the price of Bitcoin has fluctuated hugely over our sample period, and therefore there is the possibility that introducing structural breaks into our HAR models may improve the in-sample and, more importantly, the out-of-sample performance. Therefore, we employ the Inclàn and Taio (1994) iterative cumulative sum of squares (ICSS) algorithm to determine the number of break points; Figure 4 plots the Bitcoin returns with break points and $\pm$ 3 standard deviations, and Table 7 reports the structural break periods which are identified by the ICSS algorithm.[9] We find 35 break points during our sample period, which is quite considerable given our sample period, suggesting that the inclusion of structural breaks may improve the performance of the HAR models. Therefore, we re-estimate each HAR model but include a

[9]The ICSS algorithm is a popular method for detecting multiple structural breaks and has been employed by Vivian and Wohar (2012), Charles and Darné (2014), Yarovaya, Brzeszczyński, and Lau (2016), and Shahzad, Mensi, Hammoudeh, Balcilar, and Shahbaz (2018).

TABLE 6 CW statistics for HAR models for 1-month horizon

This table reports the CW statistics for the 1-month forecasting horizon. A positive result indicates that the model whose name is in the first row outperforms the model whose name is provided in the first column. The statistic is computed using Newey and West (1987) HAC. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HAR-RV-J	HAR-RV-CJ	HAR-PS	HAR-PSL	HAR-RSV	HAR-RSV-J	HARQ	HARQ-J	HARQ-F	HARQ-F-J	HAR-RV-SJ	HAR-RV-CSJ	HAR-RV-SJD	HAR-RV-CSJD	HAR-J	HAR-RJ	HAR-ARJ
HAR-RV	4.196***	13.799***	1.284*	0.935	−8.216	−7.861	2.792***	4.237***	22.916***	21.909***	0.538	−8.014	−0.579	−5.546	−2.339	−1.744	6.417***
HAR-RV-J		12.697***	−1.513	−1.477	−8.184	−8.517	−2.186	2.518***	22.792***	22.238***	−1.239	−7.863	−1.767	−5.719	−2.912	−3.006	7.160***
HAR-RV-CJ			−10.526	−10.338	−11.407	−11.570	−10.816	−6.183	20.295***	19.509***	−9.508	−11.057	−9.817	−7.970	−11.429	−11.474	−9.239
HAR-PS				−0.098	−8.545	−8.006	1.171	4.130***	22.628***	21.814***	−1.002	−8.410	−2.378	−5.881	−2.216	−2.127	5.553***
HAR-PSL					−8.119	−7.562	2.407***	4.706***	22.958***	22.135***	1.483*	−8.008	0.021	−5.701	0.605	0.319	5.914***
HAR-RSV						6.302***	13.986***	14.317***	20.250***	20.114***	13.961***	−6.938	13.832***	0.378	13.843***	13.849***	14.449***
HAR-RSV-J							13.705***	14.089***	20.347***	20.190***	13.387***	−4.548	13.220***	−1.112	13.475***	13.495***	14.285***
HARQ								3.719***	22.922***	21.891***	−0.209	−7.968	−1.003	−5.569	−2.722	−2.998	5.568***
HARQ-J									21.607***	22.906***	−0.972	−6.591	−1.220	−6.089	−1.837	−1.902	−0.074
HARQ-F										1.805**	−20.325	−13.679	−20.139	−10.535	−22.361	−22.369	−21.550
HARQ-F-J											−11.614	−10.318	−11.511	−10.286	−14.896	−15.064	−17.344
HAR-RV-SJ												−8.386	−2.315	−6.040	−0.740	−0.777	5.695***
HAR-RV-CSJ													13.769***	1.580*	13.781***	13.787***	14.367***
HAR-RV-SJD														−5.758	1.478*	1.474*	6.714***
HAR-RV-CSJD															7.206***	7.243***	8.163***
HAR-J																0.082	6.284***
HAR-RJ																	6.499***

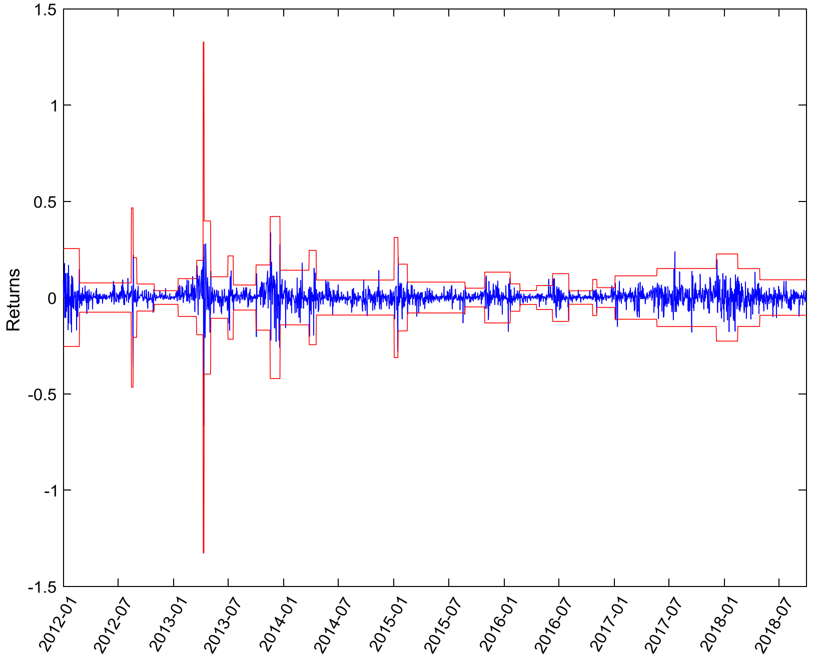

FIGURE 4 The structural breaks with ±3 standard deviation bounds
This figure shows a time-series graph of Bitcoin returns and the ±3 standard deviation bands around the structural break points. The standard deviation is calculated using the daily Bitcoin returns in the corresponding time period

structural break as in Wen et al. (2016) and report the CW statistics of the HAR models with structural breaks over different forecasting horizons in Tables 8,9.[10] Over the 1-day forecasting horizon, we find that the best model is the HARQ-F model which outperforms all other models, while over the 1-week forecasting horizon the relatively superior model is the HAR-J model although it does not outperform all models but outperforms the majority of models. Interestingly, over the longer forecasting horizon of 1 month, we find that the HARQ-F model is superior to all others. Therefore, the inclusion of the structural breaks into our HAR models changes the superior out-of-sample models quite considerably.

Tables 8,9 compare the fit of the HAR models with structural breaks against the HAR models without structural breaks. However, we need to assess whether HAR models with structural breaks outperform HAR models without structural breaks included in the estimation to determine whether models with or without structural breaks are superior in the forecasting of volatility of Bitcoin. Therefore in Table 10, we present the CW statistics when comparing HAR models with structural breaks (alternative models) and HAR models excluding structural breaks (null models). In this table, a positive statistic indicates that the corresponding HAR model with structural breaks is preferred over the HAR model without structural breaks, and we find that over the 1-day forecasting horizon all of the statistics are positive, indicating that the HAR models with structural breaks offer the best forecast of future volatility of Bitcoin.

[10]We do not report the in-sample results of the structural break HAR models to save space and because the out-of-sample analysis is more insightful, but the in-sample results are available upon request from the corresponding author.

TABLE 7 Breakpoint dates and standard deviations

This table reports the total break points and time periods of structural breaks in Bitcoin volatility identified by the ICSS algorithm. The standard deviation is calculated using the daily Bitcoin returns in the corresponding time period.

Total breaks points	Time period	Standard deviation
1	January 1, 2012–March 19, 2012	0.070
2	March 20, 2012–August 14, 2012	0.025
3	August 15, 2012–August 20, 2012	0.155
4	August 21, 2012–September 1, 2012	0.069
5	September 2, 2012–October 29, 2012	0.023
6	October 30, 2012–January 16, 2013	0.012
7	January 17, 2013–March 18, 2013	0.033
8	March 19, 2013–April 9, 2013	0.064
9	April 10, 2013–April 12, 2013	0.442
10	April 13, 2013–May 4, 2013	0.133
11	May 5, 2013–June 30, 2013	0.036
12	July 1, 2013–July 18, 2013	0.072
13	July 19, 2013–October 1, 2013	0.022
14	October 2, 2013–November 17, 2013	0.056
15	November 18, 2013–December 20, 2013	0.140
16	December 21, 2013–March 26, 2014	0.047
17	March 27, 2014–April 19, 2014	0.082
18	April 20, 2014–January 2, 2015	0.030
19	January 3, 2015–January 15, 2015	0.104
20	January 16, 2015–February 15, 2015	0.058
21	February 16, 2015–August 25, 2015	0.027
22	August 26, 2015–October 29, 2015	0.016
23	October 30, 2015–January 22, 2016	0.044
24	January 23, 2016–February 23, 2016	0.024
25	February 24, 2016–April 18, 2016	0.012
26	April 19, 2016–June 10, 2016	0.021
27	June 11, 2016–August 3, 2016	0.041
28	August 4, 2016–October 21, 2016	0.012
29	October 22, 2016–November 4, 2016	0.031
30	November 5, 2016–January 3, 2017	0.017
31	January 4, 2017–May 22, 2017	0.038
32	May 23, 2017–December 5, 2017	0.050
33	December 6, 2017–February 14, 2018	0.075
34	February 15, 2018–April 28, 2018	0.050
35	April 29, 2018–September 30, 2018	0.031

TABLE 8 CW statistics for HAR models with structural breaks for 1-day and 1-week horizons

This table reports the CW statistics for forecasting horizons of 1 day (Panel A) and 1 week (Panel B) for our models with structural breaks. A positive result indicates that the model whose name is in the first row outperforms the model whose name is provided in the first column. The statistic is computed using Newey and West (1987) HAC. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HAR-RV-J	HAR-RV-CJ	HAR-PS	HAR-PSL	HAR-RSV	HAR-RSV-J	HARQ	HAR-Q-J	HAR-Q-F	HAR-Q-F-J	HAR-RV-SJ	HAR-RV-CSJ	HAR-RV-SJD	HAR-RV-CSJD	HAR-J	HAR-RJ	HAR-ARJ
Panel A: 1-day horizon																	
HAR-RV	−0.376	−0.157	−0.240	4.448***	0.209	−0.250	7.650***	7.319***	7.965***	7.571***	1.131	1.539*	0.677	0.111	2.382***	−1.537	−0.223
HAR-RV-J		0.032	0.792	3.816***	1.624*	−0.518	7.573***	7.226***	7.815***	7.599***	3.014***	2.594***	1.086	−0.131	1.551*	−0.783	0.574
HAR-RV-CJ			1.405*	3.876***	2.359***	−0.327	7.354***	7.095***	7.524***	7.397***	3.336***	2.939***	1.285*	−0.084	1.798**	0.233	1.177
HAR-PS				3.686***	1.140	−0.643	7.714***	8.225***	7.647***	8.742***	1.658**	1.371*	0.852	−0.431	1.167	0.404	1.044
HAR-PSL					1.874**	2.579***	2.262**	3.035***	2.222**	3.031***	1.815**	1.767**	1.569*	−0.290	1.609*	1.695**	1.633*
HAR-RSV						−0.860	8.860***	8.616***	9.193***	9.238***	4.170***	1.505*	1.792**	−0.366	2.274**	2.355***	2.901***
HAR-RSV-J							4.043***	5.502***	3.706***	5.268***	1.623*	1.216	1.317*	−0.350	1.448*	1.423*	1.469*
HARQ								−0.319	0.630	0.050	−0.003	0.252	0.432	−0.879	0.210	−0.412	−0.309
HARQ-J									2.771***	0.722	2.614***	2.400***	2.007**	−0.309	2.146**	2.213**	2.177**
HARQ-F										−0.108	−0.819	−0.588	−0.030	−0.907	−0.584	−1.595	−1.424
HARQ-F-J											2.724***	2.481***	2.009**	−0.259	2.076**	2.120**	2.090**
HAR-RV-SJ												0.627	0.251	−0.458	0.737	−2.138	−0.499
HAR-RV-CSJ													1.780**	−0.350	2.815***	1.320*	2.403***
HAR-RV-SJD														1.445*	2.586***	1.673**	1.998**
HAR-RV-CSJD															1.146	1.147	1.152
HAR-J																−2.273	−1.054
HAR-RJ																	2.446***
Panel B: 1-week horizon																	
HAR-RV	0.263	1.060	−2.074	1.872**	0.041	−0.155	1.873**	1.873**	1.871**	1.870**	−1.321	−0.007	−2.015	0.039	−1.220	−1.684	−1.900
HAR-RV-J		1.136	−1.080	1.851**	−0.056	−0.278	1.853**	1.852**	1.850**	1.849**	−0.729	−0.109	−2.021	−0.060	−0.694	−1.226	−1.979
HAR-RV-CJ			−2.145	1.751**	−1.711	−1.709	1.753**	1.753**	1.750**	1.749**	−1.917	−1.717	−1.802	−1.689	−1.863	−2.053	−1.806
HAR-PS				1.876**	0.141	−0.060	1.878**	1.878**	1.875**	1.874**	−0.318	0.094	−2.016	0.150	−0.156	−1.207	−1.873
HAR-PSL					1.839**	1.832**	4.577***	4.378***	4.664***	3.520***	1.752**	1.852**	1.793**	1.860**	1.719**	1.721**	1.725**
HAR-RSV						−1.655	1.766**	1.766**	1.762**	1.761**	1.621*	−2.238	−1.844	0.596	1.636*	1.627*	−1.377
HAR-RSV-J							1.763**	1.763**	1.759**	1.758**	1.668**	1.617*	−1.823	1.684**	1.677**	1.675**	0.601
HARQ								−0.291	3.048***	2.276**	1.715**	1.829**	1.755**	1.850**	1.688**	1.692**	1.701**
HARQ-J									3.322***	2.311**	1.720**	1.828**	1.762**	1.840**	1.690**	1.689**	1.699**
HARQ-F										−2.489	1.636*	1.665**	1.691**	1.707**	1.620*	1.614*	1.626*
HARQ-F-J											1.691**	1.707**	1.743**	1.738**	1.671**	1.659**	1.666**
HAR-RV-SJ												0.161	−2.010	0.213	1.509*	−1.527	−1.780
HAR-RV-CSJ													−1.869	1.335*	1.586*	1.574*	−1.280
HAR-RV-SJD														1.385*	2.035**	2.050**	1.846**
HAR-RV-CSJD															1.626*	1.610*	−1.253
HAR-J																−1.783	−1.798
HAR-RJ																	−1.784

TABLE 9 CW statistics for HAR models with structural breaks for 1-month horizon

This table reports the CW statistics for forecasting horizons of 1 month for our models with structural breaks. A positive result indicates that the model whose name is in the first row outperforms the model whose name is provided in the first column. The statistic is computed using Newey and West (1987) HAC. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HAR-RV-J	HAR-RV-CJ	HAR-PS	HAR-PSL	HAR-RSV	HAR-RSV-J	HARQ	HAR-Q-J	HAR-Q-F	HAR-Q-F-J	HAR-RV-SJ	HAR-RV-CSJ	HAR-RV-SJD	HAR-RV-CSJD	HAR-J	HAR-RJ	HAR-ARJ
HAR-RV	1.858^{**}	3.569^{***}	0.864	4.250^{***}	1.083	2.024^{**}	4.252^{***}	4.252^{***}	4.232^{***}	4.231^{***}	0.589	0.856	−4.809	1.450*	−4.471	0.542	−0.977
HAR-RV-J		2.848^{***}	−5.284	3.976^{***}	1.098	1.622*	3.978^{***}	3.977^{***}	3.959^{***}	3.958^{***}	−5.027	0.972	−5.036	1.247	0.011	−4.913	−4.164
HAR-RV-CJ			2.415^{***}	4.555^{***}	−3.131	−0.461	4.557^{***}	4.556^{***}	4.573^{***}	4.571^{***}	2.418^{***}	−3.124	1.452*	−0.413	2.172^{**}	2.372^{***}	1.885^{**}
HAR-PS				4.282^{***}	1.436*	2.042^{**}	4.284^{***}	4.283^{***}	4.261^{***}	4.260^{***}	−4.011	1.237	−4.763	1.470*	0.751	−3.706	−0.271
HAR-PSL					6.103^{***}	6.110^{***}	3.826^{***}	4.000^{***}	7.244^{***}	6.978^{***}	5.769^{***}	6.239^{***}	5.657^{***}	6.469^{***}	5.595^{***}	5.678^{***}	5.465^{***}
HAR-RSV						2.131^{**}	5.093^{***}	5.092^{***}	5.074^{***}	5.073^{***}	3.054^{***}	−5.327	2.063^{**}	0.831	2.938^{***}	3.060^{***}	3.046^{***}
HAR-RSV-J							5.006^{***}	5.006^{***}	4.986^{***}	4.986^{***}	3.383^{***}	1.662^{**}	2.684^{***}	−5.738	3.410^{***}	3.388^{***}	3.274^{***}
HARQ								0.614	6.830^{***}	6.545^{***}	5.665^{***}	6.157^{***}	5.567^{***}	6.384^{***}	5.502^{***}	5.578^{***}	5.368^{***}
HARQ-J									6.740^{***}	6.312^{***}	5.676^{***}	6.196^{***}	5.575^{***}	6.410^{***}	5.513^{***}	5.581^{***}	5.362^{***}
HARQ-F										−1.644	0.242	0.345	−0.146	−0.638	0.714	0.285	0.359
HARQ-F-J											0.018	0.149	−0.399	−0.760	0.446	0.030	0.071
HAR-RV-SJ												1.389*	−4.502	1.609*	1.061	−1.741	1.090
HAR-RV-CSJ													2.288^{**}	1.237	3.201^{***}	3.310^{***}	3.346^{***}
HAR-RV-SJD														2.715^{***}	3.188^{***}	3.187^{***}	5.094^{***}
HAR-RV-CSJD															4.052^{***}	4.045^{***}	4.072^{***}
HAR-J																0.942	0.871
HAR-RJ																	1.246

TABLE 10 CW statistics for comparison between models with structural breaks and ones without structural breaks

This table reports the CW statistics for forecasting horizons of 1 day, 1 week, 1 month for our models with and without structural breaks. Positive statistics indicate that HAR models with structural breaks (alternative models) outperform the corresponding HAR model without structural breaks (null models). The statistic is computed using Newey and West (1987) HAC. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	$h = 1$	$h = 7$	$h = 30$
HAR-RV	1.794**	2.497***	0.731
HAR-RV-J(BPV)	1.332*	2.448***	1.656**
HAR-RV-CJ(BPV)	1.270	2.350***	4.248***
HAR-PS	1.654**	2.480***	0.424
HAR-PSL	7.087***	16.058***	21.680***
HAR-RSV	2.573***	4.895***	3.373***
HAR-RSV-J(BPV)	1.709**	4.352***	4.319***
HARQ	2.114**	−2.432	7.551***
HARQ-J(BPV)	1.353*	4.796***	15.937***
HARQ-F	7.548***	16.319***	21.556***
HARQ-F-J(BPV)	4.377***	17.191***	22.244***
HAR-RV-SJ(BPV)	1.975**	2.511***	0.108
HAR-RV-CSJ(BPV)	3.398***	5.411***	2.717***
HAR-RV-SJD(BPV)	1.569*	2.365***	−0.986
HAR-RV-CSJD(BPV)	1.920**	4.691***	2.415***
HAR-J(BPV)	2.100**	2.538***	0.339
HAR-RJ	2.040**	2.656***	0.357
HAR-ARJ	1.529*	2.484***	0.704

When we look at the 1-week forecasting horizon, we find that the vast majority of the statistics are positive and statistically significant, indicating that HAR models with structural breaks significantly outperform HAR models without structural breaks. Interestingly, when we study the 1-month and 2-month forecasting horizons, we find that the HAR models without structural breaks are sometimes again preferred although the majority of the statistics are positive, indicating that the inclusion of structural breaks improves the out-of-sample forecasting power of the HAR models. These results indicate the importance of including structural breaks in forecasting future volatility at forecasting horizons of 1 day and 1 week, but the results are mixed when including structural breaks in HAR model forecasts over 1-month and 2-month horizons.

5.2 | Alternative jump estimators

Andersen et al. (2012) point out that the standard multipower variations may be biased in finite samples and propose the "median realized variance estimator" (MedRV) which they show is more efficient and robust to jumps than its main rivals. Therefore, we re-estimate our HAR models and replace the BPV with the MedRV variation estimator, which can be defined as

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi}\left(\frac{M}{M - 2(k+1)}\right)\sum_{j=2k+3}^{M} \mathrm{med}\left(|r_{t\,j-2(k+1)}|, |r_{t\,j-(k+1)}|, |r_{t\,j}|\right)^2, \quad (30)$$

where med(·) stands for the median operator; for further details, see Andersen et al (2012). Here again we do not report all of the in-sample results to save space and focus instead on the out-of-sample analysis.[11] Table 11 reports the results, which are consistent with our previous analysis.

5.3 | Alternative estimation method

A potential issue with our previous analysis, as pointed out by Patton and Sheppard (2015), is that OLS estimation may put too much weight on highly volatile periods since volatility is the dependent variable in the model. To ensure that this is not the driving force behind our results, we re-estimate each model with WLS which employs the inverse of the fitted values as weights for the estimations. We do not report the results to save space, but the in-sample and out-of-sample results are very similar to before and our conclusions are not affected.[12]

6 | SUMMARY AND CONCLUSIONS

The volatility of Bitcoin has been a source of great interest, debate, and worry for investors since Bitcoin is one of the most volatile financial assets. Therefore, forecasting the volatility of Bitcoin is of great interest if investors are considering including Bitcoin in their investment portfolios. This paper presents a comprehensive analysis of the forecasting ability of 18 predictive HAR-type time-series models for the Bitcoin market. Having collected 5-minute high-frequency data on Bitcoin from Bitstamp, we employ OLS regression with Newey–West standard errors to estimate the model parameters. The in-sample results show that HAR models that include jumps offer more explanatory power than models that exclude jumps, with the HAR-ARJ model superior over the 1-day horizon and the HARQ-F-J model the best model over longer horizons. More importantly, the out-of-sample results also indicate that the inclusion of jumps improves the forecasting ability of HAR models, and that the HARQ-F-J model, which considers the temporal variation and the jump component, is the best model in the out-of-sample setting.

Since the volatility of Bitcoin is so high, and there is the strong possibility of structural breaks, we follow Wen et al. (2016) and re-estimate our HAR models but include structural breaks. We find that the inclusion of structural breaks improves the forecasting ability of our HAR models, especially over the 1-day and 1-week forecasting horizons. We also show that our results are robust to alternative jump estimators as well as using WLS as the estimation method rather than the OLS estimation method. Therefore, our findings suggest that modelling jumps,

[11]Again, the in-sample results are available upon request from the corresponding author.

[12]The full results of the WLS regression are available upon request from the corresponding author.

TABLE 11 CW statistics for models using MED jumps estimator

This table reports the CW statistics for forecasting horizons of 1 day (Panel A), 1 week (Panel B), and 1 month (Panel C) for HAR-typemodels where jumps are esitmated using median realized variance estimators. A positive result indicates that the model whose name is in the first row outperforms the model whose name is provided in the first column. The statistic is computed using Newey and West (1987) HAC. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HAR-RV-CJ	HAR-RSV-J	HARQ-J	HARQ-F-J	HAR-RV-SJ	HAR-RV-CSJ	HAR-RV-SJD	HAR-RV-CSJD
Panel A: 1-day horizon								
HAR-RV-J	−1.907	−0.880	0.789	2.492***	1.533*	1.298*	1.396*	−0.713
HAR-RV-CJ		−0.661	1.498*	3.298***	2.296**	1.815**	2.162**	−0.540
HAR-RSV-J			1.432*	1.508*	1.202	1.050	1.166	−0.723
HARQ-J				2.888***	1.394*	1.308*	1.300*	−0.731
HARQ-F-J					1.544*	1.433*	1.487*	−0.671
HAR-RV-SJ						−0.362	−0.136	−0.809
HAR-RV-CSJ							4.470***	−0.251
HAR-RV-SJD								−0.327
Panel B: 1-week horizon								
HAR-RV-J	−1.299	−0.353	2.286**	17.977***	1.084	0.681	0.942	−0.338
HAR-RV-CJ		0.422	3.593***	18.565***	1.546*	1.072	1.407*	0.051
HAR-RSV-J			3.564***	9.154***	1.880**	0.985	1.797**	−0.356
HARQ-J				18.449***	1.215	0.753	1.050	−0.199
HARQ-F-J					−2.776	−1.462	−2.951	−1.671
HAR-RV-SJ						−1.705	−0.098	1.573*
HAR-RV-CSJ							10.337***	2.069**
HAR-RV-SJD								1.803**
Panel C: 1-month horizon								
HAR-RV-J	20.201***	−8.259	2.057**	19.325***	1.204	−6.418	0.444	−4.709
HAR-RV-CJ		−12.635	−1.813	17.497***	−7.729	−10.765	−8.425	−8.206
HAR-RSV-J			12.645***	19.366***	13.938***	7.945***	13.801***	3.972***
HARQ-J				20.136***	0.442	−0.961	0.362	−6.624
HARQ-F-J					−3.547	−3.588	−3.660	−12.811
HAR-RV-SJ						−7.917	−1.843	−5.903
HAR-RV-CSJ							13.866***	4.500***
HAR-RV-SJD								−5.367

especially the temporal variation and structural breaks, significantly improves the accuracy of volatility forecasts of Bitcoin through popular HAR models.

Therefore, our results are consistent with the notion that the Bitcoin market has large "whales", large players who have a big price impact within Bitcoin markets (Scaillet et al., 2018). This helps explain the large number of jumps and structural breaks in the Bitcoin price. Further, the 24 hours a day, seven days a week trading structure ensures that the Bitcoin market does not always have the attention of all traders and that there are episodes of illiquidity (as shown in Eross, McGroarty, Urquhart, & Wolfe, 2019). With the introduction of Bitcoin futures in December 2017, investors are now able to speculate on falling Bitcoin prices and may reduce the number of jumps and structural breaks found in the Bitcoin market.

REFERENCES

Andersen, T. G., Benzoni, L., & Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance*, *57*, 1239–1284.

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*, 115–158.

Andersen, T. G., & Bollerslev, T. (1998). Answering the sceptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*, 885–905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics*, *89*, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of stock return volatility. *Journal of Financial Econometrics*, *61*, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*, 579–625.

Andersen, T. G., Bollerslev, T., & Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, *160*, 220–234.

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbour truncation. *Journal of Econometrics*, *169*, 75–93.

Bajgrowicz, P., Scaillet, O., & Treccani, A. (2015). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management science*, *62*(8), 2198–2217.

Barndorff-Nielson, O. E., Kinnebrock, S., & Shephard, N. (2010). Measuring downside risk – realized semivariance. In T. Bollerslev, J. Russell & M. Watson (Eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle*. Oxford: Oxford University Press.

Barndorff-Nielson, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, *2*, 1–37.

Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking and Finance*, *96*, 355–367.

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1–18.

Borri, N. (2019). Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, *50*, 1–19.

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 Research Letters*, *20*, 192–198.

Catania, L., Grassi, S., & Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting*, *35*, 485–501.

Çelik, S., & Ergin, H. (2014). Volatility forecasting using high frequency data: Evidence from stock markets. *Economic Modelling*, *36*, 176–190.

Chaim, P., & Laurini, M. P. (2018). Volatility and return jumps in Bitcoin. *Economics Letters*, *173*, 158–163.

Charles, A., & Darné, O. (2014). Large shocks in the volatility of the Dow Jones Industrial Average index: 1928-2013. *Journal of Banking and Finance*, *43*, 188–199.

Cheah, E. T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, *130*, 32–36.

Chen, X., & Ghysels, E. (2011). News – good or bad – and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, *24*, 46–81.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Corbet, S., & Katsiampa, P. (2018). Asymmetric mean reversion of Bitcoin price returns. *International Review of Financial Analysis*, https://doi.org/10.1016/j.irfa.2018.10.004.

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, *62*, 182–199.

Corbet, S., Lucey, B., & Yarovaya, L. (2018). Datestamping the Bitcoin and Ethereum bubbles. *Finance Research Letters*, *26*, 81–88.

Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, *165*, 28–34.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174–196.

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, *30*(3), 368–380.

Corsi, F., Pirino, D., & Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, *159*(2), 276–288.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*, 253–263.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, *59*, 1367–1403.

Eraker, B., Johannes, M. S., & Polson, N. G. (2003). The impact of jumps in returns and volatility. *Journal of Finance*, *53*, 1269–1300.

Eross, A., McGroarty, F., Urquhart, A., & Wolfe, S. (2019). The intraday dynamics of Bitcoin. *Research in International Business and Finance*, *49*, 71–81.

Giot, P., & Laurent, S. (2007). The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets*, *27*, 337–359.

Gronwald, M. (2019). Is Bitcoin a commodity? On price jumps, demand shocks, and certainty of supply. *Journal of International Money and Finance*, *97*, 86–92.

Huang, X., & Tauchen, G. (2005). The relative price contribution of jumps to total price variance. *Journal of Financial Econometrics*, *3*, 456–499.

Hudson, R., & Urquhart, A. (2019). Technical trading and cryptocurrencies. *Annals of Operations Research*, https://doi.org/10.1007/s10479-019-03357-1.

Inclàn, C., & Taio, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of American Statistical Association*, *89*, 913–923.

Kajtazi, A., & Moro, A. (2019). The role of Bitcoin in well diversified portfolios: A comparative global study. *International Review of Financial Analysis*, *61*, 143–157.

Kalyas, A., Papakyriakou, P., Sakkas, A., & Urquhart, A. (2019). What drives Bitcoin's price crash risk? *Economics Letters*, 108777. https://doi.org/10.1016/j.econlet.2019.108777

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

Katsiampa, P. (2018). Volatility co-movement between Bitcoin and Ether. *Finance Research Letters*, *30*, 221–227.

Katsiampa, P., Corbet, S., & Lucey, B. (2019). High frequency volatility co-movements in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money*, *62*, 35–52.

Khuntia, S., & Pattanayak, J. (2018). Adaptive market hypothesis and evolving predictability of Bitcoin. *Economics Letters*, *167*, 26–28.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, *12*, 445–475.

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*, 293–311.

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, *150*, 6–9.
Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Available at https://bitcoin.org/bitcoin.pdf
Newey, W. K., & West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.
Patton, A. J., & Sheppard, K. (2011). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, *97*(3), 683–697.
Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, *97*, 683–697.
Phillip, A., Chen, J., & Peiris, S. (2018). A new look at cryptocurrencies. *Economics Letters*, *163*, 6–9.
Platanakis, E., & Urquhart, A. (2019). Should investors include Bitcoin in their portfolios? A portfolio theory approach. *British Accounting Review*, 100837. https://doi.org/10.1016/j.bar.2019.100837
Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2016). Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets*, *36*, 758–792.
Scaillet, O., Treccani, A., & Trevisan, C. (2018). High-frequency jump analysis of the Bitcoin market. *Journal of Financial Econometrics*, https://doi.org/10.1093/jjfinec/nby013
Shahzad, S. J. H., Mensi, W., Hammoudeh, S., Balcilar, M., & Shahbaz, M. (2018). Distribution specific dependence and causality between industry-level U.S. credit and stock markets. *Journal of International Financial Markets, Institutions and Money*, *52*, 114–133.
Shen, D., Urquhart, A., & Wang, P. (2019). Does Twitter predict Bitcoin? *Economics Letters*, *174*, 118–122.
Shiller, R. J. (2017). What is Bitcoin really worth? Don't even ask. *New York Times*, December 15. Retrieved from https://www.nytimes.com/2017/12/15/business/bitcoin-investing.html
Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, *235*, 643–659.
Tauchen, G., & Zhou, H. (2011). Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics*, *160*, 102–118.
Tiwari, A. K., Jana, R. K., Das, D., & Roubaud, D. (2018). Informational efficiency of Bitcoin – An extension. *Economics Letters*, *163*, 106–109.
Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, *148*, 80–92.
Urquhart, A. (2017). Price clustering in Bitcoin. *Economics Letters*, *159*, 145–148.
Urquhart, A. (2018). What causes the attention of Bitcoin? *Economics Letters*, *166*, 40–44.
Urquhart, A., & Zhang, H. (2018). Is Bitcoin a hedge or safe-haven for currencies? An intraday analysis. *International Review of Financial Analysis*, *63*, 49–57.
Vivian, A., & Wohar, M. E. (2012). Commodity volatility breaks. *Journal of International Financial Markets, Institutions and Money*, *22*, 395–422.
Wen, F., Gong, X., & Cai, S. (2016). Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. *Energy Economics*, *59*, 400–413.
West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*, 1067–1084.
Yarovaya, L., Brzeszczyński, J., & Lau, C. K. M. (2016). Intra- and inter-regional return and volatility spillovers across emerging and developed markets: Evidence from stock indices and stock index futures. *International Review of Financial Analysis*, *43*, 96–114.

How to cite this article: Shen D, Urquhart A, Wang P. Forecasting the volatility of Bitcoin: The importance of jumps and structural breaks. *Eur Financ Manag*. 2019;1–30. https://doi.org/10.1111/eufm.12254