Forecasting realized volatility of Bitcoin: The role of the trade war

Elie Bouri[a], Konstantinos Gkillas[b], Rangan Gupta[c], Christian Pierdzioch[d]

January 2020

Abstract

We analyze the role of the US-China trade war in predicting, both in- and out-of-sample, daily realized volatility of Bitcoin returns. We study intraday data spanning from 1st July 2017 to 30th June 2019. We use the heterogeneous autoregressive realized volatility model (HAR-RV) as the benchmark model to capture stylized facts such as heterogeneity and long-memory. We then extend the HAR-RV model to include a metric of US-China trade tensions. This is our primary predictor of interest, and it is based on Google Trends. We also control for jumps, realized skewness, and realized kurtosis. For our empirical analysis, we use a machine-learning technique which is known as random forests. Our findings reveal that US-China trade uncertainty does improve forecast accuracy for various configurations of random forests and forecast horizons.

JEL classification: G17; Q02; Q47

Keywords: Bitcoin; Realized volatility; Trade war; Random forests

[a] USEK Business School, Holy Spirit University of Kaslik, Jounieh, Lebanon; Email address: eliebouri@usek.edu.lb.

[b] Department of Business Administration, University of Patras – University Campus, Rio, P.O. Box 1391, 26500 Patras, Greece; Email address: gillask@upatras.gr.

[c] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; E-mail address: rangan.gupta@up.ac.za.

[d] Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany; Email address: c.pierdzioch@hsu-hh.de.

Acknowledgments:

The research of C. Pierdzioch was supported by the German Science Foundation (Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

1 Introduction

In the wake of the recent United States-China trade war (hereafter, US-China trade war), and soaring Bitcoin prices, there have been claims made by both financial practitioners, covered in the financial press, and academics that these two facts are not necessarily independent, but rather can be considered as an indication of Bitcoin's hedging ability. To give an example, in an recent interview in Fortune's entitled "Balancing the Ledger", the founder of the Digital Currency Group, Barry Silbert, claimed that Bitcoin behaves as an asset which seems to be independent from various uncertainties that exist in the traditional financial system. In other words, according to Barry Silbert, the digital currency can be considered as a "flight to safety". Such behavior, however, is not new. Bitcoin acted as a "flight to safety" quite earlier, for example, during the begging of "Brexit" negotiations and at the peak of "Grexit" debates in Europe. Not surprisingly, using Google to search the terms "Bitcoin" and "Trade war" gave 14,000,000 results (Google.com accessed on December 20, 2019). In this vein, as the trade war intensified, several market watchers thought that indeed Bitcoin benefited from jitters in international financial markets, which, in turn, gave rise to a downward pressure on stocks and China's currency.

The claims of the hedging and "flight to safety" property of Bitcoin mentioned in the preceding paragraph are based on anecdotal evidence, while academic research on this matter is relatively scarce. Formal empirical evidence of Bitcoin acting as a hedge against trade-related uncertainties can be found in the recent works by Gozgor et al. (2018) and Bouri et al. (2019a). On the one hand, the former researchers, based on a wavelet analysis, claim that, with some exceptions, there is a positive correlation between Bitcoin returns and a newspaper-based measure of trade policy uncertainty. Bouri et al. (2019a), in turn, report that the (realized) correlation between US equities and Bitcoin returns is negatively affected by the same measure of trade uncertainty.

Taking into account the claims discussed above along with the general importance of volatility for risk management and portfolio choice, a relevant question to ask is: “How has the US-China trade war impacted the volatility of Bitcoin returns?", and, particular, “Does uncertainty caused by the US-China trade add incremental predictive value useful for forecasting the volatility of Bitcoin returns?". Volatility is a widely used measure of risk and, therefore, its accurate modeling and forecasting play a key role in investment decisions and portfolio choice (Poon and Granger 2003). In light of this, as pointed out by Baur and Dimfl (2018) for the cryptocurrency market, positive returns are closely associated with higher levels of volatility. One explanation of this returns-volatility nexus points to the influence of uninformed investors' herding, because such investors buy due to their fear of missing out on rising cryptocurrency valuations and pump and dump (known also as P&D) scams. In other words, one can argue that increases in Bitcoin returns associated with the US-China trade war, and uncertainty in general (see e.g., Bouri et al. 2017, 2018, Fang et al. 2018, Aysan et al. 2019, Bouri and Gupta 2019, Wu et al. 2019), are likely to result in heightened volatility. However, Baur and Dimfl (2018) find this evidence to be (statistically) weak for Bitcoin and Ethereum among 20 cryptocurrencies considered, which can be interpreted to indicate that these two cryptocurrencies are possibly dominated by informed investors.

Against this backdrop, we study the role of the US-China trade war in predicting, both in- and out-of-sample, daily (realized) volatility of Bitcoin returns. To this end, we use intraday Bitcoin returns as measured in 60 minute-intervals (hourly basis) covering the period from 1st July 2017 to 30th June 2019. We focus on realized volatility estimated by non-parametric techniques, which provides an accurate estimator of volatility on the basis of the actual variance if intraday returns.[1]

[1]In the empirical-finance literature, the term volatility is often used for the for the standard

In particular, any non-parametric estimator is based on quadratic variation, which, in turn, is considered to be the best estimator of (latent) volatility. Moreover, usage of the realized version of daily volatility, which is basically defined as the sum of non-overlapping squared intraday returns for a given interval (e.g., within a day), transforms volatility into an observable process. According to McAleer and Medeiros (2008), intraday data contain rich information about market conditions (such as the microstructure of the market), producing more accurate estimates of daily realized volatility.

As for the econometric framework concerned in this research for forecasting realized volatility, we make use of the heterogeneous autoregressive realized-volatility model (HAR-RV). This model, as proposed by Corsi (2009), allows stylized facts (such as multi-scaling behavior and long-memory, as detected for Bitcoin by Bouri et al. (2019b) and Takaishi (2018) respectively) of the volatility process to be captured in a straightforward and simple way. We also control for the discontinuity property of realized volatility, known as jumps, given that jumps are well-known to improve the overall fit of realized volatility models (see, e.g., Andersen et al. 2007, Bollerslev et al. 2009, Corsi et al. 2010, Neuberger and Payne 2018, Gkillas et al. forthcoming), and observed for cryptocurrencies (Bouri et al. 2019c). In order to capture left-tail events far away from the mean, following Amaya et al. (2015), we use realized skewness and realized kurtosis (see also Mei et al. 2017, and Gkillas et al. 2019 in particular in relation to the Bitcoin market). Importantly, we include in the HAR-RV model our primary predictor of interest, namely a metric for the US-China trade tension, for which we rely on Google Trends (as has been done by studies such as (Kristoufek 2013, Panagiotidis et al. 2018, 2019, Nasir et al. 2019, Subramaniam and Chakraborty 2019) for search-terms like: “US-China Trade War" (primarily), “US-China Tariffs", “Tariffs War",

deviation of returns. In our research, because there is no risk of confusion, we use the term volatility to denote the realized variance of Bitcoin returns (as defined in Equation (1) below).

“Tariffs War US-China", “Trump Trade War", and “New Trade War Tariffs", and then consider a composite of these terms based on principal component analysis. As an alternative measure, we also consider the news-based measure of US trade policy uncertainty as developed by Caldara et al. (2019).

In terms of estimation strategy, we rely on a machine-learning technique that is known in the statistical-learning literature as random forests. Recent applications of this technique in the empirical finance literature include Gupta et al. (2019), Demirer et al. (2019), and Pierdzioch and Risse (forthcoming), among others. As compared to the ordinary-least squares technique (OLS) commonly applied in earlier literature to estimate the HAR-RV model and its various extensions, random forests have several advantages (for a discussion of random forests and a comparison with other machine-learning techniques, see Hastie et al. 2009). One advantage is that random forests can be interpreted as a data-driven modeling environment that renders it possible to study, in a unified framework, the predictive ability of a large number of predictor variables for realized volatility. Such a modeling environment is ideally suited for our investigation (which concerns the incremental predictive value of trade-related uncertainty for realized volatility) due to the fact that we control for the impact of various measures of jumps (i.e., upside/downside, asymmetric, truncated large/small jumps), higher-order moments (i.e., realized skewness and realized kurtosis), returns, and a leverage effect in addition to the usual components of the benchmark HAR-RV model. Another advantage of random forests is that this modeling environment makes it not only possible to capture in a data-driven way the interdependencies between the predictor variables but also, at the same time, renders it possible to account for potentially nonlinear links between realized volatility and its predictor variables. This advantage is particularly important in our study given the extremely volatile behavior of cryptocurrency returns and their nonlinear dynamics (Gkillas and Longin 2019). Finally, random forests, by their very construction, guarantee that forecasts of realized volatility, even when we study relatively complex models that include simultaneously several predictor variables, are always nonnegative, a feature not shared by the ordinary-least squares technique. Yet another advantage of random forests is that they permit numerous realized volatility out-of-sample predictions to be computed with ease given that random forests can be efficiently estimated in terms of computing time.

Finally, it must be noted that a large literature has emerged that has aimed to predict (in- and out-of sample) daily price movements in the volatility of cryptocurrencies and in particular Bitcoin, based on same-frequency or mixed-frequency variants of the popular Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, using various types of financial and macroeconomic variables as predictors (see, e.g., Chu et al. 2017, Conrad et al. 2018, Ardia et al. 2019, and Walther et al. 2019, for detailed reviews of this literature). However, our study is the first study that uses random forests to analyze the role of US-China trade war in predicting Bitcoin's daily realized volatility by accounting for volatility jumps, realized skewness and realized kurtosis, all constructed from intraday data.

The remainder of the paper is organized as follows. Section 2 presents the empirical methods used. Section 3 describes the data. Section 4 summarizes the empirical results. Section 5 concludes.

2 Method and predictors

2.1 Random forests

Following Andersen and Bollerslev (1998), we consider a realized volatility measure constructed by using realized variance, which we denote by RV_t. Realized variance, RV_t, is a benchmark volatility estimator because it is considered as a consistent estimator of the integrated variance, including the jump contribution to the latter. In order to put it differently, RV_t is a non-parametric estimator of volatility and, therefore, provides an accurate estimator of volatility on the basis of the actual variance. As we already mentioned in the introductory section, any non-parametric estimator is based on quadratic variation, which is considered to be the best estimator of (latent) volatility. In this research, for each day t, we construct a daily point estimate of RV_t using intraday Bitcoin returns. A daily point estimate of (RV_t) is constructed by summing up all successive intraday squared Bitcoin log-returns as follows

$$RV_t = \sum_{i=1}^{N} X_{t,i}^2, \quad (1)$$

where $X_{t,i}$ denotes the intraday Bitcoin log-return for hour i within day t, while $i = 1,..,N$ denotes the total number of intraday Bitcoin log-returns within a day.

As noted in the introductory section of this paper, we apply a technique known in the literature on machine learning and statistical learning as random forests to model realized volatility of Bitcoin returns. In contrast to the common approach to estimate the widely-studied HAR-RV model, which is estimated by means of the OLS technique, random forests are a modeling environment that allows the links between realized volatility and a large number of predictors, including US-Chine trade-war related uncertainty, to be modeled in a fully data-driven way.

A random forest consists of several individual regression trees (for an introduction, we refer to the textbook by Hastie et al. 2009, Chapter 9; our notation follows the notation they use in their textbook). A regression tree, T, consists of brances that recursively partition the space of predictor variables, $\mathbf{x} = (x_1, x_2, ...)$, of realized volatility of Bitcoin returns into l non-overlapping regions, R_l. The list of predictor variables we use in this research comprises the usual HAR-RV predictors (i.e., daily RV_t, weekly $RV_{w,t}$, and monthly $RV_{m,t}$), US-Chine trade-war related uncertainty, and several other predictors commonly studied in the literature on realized volatility (see Section 2.2 for details). A simple greedy algorithm governs the recursive partitioning of the the space of predictor variables in a top-down and binary way. At the top level of a regression tree, the algorithm selects the first partition in such a manner that the partitioning predictor variable, s, and the realization of this predictor variable that is selected as a partitioning point, p, define the half-planes $R_1(s,p) = \{x_s | x_s \leq p\}$ and $R_2(s,p) = \{x_s | x_s > p\}$ in a way such that a standard squared-error loss function is minimized:

$$\min_{s,p} \left\{ \min_{\bar{RV}_1} \sum_{x_s \in R_1(s,p)} (RV_i - \bar{RV}_1)^2 + \min_{\bar{RV}_2} \sum_{x_s \in R_2(s,p)} (RV_i - \bar{RV}_2)^2 \right\}, \tag{2}$$

where i stands for the data on realized volatility that belong to the half-planes, $\bar{RV}_k = \text{mean}\{RV_i \,| x_s \in R_k(s,p)\}, k = 1,2$ stands for the half-plane-specific mean of realized volatility of Bitcoin returns. Note that the inner minimization is carried out so as to to minimize the region-specific squared error loss by means of an optimal choice of the half-plane-specific means.[2] The outer minimization consists of a brute-force search over all combinations of s and p. This search is done so as to identify the first optimal partitioning predictor variable, the first optimal partitioning point, and the two region-specific means of Bitcoin realized volatility.

Next, the minimization given in Equation (2) is applied to the new regression tree, which now has two-terminal-nodes. The minimization is applied separately to the two optimal top-level half-planes, $R_1(s,p)$ and $R_2(s,p)$, chosen in the first step. This minimization yields up to two second-level optimal partitioning predictor variables and optimal partitioning points, and four second-level region-specific means of realized volatility of Bitcoin returns. We repeat this search-partition

[2]To be precise, we should use the notation RV_{t+h} in Equation (2) because we forecast realized volatility, where h is an index for the forecast horizon. For ease of notation, we use a somewhat simpler but less precise notation in Equation (2).

process until the regression tree has a preset maximum number of terminal nodes or every terminal node has a minimum number of observations, both of which can be specified by a researcher as hyperparameters.

When the search-partition process stops, the regression tree sends the predictor variables in a top-down way along its various optimal nodes (that is, partitioning points) and branches to its terminal nodes and then predicts the realized volatility by its region-specific mean. When the tree has L regions, we have

$$T\left(\mathbf{x}_i, \{R_l\}_1^L\right) = \sum_{l=1}^{L} \bar{RV}_l \mathbf{1}(\mathbf{x}_i \in R_l), \tag{3}$$

where $\mathbf{1}$ is the indicator function.

Growing a large enough regression tree renders it possible to compute finer and finer granular predictions of realized volatility via Equation (3). A key problem of such an approach, however, is that its hierarchical structure makes a complex regression tree very data-sensitive, typically resulting in poor forecasting performance. An efficient way to circumvent this problem is to combine a large number of individual regression trees to form a random forest (Breiman 2001). To this end, a large number of bootstrap samples is obtained from the data, and in each bootstrap sample a random regression tree is estimated. A standard regression tree is different from a random regression tree in that the latter selects for every partitioning step only a random subset of the predictor variables. This random selection of predictor variables mitigates the influence of influential predictor variables on tree building and, thereby, decorrelates the predictions from individual regression trees. This decorrelation of predictions stabilizes the predictions of realized volatility of Bitcoin returns.

2.2 Predictor variables

2.2.1 The Core HAR-RV Model

The heterogeneous market hypothesis (Müller et al. 1997) forms the theoretical foundation of the HAR-RV model of realized-volatility. The main idea behind this hypothesis and, thus, the HAR-RV model is to combine volatility measures from different time resolutions and, thereby, to capture key properties of the volatility process such as heterogeneity and long memory. The baseline model for realized volatility forecasting follows the widely-used HAR-RV model developed by Corsi (2009). This model stipulates that (in addition to a constant intercept term) for h-days-ahead RV denoted by RV_{t+h}, we use as predictors weekly realized volatility, denoted by $RV_{w,t}$, and monthly realized volatility, denoted by $RV_{m,t}$, where h is the forecast horizon, $RV_{w,t}$ is the average RV from day $t-5$ to day $t-1$, and $RV_{m,t}$ is the average RV from day $t-22$ to day $t-1$. In this study, following the earlier literature, we define RV_{t+h} as mean$\{RVt+1,...,RVt+h\}$ studying three forecast horizons: the short daily forecasting horizon (where h is equal to 1), the medium weekly forecasting horizon (where h is equal to 5), and the long monthly forecasting horizon (where h is equal to 22).

2.2.2 Detecting total jumps

In addition to the predictor variables employed in the core HAR-RV model, we make use of several other predictor variables for growing random forests, including several measures that disentangle the continuous and jumps components of realized volatility. In doing so, we detect various types of volatility jumps (hereafter jumps) and signify upside/downside and asymmetric jumps and truncated large/small jumps. To this end, we apply a benchmark jump-detection scheme proposed by Huang and Tauchen (2005), Barndorff-Nielsen and Shephard (2006), and Andersen et al. (2007). In particular, we consider the following general stochastic volatility jump-diffusion model for the log-price P_t of a risky asset:

$$dP_t = \mu_t dt + \sigma_t dW_t + k_t dq_t, \quad 0 \leq t \leq N, \tag{4}$$

where μ_t denotes the drift term with a continuous variation sample path, σ_t dentes a stochastic volatility process (which is strictly positive), W_t denotes a standard Brownian motion, and $k_t dq_t$ denotes the random jump size.

For a discrete price process, the volatility at a given day t includes jump variation. In others words, a daily point estimate of integrated volatility can be derived by the quadratic variation, denoted by QV_t, which is the best estimator of total variation (that is, of integrated volatility). QV_t is given by

$$QV_t = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s<t} \kappa_s^2 t, \tag{5}$$

where $\int_{t-1}^{t} \sigma_s^2 ds$ is the continuous sample path variation, while $\sum_{t-1<s<t} \kappa_s^2 t$ denotes the discontinuous jump variation. Under weak regularity conditions and $N \to \infty$, RV_t can be considered as a consistent estimator of QV_t:

$$RV_t \xrightarrow{N\to\infty} QV_t = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s<t} \kappa_s^2 t, \tag{6}$$

where QV_t is decomposed into its continuous and jump components.

As we already mentioned, the methodology for detecting jumps we use in this research is based on the decomposition of volatility into a jump and a continuous component of RV_t. Hence, building on our discussion above, we need an estimator to separate the continuous prices moves, $\int_{t-1}^{t} \sigma_s^2 ds$, from total variation, that is, we need an estimator that excludes jumps. Barndorff-Nielsen and Shephard (2006) propose an accurate estimator of the integrated variance which excludes jumps, the so-called realized bipower variation (RBV). RBV is defined as follows:

$$RBV_t = \mu_1^{-2} \sum_{i=2}^{N} |X_{t,i}||X_{t,i-1}|, \tag{7}$$

where $X_{t,i}$ is the intraday Bitcoin log-return for i hour within day t, $i = 1,..,N$ denotes the total number of intraday Bitcoin log-returns within a day, and μ_1 is given by $E(|Z|^{\alpha})$ for α equal to 1 and $Z \sim N(0,1)$. In other words, μ_1 is the mean of a standard Gaussian random variable, Z, in absolute values, while for α equal to 1, $E(|Z|^1)$ is equal to $\sqrt{(2/\pi)}$.

Like Andersen et al. (2007), we employ the logarithmic version of the Huang and Tauchen (2005) ratio-statistic (also called jump-ratio test), denoted by $J_t^{(RBV)}$, to detect the discontinuous jump variation. Under the null hypothesis of no jumps, $J_t^{(RBV)}$ along with its logarithmic version, is a useful pre-test, prior jump detection. It should be noted that, in line with Andersen et al. (2007), we did not find any statistically significant difference between the plain jump statistic and its logarithmic version. More specifically, the logarithmic version $J_t^{(RBV)}$is given by:

$$J_t^{(RBV)} = \frac{(RV_t - RBV_t)\,RV_t^{-1}}{\left(\left(\mu_1^{-4} + 2\mu_1^{-2} - 5\right) max\left\{1, RTQ_t RBV_t^{-2}\right\}\right)^{1/2}} \rightarrow N(0,1), \qquad (8)$$

where RV_t is the realized variance defined in Equation (1) and is an asymptotically consistent estimator of the integrated variance including also the jump component. RTQ_t is an estimator of integrated quarticity, which is asymptotically unbiased in the absence of microstructure noise, and converges in probability to integrated quarticity. The RTQ_t is given by:

$$RTQ_t = N\xi_{4/3}^{-3} \sum_{i=3}^{N} |X_{t,i}|^{4/3} |X_{t,i-1}|^{4/3} |X_{t,i-2}|^{4/3}, \qquad (9)$$

where $X_{t,i}$ again denotes the intraday Bitcoin log-return for i hour within day t and $i = 1,..,N$ denotes the total number of intraday Bitcoin log-returns within a day.

Applying the $J_t^{(RBV)}$ test to a jump-detection scheme, we detect total jumps, denoted by RJ_t, by the following condition:

$$RJ_t = \mathbf{1}(J_t^{(RBV)} > \Phi_{\alpha})|RV_t - RBV_t|, \qquad (10)$$

where $\mathbf{1}$ is the indicator function. A jump is statistically different from zero when the $J_t^{(RBV)}$ test exceeds a critical value of a Gaussian distribution, denoted by Φ_α, for a given level of significance α.

We use this jump-detection scheme to detect upside/downside and asymmetric jumps as well as truncated large/small jumps.

2.2.3 More jumps: Upside, downside and asymmetric jumps

Guo et al. (2019), among others, suggest that considering potential asymmetry in the volatility process can improve the predictability of models for future excess returns. In this vein, under a realized estimation framework, Barndorff-Nielsen et al. (2010) proposed upside and downside realized semi-variances, denoted by (RV^+ and RV^-), as realized measures that are based on entirely on upward and downward intraday price movements. RV^+ and RV^- measures are defined as

$$RV_t^+ = \sum_{i=1}^{N} X_{t,i}^2 \mathbf{1}(X_{t,i} > 0), \tag{11}$$

$$RV_t^- = \sum_{i=1}^{N} X_{t,i}^2 \mathbf{1}(X_{t,i} < 0), \tag{12}$$

where $\mathbf{1}$ is the indicator function. Following Bollerslev et al. (forthcoming), we consider the realized up term as daily RV^+ (hereafter "good" realized volatility) constructed using the variability of Bitcoin upside price movements only from positive intraday returns. We also consider the realized down term as daily RV^- (hereafter "bad" realized volatility) constructed using the variability of Bitcoin downside price movements only from negative intraday returns. Good and bad realized volatility can be considered as measures of upside and downside risk, capturing the sign asymmetry in the volatility process. Several studies have identified the importance of upside and downside risk in portfolio-risk assessments and management. Realized volatility, and its downside and upside components, as a measures of risk is more accurately estimated with the use of intraday data at a daily basis (Hansen and Huang 2016). Intraday data reveal various information about, for example, the microstructure of the market, which are not easily observed at lower frequencies. In light of this, several studies have used HAR-type models along with RV^+ and RV^- to forecast realized volatility (see, e.g., Patton and Sheppard 2011, Sévi 2014, Chen et al. 2019).

Building on the concept of RV^+ and RV^-, we study an additional list of predictors to capture the sign asymmetry of the jumpy component of the volatility process, and therefore, the incremental predictive content of the US-Chine trade war for forecasting Bitcoin volatility to be isolated. In doing so, by applying the jump-detection scheme defined in Equation (8), we estimate upside and downside jumps, denoted by RJ_t^+ and RJ_t^-, receptively. Following the research by Duong and Swanson (2011, 2015), RJ_t^+ and RJ_t^- are formulated by power transformation as follows:

$$RJ_t^+ = \mathbf{1}(J_t^{(RBV)} > \Phi_\alpha) \sum_{i=1}^{T} |X_{t,i}|^q \mathbf{1}(X_{t,i} > 0), \tag{13}$$

$$RJ_t^- = \mathbf{1}(J_t^{(RBV)} > \Phi_\alpha) \sum_{i=1}^{T} |X_{t,i}|^q \mathbf{1}(X_{t,i} < 0), \tag{14}$$

where again $\mathbf{1}$ denotes the indicator function. We consider RJ_t^+ and RJ_t^+ as "good" and "bad" jumps, respectively. RJ_t^+ is constructed using only Bitcoin's positive intraday returns, while RJ_t^- is constructed using only Bitcoin's negative intraday returns. The parameter q defines the asymmetry parameter, which ranges from 2 to 6 and effects the limiting behavior of the RJ_t estimator. Very low values (at the lower bound of 2) cannot guarantee finite jump variation. Like Duong and Swanson (2011), we set q equal to 2.5. From RJ_t^+ and RJ_t^+, we estimate asymmetric jumps (the so-called signed jump), denoted by RJA_tQ, as follows:

$$RJA_t = \mathbf{1}(J_t^{(RBV)} > \Phi_\alpha)\, RJ_t^+ - RJ_t^-. \tag{15}$$

The interpretation of RJA_t is straightforward. On the one hand, a positive value means that when a jump occurs at day t it is related with events that mainly impact upside (good) volatility. On the other hand, a negative sign means that the jump is related with events that mainly impact downside (bad) volatility.

2.2.4 Truncated large and small jumps

We further estimate another set of jump covariates using decomposition based on truncation levels instead of power transformation. In particular, following Duong and Swanson (2011), we estimate the realized measure of truncated large jumps denoted by(RLJ_t), and truncated small jumps, denoted by (RSJ_t), via a decomposition by selecting a suitable threshold (fixed truncation level) that can separate the (RLJ_t) from RSJ_t). In the existing literature, HAR-type models including both RLJ_t and RSJ_t can provide significantly superior forecasting accuracy at longer forecasting horizons (such as weekly or monthly) for traditional assets (see e.g., Liu et al. 2015). By using the jump detection scheme defined in Equation (8), the $RVLJ_t$ and $RVSJ_t$ are estimated by the following:

$$RLJ_t{=}min\{RJ_t, \mathbf{1}(J_t^{(RBV)} > \Phi_\alpha)\ (\sum_{i=1}^{N} X_{t,i}^2 \mathbf{1}(|X_{t,i}| \geq \gamma), \tag{16}$$

$$RSJ_t{=}RJ_t - RVLJ_t, \tag{17}$$

where $\mathbf{1}$ is the indicator function. RJ_t is defined in Equation (10). γ is the fixed truncation level which is selected by a data-driven procedure. In line with Duong and Swanson (2011), the γ is selected to be equal to 2.

2.2.5 Realized skewness and kurtosis

As noted before, one of the main advantages of random forests is the flexibility of using a large number of predictor variables (covariates) for forecasting realized volatility. Therefore, we also compute realized skewness (*RSK*) and realized kurtosis (*RKU*) of Bitcoin returns (see, e.g., Mei et al. 2017). In line with Amaya et al. (2015), we consider *RSK* and *RKU* as realized measures of the higher-moments of the daily Bitcoin return distribution, and construct them using intraday data. *RSK* is considered as a measure of the asymmetry of the Bitcoin daily returns distribution (associated with the third moment). For a univariate price process, the third moment captures the conditional skewness. As in Barndorff-Nielsen and Shephard (2004), *RSK* captures the asymmetry risk and can be interpreted as a proxy for crash risk. Its interpretation is as follows: (i) a value equal to zero implies that the tails of the daily returns distribution on both sides of the mean balance out overall (i.e., symmetric distribution) (ii) when the left tail is longer or fatter than the right tail (i.e., left-skewed distribution), a negative value arises, and, (iii) when the right tail is longer or fatter than the left tail (i.e., right-skewed distribution), a positive value arises. *RKU*, in turn, is defined as a realized measure that captures extreme occurrences of the daily returns of Bitcoin returns (associated with the fourth moment). As in Barndorff-Nielsen and Shephard (2004), the fourth moment captures the kurtosis risk of a univariate price process with tailedness around the mean. We use realized kurtosis, *RKU*, in order to take into account whether, and how much, there are extreme deviations in Bitcoin's daily return distribution from the Gaussian distribution. The interpretation of *RKU* is also straightforward and is given as follows: (i) when extreme deviations of Bitcoin's daily return distribution are similar to those that would arise under a Gaussian distribution then kurtosis equals zero (and the distribution is a mesokurtic distribution), (ii) when there are fewer and less extreme deviations than under a Gaussian distribution then kurtosis is negative (and the distribution is a platykurtic distribution), and, (iii) when there are more extreme deviations than under a Gaussian distribution then kurtosis is positive (and is a leptokurtic distribution). RSK_t and RKU_t are defined as follows:

$$RSK_t = \left(\frac{\sqrt{N}\sum_{i=1}^{N} X_{t,i}^3}{(\sum_{i=1}^{N} X_{t,i}^2)^{3/2}}\right), \tag{18}$$

$$RKU_t = \frac{N\sum_{i=1}^{N} r_{t,i}^4}{(\sum_{i=1}^{N} X_{t,i}^2)^2} - 3, \tag{19}$$

where RSK_t is standardized by RV_t defined in Equation (1), while the scaling of $\sqrt{N}$ for RSK_t and N for RKU_t ensures that their magnitudes correspond to Bitcoin's daily skewness and kurtosis. It should be remembered that $X_{t,i}$ is the intraday Bitcoin log-return for i hour within day t and $i = 1, .., N$ is the total number of intraday Bitcoin log-returns within a day.

2.2.6 Returns and leverage

Finally, we use Bitcoin daily returns denoted by X_t, and a leverage term as additional predictors. We define the leverage term as $\mathbf{1}(X_t > 0)$.

3 Data

We use intraday (high-frequency) data for Bitcoin prices covering the period from 1st July 2017 to 30th June 2019 to construct daily measures of realized volatility, and its various covariates. We select Bitcoin prices every sixty minutes (60-minutes, hourly basis) and construct 60-minutes log-returns. We select the starting day of the sample period and the 60-minutes frequency of the data because we also present results, for comparison purposes, for other major cryptocurrencies in addition to Bitcoin (see Section 4). It should also be noted that a 60-minutes frequency renders it possible to circumvent liquidity issues (or the lack thereof), extreme high-frequency noise from no-activity periods observed mainly in very small-time windows, and zero prices. We define a trading day from Monday to Sunday from 00:00 EST to 23:59 EST, which renders it possible to have a higher number of observations compared to an 8-hourly and 12-hourly bases. We also employ intraday (high-frequency) data for EOS, Ethereum (ETH), Litecoin (LTC) and Ripple (XRP) prices covering the same period for the robustness analysis. Data for Bitcoin and other cryptocurrencies are from CryptoCompare.com (https://www.cryptoCompare.com), which provides data on a number of liquid Bitcoin markets and other major cryptocurrencies.

The entire dataset has been cleaned following the suggestions of Barndorff-Nielsen et al. (2009). First, we employ the estimators to Bitcoin prices (and for other major cryptocurrencies) of mid-quotes after filtering out spread outliers (less than 0.05 percent). Second, we omit days with recorded prices for less than 40 percent of the expected observations on operating time. Third, following Andersen et al. (2007), in order to take into consideration the existence of small values of RV_t estimates, we employ a double-stabilizing transformation of logarithm and square root (standard deviation), that is, the logarithmic standard deviation of RV_t estimates. This transformation passes all tests for structural stability. As shown by the Jarque-Bera test, this transformation follows a close-to Gaussian distribution. The reliability of the asymptotics for kind of transformation is ensured when twelve or more observations are used to construct a daily one-point estimate of RV_t (see Barndorff-Nielsen and Shephard, 2004a). In this paper, given that twenty-four observations (trading hours per day) are employed for a daily point estimate of RV_t, this condition is met. Daily RV_t in its logarithmic standard deviation form, is log-normally distributed. Finally, as in Diebold et al. (1999) and Andersen et al. (2001), we construct Bitcoin's intraday returns on day t for the i-th intraday observation as the logarithmic difference between two consecutive observed Bitcoin intraday prices within a day t, as follows:

$$X_{t,i} = log(P_{t,i}) - log(P_{t,i-1}), \tag{20}$$

where $P_{t,i}$ is Bitcoin intraday price, $X_{t,i}$ is its intraday log-return for i hour within day t, while $i = 1,..,N$ denotes the total number of intraday Bitcoin log-returns within a day.

To obtain a daily measure for the concern surrounding the US-China trade war, we rely on Google Trends (`https://trends.google.com/trends/`) data. While, the literature (as pointed out in the introduction) has used Google Trends data to obtain metrics for investor attention associated with Bitcoin, we are not aware of any study that has used this approach to derive measures of trade-related uncertainty. Specifically, we extract Google search volume intensity associated with the following items: "US-China Trade War", "US-China Tariffs", "Tariffs War", "Tariffs War US-China", "Trump Trade War", and "New Trade War Tariffs".[3] While we primarily base our analysis on the search volume derived under the "US-China Trade War" term, we also build an indicator which is a composite of these search terms based on the first principal component of (the standardized values) involving the search volume associated with these terms.

[3]Notably, Google Trends provide data for up to 5 years on a weekly basis, whereas daily data can be download for 90 days only. To address this limitation and overcome the individual scaling that might prevent us from stringing several 90-day data together (Baur and Dimpfl, 2016), we follow the procedure described below. We extracted five 90-day periods at once, and fortunately the search data from these periods will have the same maximum scale of 0-100. Given our sample period (1st July 2017 to 30th June 2019) includes at least 8 quarters, we repeat this process to connect the series together while using the overlapping quarters for chaining the series together (i.e., for 2018, we use each quarter of 2018 and the last quarter of 2017, etc.). Each daily series is based on a monthly moving window and rescaled with respect to monthly series over the whole period. Each term has been searched five times with random alphanumerical codes added to them and averaged over the whole period (we have run 20 repetitions for each term, and our main results are not be biased by this issue). Data are then converted to a 0-100 scale as they are typically given by Google Trends.

4 Empirical Results

We present in Table 1 the results for the Diebold and Mariano (1995) tests in order to examine whether trade uncertainty helps to improve out-of-sample forecast accuracy. To this end, we exclude trade uncertainty from the vector of predictors and then compare the results with those from a model that can use trade uncertainty as a predictor.[4]

– Please include Table 1 about here. –

The results in Table 1 are based on rolling-window estimates. Every rolling-estimation window is of length 350 observations, which corresponds roughly to half of the sample size, to compute out-of-sample forecast errors. We then compute the Diebold-Mariano test for three different forecast horizons ($h = 1, 5, 22$ days). We present in Table 1 results for random forests that feature a maximum number of five and ten terminal nodes, and random forests featuring a minimum node size of ten. We fix the number of trees at 500 and choose #floor(total number of covariates/3) random covariates for splitting, a choice which follows standard practice in the machine-learning literature. We present results for an absolute (L1) and a quadratic (L2) loss function, and for two types of forecast errors. As the first type, we consider unscaled standard forecast errors (that is, actuals minus forecasts) and, given that Bitcoin typically experience recurrent periods of high volatility. As the second type, we consider scaled forecast errors. We define

[4]All estimation results documented in this study were computed with the use of the R programming environment (R Core Team 2019). In particular, the following R packages were used: "RandomForest" for the estimation of random forests (see Liaw and Wiener 2002), the R package "forecast" (see Hyndman 2017, Hyndman and Khandakar 2008) for the computation of the Diebold-Mariano test's p-values (based on the well-known modified Diebold-Mariano test proposed by Harvey et al. 1997).

scaled forecast errors by dividing the forecast error by the respective actuals so as to account for heteroskedasticity (see also Bollerslev and Ghysels 1996).

The general picture that emerges is that trade uncertainty does improve forecast accuracy for various configurations of random forests and forecast horizons. The results are stronger for L1 loss than for L2 loss, which was to be expected given that the latter attaches a larger weight to large forecast errors. Large forecast errors typically occur during volatile periods of market jitters. Results for the scaled forecast error in case of L2 loss tend to be stronger in several cases than for unscaled forecast error. Results for the long forecast horizon ($h = 22$ days) are significant mainly for the scaled forecast error and when we consider L1 loss.

– Please include Table 2 about here. –

Table 2 depicts a ranking of the various covariates considered in this research according to their relative importance. Relative importance is measured by inspecting the increase in node purity (as measured by the residual sum of squares across all trees) that results from the inclusion of a variable to the list of predictors. The ranking of the covariates is based on the mean rank of a covariate across all rolling-estimation windows. As one would have expected, daily realized volatility and its weekly and monthly counterparts (that is, the standard predictors used for estimating a HAR-RV model) are among the top-ranked predictors. Trade uncertainty, in turn, has an average rank of about seven for $h = 1$, and of three and four for the two longer forecast horizons.

– Please include Table 3 about here. –

Table 3 depicts pseudo R-squared statistics for the random forest models. The R-squared statistics increase in the forecast horizon and, as one would have expected, when we increase the number of terminal nodes. The largest R-squared statistics obtain for a model that features a terminal node size of ten observations.

– Please include Table 4 about here. –

Table 4 shows results for three alternative rolling-estimation windows of length 200, 300, and 400 observations. Again, we find several significant test results. Results only become insignificant for the unscaled forecast error and L2 loss when the length of the rolling-estimation window increases, and for the short rolling-estimation window and the short forecast horizon ($h = 1$) in case of the scaled forecast error.

– Please include Table 5 about here. –

Table 5 summarizes the test results for three alternative specifications of the HAR-RV model. In the first specification, we use "good" realized volatility (that is, RV_t^+) as predictand. In the second specification, we study the predictive value of trade uncertainty for "bad" realized volatility (that is, RV_t^-) as predictand. In the second specification, we keep the standard metric of realized volatility as our predictand, but add good and bad realized volatility as additional predictors to the model. The test results show that trade uncertainty significantly improves forecast accuracy, with the results for the model that features bad realized volatility as a predictand being the only exception when we study the unscaled forecast error under a L2 loss function.

– Please include Table 6 about here. –

In Table 6, we present results for the Clark and West (2007) test, where we consider three different specifications of the number of terminal nodes and the size of the terminal nodes (and a rolling-estimation window of length 350 observations). The test results are significant for two out of the three specifications for all three forecast horizons, and for all three specifications for the long forecast horizon ($h = 22$ days).

As final extensions, we present results for an alternative measure of trade uncertainty (Table 7) and other cryptocurrencies (Table 8). As for the latter, we use intraday (high-frequency) data for EOS, Ethereum (ETH), Litecoin (LTC) and Ripple (XRP) prices covering the period from 1st July 2017 to 30th June 2019. These cryptocurrencies along with the starting day of the sample and the 60-minutes frequency employed are strictly selected by the availability of intraday data.

– Please include Table 7 about here. –

The results for an alternative news-based index of trade uncertainty, as created by Caldara et al. (2019),[5] are much weaker than for our baseline metric of trade uncertainty and almost always insignificant. The results for the other cryptocurrencies, in contrast, corroborate our findings for Bitcoin realized volatility. Both the Diebold-Mariano test and the Clark-West test yield, with few exceptions, significant results.

5 Concluding Remarks

We study the importance of the US-China trade war in predicting, both in- and out-of-sample, daily realized volatility of Bitcoin returns. We use intraday data in 5 minute-interval covering the period from 1st July 2017 to 30th June 2019. We

[5]The index is constructed by counting the frequency of joint occurrences of trade policy (tariff, import duty, import barrier, and anti-dumping) and uncertainty (uncertainty, risk, or potential) terms across major newspapers (Boston Globe, Chicago Tribune, Guardian, Los Angeles Times, New York Times, Wall Street Journal, and Washington Post). The data is downloadable from the website of Professor Matteo Iacoviello at: `https://www2.bc.edu/matteo-iacoviello/tpu.htm`.

focus on realized volatility as it is it provides an accurate estimator of volatility on the basis of the actual variance. We employ the HAR-RV model proposed by Corsi (2009) as our benchmark model, and then consider its various extensions discussed in earlier literature to capture stylized facts of the volatility process, such as heterogeneity, long-memory and discontinuity. We control realized skewness and realized kurtosis, and for various variants of jumps given that previous studies (e.g., Chaim and Laurini 2018) have indicated the importance of jumps in the Bitcoin market. Furthermore, jumps are well known for their ability to improve the overall fit of realized-volatility models. The primary predictor of interest, however, is a metric for the US-China trade tensions, and it is based on Google Trends data.

For our empirical analysis, we apply a machine-learning technique known as random forests. Random forests are ideally suited for our research due to the fact that we control for the impact of various measures of jumps, higher-order moments (i.e., realized skewness and realized kurtosis), returns, and a leverage effect in addition to the usual terms of the benchmark core HAR-RV model. As extensions, we consider a news-based measure of US trade policy uncertainty, and we analyze other cryptocurrencies. We document results based on rolling-window estimates that we use to compute out-of-sample forecast errors. We report results for an absolute and a quadratic loss function, and for two types of forecast errors (unscaled and scaled), given that Bitcoin typically experiences recurrent periods of high volatility.

Our findings show that our main Google-Trends based measure of the US-China trade war improves forecast accuracy for various configurations of random forests and forecast horizons. Our findings go beyond results reported in the earlier literature on Bitcoin and other cryptocurrencies. Specifically, our findings add to the work of Gozgor et al. (2018) and Bouri et al. (2019a) by showing, for the first time, the ability of trade uncertainty to predict the realized volatility of Bitcoin returns. This is an important finding that is useful to traders and investors in their quest to predict the volatility of Bitcoin, which continues to puzzle researchers, crypto-traders, and policy makers. Building on our findings, an interesting avenue for future research is to consider the effect of trade uncertainty on the relationship between trading volume and volatility in the cryptocurrency market.

References

Acharya, V.V. , Lochstoer, L.A. , and Ramadorai, T. (2013). Limits to arbitrage and hedging: evidence from commodity markets. *Journal of Financial Economics*, 109: 441−465.

Agnolucci, P. (2009). Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31: 316−321.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118: 135−167.

Andersen, T. G., and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4): 885−905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96: 42−55.

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

Andersen, T. G., Bollerslev, T., and 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., and Diebold, F. X. (2010). Parametric and nonparametric volatility measurement. In L.P. Hansen and Y. Ait-Sahalia (Eds.), *Handbook of Financial Econometrics: Tools and Techniques*, 67−137, Elsevier, North-Holland.

Andersen, T.G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169: 75−93.

Ardia, D., Bluteau, K., and Rede, M. (2019). Regime changes in bitcoin GARCH volatility dynamics. *Finance Research Letters*, 29: 266−271.

Arouri, M. E. H., Lahiani, A., Lévy, A., and Nguyen, D. K. (2012). Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. *Energy Economics*, 34: 283−293.

Aysan, A.F., Demir, E., Gozgor, G., and Lau, C.K.M. (2019). Effects of the geopolitical risks on Bitcoin returns and volatility. *Research in International Business and Finance*, 47: 511−518.

Barndorff-Nielsen, O.E., and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4: 1−30.

Barndorff-Nielsen, O.E., Kinnebrouk, S., and Shephard, N. (2010). Measuring downside risk: realised semivariance. In T. Bollerslev, J. Russell and M. Watson (eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle*, 117-136. Oxford University Press.

Baur, D. G., and Dimpfl, T. (2016). Googling gold and mining bad news. *Resources Policy*, 50, 306−311.

Baur, D.G., and Dimfl, T. (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters*, 173: 148−151.

Bollerslev, T., and Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business and Economic Statistics*, 14: 139−151.

Bollerslev, T., Li, S. Z., and Zhao, B. (forthcoming). Good volatility, bad volatility and the cross-section of stock returns. *Journal of Financial and Quantitative Analysis*.

Bouri, E., Gil-Alana, L.A., Gupta, R., and Roubaud, D. (2019b). Modelling long memory volatility in the Bitcoin market: Evidence of persistence and structural breaks. *International Journal of Finance and Economics*, 24: 412−426.

Bouri, E., Gkillas, K., and Gupta, R. (2019a). Trade uncertainties and the hedging abilities of Bitcoin. Department of Economics, University of Pretoria, Working Paper No. 201948.

Bouri, E., and Gupta, R. (2019). Predicting Bitcoin returns: Comparing the roles of newspaper- and internet search-based measures of uncertainty. *Finance Research Letters*. DOI: `https://doi.org/10.1016/j.frl.2019.101398`.

Bouri, E., Gupta, R., Lau, C.K.M., Roubaud, D., and Wang, S. (2018). Bitcoin and global financial stress: A copula-based approach to dependence and causality in the quantiles. *Quarterly Review of Economics and Finance*, 69: 297−307.

Bouri, E., Gupta, R., Tiwari, A.K., and Roubaud, D. (2017). Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters*, 23: 87−95.

Bouri, E., Roubaud, D., and Shahzad, S.J.H. (2019c). Do Bitcoin and other cryptocurrencies jump together? *Quarterly Review of Economics and Finance*. DOI: `https://doi.org/10.1016/j.qref.2019.09.003`.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24: 123−140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45: 5−32.

Caldara, D., Iacoviello, M., Molligo, P., Prestipino, A., Raffo, A. (2019a). The economic effects of trade policy uncertainty. *Journal of Monetary Economics*. DOI: `https://doi.org/10.1016/j.jmoneco.2019.11.002`.

Chaim, P., and Laurini, M.P. (2018). Volatility and return jumps in bitcoin. *Economics Letters*, 173: 158−163.

Chen, Y., Ma, F., and Zhang, Y. (2019). Good, bad cojumps and volatility forecasting: new evidence from crude oil and the U.S. stock markets. *Energy Economics*, 81: 52−62.

Chu, J., Chan, S., Nadarajah, S., and Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management*, 10: 17.

Clark, T.D., and West,K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138: 291−311.

Conrad, C., Custovic, A., and Ghysels, E. (2018). Long- and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, 11: 23.

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

Demirer, R., Gkillas, K. , Gupta, R., and Pierdzioch, C. (2019). Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests. Department of Economics, University of Pretoria, Working Paper No. 201972.

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

Duong, D., and Swanson, N. R. (2011). Volatility in discrete and continuous-time models: A survey with new evidence on large and small jumps. *Missing Data Methods: Time-Series Methods and Applications*, 179−233. Emerald Group Publishing Limited.

Duong, D., and Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics*, 187: 606−621.

Fang, L., Bouri, E., Gupta, R., and Roubaud, D. (2019). Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin? *International Review of Financial Analysis*, 61: 29−36.

Gkillas, K., Gupta, R., and Pierdzioch, C. (2019). Forecasting realized volatility of bitcoin returns: Tail events and asymmetric loss. Department of Economics, University of Pretoria, Working Paper No. 201905.

Gkillas, K., Gupta, R., and Pierdzioch, C. (Forthcoming). Forecasting realized oil-price volatility: The Role of financial stress and asymmetric loss. *Journal of International Money and Finance*.

Gkillas, K., and Longin, F. (2019). Is Bitcoin the new digital gold? Evidence from extreme price movements in financial markets. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3245571`.

Gozgor, G., Tiwari, A.K., Demir, E., and Akron, S. (2019). The relationship between Bitcoin returns and trade policy uncertainty. *Finance Research Letters*, 29: 75−82.

Gupta, Rangan, Pierdzioch, C., Vivian, A. J., Wohar, M. E., (2019). The predictive value of inequality measures for stock returns: An analysis of long-span UK data using quantile random forests. *Finance Research Letters*, 29: 315-322.

Guo, H., Wang, K., and Zhou, H. (2019). Good Jumps, Bad Jumps, and Conditional Equity Premium. *Asian Finance Association (AsianFA) 2014 Conference Paper*, PBCSF-NIFR Research Paper No. 14-05.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13: 281−291.

Hyndman, R.J. (2017). forecast: Forecasting functions for time series and linear models. R package version 8.0, URL: `http://github.com/robjhyndman/forecast`.

Hyndman, R.J., and Khandakar, Y (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26: 1−22.

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

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific*

Reports, 3: 3415.

Liaw, A., and Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2: 18−22.

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

McAleer, M., and Medeiros, M. C. (2008). Realized volatility: a review. *Econometric Reviews*, 27: 10−45.

Panagiotidis, T., Stengos, T., and Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. *Finance Research Letters*, 27(C): 235−240.

Panagiotidis, T., Stengos, T., and Vravosinos, O. (2019). The effects of markets, uncertainty and search intensity on bitcoin returns. *International Review of Financial Analysis*, 63: 220−242.

Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help?. *Physica A: Statistical Mechanics and its Applications*, 481: 153−159.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., and Pictet, O. V. (1997). Volatilities of different time resolutions − Analyzing the dynamics of market components. *Journal of Empirical Finance*, 4: 213−239.

Nasir, M.A., Huynh, T.L.D., Nguyen, S.P., and Duong, D. (2019). Forecasting cryptocurrency returns and volume using search engines. *Financial Innovation*, 5 (Article No. 2), DOI: 10.1186/s40854-018-0119-8.

Patton, A. J., and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97: 683−697.

Pierdzioch, C., and Risse, M. (forthcoming). Forecasting precious metal returns with multivariate random forests. *Empirical Economics*.

Poon, S.-H., and Granger, C.W.J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41: 478−539.

R Core Team (2019). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/. R version 3.6.0.

Subramaniam, S., and Chakraborty, M. (2020). Investor attention and cryptocurrency returns: Evidence from quantile causality approach. *Journal of Behavioral Finance*, 21: 103−115.

Takaishi, T. (2018). Statistical properties and multifractality of Bitcoin. *Physica A: Statistical Mechanics and its Applications*, 506: 507−519.

Walther, T., Klein, T., and Bouri, E. (2019). Exogenous drivers of Bitcoin and cryptocurrency volatility: A mixed data sampling approach to forecasting. *Journal of International Financial Markets Institutions and Money*, 63: 101−113.

Wu, S., Tong, M., Yang, Z., and Derbali, A. (2019). Does gold or Bitcoin hedge economic policy uncertainty? *Finance Research Letters*, 31: 171−178.

Table 1: Baseline Results

Panel A: Maximum number of terminal nodes is 5

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.0000	0.0112	0.0526
$L2$	0.2303	0.1542	0.1225
	Scaled forecast error		
$L1$	0.0000	0.0046	0.0242
$L2$	0.0001	0.0685	0.1276

Panel B: Maximum number of terminal nodes is 10

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.0072	0.0163	0.0540
$L2$	0.6930	0.2462	0.1273
	Scaled forecast error		
$L1$	0.0001	0.0065	0.0379
$L2$	0.0009	0.0761	0.1361

Panel C: Minimum node size is 10

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.0116	0.0145	0.0012
$L2$	0.0910	0.1558	0.0462
	Scaled forecast error		
$L1$	0.0921	0.0555	0.0318
$L2$	0.2074	0.0767	0.1396

Note: p-values of Diebold-Mariano tests. Forecasts were derived from a HAR-RV-RF model without trade uncertainty and a HAR-RV-RF model with trade uncertainty for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and was computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the trade uncertainty are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using random forests and a rolling-estimation window (350 observations). The number of trees was set to 500.

Table 2: Ranking of Covariates

Covariate	$h = 1$	$h = 5$	$h = 22$
RV	1.79	2.85	3.67
RV_w	4.05	2.76	2.65
RV_m	3.48	1.43	1.25
Asymmetric jumps	9.39	10.21	10.73
Bad jumps	4.99	5.67	5.94
Good jumps	8.92	10.16	8.71
Large jumps	6.93	7.05	5.74
Small jumps	9.11	8.55	7.97
Returns	1.39	5.26	8.24
Realized skewness	10.53	11.26	11.41
Realized kurtosis	10.43	9.42	8.97
Leverage	12.23	12.43	12.38
US China trade war	7.76	3.94	3.35

Note: The mean rank of a covariate across all rolling-estimation windows (2350 observations) was computed based on its effect on node purity as measured by the residual sum of squares. The number of trees per random forest was set to 500, and the maximum number of terminal nodes was set to five.

Table 3: Pseudo R-Squared

Specification	$h = 1$	$h = 5$	$h = 22$
Max. no. of terminal nodes = 5	0.2708	0.3852	0.5356
Max. no. of terminal nodes = 10	0.2873	0.4205	0.5863
Min. terminal node size = 10	0.2881	0.4531	0.6370

Note: The pseudo R-squared was computed as $1 - MSE_{oob}/Var$, where MSE_{oob} denotes the mean-squared error that obtained when the estimated model was used to compute the out-of-bag data (that is, the data not used for tree building when bootstrapping a random forest) and Var is an estimate of the variance of the dependent variable. The number of trees was set to 500.

Table 4: Alternative Rolling-Window Lengths

Window	$h=1$	$h=5$	$h=22$
	Unscaled forecast error (L1)		
200	0.0267	0.0116	0.0773
300	0.0000	0.0005	0.0630
400	0.0002	0.0389	0.0668
	Unscaled forecast error (L2)		
200	0.0117	0.0067	0.0949
300	0.1854	0.0094	0.0890
400	0.1976	0.1565	0.1024
	Scaled forecast error (L1)		
200	0.2746	0.0378	0.0656
300	0.0000	0.0001	0.0095
400	0.0000	0.0034	0.0282
	Scaled forecast error (L2)		
200	0.5523	0.0406	0.0635
300	0.0001	0.0284	0.0675
400	0.0001	0.0255	0.1233

Note: p-values of Diebold-Mariano tests. Forecasts were derived from a HAR-RV-RF model without trade uncertainty and a HAR-RV-RF model with trade uncertainty for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and was computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the trade uncertainty are more accurate. L1: absolute loss. L2: quadratic loss. The maximum number of nodes was set to five. The number of trees was set to 500.

Table 5: Good and Bad Realized Volatility

Loss function	$h=1$	$h=5$	$h=22$
	RV^+ as dependent variable		
Unscaled forecast error (L1)	0.0000	0.001	0.1245
Scaled forecast error (L1)	0.0000	0.0017	0.0312
Unscaled forecast error (L2)	0.023	0.0039	0.0938
Scaled forecast error (L2)	0.0016	0.0443	0.1193
	RV^- as dependent variable		
Unscaled forecast error (L1)	0.0000	0.0052	0.0158
Scaled forecast error (L1)	0.0000	0.0015	0.0158
Unscaled forecast error (L2)	0.2861	0.2470	0.1220
Scaled forecast error (L2)	0.0001	0.0191	0.1180
	RV^+ and RV^- as predictors		
Unscaled forecast error (L1)	0.0000	0.002	0.035
Scaled forecast error (L1)	0.0000	0.001	0.0192
Unscaled forecast error (L2)	0.0048	0.0574	0.0638
Scaled forecast error (L2)	0.0000	0.0441	0.1210

Note: p-values of Diebold-Mariano tests. Forecasts were derived from a HAR-RV-RF model without trade uncertainty and a HAR-RV-RF model with trade uncertainty for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and was computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the trade uncertainty are more accurate. L1: absolute loss. L2: quadratic loss. The maximum number of nodes was set to five. The number of trees was set to 500.

Table 6: Clark-West-Test

Specification	$h = 1$	$h = 5$	$h = 22$
Max. no. of terminal nodes = 5	1.8746	2.0555	2.6345
Max. no. of terminal nodes = 10	-0.2220	1.6136	2.8335
Min. terminal node size = 10	2.0200	2.2095	3.0404

Note: t-statistics of the Clark-West tests were computed using Newey-West standard erros. Critical values (one-sided test) are: 1.282 (10%) and 1.645 (5%). The number of trees was set to 500.

Table 7: Alternative Measure of Trade Uncertainty

Specification	$h = 1$	$h = 5$	$h = 22$
		Factor	
Max. no. of terminal nodes = 5	0.0287	0.2032	0.1342
Max. no. of terminal nodes = 10	0.3243	0.1561	0.1344
Min. terminal node size = 10	0.3237	0.0409	0.1441
		TPUD	
Max. no. of terminal nodes = 5	0.9599	0.9713	0.8925
Max. no. of terminal nodes = 10	0.9498	0.9479	0.8858
Min. terminal node size = 10	0.6809	0.7478	0.8596

Note: p-values of Diebold-Mariano tests (HAR-RV-RF vs. HAR-RV-RF model cum trade uncertainty) for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and was computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the risk aversion are more accurate. Loss function: $L2$. The number of trees was set to 500.

Table 8: Other Cryptocurrencies

Specification	$h = 1$	$h = 5$	$h = 22$
		EOS	
CW test	0.0294	2.7686	2.9992
DM test	0.0844	0.0787	0.0878
		ETH	
CW test	1.8065	1.5390	2.4044
DM test	0.0161	0.0482	0.1092
		LTC	
CW test	2.3241	2.7758	3.2854
DM test	0.005	0.0432	0.0493
		XRP	
CW test	1.9085	1.1216	1.6148
DM test	0.0846	0.0065	0.0741

Note: p-values of Diebold-Mariano tests (HAR-RV-RF vs. HAR-RV-RF-RA forecasts) for three different forecast horizons. CW = Clark-West test. The Diebold-Mariano (DM) was estimated using the scaled forecast error accounts for heteroscedasticity in the data computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the risk aversion are more accurate. Loss function: $L2$. The models were estimated using random forests consisting of 500 trees and a maximum of five terminal nodes.