International Journal of Forecasting 39 (2023) 1–17

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecasting Bitcoin with technical analysis: A not-so-random forest?

Nikola Gradojevic [a,b,*], Dragan Kukolj [b], Robert Adcock [a], Vladimir Djakovic [b]

[a] University of Guelph, Lang School of Business and Economics, Canada
[b] Faculty of Technical Sciences, University of Novi Sad, Serbia

ARTICLE INFO

Keywords:
Bitcoin
Deep learning
Random forest
Forecasting
Technical analysis
Market sentiment

ABSTRACT

This paper uses data sampled at hourly and daily frequencies to predict Bitcoin returns. We consider various advanced non-linear models based on a multitude of popular technical indicators that represent market trend, momentum, volume, and sentiment. We run a robust empirical exercise to observe the impact of forecast horizon, model type, time period, and the choice of inputs (predictors) on the forecast performance of the competing models. We find that Bitcoin prices are weakly efficient at the hourly frequency. In contrast, technical analysis combined with non-linear forecasting models becomes statistically significantly dominant relative to the random walk model on a daily horizon. Our comparative analysis identifies the random forest model as the most accurate at predicting Bitcoin. The estimated measures of the relative importance of predictors reveal that the nature of investing in the Bitcoin market evolved from trend-following to excessive momentum and sentiment in the most recent time period.

1. Introduction

The recent literature documents that one of the main drivers of Bitcoin fluctuations has been a series of speculative bubbles (Adcock & Gradojevic, 2019; Cheah & Fry, 2015; Corbet et al., 2018; Geuder et al., 2019; Hafner, 2018; Hencic & Gouriéroux, 2015). To illustrate the extreme nature of Bitcoin volatility, it is worthwhile to note that from 2013 to 2017 its price surged by almost 2800%, reaching a peak of US$19,783 in December 2017, and then it plunged to nearly US$3,000 in subsequent months. Surprisingly, in 2019, the Bitcoin price rebounded and climbed above US$10,000. The volatile nature of Bitcoin returns has experienced additional challenges from increased government regulation of crypto-assets (e.g., China and Russia), which inserted more uncertainty into Bitcoin trading.[1]

Such excessive price movements marked by persistence and structural breaks (Bouri et al., 2019a; Thies & Molnár, 2018) make forecasting Bitcoin/U.S. dollar (BTC/USD) exchange rates a formidable task. Moreover, given that traditional fundamental variables used in asset valuation (e.g., financial statements) are not available for Bitcoin, the panel of potential predictors in a Bitcoin forecasting model shrinks substantially. Typically, when fundamental analysis is not feasible, scholars and practitioners resort to technical analysis (Kaucic, 2010) or autoregressive time-series econometric models such as autoregressive integrated moving average (ARIMA)- or generalized autoregressive conditional heteroscedasticity (GARCH)-type models. The latter models may include external variables (e.g., the ARMAX model) and,

* Correspondence to: University of Guelph, Lang School of Business and Economics, Department of Economics and Finance, 50 Stone Road, East, Guelph, Ontario, N1G 2W1, Canada.
E-mail addresses: ngradoje@uoguelph.ca (N. Gradojevic), dragan.kukolj@rt-rk.com (D. Kukolj), radcock@uoguelph.ca (R. Adcock), v_djakovic@uns.ac.rs (V. Djakovic).

[1] Quite recently, based on the multifractal spectrum and generalized Hurst exponent analyses, Kukacka and Kristoufek (2020) emphasized higher complexity in Bitcoin prices relative to the S&P 500 index.

https://doi.org/10.1016/j.ijforecast.2021.08.001
0169-2070/

for the purpose of Bitcoin forecasting, previous studies proposed the following predictors: stock market returns, exchange rates, gold and oil returns, interest rates, internet trends (Baur et al., 2018; Dyhrberg, 2016; Kang et al., 2019; Panagiotidis et al., 2018), blockchain information (Jang & Lee, 2018), news surprises (Gronwald, 2015), money supply and internet searches (Kristoufek, 2015), market and investor sentiment (Baig et al., 2019; Karalevicius et al., 2018), trading volume (Balcilar et al., 2017; Bouri et al., 2019b), economic policy uncertainty (Demir et al., 2018), macroeconomic factors (Polasik et al., 2015), and the VIX (Adcock & Gradojevic, 2019).[2]

Only a select few papers performed technical analysis of Bitcoin prices. Recently, Corbet et al. (2019) tested the profitability of the moving average and the trading range break-out technical indicators in the Bitcoin market at the one-minute frequency. The results from this study were mixed and generally inconclusive. Similarly, Nakano et al. (2018) assessed the profitability of a deep learning artificial neural network (D-ANN) model with technical indicators over the 2016–2018 period. The Bitcoin forecasts were generated at a 15-minute horizon and the trading strategy combined more technical indicators than Corbet et al. (2019). They found that certain technical trading strategies were able to yield positive excess returns relative to the buy-and-hold strategy, while assuming realistic transaction costs. As a complementary study, Adcock and Gradojevic (2019) employed an ANN model with technical indicators and produced forecasts at a daily horizon on the 2011–2018 data span. The evidence from this work indicated that backpropagation ANNs dominated various competing models in terms of their point and density forecast accuracy.[3]

Another strand of literature concerns the application of ANNs and related non-linear techniques in Bitcoin forecasting and trading. In an influential study that was methodologically related to Adcock and Gradojevic (2019) and Nakano et al. (2018), Atsalakis et al. (2019) conceptualized a neuro-fuzzy trading system with lagged Bitcoin prices taken as the model's inputs. Their model improved upon the buy-and-hold strategy not only in Bitcoin daily trading but also for other cryptocurrencies. Related to this contribution, Lahmiri and Bekiros (2019) relied on an (autoregressive) D-ANN model to demonstrate its effectiveness at forecasting daily returns on Bitcoin, Digital Cash, and Ripple. Jang and Lee (2018) were among the first scholars to utilize ANNs for the time-series modeling and forecasting of Bitcoin. Their work used daily data for non-technical trading predictors (blockchain information, stock market indices, exchange rates, commodity prices) and identified Bayesian ANNs as the most appropriate model for predicting the Bitcoin prices. In the same vein, Kristjanpoller and Minutolo (2018) proposed a variety of hybrid ANN-GARCH-type models to forecast the volatility of BTC/USD exchange rate returns. The forecast superiority of different model variants was confirmed for the horizons of 10, 22, and 44 days.

The current paper complements and extends the previous contributions.[4] First, we focus on the statistical and directional values of Bitcoin forecasts based on an extensive panel of popular technical indicators that have not been covered before. Second, to understand how our proposed model forecasts intraday and interday, the technical predictors were constructed on both an hourly and a daily basis. For this purpose, we utilize the most recent data set of BTC/USD observations (2015–2019) and analyze the impact of market regimes—as reflected in the informativeness of individual technical indicators—on the model's forecast performance. More importantly, since we concentrate on technical indicators that represent market trend, momentum, sentiment, and volume, our approach uncovers certain time-series, stylized facts about BTC/USD exchange rate fluctuations that might be of value to investors, government policymakers, and regulators. In contrast, the complexity of other studies (e.g., Corbet et al., 2019; Jang & Lee, 2018, and Nakano et al., 2018) precludes such lessons because it is difficult to interpret their results economically. In order to identify the nature of Bitcoin "market regimes", we quantify the relative importance of technical trading predictors over sequential time periods by using the methods of pseudo-weights (Qi & Maddala, 1996) and feature importance in the random forest model. Such quantitative assessment of each predictor's non-linear importance in the models is what also distinguishes our paper from the extant work in the field. Finally, we assess the statistical significance of the forecast improvements produced by advanced non-linear models (feedforward D-ANN, support vector machine, and random forest), relative to the benchmark models (random walk, ARMAX). To the best of our knowledge, such an exhaustive and robust empirical analysis that combines technical indicators and sophisticated (deep learning) modeling techniques on both intraday and interday data has not been undertaken in the relevant hitherto literature.[5]

Our findings can be summarized as follows: (1) when data are sampled at a daily frequency, D-ANN and the random forest models based on technical indicators produce statistically significant out-of-sample forecast improvements for BTC/USD exchange rate returns, relative to the benchmark models (random walk and ARMAX); (2) when data are sampled at an hourly frequency, the Bitcoin market is generally efficient in the weak sense, irrespective of time period tested; (3) the degree of predictability of hourly BTC/USD exchange rate returns is closely related to

[2] An excellent overview of the up-to-date research on cryptocurrencies can be found in Petukhina et al. (2020).

[3] Similar recent research efforts that demonstrated the usefulness of technical analysis in BTC/USD trading include (Gerritsen et al., 2020), Huang et al. (2019), and Miller et al. (2019).

[4] This paper can also be seen as a continuation of the exchange rates literature (see, e.g., Dooley & Shafer, 1984; Menkhoff, 1997, 2010; Taylor & Allen, 1992) that explores the value of technical indicators for classical currency pairs.

[5] Our work also nicely complements the very recent contribution by Chen et al. (2021) that relies on a completely different set of predictors, such as blockchain information and macroeconomic and financial time series. Although certain non-linear methods used in Chen et al. (2021) are similar to the current paper, their research is focused only on daily data, whereas the economic interpretation of their results is not analyzed.

market volatility (i.e., market regime); (4) the statistical importance of predictors of hourly BTC/USD exchange rate returns depends on a market regime; and (5) the Bitcoin market was driven primarily by its trend and momentum during the 2015–2017 period, while the last couple of years have been dominated by the psychological tendencies of the market (fear, greed, Google searches, and volume).

The structure of this paper is as follows: Section 2 reviews the applications of ANN-based models in finance and presents the methods employed for our forecasting exercises; Section 3 describes the data; Section 4 lays out the results; and the final section concludes the paper and suggests a further research agenda.

2. Methodology

2.1. ANNs in finance

The poor forecast performance of many linear financial models (see, e.g., Hsieh, 1989) has stimulated increased scholarly interest in non-linear models such as ANNs. In a seminal paper, Hornik et al. (1989) established that multilayer feedforward ANNs represent a class of universal functional approximators. This spurred further studies that tackled financial applications of ANNs. Refenes and Azema-Barac (1994) investigated ANN applications in financial asset management. Wittkemper and Steiner (1996) studied systematic risk for German stocks with ANNs. Donaldson and Kamstra (1997) constructed a semi-parametric non-linear GARCH model based on feedforward ANNs to forecast the volatility of stock returns. In technical analysis applications, Hans and van Griensven Kasper (1998), Gençay and Stengos (1998), and Fernández-Rodríguez et al. (2000) indicated that utilizing past buy–sell signals of foreign exchange rates and stock prices in a feedforward ANN improves profitability and out-of-sample generalizations. Franses and Draisma (1997) proposed a method based on ANNs to investigate when and how seasonal patterns in macroeconomic time series change over time. Yang et al. (1999) used probabilistic ANNs in bankruptcy prediction. Blake and Kapetanios (2000) proposed a test for ARCH that uses ANNs. Heinemann (2000) investigated how adaptive learning of rational expectations may be modeled with the help of ANNs. In the pricing of option contracts, Hutchinson et al. (1994), Garcia and Gençay (2000), and Gradojevic et al. (2009) demonstrated that feedforward ANNs can be successfully used to estimate a pricing formula for options, with good out-of-sample pricing and delta-hedging performances.

In more recent works, Zhang and Qi (2005) found that ANNs were unable to model time-series seasonality directly and that ignoring seasonal or trend patterns would result in poor forecasting accuracy. Fischer and Krauss (2018) and Krauss et al. (2017) relied on D-ANNs and several other non-linear models to predict directional movements and test the profitability of the constituent stocks of the S&P 500 index. Furthermore, various trading and portfolio selection strategies that involve ANNs were examined by Bekiros and Georgoutsos (2008), Adcock (2010), Sermpinis et al. (2013), and Sermpinis et al. (2021). Also, Geppert et al. (2010) investigated the probability of the deletion of a firm from the S&P 500 index in an ANN-guided framework, while Mselmi et al. (2017) used ANNs in the context of financial distress prediction.[6]

2.2. Feedforward deep artificial neural network (FF-D-ANN)

To explain the concept of an FF-D-ANN model, we assume that Bitcoin returns at time $t+1$ (BR_{t+1}) are forecasted by k predictors (or inputs) that are represented by technical indicators ($x_{it}, i = 1, \ldots, k$), and that this mapping is a non-linear function (ϕ) whose parameters will be estimated (i.e., $BR_{t+1} = \phi(x_{1t}, x_{2t}, \ldots, x_{kt})$). The parameters are called connection weights (α_{ij} and β_j) and node biases (α_{j0} and β_0). Fig. 1 shows the structure of an FF-D-ANN model that is used to approximate the non-linear mapping. It consists of three building blocks: the input layer (where predictors are fed into the model), hidden layers (where the functional approximation or learning takes place), and the output layer (where the Bitcoin prediction is generated).

If we initially assume that the model architecture involves only one hidden layer with q computational elements (nodes or neurons), the FF-D-ANN model is estimated as

$$BR_{t+1} = \phi\left(\beta_0 + \sum_{j=1}^{q} \beta_j \psi\left(\alpha_{j0} + \sum_{i=1}^{k} \alpha_{ij} x_{it}\right)\right) \tag{1}$$

In this example, q is the number of hidden nodes, where the single hidden and the output layers are characterized by two flexible classes of non-linearities: ψ and ϕ, respectively. α_{ij} and β_j denote appropriate connection weights between the adjacent layers. Subscripts 0 for α and β stand for the biases. When it contains multiple hidden layers, this model is also called a multilayer perceptron (MLP). In such a case, Eq. (1) becomes more complex with more non-linearities nested in each other. Nevertheless, the logic of transforming inputs into each node by taking their weighted average, adding biases, and non-linearly transforming the sum (that is passed forward to the subsequent layer's nodes) still applies. It is worthwhile to stress that a multilayered model structure with at least two hidden layers is what facilitates "deep learning" in practical applications (Lahmiri & Bekiros, 2019). In our model design, we rely on complex D-ANN models that are able to extract the predictive information from a large number of inputs (lagged technical indicators), while utilizing a deep structure in the hidden layers with a large number of hidden nodes.

The training algorithm is Adam, a deep learning first-order gradient-based optimizer based on adaptive estimates of lower-order moments. It is computationally efficient and suitable for large data sets with many parameters to estimate. The default parameters follow those provided by Kingma and Ba (2015). The activation function types used in the hidden layers are sigmoid, with the linear function in the output layer. To control for

[6] Please refer to Tkáč and Verner (2016) for an extensive survey of ANNs in business.

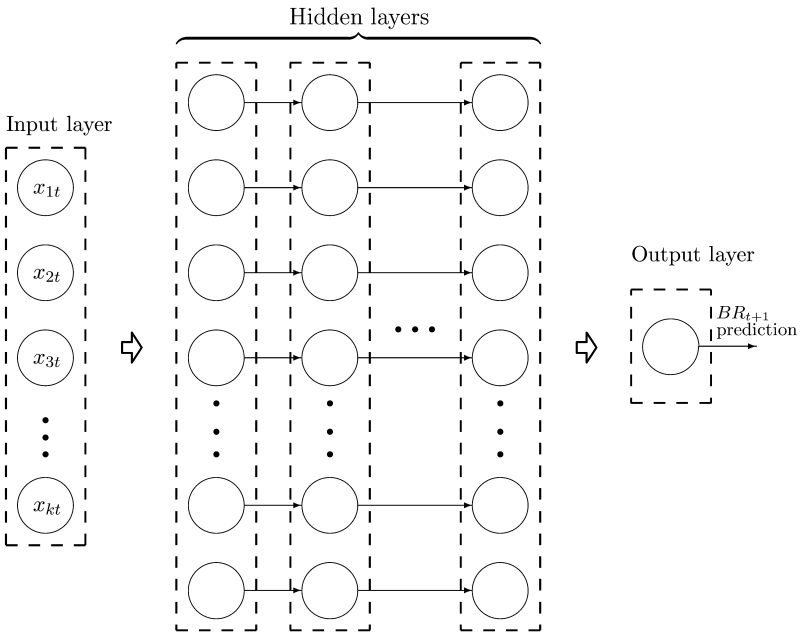


Fig. 1. Feedforward deep learning artificial neural network (FF-D-ANN) model. Notes: The FF-D-ANN forecasting model is a non-linear mapping from the inputs or predictors (k technical indicators sampled at time t: x_{kt}) to the output (BTC/USD exchange rate returns sampled at time $t+1$: BR_{t+1}). The model, also known as a multilayer perceptron (MLP), is generally composed of three segments: an input layer, multiple hidden layers, and an output layer. Each circle is a computational element or node ("neuron") that is connected to all other nodes in adjacent layers by connection weights. Computational elements take a weighted average of its inputs, add a parameter (node bias), non-linearly transform the sum, and pass it to subsequent layers.

overfitting, we use a 10% dropout. Validation data are used to select the optimal model architecture. The model will set apart 10% of the training data, will not train on it, and will evaluate the loss (i.e., mean-squared error) and any model metrics on this data at the end of each epoch. If, during the last five epochs, the monitored quantity (validation loss) has no improvement, the training will be stopped.

2.3. Support vector machine (SVM) model

An SVM is a supervised learning model for prediction and classification, introduced by Vapnik (1999). The basic idea of this algorithm applied for a non-linear regression problem (i.e., SV regression—SVR) can be described by two steps: (a) transform training samples into a space of higher dimension using a non-linear mapping function; and (b) perform linear regression in the space of higher dimension in order to separate the data samples. Transforming data from the original to the new (higher-order) feature space is performed using a predetermined kernel function. The kernel function is defined as the dot product of input vectors ($x_{it},\ i = 1, \ldots, k$) from the original space. It is assumed that the kernel function enables computations of points in the new feature space without explicitly calculating the unknown non-linear mapping ($\phi(x_{1t}, x_{2t}, \ldots, x_{kt})$). In this implementation, we employ a second-order polynomial for the form of the initial kernel function ($K(x_t, x)$). This function can be redefined by a normalization, as given by the following expression:

$$\widetilde{K}(x_t, x) = \frac{K(x_t, x)}{\sqrt{K(x_t, x_t)K(x, x)}} \tag{2}$$

The normalization of the kernel function can be viewed as a simultaneous rescaling of data rows and columns to obtain a matrix with all diagonal entries set to one (Graf & Borer, 2001). In the second step of the SVR algorithm, we perform data separation by constructing linear regression in the higher-dimensional feature space:

$$f(x) = \omega^T \phi(x) + b, \tag{3}$$

where $f(x)$ represents regression model function regarded as a hyper-plane in the new feature space, ω represents the weight vector of a hyper-plane, and b is a bias term. This SVR model is formulated as a convex programming problem. The optimization goal is to determine the trade-off between the flatness of $f(x)$ while making sure that it has at most an ϵ deviation between the obtained targets and training data outputs. This means that the loss function ignores any training data close to the model prediction. In our work, we use the SVR implementation given by Shevade et al. (2000). The ϵ parameter of the ϵ-insensitive loss function is set to 0.001.

2.4. Random forest (RF) model

A random forest (RF) is an ensemble machine learning technique introduced by Breiman (2001). It consists of a collection of regression trees whose averaged outputs determine the final prediction of the ensemble. RF learning is based on an extension of bootstrap aggregation (bagging) using random features subspace selection. Through a bagging procedure, each tree in the ensemble has randomly selected portions of training samples

(with replacement) from the original data set. However, in order to avoid possible correlations between constituent random trees and enhance the estimation performance of the RF model, the idea of a random features subspace is introduced. As a result, each tree is grown on the basis of a randomly chosen input subset. For each node, the splitting algorithm searches over a subset of selected features to determine the best split point. The algorithm continues splitting new nodes until a stop criterion is reached, e.g., when the number of instances of the node is less than a preset number. The RF growth during the learning process is determined by two parameters: the size of the features subset used in each regression tree, and the number of trees that form the forest. In our implementation, we select the features subset size as the first integer lower than $log_2 k + 1$, where k is the number of inputs ($x_{it}, i = 1, \ldots, k$), while 100 trees were adopted to build the RF structure.[7]

In the process of constructing each decision tree by splitting its nodes into descendant ones, the RF learning algorithm uses the classification and regression trees (CART) learning algorithm (Breiman et al., 1984), which adopts the Gini index as the impurity-based measure (classification), or a "goodness of fit" measure (regression). In the regression context, a typical splitting criterion is the sum of squared residuals at the node where the optimal cut-off point is determined on the basis of an evaluation of splits between every distinct pair of consecutive values. For a given input vector x, the final RF regression prediction $\hat{f}(x)$ represents the unweighted average of the predictions of all regression trees:

$$\hat{f}(x) = \frac{1}{J} \sum_{j=1}^{J} \hat{h}_j(x), \tag{4}$$

where $\hat{h}_j(x)$ is the predicted output of the jth regression tree ($j = 1, \ldots, J$). The sum of squared residuals calculated when constructing regression trees can be associated with each input feature at each split in a regression tree and used to measure the amount of variance. That is, computing the sum of the variance reductions across all splits in each tree and averaging this measure across all trees in the RF gives the value of relative importance of an input feature. The features with the largest average variance reduction are considered the most important. Feature importance in the form of variance reduction is commonly used to identify the contribution of individual predictors toward explaining the output of the RF prediction model. Such an approach will assist us in obtaining the economic interpretation of our results.

3. Data

We obtain our data from the Bloomberg Terminal and use the hourly sampling frequency for the BTC/USD exchange rate at time t (B_t). The data set spans the period from May 11, 2015 to March 7, 2019 (excluding weekends[8]). We construct hourly returns by taking the natural log-difference of B_t: $BR_t = ln(B_t) - ln(B_{t-1})$, $t = 1, \ldots, 24,517$, for a total of 24,517 observations. Fig. 2 displays the hourly B_t data, along with structural break points (denoted by solid red lines) identified by sequentially applying the Zivot–Andrews test (Zivot & Andrews, 1992) on the subsets of data. The break points between the five Bitcoin market regimes are visually sensible and are located on the following dates: November 4, 2015; June 12, 2017; December 17, 2017; and November 14, 2018. It is worthwhile to note that these breaks represent the emergence of substantial drops in the Bitcoin price that are attributed to the rising interest in Ethereum (June 12, 2017) and the possibility of stricter worldwide government regulation (December 17, 2017).

Table 1 reports the summary statistics for the hourly BTC/USD exchange rate returns across the subsamples. Several important differences between the five subsamples can be observed. First, the average return is positive and increasing over the first three subsamples, and then it becomes negative following the bubble burst on December 17, 2017. The standard deviation of returns, depicted in the bottom panel of Fig. 2, is lower for the first two subsamples, but, after it almost doubles in subsample 3, it gradually declines, remaining relatively high until the sample is exhausted. In addition, excess kurtosis can be found in all subsamples, which indicates the "extreme" nature of Bitcoin trading. The skewness coefficient increases from excess (negative) skewness in subsample 1 to a slightly positive value (0.3037) in subsample 3, which can be interpreted as a growing Bitcoin optimism in the period before the crash. The sign of the skewness coefficient reverts back to negative values in subsamples 4 and 5.

The daily data set covers the period from April 1, 2013 to June 28, 2019. The sample size of BTC/USD exchange rate returns, while controlling for the availability of technical indicators, is 1529 observations. The average daily Bitcoin return is 0.00314 (median value: 0.00227) and the standard deviation is 0.05348 or 5.348%. The skewness coefficient is close to zero (−0.0044) and the sample exhibits excess kurtosis (25.24). The hourly and daily data are divided into estimation or training data (90%) and out-of-sample or testing data (10%). The training part of each sample and subsample is, chronologically, the first 90% of the data, while the testing observations are taken as the most recent (i.e., chronologically last) 10% of the data. For model calibration (validation), we use 10% of the estimation data.

As indicated above, the predictors are a set of the most representative and popular technical indicators that are listed in Table 2.[9] They are defined as follows (Kirkpatrick & Dahlquist, 2016):

[7] We do not empirically restrict the maximum depth of the tree, because we find that the RF model's forecast performance is not sensitive to reasonable choices such as 10 or 20 levels.

[8] Bitcoin markets experience periods of very low activity on weekends (Petukhina et al., 2020). Moreover, on certain weekends, Bitcoin price movements follow the so-called "Bart Simpson" trading pattern, when the price climbs on Friday, remains stable for the weekend, then dips on Monday. Source: BNN Bloomberg.

[9] Searching for an "optimal" non-linear model specification based on technical indicators is beyond the scope of the current paper and is thus left for future research. Rather, the intention is to demonstrate that technical analysis combined with complex non-linear models

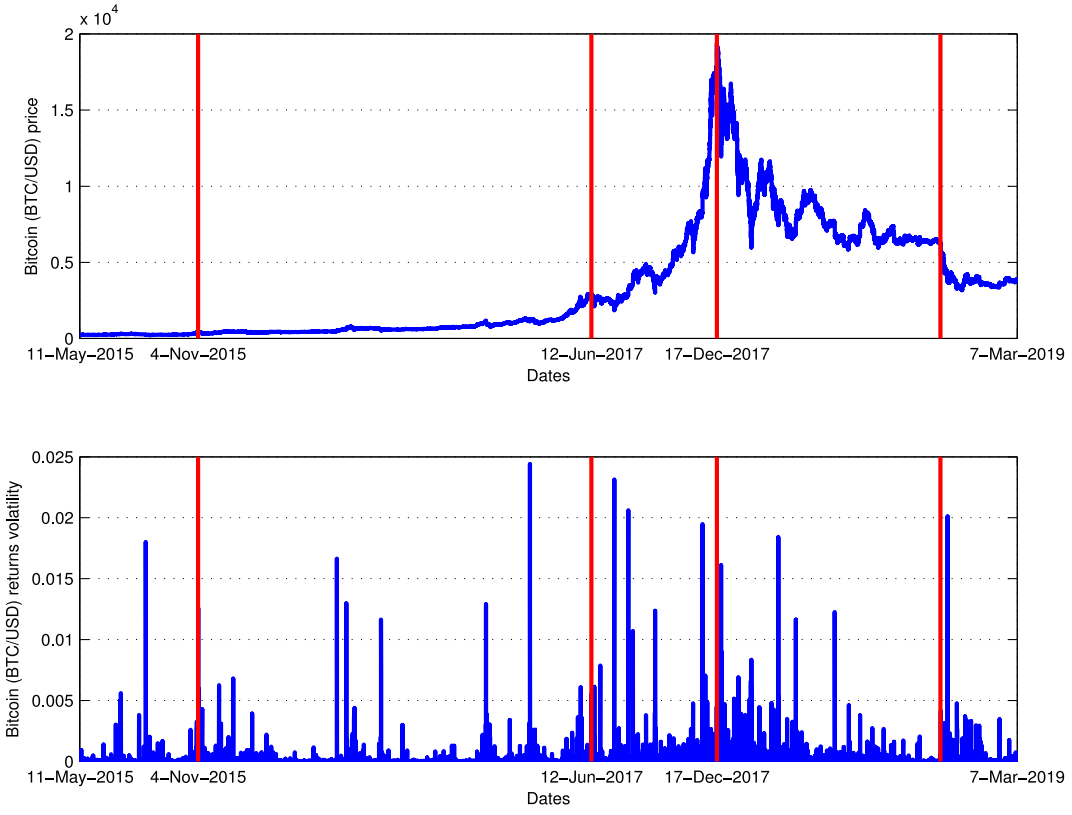


Fig. 2. BTC/USD exchange rate and volatility of returns. Notes: The data are hourly (May 11, 2015 to March 7, 2019). Top panel: BTC/USD exchange rates (24,523 observations). Bottom panel: Volatility of the BTC/USD exchange rate returns (24,523 observations). Vertical red lines indicate break points found sequentially according to the Zivot–Andrews test for structural breaks (Zivot & Andrews, 1992). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary statistics for hourly BTC/USD exchange rate returns.

Variable	Obs.	Mean	Median	Std. Dev.	Min	Max	Skewness	Kurtosis
BTC/USD returns								
Subsample 1	3,091	0.00021	0.00008	0.00670	−0.1342	0.0708	−2.3139	80.63
Subsample 2	10,288	0.00017	0.00018	0.00772	−0.1562	0.1290	−1.0011	64.96
Subsample 3	3,286	0.00056	0.00068	0.01284	−0.1521	0.1435	0.3037	25.50
Subsample 4	5,845	−0.00018	−0.00002	0.01128	−0.1356	0.1078	−0.3250	21.91
Subsample 5	2,007	−0.00023	0.00005	0.00954	−0.1418	0.0690	−1.3817	38.40

Notes: Subsample 1 is from May 11, 2015 to November 4, 2015 (3098 observations), subsample 2 is from November 4, 2015 to June 12, 2017 (10,288 observations), subsample 3 is from June 12, 2017 to December 17, 2017 (3286 observations), subsample 4 is from December 17, 2017 to November 14, 2018 (5,845 observations), and subsample 5 is from November 14, 2018 to March 7, 2019 (2003 observations).

1. Moving Average Crossover. Let m_t^n denote the time t value of a moving average rule of length n. Consequently, m_t^n is calculated by $m_t^n = (1/n)\sum_{i=0}^{n-1} B_{t-i}$, when $t \geq n$. Then, the moving average crossover indicator is $MA(n_1, n_2) = m_t^{n_1} - m_t^{n_2}$, where n_1 and n_2 are the short and the long moving averages, respectively. When the crossover indicator becomes positive (negative), a potential positive (negative) trend change is signaled. The rules used in this paper are given in Table 2. Such combinations of n_1 and n_2 were found to be profitable in the technical trading literature (e.g., Gençay & Stengos, 1998).

2. Moving Average Convergence-Divergence: MACD2. The MACD represents the difference between the 12- and 26-period exponential moving averages (EMAs). The Signal is a 9-period exponential moving average of the MACD. MACD2(12,26,9) is the difference between the MACD and the Signal. Essentially, as the Signal line crosses above (below) the differential, the interpretation is made that the price pattern is entering a new phase of positive (negative) acceleration. The EMA of length n at time t is recursively calculated as $EMA_t^n = k(B_t - EMA_{t-1}^n) +$

can be useful for forecasting and identifying the underpinnings of Bitcoin fluctuations. In some preliminary exercises, recursive forecasting schemes and more parsimonious non-linear model architectures already showed promise.

Table 2
Technical indicators (predictors).

Technical indicator	Interpretation
Moving Average Crossover: MA(1,50)	Trend
Moving Average Crossover: MA(1,200)	Trend
Moving Average Crossover: MA(5,200)	Trend
Moving Average Crossover: MA(2,200)	Trend
Moving Average Crossover: MA(1,150)	Trend
Moving Average Convergence-Divergence: MACD2(12,26,9)	Momentum
Rate of Change: ROC(10)	Momentum
Rate of Change: ROC(50)	Momentum
Rate of Change: ROC(200)	Momentum
Relative Strength Index: RSI(14)	Momentum
Relative Strength Index: RSI(20)	Momentum
Relative Strength Index: RSI(50)	Momentum
Stochastics/Williams%R: WLPR(14)	Momentum
Stochastics/Williams%R: WLPR(20)	Momentum
Stochastics/Williams%R: WLPR(50)	Momentum
On-Balance-Volume: OBV(20)	Volume/market sentiment
On-Balance-Volume: OBV(50)	Volume/market sentiment
On-Balance-Volume: OBV(100)	Volume/market sentiment
Fear & Greed: FG(4)	Market sentiment
Fear & Greed: FG(10)	Market sentiment

Notes: The first column specifies the predictors (inputs) that are used in the forecasting models. The numbers in the parentheses represent the total number of past periods (data points) used in the calculation of technical indicators. The second column explains what each technical indicator measures. "Trend" is a directional movement of prices that remains in effect long enough to be identified and exploited; "momentum" measures percentage changes in the price movements over a period of time; "volume" measures how anxious traders are to establish or close out their positions; "market sentiment" assesses the prevailing psychological tendencies of the market.

EMA^n_{t-1}, where k is the smoothing constant and $t \geq n$. A simple m^n_t is used for the first period's EMA calculation.

3. Rate of Change. Similar to the MACD2, the rate of change (ROC) is a momentum indicator. It measures how much a price has changed over a given number of past (n) periods. It is calculated at time t as $ROC^n_t = (B_t - B_{t-n})/B_{t-n}$ (for $t \geq n$). A steadily increasing ROC above the zero line indicates increasing momentum. We utilize short- (ROC^{10}_t), intermediate- (ROC^{50}_t), and long-term (ROC^{200}_t) momentum measures as predictors.

4. Relative Strength Index. This is another momentum indicator that is based on the ratio of the average "up" movements to the average "down" movements over a period of time (n periods). It is defined (at time t) as $RSI^n_t = 100 - [100/(1 + [AUPS^n_t/ADOWNS^n_t])]$, where $AUPS^n_t$ ($ADOWNS^n_t$) is a simple m^n_t of the up (down) movements in B_t over the past n periods (for $t \geq n$). When $RSI > 70$, this indicates an overbought condition, and when $RSI < 30$, this is an oversold condition. Based on Kirkpatrick and Dahlquist (2016), we use the ROC^{14}_t, ROC^{20}_t, and ROC^{50}_t that are supposed to perform well over the short to medium term.

5. Stochastics/Williams%R. This is a momentum indicator that can be thought of as an inverted stochastic oscillator. The Williams%R compares the current close with the high that occurred during the time window of n periods. Thus, the formula for the Williams%R at time t (for $t \geq n$) is as follows: $WLPR^n_t = [(max(B_t, \ldots, B_{t-n+1}) - B_t)/(max(B_t, \ldots, B_{t-n+1}) - min(B_t, \ldots, B_{t-n+1}))] \times 100$. When the indicator is between -20 and zero the price is overbought, and when the indicator is between -80 and -100, the price is oversold. We apply the same logic as for the RSI and construct $WLPR^{14}_t$, $WLPR^{20}_t$, and $WLPR^{50}_t$ for the Bitcoin predictors.

6. On-Balance-Volume. This is a volume indicator that measures how anxious traders are to establish or close out their positions (Neely et al., 2014). The on-balance-volume (OBV^n_t) indicator at time t (for the previous n time periods, when $t \geq n$) is calculated in the following manner: if $B_t > B_{t-1}$ ($B_t < B_{t-1}$), the current Bitcoin volume is added to (subtracted from) the cumulative total. In theory, OBV^n_t changes should precede price changes. We focus on the short- to medium-term Bitcoin movements and add OBV^{20}_t, OBV^{50}_t, and OBV^{100}_t to the set of predictors.

7. Fear & Greed. This technical indicator assesses the prevailing psychological tendencies of the market. Fear & Greed (FG) is a proprietary indicator from Bloomberg calculated in a way that it oscillates around a zero base line. Positive (negative) values of the indicator identify when price volatility is supporting a bullish (bearish) trend. There is a sensitivity factor used to smooth out the oscillator. Higher numbers (7–10) are used for longer-term analysis, and lower numbers (2–4) as a shorter-term indication. We use FG(4) and FG(10) as our predictors.

We complement the set of 20 technical indicators from Table 2 with two additional predictors: global spot Bitcoin volume, and the Google search trend index. The volume of Bitcoin transactions is defined as all worldwide transactions valued in U.S. dollars. The Google search trend index was constructed from the worldwide Google search trend history with the keyword "Bitcoin". Both inputs

Table 3
Augmented Dickey–Fuller (ADF) unit root t-tests.

Variable (hourly)	t-stat	Lags(f)	Variable (daily)	t-stat	Lags(f)
BR_t (BTC/USD returns)	−49.420*	9	BR_t (BTC/USD returns)	−11.342*	7
MA(1,50)	−21.946*	19	MA(1,50)	−5.567*	18
MA(1,200)	−8.197*	8	MA(1,200)	−3.807**	16
MA(5,200)	−11.596*	19	MA(5,200)	−3.768**	48
MA(2,200)	−9.616*	20	MA(2,200)	−3.858**	15
MA(1,150)	−9.268*	29	MA(1,150)	−3.174***	16
MACD2(12,26,9)	−35.489*	9	MACD2(12,26,9)	−8.681*	19
ROC(10)	−22.350*	49	ROC(10)	−5.068*	32
ROC(50)	−12.418*	59	ROC(50)	−5.035*	9
ROC(200)	−7.987*	19	ROC(200)	−3.974*	41
RSI(14)	−23.147*	9	RSI(14)	−6.354*	1
RSI(20)	−20.096*	9	RSI(20)	−5.380*	1
RSI(50)	−13.560*	9	RSI(50)	−3.571**	1
WLPR(14)	−27.634*	25	WLPR(14)	−9.562*	1
WLPR(20)	−25.946*	25	WLPR(20)	−8.106*	1
WLPR(50)	−20.543*	24	WLPR(50)	−4.883*	1
FG(4)	−34.562*	9	FG(4)	−9.451*	18
FG(10)	−23.056*	9	FG(10)	−7.388*	19
Google search trend	−4.168*	70	Google search trend	−4.499*	9
Bitcoin volume	−17.968*	46	Bitcoin volume	39.090*	0
OBV(20) changes	−29.055*	39	OBV(20) changes	−11.197*	9
OBV(50) changes	−28.681*	70	OBV(50) changes	−10.718*	19
OBV(100) changes	−22.768*	70	OBV(100) changes	−12.076*	9

Notes: The critical values are: *−3.960 (1%), **−3.410 (5%), ***−3.120 (10%).

were measured over a period of time (hourly and daily) and they were included based on their predictive power documented by Panagiotidis et al. (2018) and Balcilar et al. (2017). We consider them global market sentiment indicators.

Lastly, following a referee's recommendation,[10] we perform standard Dickey–Fuller unit roots tests on all variables, based on the entire data set. Some technical indicators are computed from the levels of the BTC/USD exchange rate and this might lead to troublesome non-stationary predictors. If some of the predictors are found to be non-stationary, transforming them to achieve stationarity could be a viable option to solve the problem.

Table 3 displays the augmented Dickey–Fuller (ADF) unit root t-test statistics, along with their optimal lag orders for each predictor. For the majority of cases, the null hypothesis of non-stationarity is rejected at the 1% significance level and there is no need to transform any of the predictors. It is important to stress that the OBV technical indicator is by definition considered in first differences and is stationary as such. Also, to our knowledge, the predictors did not undergo any revisions during their construction and visual data inspection did not show any jumps or inconsistencies.

4. Empirical results

4.1. Hourly data

We run our predictive models based on the data breakdown from Table 1. Our goal is to compare the forecast performance of the competing models (FF-D-ANN, SVM, RF, RW, and ARMAX) on each subsample; that is, we attempt to link the accuracy of out-of-sample predictions of BTC/USD exchange rate returns to the temporal regimes of the market. The random walk (RW) model is specified as follows:

$$BR_t = \alpha + \epsilon_t, \quad \epsilon_t \sim ID(0, \sigma^2), \tag{5}$$

where BR_t denotes the BTC/USD exchange rate log-returns. We generally set $\alpha = 0$, except when we have to assess the directional accuracy of the RW model, because a "driftless" RW model always predicts a zero return.

The optimal model architecture (obtained from the validation data) for the FF-D-ANN model is "22–100–30–10–1", i.e., three hidden layers with 100, 30, and 10 neurons, respectively.[11] The RF and SVM models were specified as explained in Section 2. Table 4 shows the statistical accuracy reflected in the mean squared prediction error (*MSPE*) values as well as the directional accuracy (*Sign*)[12] of the models that were tested on out-of-sample data.

What immediately stands out in Table 4 is the relative inability of all non-linear models to statistically significantly improve upon the random walk model. This is generally the case when both forecast performance criteria are considered in all subsamples. Hence, it can be concluded that the Bitcoin market is efficient in the weak sense at the hourly time horizon. We conjecture that intraday data are potentially subject to jumps, intraday periodicity, and microstructure noise, leading to poor forecasts of the BTC/USD exchange rate returns (Banulescu-Radu et al., 2016). A few interesting stylized

[10] The authors are grateful to Amir Atiya (Editor), the Associate Editor, and an anonymous referee for their very useful comments.

[11] This choice of architecture also worked well on the daily data. Additional adjustments to the number of neurons and layers did not produce any significant forecast improvements.

[12] We calculate the *Sign* statistic ex-post as the percentage of correctly forecasted directions of exchange rate movements, on the basis of the forecasts for the returns and the actual out-of-sample returns.

Table 4
Comparison of predictive models (hourly data).

	Subsample: (*MSPE*) (*Sign*)					
	1	2	3	4	5	All data
FF-D-ANN	13.123 52.75	11.616 46.64	32.379 47.25	1.185 49.48	1.525 49.25	11.276 48.38
SVM	14.060 24.59	11.874 40.33	34.422 49.08	1.136 22.94	1.396 24.37	11.754 34.06
RF	17.839 37.21	52.847 42.85	32.379 43.90	1.450 35.78	1.651 35.32	29.246 39.98
RW	13.200 59.54	11.597 56.85	31.988 55.62	1.109 48.54	1.320 50.24	11.190 54.50
ARMAX(1,1)	14.035 51.78	13.118 52.77	31.918 60.18	1.137 46.66	1.411 46.76	11.939 51.69

Notes: The predictive performance is observed for the last 10% hourly observations of each subsample as specified in Table 1. The *Sign* [%] and *MSPE* ($\times 10^{-5}$) statistics are used for the out-of-sample (testing) data. The top entry in each cell is the *MSPE*, with the *Sign* statistic reported below it. We list both statistics for each competing model across five subsamples $(1, 2, \ldots, 5)$ and the total sample (All data). The models that are considered are (1) Feedforward Deep Artificial Neural Network (FF-D-ANN), (2) Support Vector Machine (SVM), (3) Random forest (RF), (4) the random walk (RW) model, and (5) ARMAX(1,1) model.

Table 5
Relative contribution of inputs (hourly data).

Predictors (x_{it})	Subsample: (Feature importance-RF)				
	1	2	3	4	5
MA(1,50)	0.0405	**0.0558**	0.0455	0.0450	0.0447
MA(1,200)	0.0433	0.0442	0.0388	0.0470	0.0443
MA(5,200)	**0.0501**	0.0460	0.0466	0.0468	0.0472
MA(2,200)	0.0475	0.0435	0.0416	0.0432	0.0459
MA(1,150)	0.0453	0.0416	0.0418	0.0441	0.0447
MACD2(12,26,9)	0.0485	0.0490	0.0485	0.0442	0.0448
ROC(10)	0.0468	**0.0544**	**0.0584**	0.0436	0.0420
ROC(50)	0.0493	**0.0570**	**0.0598**	0.0451	**0.0568**
ROC(200)	**0.0546**	**0.0508**	**0.0497**	0.0484	**0.0562**
RSI(14)	0.0410	0.0391	0.0395	0.0377	0.0364
RSI(20)	0.0405	0.0377	0.0411	0.0359	0.0359
RSI(50)	0.0398	0.0387	0.0443	0.0373	0.0384
WLPR(14)	0.0466	0.0428	0.0449	0.0451	0.0406
WLPR(20)	0.0434	0.0428	0.0422	0.0411	0.0407
WLPR(50)	0.0446	0.0436	0.0413	0.0399	0.0460
FG(4)	**0.0498**	0.0481	0.0479	0.0481	0.0439
FG(10)	0.0445	0.0484	0.0474	0.0482	0.0423
Google search trend	0.0457	**0.0499**	0.0432	**0.0787**	**0.0658**
Bitcoin volume	0.0465	0.0460	**0.0504**	**0.0516**	**0.0535**
OBV(20)	0.0440	0.0404	0.0433	0.0450	0.0420
OBV(50)	0.0446	0.0395	0.0418	0.0414	0.0454
OBV(100)	0.0459	0.0394	0.0407	0.0417	0.0415

Notes: Feature importance is extracted from the estimated RF model (values that are greater than 0.0495 are reported in bold). Larger values denote more significant predictors. Subsamples $(1, \ldots, 5)$ are specified in Table 1.

facts, however, emerge from Table 4. The predictive performance of all models in subsample 3 (which exhibits the highest mean return and volatility) is the worst. In the subsequent two subsamples covering the 2017–2019 data, the volatility decreases, while the mean return becomes negative. This substantially improves the forecast performance of all models. It can be noted that the pessimism and negative momentum contained in hourly returns after the bubble peaked in December, 2017 was more predictable when compared to the price surge that was in effect from 2015–2017.

Next, to understand the origins of Bitcoin fluctuations across the subsamples, we calculate the feature importance for each predictor in the RF model. The first column of Table 5 lists the predictors, and the remaining columns show the corresponding feature importance values in each time period $(1, \ldots, 5)$. First, it is apparent that the trend and momentum indicators drive the BTC/USD exchange rate returns in the first two subsamples (from mid-2015 to mid-2017). The two most important features in subsample 1 are MA(5,200) and ROC(200), which suggests that longer trend and momentum (and, to a smaller extent, "greed") were responsible for the quick rise of Bitcoin. In the second subsample, short- to medium-term trend (MA(1,50)) and momentum (ROC(10), ROC(50)) gained more importance. Moreover, there are indications that market sentiment came into play during this period (FG and Google search trend). In addition to the momentum, in the third subsample (from mid- to late-2017), the attention switched more towards market sentiment indicators (FG and volume), while the trend became unimportant. This phenomenon was reinforced in the last two subsamples (2018–2019) where Google search trend and volume were the dominant features. Clearly, the burst of the Bitcoin bubble introduced an additional layer of uncertainty to the market. We attribute the source of the uncertainty to psychological forces that drove Bitcoin movements in the past couple of years.

4.2. Daily data

As mentioned in the previous section, the daily data set contains 1529 observations (22 inputs and 1 output). To assess the forecast performance of the competing models, we hold the last 10% or 153 observations out of sample. We consider whether the competing models[13] can outperform the RW model in terms of *MSPE* and the percentage of correctly predicted directions of Bitcoin price changes (*Sign*). We use the Diebold–Mariano test (Diebold & Mariano, 1995) to test the null hypothesis that there is no difference in the *MSPE* of two alternative models (in our case of the RW and the four alternative models).[14]

Table 6 shows the forecast performance of the competing models. We find strong evidence of the one-day-ahead predictability of BTC/USD exchange rate returns. All competing models are able to statistically significantly improve upon a simple random walk model in terms of both (*MSPE* and *Sign*) measures. The best-performing

[13] We do not include GARCH-type specifications in the set of competing models, as they are typically used for modeling price volatility. In fact, when we used the GARCH-M(1,1) model with technical indicators for (daily) point BTC/USD exchange rate forecasting, its performance was statistically similar to the RW model. The results are available by request from the authors.

[14] More information on this forecasting framework can be found in Gradojevic and Yang (2006).

Table 6
Comparison of predictive models (daily data).

	Statistic	
	Sign [%]	*MSPE* (DM)
FF-D-ANN	100.00	1.337 (−5.054*)
SVM	80.39	1.949 (−1.674**)
RF	100.00	1.252 (−4.003*)
RW	58.17	2.298
ARMAX(1,1)	97.38	1.762 (−4.824*)

Notes: For the entire data range (22 inputs, 1 output), 1,529 observations, we allocate 80% training, 10% validation, 10% testing (last 153 days). The *Sign* [%] and *MSPE* ($\times 10^{-3}$) statistics are used for the out-of-sample (testing) data. The models that are considered are (1) Feedforward Deep Artificial Neural Network (FF-D-ANN), (2) Support Vector Machine (SVM), (3) Random forest (RF), (4) the random walk (RW) model, and (5) ARMAX(1,1) model. The Diebold–Mariano (DM) test statistics (Diebold & Mariano, 1995) are reported in the parentheses below *MSPEs*, where applicable. The critical values are ±1.64 and ±2.58 for a confidence level of 90% and 99%, respectively. (*) and (**) indicate the DM statistic is significant at 1% and 10% significance level, respectively.

model is an RF that produces almost 50% lower out-of-sample *MSPE* than the RW model. Additionally, its sign prediction is remarkable, with 100% accuracy (Fig. 3). Similar outstanding performance is achieved by the FF-D-ANN model. A surprising finding is that the SVM underperformed relative to the ARMAX model. In all, the results demonstrate the empirical validity of the choice of technical indicators as predictors. Also, in relation to the results obtained from hourly data, it appears that lowering the data frequency to daily data effectively resolves the problem of poor intraday forecast performance that is likely due to microstructure frictions and other market shocks occurring at high frequencies.[15] The predictors—in combination with sophisticated non-linear methods—become very informative at the daily forecast horizon, resulting in extremely high forecast accuracy.

In this part of our empirical analysis, the method of pseudo-weights proposed by Qi and Maddala (1996) is adapted to identify the non-linear economic importance of the inputs of the FF-D-ANN model. More precisely, the weighted average of the input weights is used to find the marginal contribution of each input variable to the output. The formula for a pseudo-weight for the ith input ($PW_i, i = 1, \ldots, k$) of an FF-D-ANN with n hidden layers can, in general, be written as

$$PW_i = \sum_{j_1=1}^{q_1} \sum_{j_2=1}^{q_2} \ldots \sum_{j_n=1}^{q_n} \alpha_{ij_1} \alpha_{j_1 j_2} \ldots \alpha_{j_{(n-1)} j_n} \beta_{j_n}, \quad i = 1, \ldots, k \tag{6}$$

where $q_1, q_2, \ldots, q_n$ is the number of nodes in each of the n hidden layers; α_{ij_1} denotes connection weights between the input layer and the first hidden layer (i.e., between the ith input node and the j_1^{th} first hidden layer node); $\alpha_{j_1 j_2}, \ldots, \alpha_{j_{(n-1)} j_n}$ denote connection weights between the adjacent hidden layers; and β_{j_n} represents the connection weights between the last (nth) hidden layer (j_n^{th} hidden layer node) and the output layer (single node).

The first two columns of Table 7 display all 22 inputs along with their corresponding estimated (in-sample) pseudo-weights ($PW_i, i = 1, \ldots, 22$). On the whole data range of more than six years of observations, it transpires that Bitcoin fluctuations have been in large part driven by market sentiment, followed by price momentum. Fear & Greed (FG) appears to be the most dominant predictor of BTC/USD exchange rate returns on a daily horizon.

Feature importance values, extracted from the RF model and listed in the last column of Table 7, complement the evidence from pseudo-weights. In general, feature importance also attributes the most predictive power to the price momentum (ROC) and market sentiment contained in volume (OBV). However, both measures of the relative importance of inputs suggest that trend indicators are not very informative for Bitcoin forecasting.[16] This is in accord with the speculative nature of Bitcoin trading, which is very sensitive to investors' pessimism and optimism about the future.

4.3. Robustness exercises

We begin our robustness analysis with evidence on the comparative predictive performance of daily and hourly models over the same time period. Then, we turn to expanding the set of competing models in terms of both methodology and additional predictors. Overall, the results demonstrate the predictive dominance of the RF model with technical indicators and the effective federal funds rate as the appropriate inputs.

Considering that the hourly and daily data start and end at different points in time, it is not obvious whether using the same time periods for training and testing samples would produce different results. Also, by defining precise dates for both data frequencies, we would be able to directly compare the forecast performance of the models and improve the reading experience. We perform such a comparison on the maximum possible sample overlap for both data frequencies, that is, for the period from May 11, 2015 to March 7, 2019. This amounts to 24,517 observations at the hourly horizon and 999 observations at the daily horizon. We fix the training set (which includes the validation set) to the period from May 11, 2015 to October 19, 2018, while the testing set runs from October 20, 2018 to March 7, 2019. Based on this data breakdown, the daily (hourly) training set is 899 (22,079) observations and the daily (hourly) out-of-sample testing set is 100 (2,438) observations. Table 8 lists the forecast performance measures pertaining to the RF and RW models.

Before we delve into further discussion, it is important to note that the *MSPE* figures are of a different order of magnitude for daily and hourly data, whereas those for the hourly data are naturally smaller. Hence, we compare the relative (*MSPE*) performance of the RF models to the

[15] Banulescu-Radu et al. (2016) tackle this problem by pre-filtering high-frequency data to account for the presence of jumps, intraday periodicity, and microstructure noise.

[16] Gradojevic and Gençay (2013) show that moving average indicators are imperfect filters with a "nonzero phase shift", which means they are typically late at identifying turning points.

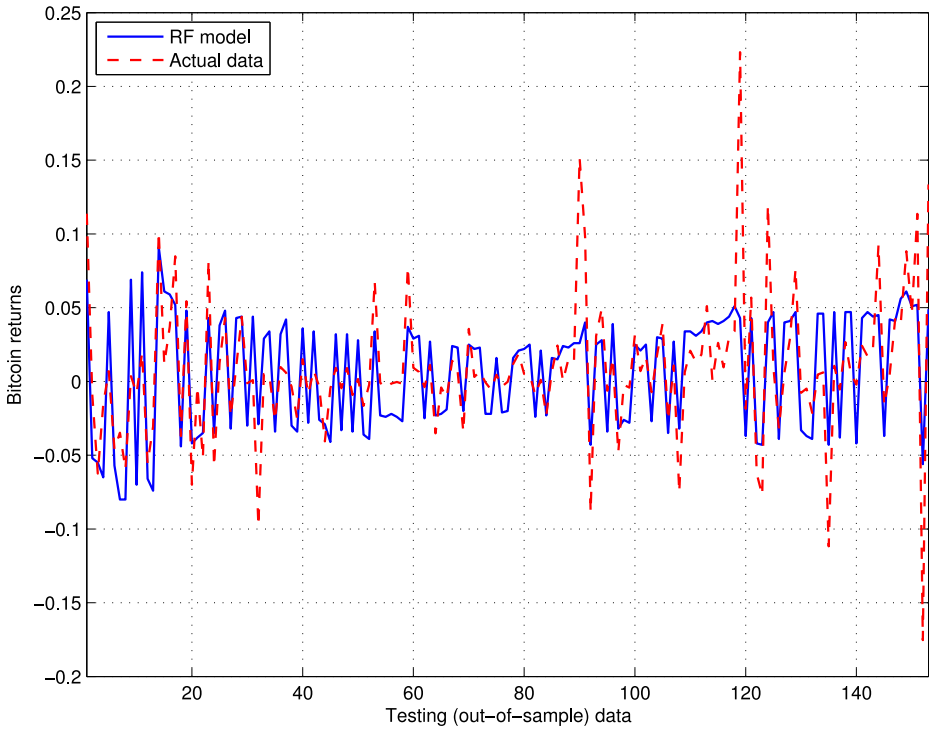


Fig. 3. RF model's forecast performance. Notes: The RF model uses 22 inputs (lagged technical indicators) and predicts BTC/USD exchange rate returns. The in-sample size is 1529 observations, and we hold the last 153 days out of sample.

Table 7
Relative contribution of inputs (daily data).

Predictors (x_{it})	Pseudo-weights (PW_i)	Feature importance (RF)
MA(1,50)	−0.0007	0.0386
MA(1,200)	−0.0014	0.0360
MA(5,200)	−0.0014	0.0440
MA(2,200)	−0.0016	0.0419
MA(1,150)	−0.0013	0.0356
MACD2(12,26,9)	0.0053	0.0399
ROC(10)	−0.0102	**0.0566**
ROC(50)	−0.0146	0.0448
ROC(200)	−0.0166	**0.0501**
RSI(14)	0.0477	0.0487
RSI(20)	0.0052	0.0458
RSI(50)	0.1105	0.0487
WLPR(14)	0.1101	**0.0499**
WLPR(20)	0.1339	**0.0496**
WLPR(50)	0.3054	0.0394
FG(4)	0.3053	0.0454
FG(10)	0.3050	0.0380
Google search trend	0.2898	0.0437
Bitcoin volume	0.2898	0.0453
OBV(20)	0.2898	**0.0541**
OBV(50)	0.2898	**0.0540**
OBV(100)	0.2898	0.0488

Notes: The PW_i $(i = 1, \ldots, 22)$ values are found by using Eq. (6), given the estimated (trained) FF-D-ANN model (in-sample). Feature importance is extracted from the estimated RF model (values that are greater than 0.0495 are reported in bold). Larger values denote more significant predictors.

RW models for daily and hourly data. However, the *Sign* statistics are directly comparable.

For the daily data, the results resemble those from Table 6. The *Sign* performance of the RF model is astounding again, with accuracy of 98%, whereas the sign accuracy of the RW model is below 50%. Moreover, the RF model improves the RW model's *MSPE* statistics by about 46%. In contrast, the evidence for the hourly data shows that the RF model performs poorly and its *MSPE* is about 43% worse relative to that of the RW model. In

Table 8
Comparison of daily and hourly models.

	Statistic	
	Sign [%]	*MSPE*
RF (daily)	98.00	9.42
RW (daily)	45.00	17.19
Ratio RF/RW	2.17	0.54
RF (hourly)	28.34	1.08
RW (hourly)	41.73	0.75
Ratio RF/RW	0.68	1.43

Notes: For the data range from May 11, 2015 to March 7, 2019 (22 inputs, 1 output), which amounts to 24,517 hourly observations and 999 daily observations, we allocate 80% training, 10% validation, 10% testing (last 100 days or 2,438 h). The *Sign* [%] and *MSPE* ($\times 10^{-4}$) statistics are used for the out-of-sample (testing) data. The models that are considered are (1) Random forest (RF) and (2) the random walk (RW) model. "Ratio RF/RW" denotes the relative performance of the RF model to the RW model.

addition, the pricing accuracy of both RW and RF models is weak, with the *Sign* statistics lower than 50%. Notably, the relative performance based on the "Ratio RF/RW" statistics resembles Table 4, although the performance of both the RW and RF models in Table 8 is significantly worse. This can be explained by the fact that the hourly model from Table 8 forecasts a substantially greater number of data points than the daily model (and also greater than in Table 4), which is computationally more demanding. Moreover, both in-sample and validation data sets are much larger for the hourly model over the whole period from May 11, 2015 to March 7, 2019. Forecasting based on a long in-sample data period of hourly observations might include potentially uninformative historical data from obsolete market regimes and that could deteriorate the predictive performance of the hourly model.

Next, we consider alternative competing models for the original model specifications with 22 predictors. Given the dominant forecast accuracy at the daily horizon, the focus is on this data frequency. We estimate the following models:

- AR(1) model;
- lasso regression;
- ridge regression;
- elastic net regression;
- unobserved components model with stochastic volatility (Stock & Watson, 2007).

The top panel of Table 9 lists the *Sign* and *MSPE* statistics for the testing data. We also test the null hypothesis that there is no difference in *MSPE* between the RF and the alternative models, and we display the DM statistics (Diebold and Mariano, 1995) in parentheses. We estimate lasso, ridge, and elastic net regressions by using the coordinate descent algorithm, which involves two stages (Friedman et al., 2007). In the first stage, we estimate the tuning parameters (α and/or λ) with ten-fold cross-validation on the in-sample data.[17] In the second stage, from the estimates of the tuning parameters, we estimate the vector of regression coefficients, and then generate out-of-sample forecasts.

The results show that the RF model produces the most accurate one-day-ahead forecasts of the BTC/USD exchange rate returns at the 1% significance level. In other words, the RF model is able to statistically significantly improve upon the competing models. In addition, the RF model is dominant in predicting the direction of BTC/USD exchange rate changes. Of the alternative model specifications, the lasso and elastic net models show the most promise. Their forecast performance is comparable to that of the SVM model in Table 6. The top panel (A) of Fig. 4 visually confirms the evidence in Table 9. Indeed, the forecasts generated by the RF model follow the actual out-of-sample data relatively closely. Meanwhile, the lasso and the AR(1) models seem to be less adaptable and produce a smoother approximation of the data.

Following this, we expand the panel of predictors with lagged BTC/USD log-returns and fundamentals (i.e., macro-financial data specific to the U.S. and global indicators to proxy the world risk appetite). The fundamental predictors available on a daily basis include the effective federal funds rate (EFFR), the nominal FRB dollar index (NFRB), and the VIX.[18] Also, based on several lag-order selection statistics, we include seven lags of the BTC/USD exchange rate returns into our expanded set of predictors (denoted LAG1,…,LAG7). Henceforth, we assess the forecast performance of the expanded model with 32 predictors. Measures of the forecast accuracy of the competing models are presented in the bottom panel of Table 9. We find it important to stress that the size of our data set slightly shrank (from 1529 to 1370 observations), due to the unavailability of market fundamentals on certain days.

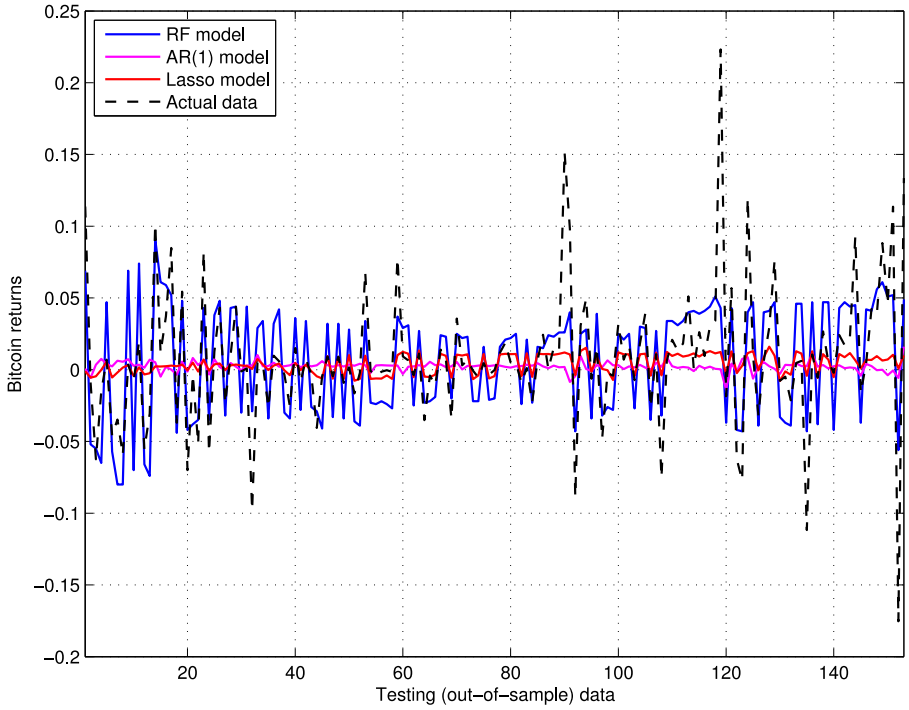
The bottom panel of Table 9 suggests that the expanded model with 32 predictors estimated by the RF model easily delivers superior forecast accuracy relative to the competing models. The *MSPE* (*Sign*) figures generated by the RF model are statistically significantly below (above) the respective values for the AR(1) and the unobserved components with stochastic volatility (UC-SV) models, as well as for the ridge (RIDGE), lasso (LASSO), and elastic net (ELASTIC) regressions with 32 predictors. More insight is obtained when the RF model with 22 predictors (RF22) is compared to the RF model with 32 predictors (RF32). It is shown that the additional variables improve the out-of-sample predictive power of the RF model in terms of the *MSPE* by about 30%.

The bottom panel (B) of Fig. 4 contains the out-of-sample forecasts from the best models (RF32 and ELASTIC) versus the naive benchmark (AR(1)) and actual data (marked by a black dashed line). The AR(1) model's forecasts exhibit more "smoothing" than the forecasts of other models that are more successful in keeping with the pattern of actual BTC/USD exchange rate returns. Again, the RF model is the most accurate in tracking fluctuations in the target variable without overfitting.

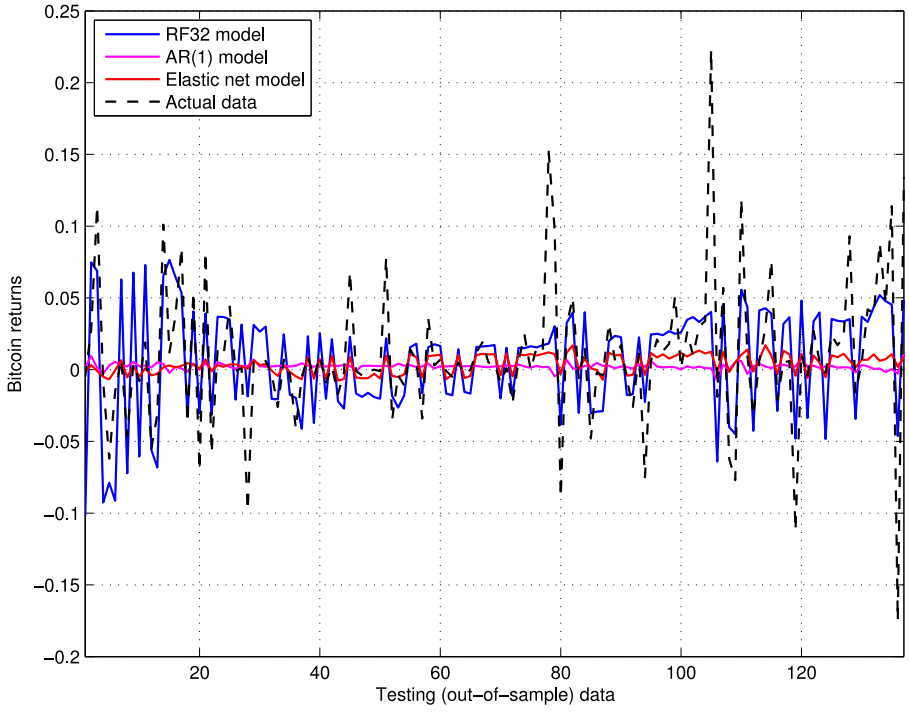
In regards to the relative importance of predictors, the values in Table 10 reveal that the effective federal funds rate (EFFR) and on-balance-volume (OBV) are the top ranked features. They are followed by the feature

[17] We generally follow Friedman et al. (2010) by allowing the tuning parameters to be found using cross-validation.

[18] The null hypothesis of non-stationarity for the log-differences of these additional time series was rejected at the 99% significance level.



Panel A



Panel B

Fig. 4. Comparative forecast performance. Notes: In Panel A: The models with external variables use 22 inputs (lagged technical indicators) and predict BTC/USD exchange rate returns. The in-sample size is 1529 observations, while we hold the last 153 days out of sample. In Panel B: The models with external variables use 32 inputs (lagged technical indicators and additional inputs) and predict BTC/USD exchange rate returns. The in-sample size is 1370 observations, while we hold the last 137 days out of sample.

Table 9
Comparison of additional models (daily data).

		Statistic	
		Sign [%]	*MSPE* (DM)
22 inputs:			
	RF	100.00	1.252
	AR(1)	62.09	2.236 (−3.951*)
	RIDGE	70.58	2.176 (−3.666*)
	LASSO	93.46	1.979 (−3.177*)
	ELASTIC	93.46	1.983 (−3.188*)
	UC-SV	64.05	2.248 (−3.720*)
32 inputs:			
	RF32	100.00	1.500
	RF22	100.00	1.987 (−2.486**)
	AR(1)	63.50	2.412 (−4.898*)
	RIDGE	68.61	2.356 (−4.619*)
	LASSO	91.97	2.169 (−4.164*)
	ELASTIC	90.51	2.149 (−4.081*)
	UC-SV	64.23	2.584 (−4.759*)

Top panel: For the entire data range (22 inputs, 1 output), 1,529 observations, we allocate 80% training, 10% validation, 10% testing (last 153 days). The models that are considered are (1) AR(1), (2) Ridge regression (RIDGE), (3) Random forest (RF), (4) Lasso regression (LASSO), (5) Elastic net regression (ELASTIC) and (5) unobserved components model with stochastic volatility (UC-SV). Bottom panel: For the entire available data range (32 inputs, 1 output), 1,370 observations, we allocate 80% training, 10% validation, 10% testing (last 137 days). The models that are considered are (1) AR(1), (2) Ridge regression (RIDGE), (3) Random forest (RF32) with 32 predictors, (4) Lasso regression (LASSO), (5) Elastic net regression (ELASTIC), (5) unobserved components model with stochastic volatility (UC-SV), and (6) Random forest with the original 22 predictors (RF22). The *Sign* [%] and *MSPE* ($\times 10^{-3}$) statistics are used for the out-of-sample (testing) data. The Diebold–Mariano (DM) test statistics (Diebold & Mariano, 1995) are reported in the parentheses below *MSPEs*, where applicable. The critical values are ± 1.96 and ± 2.58 for a confidence level of 95% and 99%, respectively. (*) and (**) indicate the DM statistic is significant at 1% and 5% significance level, respectively.

importance values for the ROC input. We conclude that in addition to price momentum and market sentiment predictors, macro-finance variables such as the effective federal funds rate have been informative for predicting Bitcoin price movements in recent years. The finding that interest rates are important predictors of Bitcoin returns is in line with, e.g., Koutmos (2020).

5. Conclusions

Accurate forecasting in an excessively volatile and regime-switching market such as the Bitcoin market can be a daunting task, especially at high frequencies. By using advanced non-linear models and technical analysis, this paper examined the predictability of BTC/USD exchange rate returns on hourly and daily forecast horizons. First, we defined five temporal Bitcoin market regimes and analyzed the out-of-sample hourly forecast performance of the competing models on each regime. The results indicated that the Bitcoin market is weakly efficient at the hourly frequency. Within the context of subsamples and market regimes, we revealed that the Bitcoin market evolved from one dominated by trend-following and momentum to a market that is driven by sentiment indicators such as fear, greed, volume, and Google search trends.[19] Essentially, in addition to the evidence of cross-sectional momentum effects in cryptocurrencies (see, e.g., Liu et al., 2020), we emphasized the increasing importance of time-series momentum effects in recent years. Finally, we documented that the degree of market efficiency varies over time as predicted by the adaptive markets hypothesis (Khuntia & Pattanayak, 2018; Lo, 2004).

When moving from hourly to daily data, we uncovered strong predictability based on both statistical and directional accuracy measures. Our horse race for forecast performance resulted in a non-linear RF model as the absolute winner. It managed to statistically significantly improve upon a simple random walk model, yielding the astounding out-of-sample directional accuracy of 100%. In general, all non-linear predictive models performed exceptionally well, which testifies to the validity of our particular choice of technical indicators for predictors. Naturally, in future research it would be worthwhile to expand the panel of technical indicators, as well as alternative cryptocurrencies, and cover more data frequencies, subject to data availability at high frequencies. Furthermore, we would like to investigate not only the time-series properties of the Bitcoin market but also how traders interact in the frequency domain. It is of the utmost importance to understand the exchange of information and microstructure interactions between low-frequency and high-frequency Bitcoin traders.

An intriguing extension of the current paper could be to gain an additional glimpse inside the machine learning "black-box" through marginal predictor impact on the

[19] A similar pattern was found by Cheng et al. (2019), who showed how investors in the last quarter of 2017 overreacted to a firm's speculative disclosure of a potential foray into Blockchain technology.

Table 10
Relative contribution of inputs (daily data, 32 predictors).

Predictors (x_{it})	Feature importance (RF)
EFFR	0.1653
DOBV(100)	0.1576
DOBV(50)	0.1168
DOBV(20)	0.1156
ROC(200)	0.0419
ROC(50)	0.0350
RSI(14)	0.0323
ROC(10)	0.0312
RSI(20)	0.0294
MA(5,200)	0.0259
RSI(50)	0.0255
LAG1	0.0186
LAG6	0.0174
LAG7	0.0173
Google search trend	0.0162
NFRB	0.0138
LAG2	0.0134
VIX	0.0134
Bitcoin volume	0.0129
LAG4	0.0123
LAG3	0.0122
FG(4)	0.0110
WLPR(14)	0.0093
WLPR(20)	0.0080
MA(1,150)	0.0073
WLPR(50)	0.0066
FG(10)	0.0064
MA(2,200)	0.0062
MA(1,200)	0.0055
MA(1,50)	0.0053
MACD2(12,26,9)	0.0053
LAG5	0.0034

Notes: Feature importance is based on variance reduction, ranked in descending order, extracted from the estimated RF model with 32 predictors. Larger values denote more significant predictors.

model's output, as in Gu et al. (2020). It would be helpful to understand how a change in one of the predictors affects the output. This would also shed new light on the economic interpretation of the findings.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adcock, C. J. (2010). Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution. *Annals of Operations Research*, *176*(1), 221–234.

Adcock, R., & Gradojevic, N. (2019). Non-fundamental, non-parametric Bitcoin forecasting. *Physica A*, *531*, Article 121727.

Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F., & Zopounidis, C. (2019). Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research*, *276*(2), 770–780.

Baig, A., Blau, B. M., & Sabah, N. (2019). Price clustering and sentiment in Bitcoin. *Finance Research Letters*, *29*, 111–116.

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, *64*, 74–81.

Banulescu-Radu, D., Hurlin, C., Candelon, B., & Laurent, S. (2016). Do we need high frequency data to forecast variances? *Annals of Economics and Statistics*, (123/124), 135–174.

Baur, D. G., Dimpfl, T., & Kuck, K. (2018). Bitcoin, gold and the dollar — A replication and extension. *Finance Research Letters*, *25*, 103–110.

Bekiros, S. D., & Georgoutsos, D. A. (2008). Direction-of-change forecasting using a volatility-based recurrent neural network. *Journal of Forecasting*, *27*(5), 407–417.

Blake, A. P., & Kapetanios, G. (2000). A radial basis function artificial neural network test for ARCH. *Economics Letters*, *69*(1), 15–23.

Bouri, E., Gil-Alana, L. A., Gupta, R., & Roubaud, D. (2019). Modelling long memory volatility in the Bitcoin market: Evidence of persistence and structural breaks. *International Journal of Finance & Economics*, *24*(1), 412–426.

Bouri, E., Lau, C. K. M., Lucey, B., & Roubaud, D. (2019). Trading volume and the predictability of return and volatility in the cryptocurrency market. *Finance Research Letters*, *29*, 340–346.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks.

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, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting* *37*(3), 1300–1301. http://dx.doi.org/10.1016/j.ijforecast.2020.02.008.

Cheng, S. F., De Franco, G., Jiang, H., & Lin, P. (2019). Riding the blockchain mania: Public firms' speculative 8-K disclosures. *Management Science*, *65*(12), 5901–5913. http://dx.doi.org/10.1287/mnsc.2019.3357.

Corbet, S., Eraslan, V., Lucey, B., & Sensoy, A. (2019). The effectiveness of technical trading rules in cryptocurrency markets. *Finance Research Letters*, *31*, 32–37.

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

Demir, E., Gozgor, G., Lau, C. K. M., & Vigne, S. A. (2018). Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters*, *26*, 145–149.

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

Donaldson, R., & Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance*, *4*(1), 17–46.

Dooley, M., & Shafer, J. (1984). Analysis of short-run exchange rate behavior: March 1973 to november 1981. In D. Bigman, & T. Taya (Eds.), *Floating exchange rates and the state of world trade and payments* (pp. 43–70). Ballinger Publishing Company Cambridge, MA.

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar – A GARCH volatility analysis. *Finance Research Letters*, *16*, 85–92.

Fernández-Rodríguez, F., González-Martel, C., & Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. *Economics Letters*, *69*(1), 89–94.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669.

Franses, P. H., & Draisma, G. (1997). Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics*, *81*(1), 273–280.

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, *1*(2), 302–332.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, Articles*, *33*(1), 1–22.

Garcia, R., & Gençay, R. (2000). Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, *94*(1–2), 93–115.

Gençay, R., & Stengos, T. (1998). Moving average rules, volume and the predictability of security returns with feedforward networks. *Journal of Forecasting*, *17*(5–6), 401–414.

Geppert, J. M., Ivanov, S. I., & Karels, G. V. (2010). Analysis of the probability of deletion of S&P 500 companies: Survival analysis and neural networks approach. *The Quarterly Review of Economics and Finance*, *50*(2), 191–201.

Gerritsen, D. F., Bouri, E., Ramezanifar, E., & Roubaud, D. (2020). The profitability of technical trading rules in the Bitcoin market. *Finance Research Letters*, *34*, Article 101263.

Geuder, J., Kinateder, H., & Wagner, N. F. (2019). Cryptocurrencies as financial bubbles: The case of Bitcoin. *Finance Research Letters*, *31*.

Gradojevic, N., & Gençay, R. (2013). Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking & Finance*, *37*(2), 578–586.

Gradojevic, N., Gençay, R., & Kukolj, D. (2009). Option pricing with modular neural networks. *IEEE Transactions on Neural Networks*, *20*(4), 626–637.

Gradojevic, N., & Yang, J. (2006). Non-linear, non-parametric, non-fundamental exchange rate forecasting. *Journal of Forecasting*, *25*(4), 227–245.

Graf, A. B., & Borer, S. (2001). Normalization in support vector machines. In B. Radig, & S. Florczyk (Eds.), *Pattern recognition* (pp. 277–282). Berlin, Heidelberg: Springer Berlin Heidelberg.

Gronwald, M. (2015). The economics of Bitcoins: News, supply vs demand and slumps. In *Working paper. Discussion paper in economics* (no. 17). University of Aberdeen Business School.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Hafner, C. M. (2018). Testing for bubbles in cryptocurrencies with time-varying volatility. *Journal of Financial Econometrics*, *18*(2), 233–249.

Hans, F. P., & van Griensven Kasper (1998). Forecasting exchange rates using neural networks for technical trading rules. *Studies in Nonlinear Dynamics & Econometrics*, *2*(4), 1–8.

Heinemann, M. (2000). Adaptive learning of rational expectations using neural networks. *Journal of Economic Dynamics and Control*, *24*(5), 1007–1026.

Hencic, A., & Gouriéroux, C. (2015). Noncausal autoregressive model in application to Bitcoin/USD exchange rates. In *Econometrics of risk* (pp. 17–40). Cham: Springer International Publishing.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Hsieh, D. A. (1989). Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business*, *62*(3), 339–368.

Huang, J.-Z., Huang, W., & Ni, J. (2019). Predicting Bitcoin returns using high-dimensional technical indicators. *The Journal of Finance and Data Science*, *5*(3), 140–155.

Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, *49*(3), 851–889.

Jang, H., & Lee, J. (2018). An empirical study on modeling and prediction of Bitcoin prices with Bayesian neural networks based on blockchain information. *IEEE Access*, *6*, 5427–5437.

Kang, S. H., McIver, R., & Hernandez, J. A. (2019). Co-movements between Bitcoin and Gold: A wavelet coherence analysis. *Physica A. Statistical Mechanics and its Applications*, Article 120888.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Economics*, *19*, 56–75.

Kaucic, M. (2010). Investment using evolutionary learning methods and technical rules. *European Journal of Operational Research*, *207*(3), 1717–1727.

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

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd international conference on learning representations* (pp. 1–13).

Kirkpatrick, C., & Dahlquist, J. (2016). *Technical analysis: the complete resource for financial market technicians*. Pearson Education, Incorporated.

Koutmos, D. (2020). Market risk and Bitcoin returns. *Annals of Operations Research*, *294*(1), 453–477.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689–702.

Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications*, *109*, 1–11.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS One*, *10*, Article e0123923.

Kukacka, J., & Kristoufek, L. (2020). Do 'complex' financial models really lead to complex dynamics? Agent-based models and multifractality. *Journal of Economic Dynamics and Control*, *113*, Article 103855.

Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, *118*, 35–40.

Liu, W., Liang, X., & Cui, G. (2020). Common risk factors in the returns on cryptocurrencies. *Economic Modelling*, *86*, 299–305.

Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, *30*(5), 15–29.

Menkhoff, L. (1997). Examining the use of technical currency analysis. *International Journal of Finance & Economics*, *2*(4), 307–318.

Menkhoff, L. (2010). The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance*, *34*(11), 2573–2586.

Miller, N., Yang, Y., Sun, B., & Zhang, G. (2019). Identification of technical analysis patterns with smoothing splines for Bitcoin prices. *Journal of Applied Statistics*, *46*(12), 2289–2297.

Mselmi, N., Lahiani, A., & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*, *50*, 67–80.

Nakano, M., Takahashi, A., & Takahashi, S. (2018). Bitcoin technical trading with artificial neural network. *Physica A*, *510*, 587–609.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, *60*(7), 1772–1791.

Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of Bitcoin returns: A LASSO approach. *Finance Research Letters*, *27*, 235–240.

Petukhina, A. A., Reule, R. C. G., & Härdle, W. K. (2020). Rise of the machines? Intraday high-frequency trading patterns of cryptocurrencies. *The European Journal of Finance*, *27*(1-2), 8–30.

Polasik, M., Piotrowska, A. I., Wisniewski, T. P., Kotkowski, R., & Lightfoot, G. (2015). Price fluctuations and the use of Bitcoin: An empirical inquiry. *International Journal of Electronic Commerce*, *20*(1), 9–49.

Qi, M., & Maddala, G. S. (1996). Option pricing using artificial neural networks: The case of S&P 500 index call options. In A. P. N. Refenes, Y. Abu-Mostafa, J. Moody, & A. Weigend (Eds.), *Neural networks in financial engineering: proceedings of the third international conference on neural networks in the capital markets* (pp. 78–91). New York: World Scientific.

Refenes, A. N., & Azema-Barac, M. (1994). Neural network applications in financial asset management. *Neural Computing & Applications*, *2*(1), 13–39.

Sermpinis, G., Karathanasopoulos, A., Rosillo, R., & de la Fuente, D. (2021). Neural networks in financial trading. *Annals of Operations Research*, *297*, 293–308. http://dx.doi.org/10.1007/s10479-019-03144-y.

Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. F., & Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization. *European Journal of Operational Research*, *225*(3), 528–540.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*, *11*(5), 1188–1193.

Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*, 3–33.

Taylor, M. P., & Allen, H. (1992). The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance*, *11*(3), 304–314.

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

Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, *38*, 788–804.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988–999.

Wittkemper, H.-G., & Steiner, M. (1996). Using neural networks to forecast the systematic risk of stocks. *European Journal of Operational Research*, *90*(3), 577–588.

Yang, Z., Platt, M. B., & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research*, *44*(2), 67–74.

Zhang, G., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501–514.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the Great Crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, *10*(3), 251–270.