Department of
Economics and Finance

Economics and Finance Working Paper Series

Working Paper No. 2011

Guglielmo Maria Caporale, Alex Plastun, and Viktor Oliinyk

Bitcoin Returns and the Frequency of Daily Abnormal Returns

May 2020

http://www.brunel.ac.uk/economics

BITCOIN RETURNS

AND THE FREQUENCY OF DAILY ABNORMAL RETURNS

Guglielmo Maria Caporale*
Brunel University London, CESifo and DIW Berlin

Alex Plastun
Sumy State University

Viktor Oliinyk
Sumy State University

May 2020

Abstract

This paper investigates the relationship between Bitcoin returns and the frequency of daily abnormal returns over the period from June 2013 to February 2020 using a number of regression techniques and model specifications including standard OLS, weighted least squares (WLS), ARMA and ARMAX models, quantile regressions, Logit and Probit regressions, piecewise linear regressions and non-linear regressions. Both the in-sample and out-of-sample performance of the various models are compared by means of appropriate selection criteria and statistical tests. These suggest that on the whole the piecewise linear models are the best but in terms of forecasting accuracy they are outperformed by a model that combines the top five to produce "consensus" forecasts. The finding that there exist price patterns that can be exploited to predict future price movements and design profitable trading strategies is of interest both to academics (since it represents evidence against the EMH) and to practitioners (who can use this information for their investment decisions).

Keywords: *cryptocurrency, Bitcoin, anomalies, abnormal returns, frequency of abnormal returns, regression analysis*

JEL classification: G12, G17, C63

*Corresponding author. Department of Economics and Finance, Brunel University, London, UB8 3PH.

Email: Guglielmo-Maria.Caporale@brunel.ac.uk

1. Introduction

According to the Efficient Markets Hypothesis (EMH – see Fama, 1965), which remains the dominant paradigm in financial economics, asset prices should follow a random walk, and therefore it should not be possible to design trading strategies that exploit predictable patterns to generate abnormal profits. However, there is a large body of empirical evidence indicating that there exist various market anomalies resulting in identifiable price patterns such as contrarian and momentum effects; these include calendar anomalies, price over- and under-reactions, other types of anomalies associated with trading volumes and so on. In the case of the newly emerged cryptocurrency markets various studies have been carried out which have provided mixed evidence on price predictability (Ciaian et al., 2016; Balcilar et al., 2017; Khuntia and Pattanayak, 2018; Al-Yahyaee et al., 2019 and many others).

The current paper contributes to this literature by investigating the relationship between Bitcoin returns and the frequency of daily abnormal returns over the period from June 2013 to February 2020. It extends previous studies by Angelovska (2016) and Caporale et al. (2019) by considering a much wider range of econometric models and approaches over a longer sample, assessing the role of an additional regressor (namely the difference between the frequency of positive and negative abnormal returns), and evaluating the in-sample as well as the out-of-sample performance of the rival models. These include standard OLS, weighted least squares (WLS), ARMA and ARMAX models, quantile regressions, Logit and Probit regressions, piecewise linear regressions and non-linear regressions.

The remainder of the paper is organised as follows. Section 2 contains a brief review of the relevant literature. Section 3 describes the methodology. Section 4 discusses the empirical results. Section 5 provides some concluding remarks.

2. Literature Review

Cryptocurrencies have established themselves in recent years both as an alternative to fiat money and as a tradable asset used for risk-hedging purposes. Various papers have analysed the properties of these newly created markets. For instance, Bartos (2015) and Urquhart (2016) analysed their efficiency; Corbet et al. (2018) focused on price bubbles; other market anomalies were explored by Kurihara and Fukushima (2017) and Caporale and Plastun (2019); Bariviera et al. (2017) and Caporale et al. (2018) investigated their persistence and long-memory properties; Bouri et al (2019) examined price predictability.

Of particular interest is the issue of whether or not abnormal returns generate stable patterns in price behaviour. This has been a popular topic for investigation since De Bondt and Thaler (1985) developed the overreaction hypothesis. The evidence is mixed: some papers find price reversals after abnormal price changes (Bremer and Sweeny, 1991; Larson and Madura, 2001), whilst others detect momentum effects (Schnusenberg and Madura, 2001; Lasfer et al., 2003). In the specific case of the cryptocurrency markets, Chevapatrakul and Mascia (2019) estimated a quantile autoregressive model and concluded that days with extremely negative returns are likely to be followed by periods characterised by weekly positive returns as Bitcoin prices continue to rise. Caporale and Plastun (2019) used a variety of statistical tests and trading simulation approaches and found that after one-day abnormal returns price changes in the same direction are bigger than after "normal" days (the so-called momentum effect). Caporale et al. (2019) provided evidence on the role played by the frequency of overreactions. Qing et al. (2019) applied DFA and MF-DFA methods and found momentum effects in Bitcoin and Ethereum prices after abnormal returns. Momentum effects were also detected by Panagiotis et al. (2019) and Yukun and Tsyvinski (2019). The present study extends the previous one by Caporale et al. (2019) as detailed below.

3. Methodology

The selected sample includes daily and monthly BitCoin data over the period 06.2013-02.2020. The data source is CoinMarketCap:

https://coinmarketcap.com/currencies/bitcoin/. For forecasting purposes two subsamples are created, namely 01.06.2013-30.12.2018 and 01.01.2019-28.02.2020 at the daily frequency, and June 2013-December 2018 and January 2019-February 2020 at the monthly frequency; various models are estimated over the first subsample, forecasts are then generated in each case for the second subsample using the estimated parameters and their accuracy is evaluated by means of various statistical criteria.

As a first step abnormal returns are computed using the daily series. The dynamic trigger approach is based on relative values, specifically abnormal returns are defined on the basis of the number of standard deviations to be added to average returns (Wong, 1997). By contrast, the static approach requires setting a threshold; for example, Bremer and Sweeney (1991) use a 10% price change as a criterion for abnormal returns. Caporale and Plastun (2019) compared the suitability of these methods in the case of the cryptocurrency markets and concluded that the latter is preferable; therefore this will be applied here.

Returns are defined as:

$$R_t = \ln(P_t) - \ln(P_{t-1}) \quad (1)$$

where R_t stands for returns, and P_t and P_{t-1} are the close prices of the current and previous day. To analyse their frequency distribution histograms are created. Values 10% above or below those of the population are plotted. Thresholds are then obtained for both positive and negative abnormal returns, and periods can be identified when returns were above or equal to the threshold. Such a procedure generates a data set for daily abnormal returns. We then calculate their frequency, namely the cumulative number of positive / negative abnormal returns detected during a month (which is a time-varying parameter changing on a daily basis) and use the end-of-the-month values for the following regression analysis.

Next the data set for the frequency of abnormal returns is divided into 3 subsets including respectively the frequency of negative and positive abnormal returns, and their difference, known as delta. The relationship between the frequency of one-day abnormal returns and Bitcoin returns is investigated by using a number of regression techniques and model specifications including standard OLS, weighted least squares (WLS), ARIMA and ARMAX models, quantile regressions, Logit and Probit regressions, piecewise linear regressions and non-linear regressions.

The specification of the standard OLS regression is the following (2):

$$\mathrm{Y_t} = \mathrm{a_0} + \mathrm{a_1 F_t^+} + \mathrm{a_2 F_t^-} + \varepsilon_\mathrm{t} \tag{2}$$

where Y_t – BitCoin log returns in period (month) *t;*

$\mathrm{a_0}$– BitCoin mean log return;

$\mathrm{a_1}$ ($\mathrm{a_2}$) – coefficients on the frequency of positive and negative one-day abnormal price respectively;

$\mathrm{F_t^+}$ ($\mathrm{F_t^-}$) – the frequency of positive (negative) one-day abnormal price days during period *t*;

ε_t – Random error term at time *t*.

An OLS regression including the single parameter *Delta* (*Delta*=F^+ - F^-) instead of $\mathrm{F_t^+}$ ($\mathrm{F_t^-}$) is also run:

$$\mathrm{Y_t} = \mathrm{a_0} + \mathrm{a_1 Delta_t} + \varepsilon_\mathrm{t} \tag{3}$$

The size, sign and statistical significance of the estimated coefficients provide information about the possible effects of the frequency of daily abnormal returns on BitCoin log returns. The weighted least squares regressions are similar, but instead of treating all observations equally they are weighted to increase the accuracy of the estimates.

To obtain further evidence an ARMA(p,q) model is also estimated (4):

$$Y_t = a_0 + \sum_{i=1}^{p} \psi_{t-i} Y_{t-i} + \sum_{i=0}^{q} \theta_{t-i} \varepsilon_{t-i} \tag{4}$$

where Y_t – BitCoin log returns in month *t;*

a_0 – constant;

$\psi_{t-i}; \theta_{t-j}$ – coefficients the lagged log returns and random error terms respectively;

ε_t – random error term at time *t*;

This is a special case of an ARIMA(p,d,q) specification with d=0, which is appropriate in our case since all series are stationary, as indicated by a variety of unity root tests which imply that differencing is not required (the test results are not reported for reasons of space but are available from the authors upon request).

Next, in order to improve the basic ARMA(p,q) specification exogenous variables are added, namely the frequency of negative and positive one-day abnormal returns in (5) and Delta in (6), to obtain the following ARMAX(p,q,2) and ARMAX(p,q,1) models:

$$Y_t = a_0 + \sum_{i=1}^{p} \psi_{t-i} Y_{t-i} + \sum_{i=0}^{q} \theta_{t-i} \varepsilon_{t-i} + a_1 F_t^{+} + a_2 F_t^{-} \tag{5}$$

$$Y_t = a_0 + \sum_{i=1}^{p} \psi_{t-i} Y_{t-i} + \sum_{i=0}^{q} \theta_{t-i} \varepsilon_{t-i} + a_1 Delta_t \tag{6}$$

A non-parametric method not requiring normality is also used; specifically, quantile regressions are run to estimate the conditional median instead of the conditional mean. More precisely, the quantile regression model for the τ-th quantile is specified as follows (7-8):

$$Y_t = a_0(\tau) + a_1(\tau) F_t^{+} + a_2(\tau) F_t^{-} + \varepsilon_t(\tau) \tag{7}$$

$$Y_t = a_0(\tau) + a_1(\tau) Delta_t + \varepsilon_t(\tau) \tag{8}$$

where τ – the τ-th quantile and $\tau \in (0,1)$;

Next, Probit and Logit regression models are estimated. These are specific cases of binary choice models that provide estimates of the probability that the dependent variable will take the value 1. In a Logit regression, it is assumed that $P\{y=1|x\}=f(z)$, where $f(z)=\frac{1}{1+\exp(-z)}$ - is the logistic function, and the parameter z is obtained from the regression (9-10):

$$z_t = a_0 + a_1 F_t^{+} + a_2 F_t^{-} + \varepsilon_t \quad (9)$$

$$z_t = a_0 + a_1 \mathrm{Delta}_t + \varepsilon_t \quad (10)$$

where z_t is a binary variable equal to 1 if the return in month *t* increased compared to day *t-1*, and 0 otherwise.

To allow for the possibility that the linear relationship between the dependent variable and the independent ones changes between subsamples a piecewise linear regression is then run to obtain estimates of the coefficients of interest before and after a given breakpoint, specifically:

$$Y=\begin{cases} a_0 + a_1F^{+} + a_2F^{-} + \varepsilon_1, Y \le C_1 \\ b_0 + b_1F^{+} + b_2F^{-} + \varepsilon_2, Y > C_1 \end{cases} \quad (11)$$

$$Y=\begin{cases} a_0 + a_1 Delta + \varepsilon_1, Delta \le C_2 \\ a_0 + a_1 Delta + +\varepsilon_2, Delta > C_2 \end{cases} \quad (12)$$

where C_1 and C_2 are the breakpoints.

Possible non-linearities are also considered by estimating a non-linear regression model (NLS) such as:

$$Y = f(x_i)\ (i=\overline{1,n}) \quad (13)$$

where *Y* – dependent variable;

x_i – regressors.

Specifically, we run the following regression:

$$Y = a_0 + b(F^{+})^{\wedge} p + c(F^{-})^{\wedge} q + \varepsilon \quad (14)$$

where a_0,*b*,*c*, *p*, *q* are the model parameters.

Information criteria, namely AIC (Akaike, 1974) and BIC (Schwarz, 1978), are used to select the best model specification for Bitcoin log returns. To compare the forecasting performance of different models various measures such as the Mean Absolute Error (MAE) and Theil's statistic are computed instead.

4. Empirical Results

As a first step, thresholds are calculated by analysing the frequency distribution of log returns to detect abnormal returns (see Table A.1 and Figure A.1). As can be seen, two symmetric fat tails are present in the distribution for log returns: -0.04 for negative returns and 0.05 for positive ones; these are then used as the thresholds to detect negative and positive abnormal returns respectively.

Next we carry out correlation analysis for negative and positive abnormal returns and Bitcoin log returns as in Caporale et al. (2019). Specifically, we compute the correlation between Delta and Bitcoin log returns, which is equal to 0.87, and to make sure that there is no need to shift the data we calculate the cross-correlations at the time intervals t and $t+i$, where $I = \{-10, . . . , 10\}$. Figure 1 shows them over the whole sample period for different leads and lags. The highest coefficient corresponds to lag length zero, which means that there is no need to shift the data.

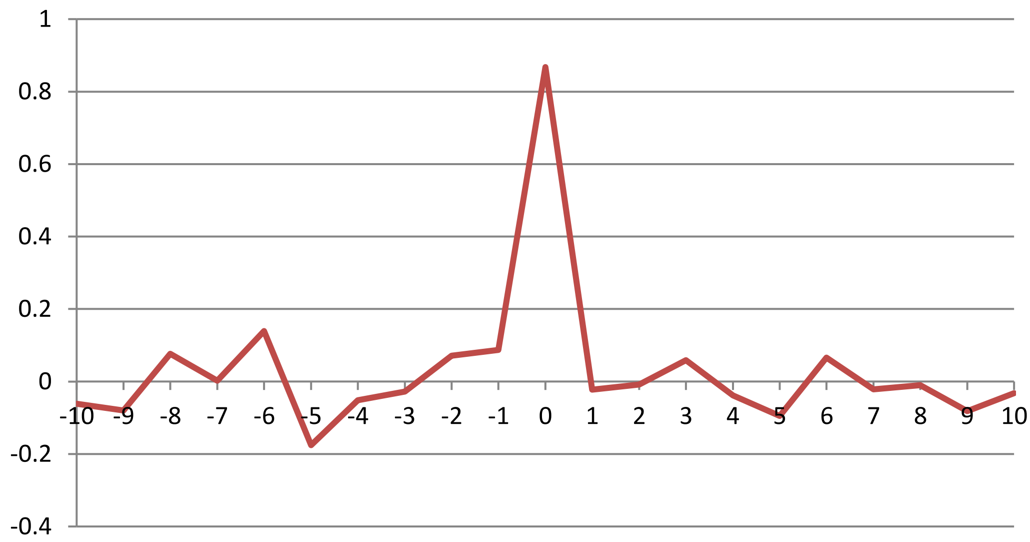


Figure 1: Cross-correlation between Bitcoin log returns and the frequency of the Delta parameter over the whole sample period for different leads and lags

This figure displays the correlation coefficients between BitCoin log returns and Delta over the whole sample period with lags in the interval [-10…+10].

The OLS and WLS regression results are reported in Table 1. Models 1 and 2 are the standard OLS regressions given by (2) and (3), whilst models 1.1 and 2.1 are the WLS ones, where the weights are the inverse of the standard error for each observation used.

Table 1: Regression analysis results: BitCoin log returns

Parameter	Model 1	Model 1.1	Model 2	Model 2.1
	Delta	Delta	Frequency of negative and positive abnormal returns as separate variables	Frequency of negative and positive abnormal returns as separate variables
a_0	0.0901 (0.000)	0.0777 (0.000)	0.0650 (0.024)	0.0626 (0.023)
Coefficient on abnormal returns ($case\ of\ Delta$)	0.0953 (0.000)	0.0868 (0.000)	-	-
Coefficient on the frequency of negative abnormal returns	-		-0.0904 (0.000)	-0.0849 (0.000)
Coefficient on the frequency of positive abnormal returns	-		0.0993 (0.000)	0.0916 (0.000)
R^2	0.7721	0.7652	0.7767	0.7722
p-value	0.0000	0.0000	0.0000	0.0000
Log Likelihood	34.3527	33.3493	35.0369	34.3603
Model Standard Error	0.1471	0.1493	0.1467	0.1482
AIC	-64.7054	-62.6986	-64.0739	-62.7206
BIC	-60.2960	-58.2892	-57.4598	-56.1066

* P-values are in parentheses

This table presents coefficient estimates and p-values (in parentheses) from the regression models. The first column reports parameter estimates for BitCoin log returns, the second and the third for Delta (cases of Model 1 and 1.1 respectively); the fourth the frequency of negative and positive abnormal returns as separate variables in Model 2, the fifth the frequency of negative and positive abnormal returns as separate variables in Model 2.1.

As can be seen the two sets of estimates are very similar. The selected specification, on the basis of the R-squared for the whole model, the p-values for the individual estimated coefficients as well as AIC and BIC criteria, is the following:

$$\text{Bitcoin}\log\text{return}_{\text{i}} = 0.0650 + 0.0993 \times \text{F}_{\text{i}}^{+} - 0.0904 \times \text{F}_{\text{i}}^{-} \qquad (15)$$

which implies a significant positive (negative) relationship between Bitcoin log returns and the frequency of positive (negative) abnormal returns. Any difference between the actual and estimated values suggests that Bitcoin is over- or under-valued, and therefore that it should be sold or bought till the observed difference disappears, at which stage positions should be closed.

The estimates from the selected ARMA(p,q) models on the basis of the AIC and BIC information criteria, namely ARMA(2,2) and ARMA(3,3), are presented in Table 2. As can be seen, although most coefficients are significant, the explanatory power of these models is rather low.

Table 2: Parameter estimates for the best ARMA models

Parameter	Model 3: ARMA(2,2)	Model 4: ARMA(3,3)
a_0	0.0516(0.2103)	0.0513(0.1887)
ψ_{t-1}	0.3486(0.006)	-
ψ_{t-2}	-0.7381(0.000)	-0.3874(0.000)
ψ_{t-3}	-	-0.6209(0.000)
θ_{t-1}	-0.3418(0.000)	-
θ_{t-2}	1.000(0.000)	0.5790(0.000)
θ_{t-3}	-	0.6487(0.000)
R^2	0.0562	0.0373
Log Likelihood	-12.3733	-13.3259
Model Standard Error	0.2831	0.2885
AIC	36.7466	38.6518
BIC	49.9748	51.8800

This table presents the coefficient estimates and p-values (in parentheses) from the ARMA models. The first column reports the parameter estimates for BitCoin log returns (*Y*), the second column shows the parameter estimates for Model 3: ARMA (2,2); the third column for Model 4: ARMA (3,3).

To establish whether it can be improved by taking into account information about the frequency of abnormal returns, ARMAX models (4) are estimated. First F_t^+ (the frequency of positive abnormal returns) and F_t^- (the frequency of negative abnormal returns) are added as regressors. The estimated parameters are reported in Table 3. Model 6 and 7 correspond respectively to Model 3 and 4 with the frequency of negative and positive abnormal returns as additional regressors. They outperform Model 5, namely the best ARMAX specification with p=1. Table 4 reports instead the estimates from the ARMAX models with Delta as a regressor.

Table 3: Estimated parameters for the ARMAX models:

regressors F^+ and F^-

Parameter	Model 5 ARMAX(1,1,2)	Model 6 ARMAX(2,2,2)	Model 7 ARMAX(3,3,2)
a_0	0.0710(0.0674)	0.0678(0.0193)	0.0653(0.0185)
ψ_{t-1}	0.9488(0.000)	-1.3021(0.000)	-
ψ_{t-2}	-	-0.7734(0.000)	-0.1899(0.0932)
ψ_{t-3}	-	-	-0.8078(0.000)
θ_{t-1}	-0.8963(0.000)	0.06834(0.000)	-
θ_{t-2}	-	1.0000(0.000)	0.3585(0.000)
θ_{t-3}	-	-	0.8009(0.000)
a_1	0.0996(0.000)	0.1020(0.000)	0.0973(0.000)
a_2	-0.0927(0.000)	-0.0936(0.000)	-0.0886(0.000)
R^2	0.7817	0.7912	0.7916
Log Likelihood	35.8117	37.6596	37.5804
Model Standard Error	0.1416	0.1342	0.1348
AIC	-59.6234	-59.3193	-59.1608
BIC	-46.3952	-41.6817	-41.5232

This table presents coefficient estimates and p-values (in parentheses) from the ARMAX models. The first column reports parameter estimates for BitCoin log returns (Y), the second column shows parameter estimates for model 5, the third column for model 6 and the fourth column for model 7.

As can be seen all coefficients in Tables 3 and 4 are statistically significant. The best model on the basis of the AIC and BIC criteria is the one with Delta as a regressor. The R^2 indicates that the ARMAX (3,3,1) is the most adequate model (Model 10).

Table 4: Estimated parameters for the ARMAX models: regressor Delta

Parameter	Model 8 ARMAX(1,1,1)	Model 9 ARMAX(2,2,1)	Model 10 ARMAX(3,3,1)
a_0	0.0914(0.005)	0.0913(0.000)	0.0926(0.007)
ψ_{t-1}	0.9445(0.000)	-1.2701(0.000)	-0.3639(0.020)
ψ_{t-2}	-	-0.7467(0.000)	0.4780(0.000)
ψ_{t-3}	-	-	0.7355(0.000)

θ_{t-1}	-0.8828(0.000)	1.402(0.000)	0.5240(0.000)
θ_{t-2}	-	1.0000(0.000)	-0.2618(0.050)
θ_{t-3}	-	-	-0.8914(0.000)
a_1	0.0966(0.000)	0.0982(0.000)	0.0992(0.000)
R^2	0.7793	0.7882	0.7942
Log Likelihood	35.4427	37.1313	38.1502
Model Standard Error	0.1424	0.1352	0.1330
AIC	-60.8857	-60.2627	-58.3005
BIC	-49.8622	-44.8298	-38.4582

This table presents coefficient estimates and p-values (in parentheses) from the ARMAX models. The first column reports parameter estimates for BitCoin log returns (Y), the second column shows parameter estimates for model 8, the third column for model 9 and the fourth column for model 10.

Tables 5, 6, and 7 report the estimates from the quantile regression models with quantiles equal to 0.4, 0.5 and 0.6 respectively, where the 0.5 quantile corresponds to the regression using the median.

Table 5: Estimated parameters for the quantile regression: case of Q=0.4

Parameter	Model 11 F_t^+, F_t^-	Model 12 Delta
a_0	0.0257(0.4261)	0.0489(0.0123)
a_1	0.0966(0.000)	0.0845(0.000)
a_2	-0.0821(0.000)	-
R^2	0.7676	0.7477
Log Likelihood	34.8065	33.7560
Model Standard Error	0.1093	0.1140
AIC	-63.6130	-63.5120
BIC	-56.9989	-59.1026

Table 6: Estimated parameters for the quantile regression: case of Q=0.5

Parameter	Model 13 F_t^+, F_t^-	Model 14 Delta
a_0	0.0414(0.1360)	0.0810(0.000)
a_1	0.1010(0.000)	0.0889(0.000)
a_2	-0.0803(0.000)	-
R^2	0.7663	0.7682

Log Likelihood	37.2054	33.5500
Model Standard Error	0.1055	0.1115
AIC	-68.4109	-62.9594
BIC	-61.7968	-58.5500

Table 7: Estimated parameters for the quantile regression: case of Q=0.6

Parameter	Model 15 F_t^+, F_t^-	Model 16 Delta
a_0	0.0578(0.0339)	0.1330(0.000)
a_1	0.1060(0.000)	0.0887(0.000)
a_2	-0.0825(0.000)	-
R^2	0.7522	0.7458
Log Likelihood	37.2322	32.8061
Model Standard Error	0.1080	0.1173
AIC	-68.4645	-61.6123
BIC	-61.8504	-57.2029

In Models 11, 13 and 15 the regressors are the frequency of negative and positive daily abnormal returns, whilst in Models 12, 14, 16 Delta is the independent variable. In the case of the quantile regression with Q=0.5 Model 13 is the most adequate according to AIC.

The Logit and Probit regression results are presented in Table 8. As a selection criterion the percentage of correctly predicted cases is used; this suggests that the best specification is Model 19 which includes the frequency of negative and positive daily abnormal returns.

Table 8: Logit and Probit regression analysis results

Parameter	Logit		Probit	
	Model 17 F_t^+, F_t^-	Model 18 Delta	Model 19 F_t^+, F_t^-	Model 20 Delta
a_0	0.7506 (0.140)	0.9782 (0.018)	0.4375(0.136)	0.5682(0.014)
a_1	1.4789 (0.000)	1.3846 (0.000)	0.8613(0.000)	0.8137(0.000)
a_2	-1.3585(0.000)	-	-0.7981(0.000)	-
McFadden R-squared	0.4759	0.4695	0.4799	0.4742
Log Likelihood	-24.2414	-24.5353	-24.0562	-24.3160
AIC	54.4829	53.0706	54.1124	52.6320
BIC	61.0970	57.4799	60.7265	57.0414
The percentage of correctly predicted cases	82.1	80.6	82.1	80.6
LR statistic	44.0253(0.000)	43.4376(0.000)	44.3958(0.000)	43.8762(0.000)

Table 9 shows the piecewise linear regression results. Model 2 includes the frequency of negative and positive daily abnormal returns and C_1=0 is used as a breakpoint: for C_1>0 BitCoin returns are positive, otherwise (C_1<0) they are negative. Model 22 includes instead the Delta parameter with C_2=0 as the breakpoint. Both R^2 and AIC imply that Model 21 should be preferred.

Table 9: Estimated parameters for the piecewise linear regression

Parameter	Model 21 F_t^+, F_t^-	Model 22 Delta
a_0	-0.0339(0.359)	0.0404(0.091)
a_1	0.0038(0.820)	0.0665(0.000)
a_2	-0.0358(0.002)	0.0504(0.003)
b_0	0.0911(0.002)	-
b_1	0.1003(0.000)	-
b_2	-0.0845(0.000)	-
R^2	0.8707	0.8012
Log Likelihood	53.3363	38.9128
Model Standard Error	0.1144	0.1385
AIC	-94.6726	-71.8255
BIC	-81.4444	-65.2115

Non-linear models of two types are estimated next: non-linear in the regressors (but linear in the parameters) and in the parameters respectively. In the first case the model can be transformed into a linear one by replacing the variables, and then the parameters can be estimated using OLS. In the second case iterative procedures have to be used instead.

The first type can be formulated as follows (13):

$$Y = a_0 + \sum_{i=1}^{n} a_i x_i + \varepsilon \qquad (13)$$

where Y_t – BitCoin log returns;

a_0 – constant;

a_i – coefficients on the *i*-th regressors;

x_i – regressors;

ε – random error.

The modified variables (selected after some experimentation) are the following:

$x_1 = Delta$; $x_2 = F^+ \times Delta$; $x_3 = \tan(F^+) \times (F^+ + F^-)$; $x_4 = \sin(F^-) \times (F^+)\hat{}2$;
$x_5 = F^- \times (F^+ + F^-)$; $x_6 = Delta \times F^- \times (F^+ + F^-)$

Table 10 reports the corresponding parameter estimates. As can be seen both models 23 and 24 have statistically significant coefficients, but according to R^2 and AIC Model 24 should be preferred.

Table 10: Non-linear regression model type 1: estimated parameters

Parameter	Model 23	Model 24
a_0	0.0853(0.000)	0.0588(0.001)
a_1	0.0755(0.000)	0.0734(0.000)
a_2	0.0029(0.004)	0.0065(0.000)
a_3	0.0022(0.013)	0.0030(0.000)
a_4	-	-0.0038(0.000)
a_5	-	0.0012(0.004)
R^2	0.8166	0.8783
Log Likelihood	41.6371	55.3723
Model Standard Error	0.1340	0.1109
AIC	-75.2742	-98.7446
BIC	-66.4554	-85.5164

The second type of non-linear model incorporates a new variable, namely $x_6 = x_1x_5$, and is specified as follows:

$$Y = a_0 + b(F^+)\hat{}\, p + c(F^-)\hat{}\, q + \sum_{i=1}^{n} a_i x_i + \varepsilon \tag{14}$$

The corresponding estimates are shown in Table 11. All coefficients are statistically significant. Model 27 is the most data congruent:

$$\begin{aligned} Y = 0.0618 + 0.0481 \times (F^+)\hat{}1.4688 - 0.0472 \times (F^-)\hat{}1.4018 + \\ + 0.0031 \times \tan(F^+) \times (F^+ + F^-) - 0.0036 \times \sin(F^+) \times (F^-)\hat{}2 - 0.0006 \times Delta \times F^- \times (F^+ + F^-) \end{aligned} \tag{15}$$

Table 11: Non-linear regression model type 2: estimated parameters

Parameter	Model 25	Model 26	Model 27
a_0	0.0739(0.047)	0.0658(0.036)	0.0618(0.026)
a_1	-	-	-
a_2	-	0.0049(0.007)	-
a_3	-	0.0032(0.000)	0.0031(0.000)
a_4	-	-0.0038(0.000)	-0.0036(0.000)
a_5	-	-	-

a_6	-	-	-0.0006(0.005)
b	0.0511(0.008)	0.0590(0.003)	0.0481(0.000)
c	-0.0589(0.030)	-0.0709(0.011)	-0.0472(0.002)
p	1.2753(0.000)	1.1776(0.000)	1.4688(0.000)
q	1.1609(0.000)	0.9531(0.000)	1.4018(0.000)
R^2	0.7919	0.8787	0.8814
Log Likelihood	37.4026	55.4726	56.2452
Model Standard Error	0.1439	0.1126	0.1113
AIC	-64.8053	-94.9453	-96.4904
BIC	-53.7818	-77.3078	-78.8529

Table 12 reports the ranking of the top five models (of the 29 considered) according to the AIC criterion. As can be seen the non-linear and piecewise linear regressions appear to be the most data congruent.

Table 12: Ranking of the models based on their in-sample performance (06.2013-12.2018)

Rank	Model #	AIC	R^2	Standard Error
1	24	-98.7446	0.8783	0.1109
2	27	-96.4904	0.8814	0.1113
3	26	-94.9453	0.7919	0.1126
4	21	-94.6726	0.8707	0.1144
5	22	-71.8255	0.8012	0.1385

Next we use the estimated models to generate forecasts over the period January 2019-February 2020; both predicted and actual values are reported in Table B.1. Table C.1 presents the following measures of their forecasting accuracy: the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Mean Percentage Error (MPE), the Mean Absolute Percentage Error (MAPE), and Theil's U. Table 13 ranks the rival models in terms of their forecasting performance using the Mean Absolute Error (MAE) and Theil's U criteria.

Table 13: Ranking of the models on the basis of the MAE and Theil's U criteria

Rank	Model #	MAE	Rank	Model #	Theil's U
1	21	0.0796	1	21	0.5485
2	22	0.0889	2	22	0.6600
3	23	0.0949	3	15	0.6639
4	25	0.0958	4	13	0.6675
5	2.1	0.0997	5	2.1	0.6767

It can be seen that Models 21 and 22 (piecewise linear regressions) are still in in the top five specifications, and therefore the overall evidence based on both in-sample and out-of-sample performance suggests that they are the best models for Bitcoin returns.

Finally, we evaluate the accuracy of the "consensus" forecasts produced by a model that combines the top five selected above and therefore is specified as follows:

$$Y = 0.0754 + 7.2578 Model2.1 - 5.9761 Model14 + 1.6021 Model22 - \\ - 10.3993 Model24 + 8.6068 Model26 \quad (15)$$

R^2=0.7211. F=4.1356 (0.0374)

where the weights have been estimated by running a standard multiple linear regression. As can be seen from the forecasting accuracy measures reported in Table C.1, this model outperforms all the individual ones.

5. Conclusions

This paper carries out a comprehensive examination of the role played by the frequency of daily abnormal returns in driving Bitcoin returns over the period from June 2013 to February 2020. It extends the work of Caporale et al. (2019) by considering a much wider range of models over a longer sample period, exploring the role of the difference between the frequency of positive and negative abnormal returns as well, and assessing the forecasting accuracy of the rival models in addition to their in-sample performance. The results indicate that, if one takes into account both in-sample and out-of-sample performance, piecewise linear models are the best for Bitcoin returns. However, in terms of forecasting accuracy they are outperformed by a model that combines the top five to produce "consensus" forecasts.

The finding that there exist price patterns that can be exploited to predict future price movements and design profitable trading strategies is of interest both to academics (since it represents evidence against the EMH) and to practitioners (who can use this information for their investment decisions).

References

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

Al-Yahyaee, K., M. Rehman, W. Mensi, and I. Al-Jarrah, (2019), Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches, The North American Journal of Economics and Finance, 49, 47-56.

Angelovska J., (2016), Large share price movements, reasons and market reaction, Management, 21, 1-17.

Balcilar, M., E. Bouri, R. Gupta, D. Rouband, (2017), Can volume predict bitcoin returns and volatility? A qunatiles based approach, Economic Modeling, 64, 74-81.

Bariviera, Aurelio F., (2017), The Inefficiency of Bitcoin Revisited: A Dynamic Approach. Economics Letters, 161, 1-4.

Bartos, J., (2015), Does Bitcoin follow the hypothesis of efficient market? International Journal of Economic Sciences, IV(2), 10-23.

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

Bremer, M. and R. J. Sweeney, (1991), The reversal of large stock price decreases. Journal of Finance 46, 747-754.

Caporale G. M. Luis Gil-Alana, Alex Plastun, (2018), Persistence in the cryptocurrency market. Research in International Business and Finance, Volume 46, Pages 141-148.

Caporale, G.M., A. Plastun and V. Oliinyk, (2019), Bitcoin fluctuations and the frequency of price overreactions. Financial Markets and Portfolio Management, 33(2), 109-131.

Caporale, G. and Plastun, A., (2019), Price overreactions in the cryptocurrency market, Journal of Economic Studies, 46(5), 1137-1155.

Ciaian, P., M. Rajcaniova, D.A. Kancs, (2016), The economics of BitCoin price formation, Applied Economics, 48 (19), 1799-1815.

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

Chevapatrakul, T., and D. Mascia, (2019), Detecting overreaction in the Bitcoin market: A quantile autoregression approach. Finance Research Letters, 30, 371-377.

De Bondt W. and R. Thaler, (1985), Does the Stock Market Overreact? Journal of Finance, 40, 793-808.

Fame, E. F., (1965), The Behavior of Stock-Market Prices. The Journal of Business 38, 34-105.

Khuntia, S. and J. Pattanayak, (2018), Adaptive market hypothesis and evolving predictability of bitcoin, Economic Letters, 167, 26-28.

Kurihara, Yutaka and Fukushima, Akio, (2017), The Market Efficiency of Bitcoin: A Weekly Anomaly Perspective Journal of Applied Finance & Banking, 7 (3), 57-64.

Larson, S. and J. Madura, (2001), Overreaction and underreaction in the foreign exchange market. Global Finance Journal, 12 (2), 153-177.

Lasfer, M. A., Melnik, A., and Thomas, D. C. (2003), Short-Term Reaction of Stock Markets in Stressful Circumstances, Journal of Banking & Finance, 27(10), 1959–1977.

Panagiotis Tzouvanas, Renatas Kizys, Bayasgalan Tsend-Ayush, (2019), Momentum trading in cryptocurrencies: Short-term returns and diversification benefits, Economics Letters, https://doi.org/10.1016/j.econlet.2019.108728.

Qing C., L. Xinyuan, and Z. Xiaowu, (2019), Cryptocurrency momentum effect: DFA and MF-DFA analysis, Physica A: Statistical Mechanics and its Applications, Volume 526, https://doi.org/10.1016/j.physa.2019.04.083

Schnusenberg, O. and Madura, J., (2001), Do US Stock Market Indexes Over-or Under React? Journal of Financial Research, 24(2), 179–204.

Schwarz, G. E., (1978), Estimating the dimension of a model. Annals of Statistics, 6 (2), 461–464.

Urquhart, A., (2016), The Inefficiency of Bitcoin, Economics Letters, 148, 80-82.

Wong, M., (1997), Abnormal Stock Returns Following Large One-day Advances and Declines: Evidence from Asian-Pacific Markets, Financial Engineering and Japanese Markets, 4, 71-177.

Yukun, Liu and Aleh, Tsyvinski, (2019), Risks and Returns of Cryptocurrency, 2019 Meeting Papers 160, Society for Economic Dynamics.

Appendix A

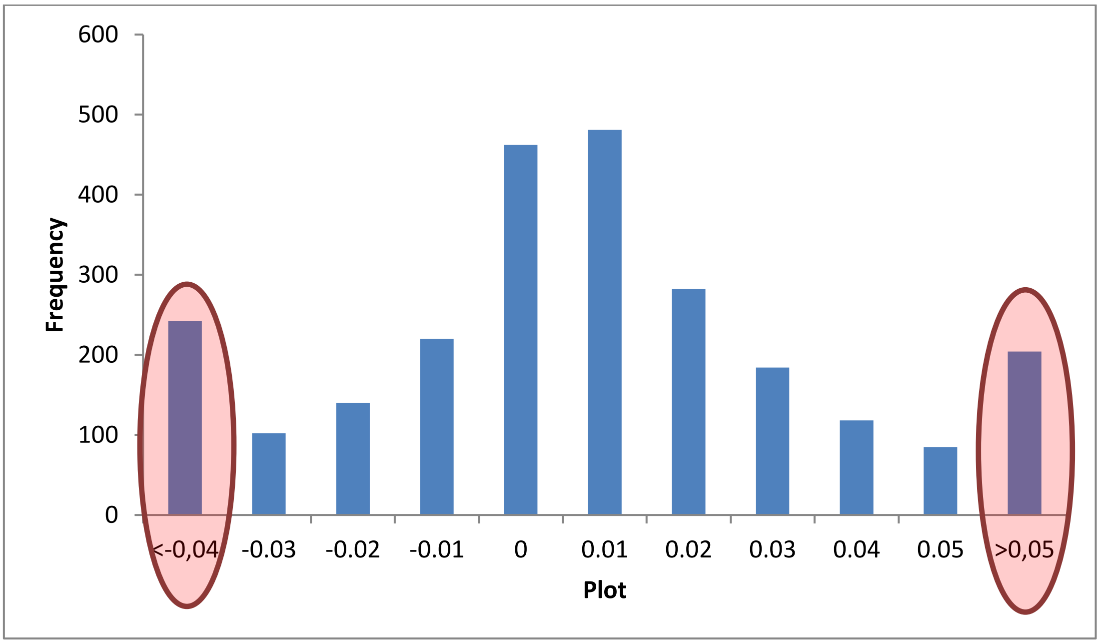
Frequency distribution of BitCoin

TableA.1: Frequency distribution of BitCoin, 05.2013-02.2020

Plot	Frequency
<-0.04	242
-0.03	102
-0.02	140
-0.01	220
0	462
0.01	481
0.02	282
0.03	184
0.04	118
0.05	85
>0.05	204

This table presents estimates of the frequency distribution for Bitcoin log returns over the period 01.05.2013-28.02.2020. The first column reports the values for Bitcoin log returns, the second column the corresponding frequency.

Figure A.1: Frequency distribution of Bitcoin, 05.2013-02.2020



This figure presents the frequency distribution estimates for Bitcoin log returns over the period 01.05.2013-28.02.2020. The plot size is displayed on the x axis; the number of log returns fitting the corresponding plot is displayed on the y axis.

Appendix B

Table B.1: Predicted vs actual values over the period 01.2019-02.2020

Period	01.2019	02.2019	03.2019	04.2019	05.2019	06.2019	07.2019	08.2019	09.2019	10.2019	11.2019	12.2019	01.2020	02.2020
Actual value	-0.0791	0.1086	0.0629	0.2649	0.4715	0.2323	-0.07	-0.0461	-0.1494	0.1036	-0.195	-0.0509	0.2622	-0.0837
Model 1	0.0901	0.1855	0.0901	-0.0052	0.2808	0.2808	-0.1006	-0.196	-0.0052	0.1855	-0.2913	0.0901	0.2808	-0.196
Model 2	0.074	0.1734	0.0651	-0.0165	0.2817	0.2995	-0.0803	-0.1886	-0.0165	0.1734	-0.2968	0.074	0.2639	-0.2064
Model 3	-0.1279	0.0685	0.19	0.0874	-0.0381	-0.0061	0.0977	0.1102	0.038	0.0036	0.0449	0.0847	0.0681	0.0329
Model 4	-0.0823	0.0201	0.1232	0.1465	0.0429	-0.0301	-0.0044	0.0882	0.1236	0.0717	0.0005	-0.0014	0.0584	0.1033
Model 5	0.1021	0.2006	0.0929	0.0058	0.3038	0.3166	-0.0693	-0.1768	0.0011	0.1927	-0.2858	0.0915	0.2833	-0.1951
Model 6	0.1515	0.1458	0.0518	0.0285	0.2411	0.3314	-0.0836	-0.2201	0.0113	0.159	-0.3041	0.0874	0.2551	-0.2
Model 7	0.1093	0.2075	0.0316	-0.0498	0.2549	0.329	-0.044	-0.1711	-0.0482	0.1427	-0.2923	0.1068	0.2839	-0.2041
Model 8	0.1187	0.2138	0.1157	0.0177	0.3065	0.3053	-0.0825	-0.1803	0.012	0.2044	-0.2798	0.106	0.2985	-0.1856
Model 9	0.1681	0.1514	0.0824	0.0329	0.2439	0.3138	-0.1054	-0.2224	0.0176	0.1727	-0.2984	0.1	0.2745	-0.1929
Model 10	0.1951	0.1855	0.1186	0.0563	0.276	0.3458	-0.0868	-0.1971	0.0398	0.1929	-0.2766	0.1172	0.2962	-0.1748
Model 11	0.0402	0.1368	0.0257	-0.0419	0.2479	0.2769	-0.0807	-0.1918	-0.0419	0.1368	-0.3029	0.0402	0.219	-0.2207
Model 12	0.049	0.1335	0.049	-0.0356	0.218	0.218	-0.1201	-0.2046	-0.0356	0.1335	-0.2892	0.049	0.218	-0.2046
Model 13	0.0621	0.1632	0.0414	-0.0183	0.2849	0.3263	-0.0366	-0.1583	-0.0183	0.1632	-0.2801	0.0621	0.2436	-0.1997
Model 14	0.081	0.17	0.081	-0.0079	0.259	0.259	-0.0969	-0.1859	-0.0079	0.17	-0.2748	0.081	0.259	-0.1859
Model 15	0.0813	0.1873	0.0578	-0.0013	0.3168	0.3636	-0.0136	-0.143	-0.0013	0.1873	-0.2725	0.0813	0.2699	-0.1899
Model 16	0.1331	0.2218	0.1331	0.0443	0.3106	0.3106	-0.0444	-0.1332	0.0443	0.2218	-0.222	0.1331	0.3106	-0.1332
Model 21	-0.0317	0.2072	0.0911	0.0224	0.3233	0.3549	-0.0967	-0.1369	-0.0675	0.2072	-0.1771	-0.0317	0.2917	-0.1413
Model 22	0.0404	0.1575	0.0404	-0.0262	0.2745	0.2745	-0.0927	-0.1593	-0.0262	0.1575	-0.2259	0.0404	0.2745	-0.1593
Model 23	0.0923	0.1523	0.0854	0.0172	0.2755	0.2655	-0.0638	-0.1929	0.0172	0.1523	-0.2168	0.0923	0.2386	-0.1413
Model 24	0.0678	0.1159	0.0589	-0.0027	0.2386	0.4327	-0.0115	-0.1891	-0.0027	0.1159	-0.2147	0.0678	0.2187	-0.1502
Model 25	0.0662	0.1389	0.074	-0.0066	0.242	0.2821	-0.0979	-0.1838	-0.0066	0.1389	-0.2206	0.0662	0.1978	-0.1369
Model 26	0.0608	0.1046	0.0659	-0.0058	0.2375	0.441	-0.0089	-0.194	-0.0058	0.1046	-0.2001	0.0608	0.2053	-0.1363
Model 27	0.0696	0.1129	0.0618	0.0005	0.2598	0.4395	-0.0214	-0.2233	0.0005	0.1129	-0.2271	0.0696	0.1813	-0.1414
Model 1.1 (w)	0.0777	0.1646	0.0777	-0.0092	0.2515	0.2515	-0.0961	-0.183	-0.0092	0.1646	-0.2699	0.0777	0.2515	-0.183
Model 2.1 (w)	0.0693	0.1609	0.0626	-0.0156	0.2592	0.2725	-0.0806	-0.1788	-0.0156	0.1609	-0.2771	0.0693	0.2459	-0.1922
Multi 1	-0.0226	0.1748	0.065	-0.0545	0.411	0.2411	-0.0362	-0.0702	-0.0545	0.1748	-0.1444	-0.0226	0.2451	-0.0749

Appendix C

Table C.1: Forecasting accuracy tests

Parameter	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	Mean Percentage Error (MPE), %	Mean Absolute Percentage Error (MAPE),%	(Theil's U)	R^2
Standard linear multiple regressions						
Model 1	0.1309	0.1113	-3.0507	107.28	0.6955	0.495
Model 1.1(w)	0.1273	0.1013	4.7343	95.0821	0.6784	0.522
Model 2	0.1285	0.1046	-0.5218	96.827	0.6870	0.513
Model 2.1(w)	0.1260	0.0997	5.4352	90.3987	0.6767	0.532
ARMA, ARMAX models						
Model 3	0.2058	0.1741	103.8938	141.4790	0.8682	-0.247
Model 4	0.1877	0.1502	94.8069	109.0351	0.9447	-0.037
Model 5	0.1291	0.1107	1.3556	104.9	0.6820	0.508
Model 6	0.1408	0.1156	7.7027	110.75	0.6868	0.416
Model 7	0.1411	0.1167	16.021	107.84	0.7054	0.413
Model 8	0.1321	0.1164	-0.1606	113.63	0.6942	0.485
Model 9	0.1429	0.1195	4.5689	117.73	0.6941	0.398
Model 10	0.1439	0.1246	8.2045	123.69	0.6978	0.390
Quantile regressions						
Model 11	0.1301	0.1025	0.9795	90.522	0.7004	0.501
Model 12	0.1313	0.1033	-0.8925	93.267	0.7036	0.476
Model 13	0.1240	0.1035	10.751	92.638	0.6675	0.546
Model 14	0.1279	0.1030	2.9263	97.844	0.6814	0.518
Model 15	0.1254	0.1075	13.893	98.023	0.6639	0.536
Model 16	0.1316	0.1134	19.273	111.13	0.7006	0.490
Logit and Probit regressions						
Model 17	0.3463	0.2310	-	-	-	-
Model 18	0.3475	0.2286	-	-	-	-
Model 19	0.3427	0.2261	-	-	-	-
Model 20	0.3443	0.2238	-	-	-	-
Piecewise linear regressions						
Model 21	0.0999	0.0796	-22.7047	63.3832	0.5485	0.706
Model 22	0.1162	0.0889	6.9187	78.8485	0.6600	0.602
Non-linear regressions (for the factors)						
Model 23	0.1222	0.0949	16.248	92.625	0.6805	0.560
Model 24	0.1347	0.1048	19.289	91.271	0.6788	0.466
Non-linear regressions (for the estimated parameters)						
Model 25	0.1236	0.0958	12.1089	85.7383	0.6838	0.550
Model 26	0.1352	0.1028	19.706	88.422	0.6858	0.461
Model 27	0.1366	0.1081	14.342	95.903	0.7165	0.450
Consensus forecast						
Multi 1	0.0973	0.0601	16.2657	43.0299	0.6472	0.721