

# On the determinants of bitcoin returns: a LASSO approach

Theodore Panagiotidis<sup>a</sup>, Thanasis Stengos<sup>b</sup> and Orestis Vravosinos<sup>c</sup>

<sup>a</sup>*University of Macedonia*

<sup>b</sup>*University of Guelph*

<sup>c</sup>*Barcelona Graduate School of Economics*

February 13, 2018

## **Abstract**

We examine the significance of twenty-one potential drivers of bitcoin returns for the period 2010 to 2017 (2,533 daily observations). Within a LASSO framework, we examine the effects of factors such as stock market returns, exchange rates, gold and oil returns, FED's and ECB's rates and internet trends on bitcoin returns for alternate time periods. Search intensity and gold returns emerge as the most important variables for bitcoin returns.

**Keywords:** bitcoin, cryptocurrency, exchange rate, returns, LASSO

**JEL classification codes:** G12, G15

## 1. Introduction

Bitcoin has been at the centre of attention by both investors and researchers since its introduction by Nakamoto (2008). The consumer base, transaction frequency in the digital currencies market, as well as the number of businesses and organisations accepting payments in bitcoin have been rapidly expanding (Polasik et al., 2015; Dyhrberg, 2016b). Along with its usefulness as a medium of exchange (Kris-toufek, 2015; Polasik et al., 2015), bitcoin also has some favourable characteristics as an asset, such as hedging, diversifying and safe haven capabilities (Dyhrberg, 2016a,b; Bouri et al., 2017). Panagiotidis et al. (2017) provide a review of the relevant literature.

The aim of this study is to identify the covariates of bitcoin returns. We examine (almost) all the potential drivers that have been suggested in the literature: stock market indices, exchange rates, gold and oil, central bank rates, internet trends and policy uncertainty on bitcoin returns employing the Least Absolute Shrinkage and Selection Operator (LASSO) framework. The advantage of LASSO regression is that we consider all potential drivers but only a subset of the covariates are selected. The rest of the paper is organised as follows: section 2 presents the data and methods used, section 3 discusses the results and the last section concludes.

## 2. Methods

The data employed are daily (7-day week) and cover the period from June 17th 2010 to June 23rd 2017 (2,533 observations). The list of all the independent variables employed and their sources are available in the Appendix (Table 3). Wikipedia trend data were retrieved using the package *'wikipediatrend'*, where they were available till the 21st of January 2016. More recent data were filled from [tools.wmflabs.org](https://tools.wmflabs.org).<sup>1</sup>

---

<sup>1</sup>Due to discrepancies in the data and different scaling between the two sources, a simple linear regression was estimated for the time period for which data from both sources was available (Pearson correlation coefficient for the common period was 0.93) and Wikipedia trend values from 21 January and on were estimated using the values from *tools.wmflabs.org*.

Kristoufek (2013) emphasizes the difference between interest due to positive versus negative events and finds significant asymmetries in the effects of Google and Wikipedia trends depending on whether the price of bitcoin is above or below trend (Kristoufek, 2013, 2015). We follow the latter intuition and decompose all internet trend variables used in the analysis as follows:

$$\text{Above-trend internet feedback}_t = \text{Internet trend}_t \cdot D$$

$$\text{Below-trend internet feedback}_t = \text{Internet trend}_t \cdot (1 - D)$$

where  $D$  a dummy variable taking the value 1, when the internet trend variable is above its 7-day simple moving average and 0, when below. Values for variables that are not in daily frequency or are in a 5-day week frequency have been linearly interpolated. All variables except central bank rates are used in logarithmic first differences, so that they are stationary and their coefficients comparable.<sup>2</sup> Following Forbes and Rigobon (2002), we account for differences in opening hours of the stock markets by using 2-day rolling averages for all variables except central bank rates.<sup>3</sup>

We consider the full sample, as well as three sub-periods of it separately: (1) July 24th 2010 – October 1st 2013, (2) October 2nd 2013 – January 3rd 2017 and (3) January 4th 2017 – June 23rd 2017. The selection of these three periods is motivated by (1) the different phases the bitcoin market has been through (we identify them by taking into account the major events of the cryptocurrency) and (2) an ADF Breakpoint unit root test (with breaks in both the intercept and in the trend). Figures 1 and 2 in the Appendix present the Coindesk Bitcoin Price Index (BPI) and the ADF Breakpoint unit root test  $t$ -statistics.<sup>4</sup> The first period corresponds to a

<sup>2</sup>All variables were found to be  $I(0)$ . Unit root tests are available upon request.

<sup>3</sup>Their results remain robust, when using daily or weekly returns. Nevertheless, the importance of accounting for non-synchronous trading has been emphasized in the literature (e.g. Becker et al., 1992; Miller, 2012; Výrost et al., 2015; Grigoryeva et al., 2017; Peresetsky and Yakubov, 2017).

<sup>4</sup>An alternative unit root test with Haar filter at unit scale discrete wavelet decomposition against trend stationary alternatives with linear trend (Fan and Gençay, 2010) on BPI for the entire sample gives test-statistics of -6.42 and -1674.29 for the level and log-returns, respectively. The critical val-

relatively less volatile era before the bitcoin boom in late 2013 and Mt. Gox's suspension of trading and filing for bankruptcy protection.<sup>5</sup> The second is the period of the crash and the gradual recovery, while the third one corresponds to the recent alleged bubble.<sup>6</sup>

We estimate LASSO models for each of the three periods and the entire sample using two R packages: 'glmnet' and 'lars'. LASSO is a selection procedure that performs both variable selection and regularization to improve prediction accuracy and interpretability. It combines the least-squares approach with a constraint on the sum of the absolute values of the coefficients. In this way, it offers coefficient shrinkage (even setting some to zero), thus, automating model selection in linear regression. Unlike other model selection criteria, the resulting optimization problem can be solved efficiently for large problems, as it is convex (Hastie et al., 2009, 2015). The LASSO estimator is:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j \right)^2 \quad \text{subject to } \|\beta\|_1 \leq t$$

where  $y_i$  the  $i$ -th observation of the dependent variable,  $\beta_0$  an intercept,  $x_{ij}$  the  $i$ -th observation of the  $j$ -th explanatory variable and  $\beta_j$  the corresponding coefficient,  $\|\beta\|_1 \equiv \sum_{j=1}^P |\beta_j|$  is the  $L1$  norm and  $t$  a user-specified tuning parameter.

The LASSO problem can be equivalently written in Lagrangian form as:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j \right)^2 + \lambda \|\beta\|_1 \right\}$$

As is common practice,  $\lambda$  is allowed to vary and a value for it is chosen through cross validation.<sup>7</sup> Variables are standardized so that each has zero mean and unit

---

ues are: -50.77, -36.54 and -30.23 at the 1%, 5% and 10% level, respectively. BPI returns are stationary. This R code has been used for implementing the test (Bartlett bandwidth used is 20).

<sup>5</sup>Forbes, December 17th 2013: [The Bitcoin Boom and Wikipedia Mt. Gox page](#)

<sup>6</sup>FT, May 14th 2017: [Bitcoin's surge fuels fears of asset bubble](#) and FT, January 3rd 2017: [Bitcoin passes \\$1000 but only number that matters is zero](#)

<sup>7</sup>For the three sub-periods, the same value that minimizes the CV error for the first period is selected for  $\lambda$ , so that the coefficients are comparable across the sub-periods. When considering

variance, as otherwise the estimate would depend on the units of measurement.<sup>8</sup> For more on LASSO see Tibshirani (1996), Tibshirani (2011), Hastie et al. (2009) and Hastie et al. (2015).

Table 1 summarises the approaches employed in this paper.

Table 1: Summary of methods used

Method	Details - specifications
glmnet LASSO	Bitcoin distribution: Gaussian; Equal observation weights of 1; number of values for tuning parameter = 100; no intercept; Convergence threshold for coordinate descent= $10^{-7}$ ; Algorithm: covariance (saves all innerproducts ever computed)
lars LASSO	Bitcoin distribution: Gaussian; no intercept; 10-fold cross validation used for the tuning parameter

Note: Least Angle Regression (LARS) was proposed in Efron et al. (2004). Glnmet is a package that fits a generalized linear model via penalized maximum likelihood (for more see here: [http://web.stanford.edu/~hastie/glmnet/glmnet\\_alpha.html](http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html)). The two packages employ different algorithms to implement LASSO; lars uses the LARS algorithm, while glmnet uses coordinate descent optimization.

### 3. Results

Table 2 presents the coefficients of the independent variables for each model ranked based on the absolute value of the coefficients in the glmnet LASSO model for the total sample. Figure 3 presents the glmnet LASSO results graphically. The magnitude of the coefficient is measured on the vertical axis and the  $L1$  norm on the horizontal. Significance of a coefficient is reflected in increases as one moves to the

the full sample,  $\lambda$  is selected, so that the CV error is minimized in the full sample.

<sup>8</sup>This applies to explanatory variable standardization. The bitcoin return is also standardized and, thus, no intercept is used. Standardization is performed for each sample separately.

right in the horizontal axis that measures the  $L1$  norm. In the case where the total sample is employed, the first covariate that diverges from 0 is Google+ followed by gold (Figure 3 and Table 2 provide the same results).

Under both LASSO methodologies (glmnet and lars), Google+ and Gold emerge as having the largest positive coefficients. The largest negative coefficients are attached to Google- the Chinese Uncertainty Index (CEPU). The third period reveals a higher negative coefficient of the European Uncertainty (EEPU) and higher positive coefficient for oil and Nikkei (0.056 and 0.15 respectively in the case of glmnet LASSO).

All uncertainty indices affect bitcoin returns negatively in the following order (more negative first): CEPU, EEPU and USEPU. All exchange rates affect bitcoin returns positively in the following order (more positive first): JPY/USD, CNY/USD, USD/EUR and GBP/USD). Positive is also the effect of interest rates with the ECB being more pronounced than the FED. Variables that capture information demand (or search intensity) have the predicted sign: positive for Google+ and Wiki+ and negative for Google- and Wiki-. The picture is mixed for stock markets: positive is the effect of DJ, SSEC and Nasdaq and negative that of SP350, NIKKEI and VXD. Both gold and oil effects are found to be positive.

## 4. Conclusion

This paper considers twenty-one variables that can affect bitcoin returns. Lasso regression allows both variable selection and regularization in the analysis. Search intensity (Google), gold returns and policy uncertainty are found to be the most important. Overall the effect on bitcoin returns is : (i) negative from uncertainty, (ii) positive from exchange rates, (iii) positive from interest rates, (iv) positive for gold and oil, (v) the expected one from information demand and (vi) mixed from stock markets. In the most recent period, European economic policy uncertainty, the NIKKEI index and negative Google trend feedback emerge as bitcoin determinants.

Table 2: Coefficients of independent variables for each period and the total sample

Rank	Variable	glmnet LASSO				lars LASSO			
		Full Sample	1st period	2nd period	3rd period	Full Sample	1st period	2nd period	3rd period
1	GOOGLE+	0.127	0.111	0.119	0.031	0.128	0.113	0.121	0.000
2	GOLD	0.072	0.068	0.005	0.000	0.075	0.070	0.010	0.000
3	GOOGLE-	-0.069	-0.083	0.000	-0.118	-0.071	-0.087	0.000	-0.024
4	CEPU	-0.060	-0.049	-0.040	0.000	-0.062	-0.051	-0.044	0.000
5	ECB DFR	0.033	-0.005	0.014	-	0.038	-0.006	0.037	-
6	DJ	0.030	0.000	0.000	-0.080	0.036	0.000	0.000	0.000
7	VXD	-0.025	-0.070	0.000	-0.086	-0.026	-0.072	0.005	0.000
8	EEMU	-0.024	-0.024	0.000	-0.172	-0.026	-0.027	0.000	-0.055
9	WIKI-	-0.023	-0.014	-0.016	0.081	-0.025	-0.017	-0.020	0.006
10	EFFR	0.019	0.021	0.009	0.051	0.024	0.023	0.032	0.000
11	JPY/USD	0.019	0.000	0.000	0.027	0.025	0.000	0.000	0.000
12	USEMU	-0.017	-0.005	-0.007	0.000	-0.019	-0.008	-0.011	0.000
13	CNY/USD	0.017	0.000	0.010	0.000	0.019	0.000	0.013	0.000
14	OIL	0.015	0.003	0.008	0.056	0.017	0.004	0.011	0.000
15	SSEC	0.013	0.007	0.000	0.000	0.016	0.009	0.000	0.000
16	SP350	-0.011	0.000	0.000	-0.041	-0.019	0.000	0.000	0.000
17	USD/EUR	0.009	0.000	0.000	-0.036	0.011	0.000	0.000	0.000
18	NIKKEI	-0.006	0.000	0.000	0.150	-0.011	0.000	0.000	0.000
19	WIKI+	0.002	0.000	0.000	0.021	0.004	0.000	0.000	0.000
20	GBP/USD	0.000	0.002	-0.011	0.040	0.000	0.003	-0.015	0.000
21	NASDAQ	0.000	0.000	0.000	-0.110	0.000	0.000	0.009	0.000
	$\lambda$ selection criterion	min CV error	min CV error for 1st period			min CV error	min CV error for 1st period		

Note: ECB DFR experienced no change in the 3rd period and, thus, is not included in the models.

## References

- Becker, K. G., Finnerty, J. E., and Tucker, A. L. (1992). The intraday interdependence structure between U.S. and Japanese equity markets. *Journal of Financial Research*, 15(1):27–37.
- Bouri, E., Molnar, P., Azzi, G., Roubaud, D., and Hagfors, L. I. (2017). On the hedge and safe haven properties of bitcoin: Is it really more than a diversifier? *Finance Research Letters*, 20:192–198.
- Dyhrberg, A. H. (2016a). Bitcoin, gold and the dollar – A GARCH volatility analysis. *Finance Research Letters*, 16:85–92.
- Dyhrberg, A. H. (2016b). Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters*, 16:139–144.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, 32(2):407–499.
- Fan, Y. and Gençay, R. (2010). Unit root tests with wavelets. *Econometric Theory*, 26(5):1305–1331.
- Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261.
- Grigoryeva, L., Ortega, J.-P., and Peresetsky, A. (2017). Volatility forecasting using global stochastic financial trends extracted from non-synchronous data. *Econometrics and Statistics*, pages –. In press.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer-Verlag New York.
- Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall/CRC.
- Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, 3.
- Kristoufek, L. (2015). What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *Plos One*, 10(4).
- Miller, S. M. (2012). Booms and busts as exchange options. *Multinational Finance Journal*, 16(3-4):189–223.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Available at: <https://bitcoin.org/bitcoin.pdf>.

- Panagiotidis, T., Stengos, T., and Vravosinos, O. (2017). On the determinants of bitcoin shocks. Mimeo.
- Peresetsky, A. A. and Yakubov, R. I. (2017). Autocorrelation in an unobservable global trend: does it help to forecast market returns? *International Journal of Computational Economics and Econometrics*, 7(1/2):152–169.
- Polasik, M., Piotrowska, A. I., Wisniewski, T. P., Kotkowski, R., and Lightfoot, G. (2015). Price fluctuations and the use of bitcoin: An empirical inquiry. *International Journal of Electronic Commerce*, 20(1):9–49.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 58:267–288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282.
- Výrost, T., Lyócsa, u., and Baumöhl, E. (2015). Granger causality stock market networks: Temporal proximity and preferential attachment. *Physica A: Statistical Mechanics and its Applications*, 427:262 – 276.

## 5. Appendix

**Table 3:** Variables employed, (7-day week) sample: June 17th 2010 to June 23rd 2017 (2,533 observations)

Variable (name used in Table 2 in brackets)	Source	Code
Coindesk Bitcoin Price Index [BPI]	coindesk.com	-
Brent oil price (in USD per barrel) [OIL]	Quandl	FRED/DCOILBRETEU
Gold price (in USD per troy ounce) [GOLD]	Quandl	WGC/GOLD_DAILY_USD
Fed Funds effective rate [EFFR]	Quandl	FED/RIFSPFF_N_D
ECB deposit facility rate [ECB DFR]	Quandl	BUNDESBANK/BBK01_SU0200
EUR/USD exchange rate [EUR/USD]	Quandl	FRED/DEXUSEU
GBP/USD exchange rate [GBP/USD]	Quandl	FRED/DEXUSUK
CNY/USD exchange rate [CNY/USD]	Quandl	FRED/DEXCHUS
JPY/USD exchange rate [JPY/USD]	Quandl	FRED/DEXJPUS
Dow Jones NYSE index [DJ]	Quandl	BCB/7809
Nasdaq index [NASDAQ]	Quandl	BCB/7810
Nikkei225 index [NIKKEI]	Quandl	NIKKEI/INDEX
S&P350 index [SP350]	us.spindices.com	-
Shanghai Composite Index [SSEC]	Quandl	SGE/CHNMKT
CBOE DJIA Volatility Index [VXD]	Quandl	CBOE/VXD
US policy uncertainty index [USEPU]	policyuncertainty.com	-
Europe policy uncertainty index [EPU]	policyuncertainty.com	-
China policy uncertainty index [CEPU]	policyuncertainty.com	-
Google trend for the term 'bitcoin' [GOOGLE+ and GOOGLE-]	R package 'gtrendsR'	-
Wikipedia trend for the article on bitcoin [WIKI+ and WIKI-]	R package 'wikipediatrend' tools.wmflabs.org	-

Figure 1: Coindesk Bitcoin Price Index

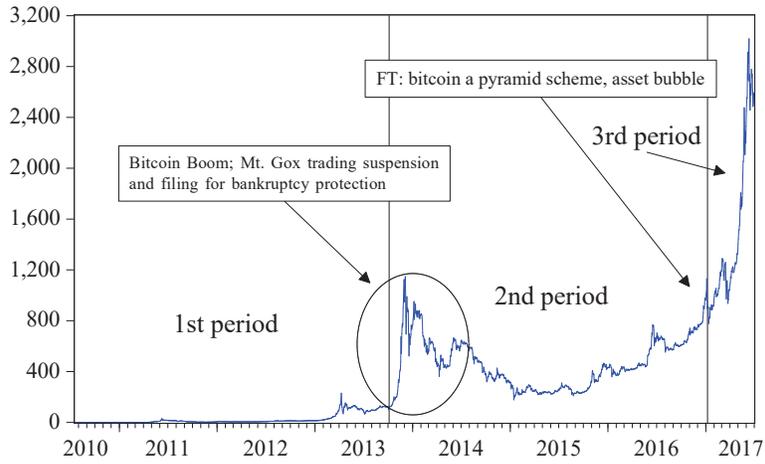


Figure 2: ADF breakpoint unit root test t-statistics

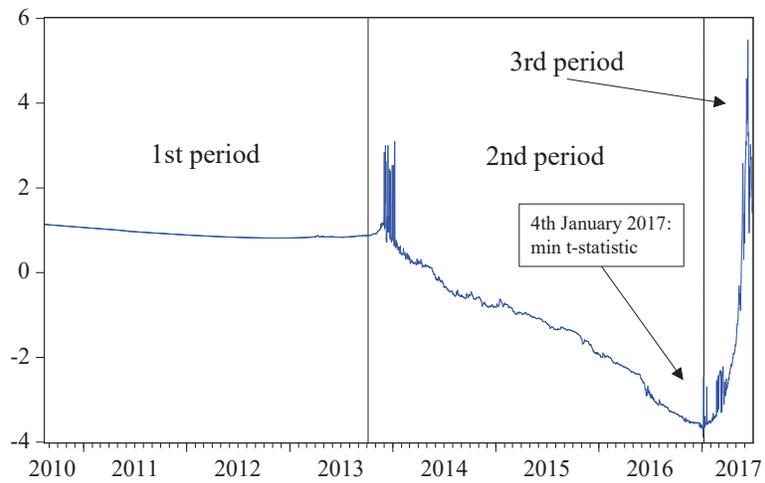


Figure 3: glmnet LASSO coefficient paths

