Journal of Economic Behavior and Organization 176 (2020) 129–144

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

Understanding risk of bubbles in cryptocurrencies

F.A. Enoksen[a], Ch.J. Landsnes[a], K. Lučivjanská[b,1], P. Molnár[a,c,2,*]

[a] University of Stavanger, Stavanger, Norway
[b] Šafárik University in Košice, Slovakia and University of Economics, Prague, Czechia
[c] Nicolaus Copernicus University in Torun, Torun, Poland and University of Economics, Prague, Czechia

ARTICLE INFO

Article history:
Received 15 September 2019
Revised 5 May 2020
Accepted 9 May 2020
Available online 5 June 2020

Keywords:
Cryptocurrencies
Bubbles
PSY test
Uncertainty

ABSTRACT

As cryptocurrencies emerged only recently, they are subject to only very limited financial regulations. In this paper we study which variables can predict bubbles in the prices of eight major cryptocurrencies, focusing on uncertainty measures as predictors. We detect multiple bubble periods for all eight cryptocurrencies, particularly in 2017 and early 2018. We find that higher volatility, trading volume and transactions are positively associated with the presence of bubbles across cryptocurrencies. Regarding the uncertainty variables, the VIX-index consistently demonstrates negative relationships with bubble occurrence, while the EPU-index mostly exhibits positive associations with bubbles. These results may assist authorities in designing appropriate regulations.

1. Introduction

Emergence of cryptocurrencies was one of the most remarkable financial innovations of the last decade. Their futuristic properties and extreme price behavior have attracted excessive media coverage, as well as regulators' and researchers' attention. Most cryptocurrencies are known to have volatile prices and have experienced dramatic price increases and collapses in the recent years. This has triggered discussions as to whether cryptocurrencies experience bubbles and how cryptocurrencies should be regulated.

Bitcoin, the first cryptocurrency, has experienced severe price fluctuations; its price reached a peak in late 2017. Bitcoin was originally intended to function as digital money: it was designed to be a reliable and trustworthy transaction system with low costs (Grinberg, 2012). Bitcoin and other cryptocurrencies have the potential to replace the intermediate role of financial third parties. Though it was intended to be utilized as money, its decentralized and unregulated market have attracted criticism (Grinberg, 2012) and experts have discussed whether it should be classified as either a speculative asset or as a means of exchange. Yermack (2015) and Glaser et al. (2014) concluded in their research that it was primarily held as a speculative asset. Given its apparent risky nature and extreme price behavior, the presence of bubbles in this currency is naturally an interesting topic for research.

* Corresponding author.
E-mail address: peto.molnar@gmail.com (P. Molnár).

1 Lučivjanská was supported by the Czech Science Foundation under grant no. 20-16786S and by the Slovak Research and Development Agency under grant no. APVV-17-0568.

2 Molnár was supported by the Czech Science Foundation under grant no. 20-16786S and by the National Science Centre, Poland under grant no. 2017/26/E/HS4/00858.

https://doi.org/10.1016/j.jebo.2020.05.005
0167-2681/

Table 1
Time Period Employed for Each Cryptocurrency Price dataset start and end dates for each of the eight cryptocurrencies.

Cryptocurrency	From	To	# of days
Bitcoin (BTC)	27.12.2013	15.02.2019	1876
Ethereum (ETH)	27.07.2016	15.02.2019	933
Ripple (XRP)	31.12.2013	15.02.2019	1872
Litecoin (LTC)	27.12.2013	15.02.2019	1876
Monero (XMR)	16.04.2015	15.02.2019	1401
Dash coin (DASH)	20.01.2015	15.02.2019	1487
Nem coin (XEM)	29.03.2016	15.02.2019	1053
Dogecoin (DOGE)	16.12.2014	15.02.2019	1522

In this paper we detect bubbles in all the major cryptocurrencies and study which factors can predict these bubbles. Such information can be useful for both investors and regulators. We consider factors specific to particular cryptocurrencies, such as trading volume, as well as global uncertainty measures such as economic policy uncertainty (EPU). The daily EPU measure reflects uncertainty about legislation and regulation. Since financial regulation of cryptocurrencies is only emerging now, uncertainty in this area could have a major impact on cryptocurrencies. Several papers have analyzed the impact of such uncertainty (measured by EPU or the VIX index) on Bitcoin (Bouri et al., 2017; Aalborg et al., 2018; Demir et al., 2018; Wang et al., 2019; Wu et al., 2019). However, these papers have only studied one cryptocurrency (Bitcoin) and none of them have looked at the presence of bubbles.

From an economic perspective, a bubble is a deviation from the fundamental value. However, where cryptocurrencies are concerned it is hard to pinpoint what the fundamental value is. We therefore define a bubble as explosive price behaviour, as proposed by Phillips et al. (2015a,b). There have been numerous attempts to develop statistical procedures to identify bubbles. Diba and Grossman (1988) applied a unit root test to detect explosive behavior in asset prices. Extensions of this method based on various forms of the augmented Dickey-Fuller test were suggested by Phillips et al. (2011) and Phillips et al. (2015a,b) to identify bubbles, and these methods become known as acronyms of the respective authors' names, PWY and PSY. Phillips et al. (2015a,b) show that the PSY method outperforms the PWY method in detecting multiple bubbles. We therefore use the PSY method.

The PSY framework was initially developed to identify price bubbles. Subsequent research by Phillips (2017) has shown that the PSY procedure can also be used as a warning device for crisis, as the method can be extended to cover market collapse dynamics. Phillips and Shi (2018) incorporated this crisis detection aspect into the PSY method presented in Phillips et al. (2015a,b) and improved the PSY procedure by optimizing the recursive evolving test algorithm.

Several papers have used the PSY framework to detect bubbles in cryptocurrencies. Cheung et al. (2015) and Su et al. (2018) date-stamp bubbles in Bitcoin price and find that the bubble periods coincide with major events that affected the Bitcoin market. Corbet et al. (2018) and Bouri et al. (2018) use the PSY framework to identify bubbles in multiple cryptocurrencies. Corbet et al. (2018) look at Bitcoin and Ethereum and detect bubble behavior in both, particularly at the end of their sample period (mid 2017). Bouri et al. (2018) identify bubbles in Bitcoin, Ripple, Ethereum, Litecoin, NEM, Dash and Stellar and find that the likelihood of bubble periods in one cryptocurrency is related to the existence of bubbles in other cryptocurrencies. Various methods have also been used to study the presence of speculative bubbles, see Cheah and Fry (2015) and Fry and Cheah (2016). However, none of these papers have attempted to predict cryptocurrency bubbles.

We set out to study which factors can predict bubbles for a larger set of cryptocurrencies. The ability to predict bubbles is valuable not only for understanding the cryptocurrencies' price dynamics, but also for market monitoring. First, we use the PSY framework to locate bubbles for the cryptocurrencies Bitcoin, Ethereum, Ripple, Litecoin, Monero, Dash coin, Nem coin and Dogecoin. Next, we study whether any of four variables related to the particular cryptocurrency (Google search queries for the cryptocurrency's name, its price volatility, number of transactions, trading volume) or three variables capturing uncertainty in general financial markets (the economic policy uncertainty (EPU) index, the VIX-index and the TED-spread) can predict those bubbles. We find that volatility and trading volume consistently exhibit a positive relationship with bubble behavior. High EPU levels imply a greater likelihood of bubbles, while high VIX-index levels imply a lower likelihood of bubbles.

The remainder of the paper is structured as follows: Section 2 describes the data on the cryptocurrencies and the variables that may predict bubbles. In Section 3 we present our analysis of the detected bubbles and their potential predictors. We offer conclusions in section 4.

2. Data

The data used in the analysis cover the time period December 27, 2013 to February 15, 2019. The starting dates vary depending on the availability of data for the individual cryptocurrencies studied, see Table 1. The cryptocurrencies to be considered were chosen based on the length of their data sets, their popularity and their total market value. The VIX-index, which we use in our analysis, is not reported on weekends or on certain holidays. These days are therefore omitted from our analysis.

Daily price and trading volume data for the cryptocurrencies were collected from CoinMarketCap through an API in R Studio. Transaction volume was collected from Coinmetrics. Though it is possible to obtain earlier data from other sources, we chose to use these data sets due to their apparent reliability compared to other available sources. The economic policy uncertainty index (EPU) data was collected from the Economic Policy Uncertainty web page. Data on the TED-spread and VIX-index were collected from the FRED database, the Federal Reserve Bank of St. Louis. For the remainder of the paper, we frequently use ticker symbols when we refer to each cryptocurrency. The tickers are displayed in parentheses in Table 1.

To measure the volatility of the cryptocurrencies we use the estimator based on trading price range during a day, as proposed by Garman and Klass (1980). The method, which offers an improvement in accuracy compared to the common method of measuring volatility by standard deviation of returns (Molnár, 2012), has recently gained popularity (Molnár, 2016; Bašta and Molnár, 2018; Fiszeder, 2018; Fiszeder and Fałdziński, 2019; Fiszeder et al., 2019). Daily volatility is calculated as follows:

$$\text{Volatility}_\text{t} = \sqrt{\frac{1}{2}(h_t - l_t)^2 - (2\log 2 - 1)c_t^2}, \tag{1}$$

where $c_t = \log(\text{close}_t)-\log(\text{open}_t)$, $l_t = \log(\text{low}_t)-\log(\text{open}_t)$ and $h_t = \log(\text{high}_t)-\log(\text{open}_t)$. In order to deal with possible weekly seasonality, we convert the preceding daily values into a 7-day arithmetic average by the following equation:

$$\overline{\text{Volatility}_t} = \frac{1}{7}\sum_{\tau=t-6}^{t}\text{Volatility}_\tau. \tag{2}$$

Cryptocurrency transfers can be classified into transfers between a user and a cryptocurrency exchange, and transfers between two users. In general, transfers between users are assumed to represent purchases of goods or services using cryptocurrency, whereas transfers with exchanges represent buying or selling cryptocurrency (in exchange for conventional currency). It is therefore useful to differentiate between these forms of transfers. In our paper, transaction volume (TV) is classified as the volume of transfers of a cryptocurrency between users. Transaction volume is standardized in the same way as in Aalborg et al. (2018), as a deviation from the average volume over the last year divided by the standard deviation over the same period:

$$\text{Transactions}_t = \frac{\text{TV}_t - \overline{\text{TV}}}{\sigma(\text{TV})}. \tag{3}$$

Trading volume, on the other hand, is classified as transfers over an exchange, and does not include direct transfers between users. The time series for Bitcoin's trading volume has historically exhibited both linear and exponential trend components (Balcilar et al., 2017). By following the procedure of Gebka and Wohar (2013) we can remove these from the series, which is necessary to make the variable stationary. The trend elements can be estimated by converting the data to logarithmic form and regressing a constant, (t/T) and $(t/T)^2$ on volume, where T is total observations. Following these estimations, each observation is corrected by subtracting the trend components. Trends exist for all cryptocurrencies, as all the estimated coefficients are statistically significant.

We employ search volume from Google trends in our analysis because it measures public interest in each specific cryptocurrency. This variable is constructed as the relative level of web searches provided by Google, and has previously been demonstrated to have predictive potential, as Choi and Varian (2009), Choi and Varian (2012), Bijl et al. (2016), Molnár and Bašta (2017) have reported. This data can be collected for various time scales and is measured as an index of relative search volume (SVI) between 0 and 100. The daily data can only be collected in samples with a maximum time span of 10 months. In order to make observations between data sets into one complete set, we apply the methodology described by Bleher and Dimpfl (2018). The search results are not case sensitive and the keywords used are: ”Bitcoin", “Litecoin", “Ripple", “Ethereum", “Monero", “Dash coin", “Nem coin" and “Dogecoin".

We standardize the data following the procedure used in Da et al. (2011) and Kim et al. (2019). Each daily observation is measured as a deviation from the median. The measure is calculated as the difference from the median of the previous 8 corresponding weekdays. For example, if the observation is on a Monday it is compared against the 8 previous Mondays. More precisely, it is given by the following equation:

$$\text{Google}_t = \log[\text{SVI}_t] - \log[\text{Median}(\text{SVI}_{t-7}, \text{SVI}_{t-14}, \ldots, \text{SVI}_{t-56})]. \tag{4}$$

The EPU-index can be considered a proxy for economic policy uncertainty in the US economy, as perceived by the public. It is constructed by measuring and standardizing the volume of news articles that contain a combination of certain key words, such as ”economy+regulation+uncertainty", from over 1000 US news outlets (Economic Policy Uncertainty, 2019). In an attempt to reduce noise in the data series and deal with possible weekly seasonality, we use the moving average of the most recent 7 days in our analysis:

$$\text{EPU}_t = \log\left[\frac{1}{7}\sum_{\tau=t-6}^{t}\text{EPU}_\tau\right]. \tag{5}$$

Table 2
Descriptive statistics for currency-specific variables.

	BTC		ETH		XRP		LTC	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Volatility	0.027	0.018	0.043	0.025	0.037	0.034	0.037	0.026
Transactions	1.078	1.133	1.289	1.572	0.644	1.380	0.276	1.435
Volume	16.639	0.804	14.799	0.835	11.932	1.345	15.446	1.214
Google	0.026	0.309	0.045	0.481	0.017	0.317	0.018	0.331
	XMR		DASH		XEM		DOGE	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Volatility	0.055	0.026	0.049	0.027	0.062	0.034	0.048	0.031
Transactions	0.450	1.385	0.539	1.494	0.944	1.470	0.222	1.281
Volume	8.213	1.072	9.891	1.055	9.479	1.334	11.263	1.255
Google	0.007	0.303	-0.688	0.805	-0.129	1.038	0.018	0.308

Table 3
Descriptivestatistics for macroeconomic variables.

Variables	N	Mean	St. Dev.	Min	Max	Skew	Kurtosis
EPU	1259	4.370	0.323	3.526	5.649	0.386	3.407
VIX	1259	2.669	0.254	2.213	3.707	0.751	3.374
TED	1259	−1.199	0.336	−1.897	−0.386	0.282	2.059

Table 4
Correlation Matrix. The average correlations between the independent variables used in the analysis are reported in the table. We apply the same methodology as Da et al. (2011). First, we estimate each correlation individually for the specific cryptocurrencies. Then, we average the results across all cryptocurrencies.

	Google	Volatility	Transactions	Volume	EPU-index	VIX-index	TED-spread
Google	1.00						
Volatility	0.22	1.00					
Transactions	0.36	0.40	1.00				
Volume	0.25	0.47	0.41	1.00			
EPU-index	0.04	0.14	0.13	0.07	1.00		
VIX-index	0.27	0.14	0.31	0.19	0.06	1.00	
TED-spread	0.09	0.16	0.16	0.25	0.20	0.12	1.00

The VIX-index is a measure of perceived short term price uncertainty in the stock market and is commonly called a fear index. It is constructed from option prices based on the S&P 500, with an expiration date of approximately one month (CBOE, 2019). For the purposes of our analysis, this variable has undergone logarithmic transformation.

The level of credit risk in the economy is proxied by the TED-spread, which is constructed as the difference between the US interbank rate and the risk-free US Treasury rate. The intuition behind this metric is that the spread between the interbank interest rate and Treasury rate increases when the possibility of counter party default increases. Historically, when the financial sector has experienced periods of uncertainty and higher default risk, the TED-spread has been high (Boudt et al., 2017). For the purposes of our analysis, this variable has undergone logarithmic transformation.

Table 2 provides the descriptive statistics of the currency-specific variables included in the analysis and Table 3 reports the descriptive statistics of the macroeconomic variables. For the remainder of this paper, in regression tables and equations, EPU, VIX and TED are used as abbreviations for the EPU-index, VIX-index and TED-spread, respectively.

The correlations between the variables are presented in Table 4. It is notable that the correlations between volume and volatility and between volume and transactions are relatively high, with coefficients of 0.47 and 0.41, respectively. Furthermore, we see that the correlation between the uncertainty variables (EPU-index, VIX-index and TED-spread) are quite low. This indicates that collinearity is not a problem and that the variables capture different aspects or forms of uncertainty.

3. Results

We begin by discussing the results from the PSY algorithm and providing an overview of the bubble periods. We then study which variables can predict cryptocurrency bubbles. Details of the PSY methodology can be found in Appendix B.

3.1. Bubble Detection - PSY Test

Fig. 1 illustrates the PSY test, when applied to the logarithm of Bitcoin price (black line). The red line represents the 95%-level critical value of the bootstrapped Dickey-Fuller test statistics generated by this framework. The explosive periods

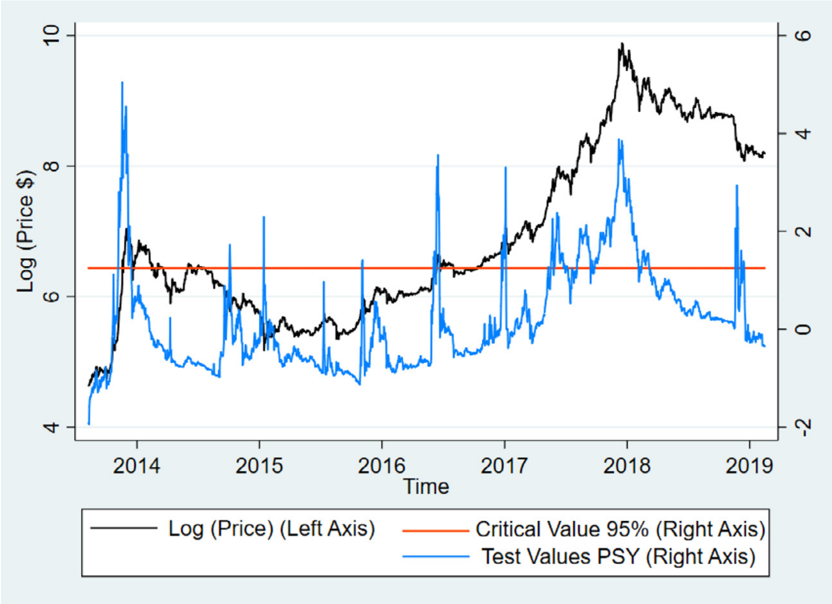

Fig. 1. PSY Test of Bitcoin Bubbles.

Table 5
Statistics of Bubble Periods. Panel A reports the number of bubble days (days when the cryptocurrency was experiencing a bubble) for the individual cryptocurrencies. Panel B reports the same data but this time expressed as percentage of the total number of days in each given year and over the whole sample period.

	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE	Sum
Panel A: Number of bubble days									
2013	0	–	0	0	–	–	–	–	0
2014	1	–	25	4	–	–	–	0	30
2015	3	–	0	11	0	0	–	0	14
2016	12	1	0	1	24	0	11	2	51
2017	129	79	57	91	44	174	66	54	694
2018	48	11	18	11	24	14	2	8	136
2019	0	0	0	0	0	0	0	0	0
Sum bubble days	193	91	100	118	92	188	79	64	925
Panel B: % of days with explosiveness									**Average**
2013	0.0 %	–	0.0%	0.0%	–	–	–	–	0.0%
2014	0.3%	–	6.8%	1.1%	–	–	–	0.0%	2.1%
2015	0.8%	–	0.0%	3.0%	0.0%	0.0%	–	0.0%	0.6%
2016	3.3%	0.3%	0.0%	0.3%	6.6%	0.0%	3.0%	0.5%	1.7 %
2017	35.3%	21.6%	15.6%	24.9%	12.1%	47.7%	18.1%	14.8%	23.8%
2018	13.2%	3.0%	4.9%	3.0%	6.6%	3.8%	0.5%	2.2%	4.7%
2019	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Total%	7.6%	6.2%	3.9%	4.6%	5.0%	10.3%	5.4%	2.9%	5.7%

occur when the PSY test values, illustrated by the blue line, exceed the critical value. Evidently, there were numerous bubble periods in Bitcoin during the observed sample period.

Fig. 2 plots the time-stamped bubble periods from the PSY test against development of the uncertainty variables (VIX-index, EPU-index and TED-spread) employed in the regression models. Most of the explosive periods lasted only for a few days, although a few were much longer-lived. The short-lived bubbles occurred at different times for different cryptocurrencies. The longer-lived bubbles coincided to a greater extent across the cryptocurrencies than the short-lived bubbles.

The prices of all the cryptocurrencies studied in this paper increased dramatically during 2017. As Fig. 2 shows, the PSY algorithm reveals that there were bubbles in most of the cryptocurrencies for large parts of 2017. Bitcoin in particular exhibits long-lived bubble periods in both 2017 and 2018. The date-stamped bubble periods for each cryptocurrency ended some time after the price collapse in January 2018. Notably, that price collapse seems to have coincided with a substantial increase in the VIX-index. By February 15, 2019, the analyzed cryptocurrencies had declined on average 90% from their peak in December 2017/January 2018.

An overview of the bubble periods we identified is provided in Table 5. Panel A presents the number of bubble days (days when the cryptocurrency was was in a bubble state). The cryptocurrencies BTC and DASH experienced the highest

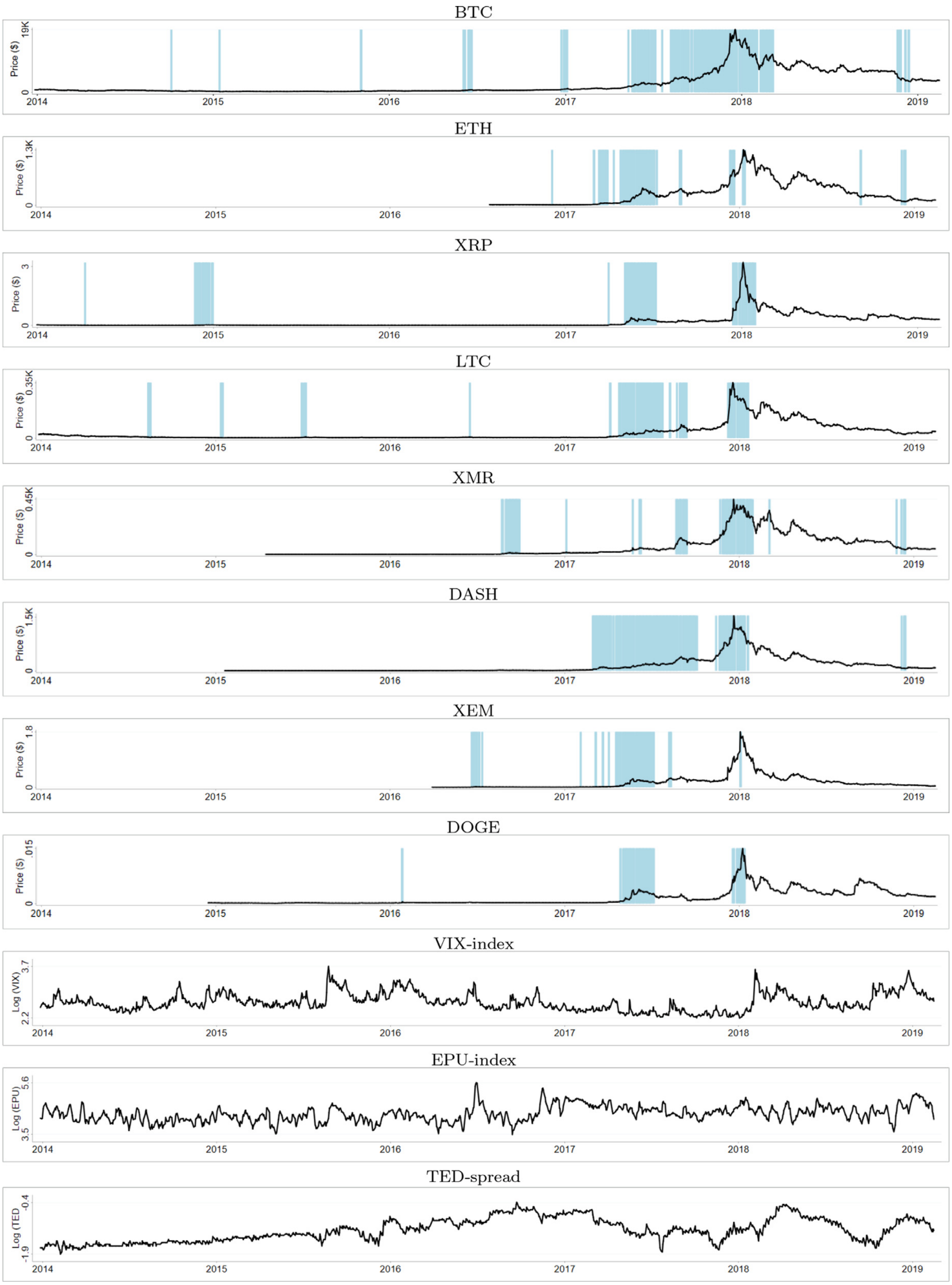

Fig. 2. Bubble Periods in Cryptocurrencies and Uncertainty Variables. The colored areas in this figure mark the explosive periods in the individual cryptocurrencies detected by the PSY framework. The black lines for the cryptocurrencies represent the price in \$. The lines start when the dataset of prices begins for each individual cryptocurrency and end on February 15, 2019. The black lines for the uncertainty variables VIX-index, EPU-index and TED-spread display their historical development.

Table 6
Summary ofregression models.

Sample	Dependent Variable	Estimator
All	Bubble dummy	Panel probit with random effects & cluster robust standard errors
All	PSY statistics	Panel Prais-Winsten with panel corrected standard errors, in Appendix
Individual	Bubble dummy	Probit with optimal cluster robust standard errors
Individual	PSY statistics	OLS with optimal lag Newey-West standard errors, in Appendix

total number of bubble days: 193 and 188 days, respectively. Most bubble days occured in 2017. DASH had the highest frequency of bubble days in 2017 (174 days). Panel B indicates that the percentage of days with explosiveness was higher in 2017 than in other years. DASH featured explosiveness most frequently (on 10.3% of days over the time period 2015–2019) and DOGE least frequently (on 2.9% of days over the time period 2014–2019).

3.2. Bubble predictors

Having applied the PSY framework, we generated the PSY statistics for each of the cryptocurrencies. We then analyzed the results by performing various regressions in order to evaluate which variables can predict cryptocurrency bubbles. We estimated both probit models and regular linear regression models (the results of the linear models are presented in Appendix A). First, we present estimates of panel models with all cryptocurrencies in the same sample. Second, we present estimates of models for each cryptocurrency separately.

The two dependent variables (the bubble dates dummy and the PSY test statistic) applied in the regressions measure the same property to some extent, as they are both derived from the PSY statistics. As we described in Section 2, previous studies have shown that there are correlations between cryptocurrency prices and variables such as Google Trends, EPU, volatility and trading volume etc. This research provides a starting point for the predictor selection in our analysis.

The dependent binary variable of the probit models, denoted $\mathrm{BUB}_{i,t}$ for panel probit regressions and BUB_t for time series regressions, takes values of one and zero. $\mathrm{BUB}_{i,t}$ and BUB_t is set to 1 when the PSY statistic for the respective observation is above the generated critical value for the considered cryptocurrency (i.e. there is a bubble), and zero when this statistic is below the critical value (i.e. there is no bubble):

$$\mathrm{BUB}_{i,t} = \begin{cases} 1, & \text{if } PSY_{i,t}(r_0) > cv_{i,t}(\beta_T) \\ 0, & \text{if } PSY_{i,t}(r_0) < cv_{i,t}(\beta_T) \end{cases}$$

The panel probit model and time series probit model can, respectively, be expressed as:

$$P(\mathrm{BUB}_{i,t} = 1) = \Phi(\beta x_{i,t-1} + \nu_i), \tag{6}$$

$$P(\mathrm{BUB}_t = 1) = \Phi(\beta x_{t-1}), \tag{7}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function. In the panel probit models, $x_{i,t-1}$ is the vector of lagged predictors in cryptocurrency $i = \mathrm{BTC}, \mathrm{ETH}, \ldots, \mathrm{DOGE}$ at time $t-1$ and $\nu_i \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$ corresponds to random effects. x_{t-1} is a vector of lagged predictors in the models for individual cryptocurrencies. The linear regression models use the generated PSY statistic as a dependent variable. The PSY statistic is the supremum of the estimated ADF statistic for the respective observation, generated by the algorithm, as defined in Appendix B in Eq. (15). The estimated probit models consider only whether the PSY test statistic is below or above the generated critical value; they do not use the actual value of PSY statistic. We therefore also consider an alternative linear model which employs the PSY statistic directly. The linear panel regression model is specified as follows:

$$\begin{aligned} PSY_{i,t}(r_0) = {} & \beta_0 + \beta_1 \mathrm{Google}_{i,t-1} + \beta_2 \mathrm{Volatility}_{i,t-1} + \beta_3 \mathrm{Transactions}_{i,t-1} + \beta_4 \mathrm{Volume}_{i,t-1} \\ & + \beta_5 \mathrm{EPU}_{t-1} + \beta_6 \mathrm{VIX}_{t-1} + \beta_7 \mathrm{TED}_{t-1} + \epsilon_{i,t}, \end{aligned} \tag{8}$$

while the linear time series regression model takes the form:

$$\begin{aligned} PSY_t(r_0) = {} & \beta_0 + \beta_1 \mathrm{Google}_{t-1} + \beta_2 \mathrm{Volatility}_{t-1} + \beta_3 \mathrm{Transactions}_{t-1} + \beta_4 \mathrm{Volume}_{t-1} \\ & + \beta_5 \mathrm{EPU}_{t-1} + \beta_6 \mathrm{VIX}_{t-1} + \beta_7 \mathrm{TED}_{t-1} + \epsilon_t. \end{aligned} \tag{9}$$

An overview of the models used is presented in Table 6. To improve this paper's readability, we report the results of the linear models only in Appendix A. The data sample is always either an individual cryptocurrency or all the cryptocurrencies together. Due to potential autocorrelation and heteroscedasticity, we apply models suitable for dealing with this issue. The panel probit models are estimated with random effects and robust standard errors, clustered by cryptocurrency. The linear panel models use a Prais–Winsten estimator with standard errors corrected for AR(1) autocorrelation, heteroscedasticity and cross-sectional correlation. Both these methods are suggested by Hoechle (2007). The linear time series models are

Table 7
Probit Regression Results - Panel Regression. The table reports average marginal effects for standardized explanatory variables. The dependent binary variable $BUB_{i,t}$ only takes the values 1 (explosive dates) and 0 (non-explosive dates). The independent variables are described in Section 2 and are standardized by substracting the sample mean and dividing by the sample standard deviation. The sample includes *all cryptocurrencies* (see Table 1 for the individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. The panel model is estimated with random effects. All the reported estimates are coefficients with corresponding cluster-robust standard errors, by cryptocurrency.

	Dependent variable: $BUB_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Google_{i,t-1}$	0.153***							0.0198
	(0.0350)							(0.0294)
$Volatility_{i,t-1}$		2.968***						0.794***
		(0.537)						(0.239)
$Transactions_{i,t-1}$			0.0590***					0.0141
			(0.0125)					(0.0109)
$Volume_{i,t-1}$				0.0796***				0.0713***
				(0.00619)				(0.00796)
$EPU_{i,t-1}$					0.115***			0.0474*
					(0.0247)			(0.0285)
$VIX_{i,t-1}$						−0.449***		-0.158***
						(0.0713)		(0.0448)
$TED_{i,t-1}$							−0.107***	0.0166
							(0.0337)	(0.0228)
Observations	8060	8060	8060	8060	8060	8060	8060	8060
McFadden R-squared	0.1083	0.1510	0.1507	0.0515	0.0198	0.152	0.0123	0.3613

estimated with Newey-West standard errors (Newey and West, 1987), treating the gaps as equally spaced, as suggested by Datta and Du (2012). Optimal lags are 5 for all models, following the lag selection procedure presented in Greene (2007).[3]

For the measure of fit metrics, regular R-squared is the share of variance in the dependent variable that can be explained by the estimated model. Interpretation of the the McFadden R-squared is not as straightforward, but still applicable when comparing the fitness of different models. It is constructed by utilizing the log-likelihood ratio of the models with and without explanatory variables (McFadden, 1973).

3.2.1. Panel regressions: All cryptocurrencies together

The regression results from the probit panel regressions are provided in Table 7 (the PSY statistic panel regressions in Table A.3). We use panel regressions to analyze the variables' predictive effects across cryptocurrencies. We estimate univariate models investigating one explanatory variable at a time and a multivariate model with all variables.

For a proper evaluation of (not only statistical significance, but also) economic significance, we standardize the explanatory variables by subtracting the sample mean and dividing it by the sample standard deviation. Furthermore, we report average marginal effects. For the probit regression, these are the most informative and similar alternative to simple beta coefficients in a classical linear regression.

Positive coefficients indicate a higher predicted probability. An increase in the variable is thus associated with a higher likelihood of bubbles. A negative coefficient would indicate a lower likelihood of bubbles. A higher absolute value of the coefficient indicates stronger economic significance.

It is important to emphasize that we utilize two types of explanatory variables: variables related to particular cryptocurrencies (volatility, transactions, volume, Google searches), and variables capturing various aspects of uncertainty in general financial markets (the EPU index, the VIX index and the TED spread).

Let us first discuss the cryptocurrency-specific variables. In both univariate and multivariate panel models, higher volatility raises the likelihood of bubble states. The research by Bekiros et al. (2017) states that herding behavior is usually more prevalent in periods of excessive volatility, which might make volatility a natural property of bubbles. Volume exhibits a positive relationship with bubble occurrence in all models. Higher trading volume is also associated with a higher likelihood of bubbles. This can possibly be explained by theories such as rational bubbles[4] or herding behavior. Trading volume is naturally related to the price dynamics of cryptocurrencies, and is thus assumed to be closely connected with bubble behavior. This differs from the research by Blau (2017), which does not find any connection between speculative trading and extreme market behavior. Google searches and transactions have positive effects on bubble behavior in all univariate panel models, although these effects are not significant when other variables are controlled for in the multivariate probit panel model. We suspect that both these variables are closely connected to the trading volume, which may explain why the effects are not significant in the multivariate probit panel model, when trading volume is included. To some extent, the volume,

[3] Optimal lag size is calculated by the smallest integer of $T^{\frac{1}{4}}$, where T is total sample size. This procedure is presented on page 463 in Greene (2007).

[4] The concept of rational bubbles was established by Blanchard and Watson (1982), who indicates that temporary price levels above the intrinsic value can be consistent with rationality, if the expected future price is higher than the current price.

Table 8
Probit Regression Results - Time Series. The table reports average marginal effects for standardized explanatory variables. The dependent binary variable BUB_t only takes the values 1 (explosive dates) and 0 (non-explosive dates). Independent variables are described in Section 2 and are standardized by subtracting the sample mean and dividing by the sample standard deviation. The sample includes *all* dates for the *respective cryptocurrency* (see Table 1 for individual time spans). *, **and *** represents significance at the 10%, 5% and 1% level, respectively. All the reported estimates are coefficients with corresponding robust standard errors.

	Dependent variable: BUB_t							
	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE
$Google_{t-1}$	0.172***	0.154***	0.00520	0.0340**	0.0765***	−0.0621***	−0.000346	0.00999
	(0.0279)	(0.0237)	(0.0109)	(0.0152)	(0.0250)	(0.0147)	(0.00645)	(0.0115)
$Volatility_{t-1}$	2.609***	0.897**	0.298***	0.452**	1.240***	1.245***	0.703***	−0.0817
	(0.488)	(0.368)	(0.112)	(0.187)	(0.246)	(0.398)	(0.272)	(0.145)
$Transactions_{t-1}$	−0.0332***	0.0470***	0.000528	0.0215***	0.0148**	0.0217**	0.0356***	−0.00563
	(0.00807)	(0.00977)	(0.00441)	(0.00539)	(0.00669)	(0.00919)	(0.00757)	(0.00444)
$Volume_{t-1}$	0.111***	0.0154	0.0606***	0.0412***	0.0520***	0.105***	0.0319***	0.0582***
	(0.0130)	(0.00978)	(0.00609)	(0.00781)	(0.00843)	(0.0162)	(0.00877)	(0.00545)
EPU_{t-1}	−0.0693**	0.0467	0.0223	−0.00279	0.0234	(0.0162)	0.0685**	0.0518***
	(0.0304)	(0.0330)	(0.0217)	(0.0228)	(0.0261)	(0.0162)	(0.0284)	(0.0143)
VIX_{t-1}	−0.192***	0.111***	−0.0675***	−0.0418	−0.0413	(0.0162)	−0.0936**	−0.117***
	(0.0363)	(0.0372)	(0.0226)	(0.0279)	(0.0430)	(0.0162)	(0.0426)	(0.0337)
TED_{t-1}	0.108***	−0.0343	−0.0557**	0.0237	0.0569*	(0.0162)	−0.0536*	0.0330**
	(0.0239)	(0.0313)	(0.0237)	(0.0172)	(0.0294)	(0.0162)	(0.0304)	(0.0156)
Observations	1258	625	1256	1258	939	998	707	1019
McFadden R-squared	0.4854	0.5383	0.5781	0.5531	0.4651	0.5674	0.3959	0.6876

Google searches and transactions variables are similar, as they are all related to the market demand for cryptocurrencies. The fact that they demonstrate the same direction of effects supports this intuition.

When it comes to uncertainty variables, the TED-spread is significant and negatively associated with bubbles in the univariate probit model, but the effect is not significant when other variables are included in the multivariate regression. Of the three variables capturing uncertainty in financial markets, TED spread matters the least for cryptocurrency bubbles.

The EPU-index is positive and significant for both the univariate and multivariate probit models. This implies that the probability of cryptocurrency bubbles is higher when economic policy uncertainty is high. This result is quite intuitive.

However, the VIX-index, which is significant in both the univariate model and the multivariate models, demonstrates negative relationships with bubbles in all panel models. This implies that even though both EPU and VIX are measures of uncertainty, these measures capture significantly different aspects of uncertainty.

The EPU index is based on the number of articles that contain at least one term from each of three sets of terms. The first set is economic or economy. The second is uncertain or uncertainty. The third set is legislation or deficit or regulation or congress or federal reserve or white house. In other words, the EPU index reflects only sources of uncertainty that are already reflected and discussed in the media. Therefore, the EPU index might be a good proxy for uncertainty as viewed by the general public. The VIX index, on the other hand, is based on option prices, which capture the market consensus and respond to new information almost immediately. The VIX index therefore mainly captures uncertainty as viewed by professional investors in the financial markets.

One possible explanation why high EPU is associated with a greater likelihood of bubble occurrence, while high VIX is associated with lower bubble occurrence, is that when the general public perceives high uncertainty, some people resort to cryptocurrencies, raising the likelihood of bubbles. However, when professional investors perceive high uncertainty they become more cautious, reducing the likelihood of bubbles. This explanation is only one possible explanation; we do not currently have empirical evidence to support or disprove this explanation.

Considering the measures of fit metrics of the panel probit models, the McFadden R-squared shows that the models display varying ability to predict bubbles. The model with the VIX-index as an explanatory variable has the highest value and the model with TED-spread as an explanatory variable has the lowest value of the McFadden R-squared.

3.2.2. Time series regressions: individual cryptocurrencies

The results from the estimated probit regressions for individual cryptocurrencies are shown in Table 8.[5] We study the cryptocurrencies separately to examine whether the predictive effects seem to be cryptocurrency-dependent or consistent across cryptocurrencies.

Similar to what we observed in the results of the panel regressions, volatility and volume exhibit positive associations with bubbles for most cryptocurrencies. This means that high volatility or volume correspond with a higher likelihood of bubbles, as demonstrated in the panel regression models. Google searches shows varying direction of effects and predictive ability depending on the particular cryptocurrency studied. Google searches are positively associated with bubbles for BTC

[5] We also estimated linear univariate and multivariate regressions for the individual cryptocurrencies, which are included in the appendix (Table A.1, Table A.2 and Table A.4).

Table 9
Models' Predictive Ability. *% True Bubble Days Predicted* is the share of the bubble days detected by the PSY framework which the respective model is able to predict. *% Correct Predictions* is the share of model-predicted bubble days that are detected by the PSY framework as bubble days.

	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE	Average
PSY Detected Bubbles Days	193	91	100	118	92	188	79	64	
Panel A: Probit regression									
Predicted Bubble Days	144	76	69	94	53	169	46	54	
% True Bubble Days Predicted	58.6%	65.9%	53.00%	62.7%	46.7%	75.5%	40.51%	70.3%	59.2%
% Correct Predictions	78.5%	79.0%	76.81%	78.7%	81.1%	84.0%	69.57%	83.3%	78.9%
Panel B: Linear regression									
Predicted Bubble Days	115	113	86	81	52	79	70	38	
% True PSY Bubble Days Predicted	46.63%	78.02%	59.00%	52.54%	34.78%	32.98%	53.16%	48.44%	50.69%
% Correct Predictions	78.26%	62.83%	68.60%	76.54%	61.54%	78.48%	60.00%	81.58%	70.98%

and ETH and negatively associated with bubbles for DASH and XMR. This variation in effects may possibly be explained by the differences in the various cryptocurrencies' total market values. The transactions variable generally demonstrates a positive relationships with bubbles, except in the case of BTC, where it displays a negative effect. One possible explanation for this exception is that BTC is one of the highest-ranking cryptocurrencies in terms of total market value. An increase in transactions for such a high-value cryptocurrency might imply a higher degree of use of that currency as means of exchange. This could lead to a weaker association with bubble behavior, as it could indicate practical utility for the owners. Overall, the cryptocurrency-specific variables mostly demonstrate the same positive associations with bubble behavior as we saw in the panel regression models.

The examined uncertainty variables EPU-index, VIX-index and TED-spread show varying relationships with bubble states when it comes to the direction of their effects. As the results from the panel model regression indicate, the EPU-index is positively associated with bubbles, although this relationship is dependent on the particular cryptocurrency studied. The VIX-index is in general negatively associated with bubbles across cryptocurrencies, as we also saw in the panel regression models. In the probit models, the VIX variable is negatively associated with bubbles for BTC, DASH and DOGE, which might explain why the panel models exhibit the same effect. TED-spread shows a positive relationship with bubbles for BTC in the time series models, but not in the panel models, where there are only weak indications of an effect. Note that TED-spread does not show any significant effects for the other cryptocurrencies.

The measures of fit metrics, R-squared and McFadden R-squared are relatively high, which demonstrates that the models have a considerable ability to predict bubbles.

3.2.3. Summary of regression results

In general, the cryptocurrency-specific variables volatility and trading volume demonstrate similar and consistent results in both the panel regressions and time series regressions. In the panel regression models, Google searches and transactions are generally positively associated with bubbles. In the time series regression models, Google searches and transactions demonstrate varying effects for the various cryptocurrencies studied.

The uncertainty variables EPU-index, VIX-index and TED-spread exhibit differing associations with bubble behavior in the panel regression models. The EPU-index shows positive relationships in the probit panel models, the VIX-index demonstrates negative relationships with bubbles in all panel models, while the TED-spread exhibits a more ambiguous relationship. The time series regressions for the uncertainty variables reveal varying effects depending on the cryptocurrency studied.

In summary, we find that several variables can predict bubbles. Overall, the panel regression results for the uncertainty variables are primarily in line with the time series regression results. In particular, we find that volatility, trading volume and the VIX-index demonstrate a general potential to predict bubble behavior across cryptocurrencies. The predictive effect of other variables is contingent on whether we look at the probit models or the linear models, and which cryptocurrency we examine.

3.2.4. Models' predictive ability

Table 9 presents a comparison of the time series models' ability to predict the bubble dates estimated using the PSY framework. The models utilized to test this predictive ability are the multivariate regressions displayed in Table 8 and Table A.4. The probit models presented in panel A predict that a bubble is expected for the next observation if the estimated probability is above a 50% threshold. The linear regression models predict the PSY statistic for the next observation. A bubble is predicted if the estimated PSY statistic exceeds the critical value (generated by the PSY framework) for the respective cryptocurrency.

The results in Table 9 indicate that the probit models are generally superior to the linear regression models. These results contradict our a priori expectation that the linear models would perform better than the probit models. We had suspected that trying to predict the underlying PSY values would result in greater predictive accuracy.

The superiority of the probit models to the linear models might be due to the binary categorization of the detected bubble days. Following the definition used in the PSY framework, bubble days are detected when the PSY values are high and above the generated critical value. Therefore, it seems that extreme values fit better into the binary structure (bubble/no bubble) of the probit models. On the other hand, the linear PSY models might be a better fit with the underlying PSY data.

4. Conclusion

In this paper, we have examined whether certain variables can predict bubbles in cryptocurrency prices. The ability to predict bubbles potentially represents an important contribution to market monitoring and to the understanding of price dynamics for cryptocurrencies. To our knowledge, this is the first study to examine predictors of bubbles in cryptocurrencies.

As cryptocurrencies emerged only recently, they are only now beginning to be financially regulated. We have therefore included economic policy uncertainty (EPU), the VIX index and the TED spread among our tested bubble predictors. The EPU index captures uncertainty about legislation and regulations, while the VIX index and TED spread capture general uncertainty in financial markets.

We have studied a set of variables with potential impacts on cryptocurrency prices and used this as a basis for our selection of predictors in the regression models.

Our results, based on the PSY test, reveal multiple bubble periods in all the studied cryptocurrencies, particularly during 2017 and 2018. This is in line with the results of Corbet et al. (2018) and Bouri et al. (2018), who also detect extensive cryptocurrency bubbles in the same periods. Furthermore, Bouri et al. (2018) find that Bitcoin in particular demonstrates extensive price explosivity, and our findings support this.

We have also looked into which factors can predict these bubbles. Where cryptocurrency-specific variables are concerned, volatility and volume are distinctly associated with bubble behavior across all the studied cryptocurrencies. Google trends and transactions mostly demonstrate positive relationships with bubbles, but the effects are dependent on the cryptocurrency studied and the type of regression model. Among the uncertainty variables we tested, the VIX-index generally exhibits a negative association with bubbles, while the EPU-index demonstrates a positive relationship with bubbles. The TED-spread exhibits a more ambiguous relationship with bubbles. Overall, many of the variables we have investigated exhibit potential to predict bubbles, and of these, trading volume, volatility and the VIX-index appear to be particularly strong predictors. These results may assist authorities in designing appropriate financial regulations for cryptocurrencies.

Declaration of Competing Interest

No conflict of interest exists.

Appendix A. Additional Tables

The regression results from the probit univariate regressions and the PSY statistic univariate regressions are provided in Table A.1 and A.2, respectively. The univariate models employ regressions between the dependent variable (PSY-statistic or bubble dates dummy) with one explanatory variable at a time, for each cryptocurrency. The models are estimated with a constant, but only the parameters of the explanatory variables and the corresponding standard errors are reported in the table. This implies that we estimate 7 univariate regression equations per cryptocurrency.

Table A.1
Probit Regression Results - Univariate Time Series Regressions. The dependent binary variable $BUB_{i,t}$ only takes the values 1 (explosive dates) and 0 (non-explosive dates). The independent variables are described in the data section. The sample includes *all* dates for the *respective cryptocurrency* (see Table 1 for individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. All the reported estimates are coefficients with corresponding Newey-West standard errors.

	Dependent variable: $BUB_{i,t}$							
	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE
$Google_{t-1}$	1.187***	2.075***	1.173***	1.725***	1.866***	0.454***	0.311***	1.742***
	(0.420)	(0.298)	(0.416)	(0.486)	(0.608)	(0.137)	(0.0799)	(0.431)
$Volatility_{t-1}$	36.23***	20.25***	17.69***	22.97***	22.65***	15.77***	16.52***	19.07***
	(4.685)	(4.939)	(4.364)	(4.806)	(3.810)	(4.515)	(3.446)	(3.317)
$Transactions_{t-1}$	−0.213**	0.700***	0.162**	0.162**	0.609***	0.450	0.445***	0.374***
	(0.0836)	(0.174)	(0.0691)	(0.0691)	(0.0909)	(0.313)	(0.108)	(0.0713)
$Volume_{t-1}$	1.204***	0.620***	1.070***	1.051***	0.847***	0.954***	0.526***	1.338***
	(0.160)	(0.162)	(0.132)	(0.140)	(0.171)	(0.160)	(0.0816)	(0.186)
EPU_{t-1}	0.464**	0.489*	0.563*	0.767***	0.0138	1.003***	0.828***	0.866***
	(0.208)	(0.291)	(0.292)	(0.234)	(0.362)	(0.248)	(0.289)	(0.229)
VIX_{t-1}	−2.117***	−2.370***	−2.659**	−3.296***	−1.936***	−4.864***	−1.572*	−4.001**
	(0.754)	(0.827)	(1.132)	(1.160)	(0.650)	(1.129)	(0.830)	(1.711)
TED_{t-1}	−0.199	−1.020***	−0.579***	−0.353**	0.113	−1.064***	−1.098***	−0.912***
	(0.193)	(0.304)	(0.200)	(0.160)	(0.427)	(0.290)	(0.311)	(0.216)
Observations	1258	1258	1256	1019	998	939	625	707

Table A.2
Linear Regression Results - Univariate Time Series Regressions. The dependent variable is the PSY-statistic. The independent variables are described in the data section. The sample includes *all* bubble dates for the *respective cryptocurrency i* (see Table 1 for individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. All the reported estimates are coefficients with corresponding Newey–West standard errors.

	Dependent variable: $PSY_{r_2^i}(r_0)_{i,t}$							
	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE
Google_{t-1}	1.069***	1.071***	0.775***	0.899***	0.588**	0.453***	0.205***	0.936***
	(0.267)	(0.174)	(0.295)	(0.233)	(0.229)	(0.150)	(0.0474)	(0.340)
Volatility_{t-1}	28.87***	19.23***	19.28***	25.11***	10.42***	13.93***	15.28***	20.53***
	(3.346)	(2.783)	(2.700)	(2.939)	(2.372)	(4.020)	(2.382)	(2.843)
$\text{Transactions}_{t-1}$	−0.250***	0.404***	−0.0344	0.491***	0.467***	0.458***	0.281***	0.392***
	(0.0458)	(0.0463)	(0.0459)	(0.0322)	(0.0359)	(0.111)	(0.0631)	(0.0558)
Volume_{t-1}	0.734***	0.478***	0.539***	0.629***	0.373***	0.561***	0.561***	0.619***
	(0.0708)	(0.107)	(0.0451)	(0.0391)	(0.0663)	(0.0827)	(0.0827)	(0.0629)
EPU_{t-1}	0.836***	−0.0322	0.267	0.292	0.725***	1.030***	0.184	0.389*
	(0.154)	(0.235)	(0.206)	(0.211)	(0.191)	(0.223)	(0.188)	(0.219)
VIX_{t-1}	−1.176***	−0.988***	−1.043***	−0.995***	−1.539***	−1.985***	−1.250***	−0.779**
	(0.267)	(0.261)	(0.275)	(0.294)	(0.215)	(0.261)	(0.263)	(0.305)
TED_{t-1}	0.263*	−1.246***	−0.559***	−0.639***	0.674***	0.281	−0.913***	−1.079***
	(0.147)	(0.221)	(0.177)	(0.179)	(0.213)	(0.242)	(0.223)	(0.213)
Observations	1258	1258	1256	1019	998	939	625	707

Table A.3
Linear Regression Results - Panel Regression. The dependent variable is the PSY statistic. The independent variables are described in Section 2. The sample includes *all cryptocurrencies* (see Table 1 for the individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. The coefficients are estimated by Prais–Winsten regression. The panel model is estimated with random effects. The standard errors are corrected for AR(1) autocorrelation, heteroscedasticity and cross-sectional correlation.

	Dependent variable: $PSY_{i,t}(r_0)$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Google}_{i,t-1}$	0.0409***							0.0413***
	(0.00628)							(0.00729)
$\text{Volatility}_{i,t-1}$		3.397***						5.076***
		(0.400)						(0.414)
$\text{Transactions}_{i,t-1}$			0.0403***					0.0351***
			(0.00472)					(0.00476)
$\text{Volume}_{i,t-1}$				0.0241***				0.0419***
				(0.00460)				(0.00405)
$\text{EPU}_{i,t-1}$					−0.00328			0.0650
					(0.0338)			(0.0414)
$\text{VIX}_{i,t-1}$						−0.148***		−0.339***
						(0.0538)		(0.0602)
$\text{TED}_{i,t-1}$							−0.0497	−0.0499
							(0.0576)	(0.0614)
Intercept	0.00759	−0.138***	−0.0164	−0.291***	0.0233	0.404***	−0.0463	−0.194
	(0.0320)	(0.0348)	(0.0302)	(0.0653)	(0.153)	(0.148)	(0.0746)	(0.277)
Observations	8060	8060	8060	8060	8060	8060	8060	8060
R-Squared	0.0043	0.0189	0.0106	0.0040	0.0000	0.0028	0.0002	0.0830

Table A.4
Linear Regression Results - Time Series Regressions. The dependent variable is the PSY statistic. The independent variables are described in Section 2. The sample includes *all* dates for the *respective cryptocurrency* (see Table 1 for individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. All the reported estimates are coefficients with corresponding Newey–West standard errors.

	Dependent variable: $PSY_t(r_0)$							
	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE
Google_{t-1}	0.885***	0.520***	0.0947	0.0152	−0.451***	−0.196***	0.0599***	0.127*
	(0.154)	(0.112)	(0.107)	(0.101)	(0.146)	(0.101)	(0.0298)	(0.139)
Volatility_{t-1}	12.07***	5.506***	8.548***	10.32***	2.961*	7.231***	8.236***	5.307**
	(2.854)	(1.757)	(2.404)	(2.695)	(1.519)	(2.297)	(1.972)	(2.062)
$\text{Transactions}_{t-1}$	−0.191***	0.304***	−0.186***	0.175***	0.355***	0.228**	0.149***	0.132***
	(0.0340)	(0.0368)	(0.0329)	(0.0521)	(0.0465)	(0.104)	(0.0519)	(0.0364)
Volume_{t-1}	0.423***	0.0953*	0.446***	0.320***	0.115**	0.320***	0.164***	0.462***
	(0.0647)	(0.0545)	(0.0454)	(0.0552)	(0.0491)	(0.0984)	(0.0281)	(0.0658)
EPU_{t-1}	0.451***	−0.0141	0.144	−0.0267	0.441***	0.569***	−0.159	0.365***
	(0.117)	(0.107)	(0.117)	(0.132)	(0.119)	(0.145)	(0.110)	(0.129)
VIX_{t-1}	−0.485***	0.722***	−0.301**	−0.0423	−0.733***	−1.057***	−0.769***	−0.0208
	(0.158)	(0.191)	(0.135)	(0.178)	(0.175)	(0.217)	(0.147)	(0.166)
TED_{t-1}	0.469***	−0.877***	−0.181	−0.006	0.218	0.745***	−0.208*	−0.16
	(0.118)	(0.133)	(0.117)	(0.137)	(0.156)	(0.155)	(0.107)	(0.127)
Intercept	−7.118***	−4.384***	−5.716***	−5.342***	−0.941	−2.349*	0.720	−7.670***
	(1.279)	(1.036)	(0.894)	(1.151)	(0.799)	(1.209)	(0.721)	(1.036)
Observations	1258	625	1256	1258	939	998	707	1019
R-Squared	0.603	0.722	0.583	0.566	0.617	0.560	0.568	0.585

Appendix B. Details on Methodology

In this paper, we detected bubbles using the psymonitor [6] package. For sake of completeness, we briefly present the PSY procedure here. First, we provide the rationale behind the identification of price explosiveness. Second, we present the PWY and PSY tests and their respective test statistics. Third, we outline how the date-stamping of bubbles is executed. Lastly, we describe how the PSY framework can be extended to identify market collapses or crisis.

B1. Identification of price explosiveness

Phillips and Magdalinos (2007) propose that explosive behavior in asset price series can be regarded as a warning signal of an expansionary phase of a bubble period. It is this assumption that lays the foundation for econometric testing of market data time series by applying recursive right-tailed unit root test procedures. Although the PWY, the sequential PWY and the PSY date-stamping strategies each use distinctive recursive algorithms, they are all based on recursive right-sided unit root tests.

Phillips et al. (2015a,b) integrate the mild drift in price processes that frequently appears over long time series by adding an asymptotically negligible drift to the martingale process. The null hypothesis (H_0) of the date stamping strategies assumes normal market behavior and takes the following form:

$$y_t = dT^{-\eta} + \theta y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right), \quad \theta = 1 \tag{10}$$

where $dT^{-\eta}$ (with constant d, and sample size T) perceive any small drift process that may occur in the price time series, but which is of lower order than the martingale element θy_{t-1} and is consequently asymptotically negligible. The localizing parameter η regulates the impact of the intercept and drift as the sample size T goes to infinity.

One can rewrite Eq. (10) to obtain $y_t = d\frac{t}{T^\eta} + \sum_{j=1}^{\infty} \epsilon_j + y_0$. The deterministic drift is represented by the component $d\frac{t}{T^\eta}$. The drift is minor in relation to a linear trend when the localizing coefficient $\eta > 0$, the drift is minor relative to the martingale element of y_t when $\eta > \frac{1}{2}$. Furthermore, the standardized output $T^{-\frac{1}{2}} y_t$ behaves like a Brownian motion with drift when $\eta < \frac{1}{2}$. The reason for the inclusion of the drift term is to separate the transient drift component and be able to perform tests for explosiveness similar to the ordinary augmented Dickey-Fuller unit root test against stationarity.

B2. Models and test statistics

Phillips et al. (2011) presented the sup augmented Dickey–Fuller test (SADF), known as the PWY test. Later Phillips et al. (2015a,b) presented the general sup augmented Dickey-Fuller test (GSADF), named the PSY test. Both tests are based on recursive approaches and contain a rolling window augmented Dickey-Fuller style regression. The window size of the rolling ADF regression is denoted r_w, defined by $r_w = r_2 - r_1$ and the set minimum window width r_0. Note that all the parameters, r_w, r_2, r_1, r_0 are defined as fractions of the overall number of periods.

[6] https://cran.r-project.org/web/packages/psymonitor/index.html.

The PWY and PSY procedures are based on the following reduced form empirical equation, to respectively obtain the SADF and GSADF test statistics:

$$\Delta y_t = \hat{\alpha}_{r_1,r_2} + \hat{\beta}_{r_1,r_2} y_{t-1} + \sum_{i=1}^{k} \hat{\psi}^{i}_{r_1,r_2} \Delta y_{t-i} + \hat{\epsilon}_t, \quad \epsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right) \tag{11}$$

where k is the transient lag order. $\hat{\alpha}_{r_1,r_2}$, $\hat{\beta}_{r_1,r_2}$ and $\hat{\psi}_{r_1,r_2}$ are parameters estimated using OLS and y_t is the logarithm of the cryptocurrency price. The numbers r_1 and r_2 represent the starting and ending points in the regression window of the total sample (T). The observation quantity in the regression is denoted by $T_w = \lfloor T r_w \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. The ADF statistic (t-ratio) from the regression, denoted by $ADF^{r_2}_{r_1}$, is given by the ratio of $\hat{\beta}_{r_1,r_2}$ and its standard error. We then apply this type of ADF rolling window regression to acquire a series of ADF statistics and detect bubbles.

To identify explosiveness (explosive behaviour) we perform a right-tailed variation of the standard Augmented Dickey-Fuller unit root test. As Caspi (2017) specifies, in both the PWY and PSY framework, we test for:

$$H_0 : \hat{\beta}_{r_1,r_2} = 1, H_1 : \hat{\beta}_{r_1,r_2} > 1. \tag{12}$$

The null and alternative hypotheses are dependent on the test statistic used. In the PWY test the null hypothesis is of a unit root, and the alternative hypothesis is of a single periodically collapsing bubble period. The PSY test's null hypothesis is also of a unit root, but the alternative hypothesis is of multiple periodically collapsing bubbles. A comparison between PWY and PSY are given in Section B.2.3.

B2.1. The PWY Test for Bubbles (SADF Test)

Phillips et al. (2012) suggest a sup ADF (SADF) process, also known as the PWY approach, to identify bubbles in asset prices. The SADF statistics series is denoted by:

$$SADF(r_0) = \sup_{r_2 \in [r_0,1]} \{ADF^{r_2}_{0}\}. \tag{13}$$

This statistic is obtained through the PWY test which, as mentioned above, relies on repeated estimation of the Augmented Dickey Fuller regression model on a forward expanding sample sequence. The window size r_w expands from r_0 (smallest window width fraction of the total sample size) to 1 (largest window width fraction of the total sample size). In the PWY test, the starting point in the data r_1 is fixed at 0. The endpoint varies with r_w and ends up in $r_2 = 1$. The non-varying starting point in the PWY test stand in contrast to the PSY test, where both the starting point r_1 and ending point r_2 in the sample window is allowed to vary.

B2.2. The PSY Test for Bubbles (GSADF Test)

Phillips et al. (2015a) suggest a generalized sup ADF (GSADF) process, also known as the PSY approach, to detect and date-stamp bubble periods. The date-stamping is done by performing a recursive backward method, which is presented in Section B.3. Similar to PWY, the PSY dating strategy applies recursive right-tailed ADF tests and accepts flexible window widths. As distinct from the SADF test of PWY, the GSADF process enables both the starting and ending point to be adjusted over a reasonable range of flexible windows. The PSY test allows the starting point in the ADF regression model (11) to vary from 0 to $r_2 - r_0$, in addition to also changing the endpoint as in the PWY test. As a consequence, the subsamples used in the recursion are substantially more comprehensive than those in the PWY test. The power of the GSADF statistic is hence larger compared to the SADF statistic. For a better grasp of the recursion in the PSY test, we refer the reader to Fig. 1 in Phillips et al. (2015a). Formally the GSADF statistic is defined as:

$$GSADF(r_0) = \sup_{r_2 \in [r_0,1], r_1 \in [0, r_2 - r_0]} \{ADF^{r_2}_{r_1}\}. \tag{14}$$

B2.3. Comparison of bubble identification tests

Phillips et al. (2015a) show that the PSY method outperforms the PWY approach, a modified sequential PWY algorithm developed in the same paper, and a procedure called the CUSUM approach. The main reason for its better performance is that the PSY approach covers more subsamples and has more flexibility when it comes to choosing and adjusting window width. The PWY approach can be unreliable when multiple bubbles appear: when the sample period includes several episodes of explosive behavior, the PWY approach may suffer from reduced power and can be unreliable when it comes to detecting the presence of bubbles. These inconsistencies become even more evident when using long time series or swiftly fluctuating market data in which more than one bubble period is expected.

The high degree of volatility in cryptocurrency prices makes the PWY method unsuitable for our study. In contrast to the PWY dating strategy, the PSY procedure is consistent in time stamping the origination and termination of multiple bubbles. The PSY approach is hence considerably more suitable for identifying bubbles in cryptocurrencies because of their rapidly changing price behavior. We therefore use the PSY approach in this paper.

B3. Date-stamping Bubbles

The PSY test allows the origination and termination points of a bubble to be date stamped. Bubble periods are found by executing a rolling window test backwards. The *psymonitor* package used in our paper employs a optimized recursion, introduced in Phillips and Shi (2018), when date-stamping the bubbles. The PSY statistic is defined as the supremum of the ADF statistic sequence, i.e.

$$PSY_{r^\dagger}(r_0) = \sup_{r_1 \in [0, r^\dagger - r_0], r_2 = r^\dagger} \{ADF_{r_1}^{r_2}\}. \tag{15}$$

The PSY framework then suggests comparing each element of the estimated $ADF_{r_1}^{r_2}$ test statistic sequence to the related right-tailed critical values of the standard ADF statistic, in order to detect explosive behaviour at time $T_{r^\dagger}$. The first chronological observation where the ADF statistic exceeds the critical value is defined as the origination point of the bubble T_{r_e}. The estimated termination point of the bubble T_{r_f} is the first chronological observation after T_{r_e} where the ADF statistic goes below the critical value from above. Denote $cv_{r^\dagger}(\beta_T)$ the $100(1-\beta_T)$ critical value of the $PSY_{r^\dagger}(r_0)$ statistic where β_T is the test size. The origination and termination of the explosiveness are respectively stated according to the following crossing time fractions:

$$\hat{r}_e = \inf_{r^\dagger \in [r_0, 1]} \left\{ r^\dagger : PSY_{r^\dagger}(r_0) > cv_{r^\dagger}(\beta_T) \right\}, \tag{16}$$

$$\hat{r}_f = \inf_{r^\dagger \in [\hat{r}_e, 1]} \left\{ r^\dagger : PSY_{r^\dagger}(r_0) < cv_{r^\dagger}(\beta_T) \right\}. \tag{17}$$

B4. The PSY Test for bubble vs. crisis identification

The PSY method presented in Phillips et al. (2015a,b) was intended to detect and time-stamp explosive behavior in asset prices. More recently, Phillips (2017) has shown that the PSY procedure also can be used as a warning device for crisis, as the algorithm can be extended to cover market collapse dynamics. Under the null hypothesis of normal market behavior, asset prices follow a martingale process with a mild drift function. In the bubble identification setting, the alternative hypothesis is a mildly explosive process. When it comes to detecting crisis, the alternative hypothesis is a random-drift martingale process.

In our paper we examine whether the asset prices follow a martingale process with a mild drift (null hypothesis - normal market conditions) or not (alternative hypothesis - either a bubble or a crisis). We do not distinguish between bubbles and crises, since the PSY algorithm doesn't separate these either.[7] In the following paragraphs we present the rationale associated with the PSY test for bubble and crisis identification, respectively.

Phillips and Magdalinos (2007) propose that explosive behavior in asset price series can be a signal of bubble behavior. In this case, asset prices can be expressed as a mildly explosive process of the form:

$$y_t = \delta_T y_{t-1} + u_t, \tag{18}$$

in which $\delta_T = 1 + cT^{-\eta}$ is a autoregressive coefficient which mildly exceeds unity, where $c > 0$ and $\eta \in (0, 1)$. Bubble identification is achieved by testing the null hypothesis of normal market conditions (martingale process with a drift) against the bubble alternative (mildly explosive process).

Phillips (2017) modeled the dynamics of asset prices during market collapses as a random drift martingale process. The logarithmic price change $y_t - y_{t-1}$ is affected by a random sequence term $(-L_t)$ and the martingale difference innovations u_t:

$$y_t - y_{t-1} = -L_t + u_t, \tag{19}$$

where u_t are the superposition of martingale differences and L_t is a random sequence independent of u_t, which follows an asymmetric scaled uniform distribution. L_t may take different forms, which cause diversity in the types of crises, and is given by:

$$L_t = Lb_t, \quad b_t \overset{iid}{\sim} U[-\epsilon, 1], \quad 0 < \epsilon < 1, \tag{20}$$

where L is a positive scale quantity which represents the shock intensity and b_t is uniform on the interval from $-\epsilon$ to 1. A crisis is identified by testing the null hypothesis of normal marked conditions (martingale process with a drift) against the alternative of crisis (random-drift martingale process). Mathematically, the null and alternative hypothesis of the empirical regression model from Eq. (11) can then be written as:

$$H_0 : \mu = dT^{-\eta} \text{ and } \rho = 0 \quad H_{1,crash} : \mu = K \text{ and } \rho = 0, \tag{21}$$

where K is the expected value of L_t and $dT^{-\eta}$ perceive any small drift process that may occur in the price time series as in Eq. (10).

[7] When using the terms “bubble”, “explosive behavior”, “crisis”, “market collapse” and so on, we have detected that there is a deviation from normal market conditions (null hypothesis of martingale process with mild drift fails) and that there is either a bubble or a crisis (alternative hypothesis of either a mildly explosive process or random-drift martingale process is valid).

References

Aalborg, H.A., Molnár, P., de Vries, J.E., 2018. What can explain the price, volatility and trading volume of Bitcoin? Finance Res. Lett..

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.

Bašta, M., Molnár, P., 2018. Oil market volatility and stock market volatility. Finance Research Letters 26, 204–214.

Bekiros, S., Jlassi, M., Lucey, B., Naoui, K., Uddin, G.S., 2017. Herding behavior, market sentiment and volatility: Will the bubble resume? North Am. J. Econ. Finance 42, 107–131.

Bijl, L., Kringhaug, G., Molnár, P., Sandvik, E., 2016. Google searches and stock returns. Int. Rev. Financ.Anal. 45, 150–156.

Blanchard, O., Watson, M., 1982. Bubbles, Rational Expectations and Financial Markets. NBER Working Papers.

Blau, B.M., 2017. Price dynamics and speculative trading in Bitcoin. Res. Int. Bus. Finance 41, 493–499.

Bleher, J., Dimpfl, T., 2018. Knitting Daily Google Trends–With an Application to Augur Cryptocurrency Returns. Available at SSRN 3126324.

Boudt, K., Paulus, E.C., Rosenthal, D.W., 2017. Funding liquidity, market liquidity and TED spread: a two-regime model. J. Empir. Finance 43, 143–158.

Bouri, E., Gupta, R., Tiwari, A.K., Roubaud, D., 2017. Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. Finance Res. Lett. 23, 87–95.

Bouri, E., Shahzad, S.J.H., Roubaud, D., 2018. Co-explosivity in the cryptocurrency market. Finance Res. Lett.

Caspi, I., 2017. Rtadf: Testing for Bubbles with EViews. J. Stat. Softw. Code Snippets 81 (1), 1–16. doi:10.18637/jss.v081.c01.

CBOE, 2019. White paper CBOE volatility index. Retrieved May 29, 2019, from https://www.cboe.com/micro/vix/vixwhite.pdf.

Cheah, E.-T., Fry, J., 2015. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. Econ. Lett. 130, 32–36.

Cheung, A., Roca, E., Su, J.-J., 2015. Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. Appl. Econ. 47 (23), 2348–2358.

Choi, H., Varian, H., 2009. Predicting initial claims for unemployment benefits. Google Inc 1–5. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.549.7927&rep=rep1&type=pdf.

Choi, H., Varian, H., 2012. Predicting the present with Google Trends. Econ. Record 88, 2–9.

Corbet, S., Lucey, B., Yarovaya, L., 2018. Datestamping the Bitcoin and Ethereum bubbles. Finance Res. Lett. 26, 81–88.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. J. Finance 66 (5), 1461–1499.

Datta, D., Du, W., 2012. Nonparametric HAC estimation for time series data with missing observations. FRB Int. Finance Discuss. Pap. (1060).

Demir, E., Gozgor, G., Lau, C.K.M., Vigne, S.A., 2018. Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. Finance Res. Lett. 26, 145–149.

Diba, B.T., Grossman, H.I., 1988. Explosive rational bubbles in stock prices? Am. Econ. Rev. 78 (3), 520–530.

Economic Policy Uncertainty, 2019. US EPU (Monthly, Daily, Categorical). Retrieved May 23, 2019, from http://www.policyuncertainty.com/us_monthly.html.

Fiszeder, P., 2018. Low and high prices can improve covariance forecasts: the evidence based on currency rates. J. Forecast. 37 (6), 641–649.

Fiszeder, P., Fałdziński, M., 2019. Improving forecasts with the co-range dynamic conditional correlation model. J. Econ. Dyn. Control 108, 103736.

Fiszeder, P., Fałdziński, M., Molnár, P., 2019. Range-based DCC models for covariance and value-at-risk forecasting. J. Empir. Finance 54, 58–76.

Fry, J., Cheah, E.-T., 2016. Negative bubbles and shocks in cryptocurrency markets. Int. Rev. Financ. Anal. 47, 343–352.

Garman, M.B., Klass, M.J., 1980. On the estimation of security price volatilities from historical data. J. Bus. 53 (1), 67–78.

Gebka, B., Wohar, M.E., 2013. Causality between trading volume and returns: Evidence from quantile regressions. Int. Rev. Econ. Finance 27, 144–159.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M., Siering, M., 2014. Bitcoin - Asset or currency? Revealing users' hidden intentions. ECIS 2014 (Tel Aviv). Available at SSRN: https://ssrn.com/abstract=2425247.

Greene, W.H., 2007. Econometric Analysis (6th edition). Pearson Education, Inc., Upper Saddle River, New Jersey.

Grinberg, R., 2012. Bitcoin: an innovative alternative digital currency. Hastings Sci. Tech. LJ 4, 159.

Hoechle, D., 2007. Robust standard errors for panel regressions with cross-sectional dependence. Stata J. 7 (3), 281–312.

Kim, N., Lučivjanská, K., Molnár, P., Villa, R., 2019. Google searches and stock market activity: evidence from Norway. Finance Res. Lett. 28, 208–220.

McFadden, D.L., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Wiley, New York, pp. 105–142 ISBN 0127761500. - 1974.

Molnár, P., 2012. Properties of range-based volatility estimators. Int. Rev. Financ. Anal. 23, 20–29.

Molnár, P., 2016. High-low range in garch models of stock return volatility. Applied Economics 48 (51), 4977–4991.

Molnár, P., Bašta, M., 2017. Google searches and Gasoline prices. In: Proceedings of the 2017 14th International Conference on the European Energy Market (EEM). IEEE, pp. 1–5.

Newey, W.K., West, K.D., 1987. Hypothesis testing with efficient method of moments estimation. Int. Econ. Rev. 28 (3), 777–787.

Phillips, P.C., 2017. Detecting Financial Collapse and Ballooning Sovereign Risk. Technical Report.

Phillips, P.C., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. J. Econom. 136 (1), 115–130.

Phillips, P.C., Shi, S., 2018. Real Time Monitoring of Asset Markets: Bubbles and Crises. Technical Report.

Phillips, P. C., Shi, S., Yu, J., 2012. Testing for multiple bubblesCowles Foundation Discussion Paper No. 1843. Yale University.

Phillips, P.C., Shi, S., Yu, J., 2015. Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. Int. Econ. Rev. 56 (4), 1043–1078.

Phillips, P.C., Shi, S., Yu, J., 2015. Testing for multiple bubbles: Limit theory of real-time detectors. Int. Econ. Rev. 56 (4), 1079–1134.

Phillips, P.C., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? Int. Econ. Rev. 52 (1), 201–226.

Su, C.-W., Li, Z.-Z., Tao, R., Si, D.-K., 2018. Testing for multiple bubbles in bitcoin markets: a generalized sup ADF test. Jpn World Econ. 46, 56–63.

Wang, G.-J., Xie, C., Wen, D., Zhao, L., 2019. When bitcoin meets economic policy uncertainty (EPU): measuring risk spillover effect from EPU to bitcoin. Finance Res. Lett. 31.

Wu, S., Tong, M., Yang, Z., Derbali, A., 2019. Does gold or bitcoin hedge economic policy uncertainty? Finance Res. Lett. 31, 171–178.

Yermack, D., 2015. Is Bitcoin a real currency? An economic appraisal. In: Handbook of digital currency. Elsevier, pp. 31–43.