University of Stavanger
Stavanger, Spring 2019

What Can Predict Bubbles in Cryptocurrency Prices?

Identification and explanation based on PSY test and regression models

Christian J. Landsnes and Fredrik A. Enoksen

Supervisor: Peter Molnar

Master thesis, Economics and Business Administration

Major: Applied Finance

UNIVERSITY OF STAVANGER

Universitetet i Stavanger

UIS BUSINESS SCHOOL

MASTER'S THESIS

STUDY PROGRAM:	THESIS IS WRITTEN IN THE FOLLOWING SPECIALIZATION/SUBJECT:
Master's Degree, Economics and Business Administration	Applied Finance
	IS THE ASSIGNMENT CONFIDENTIAL? No

TITLE: What Can Predict Bubbles in Cryptocurrency Prices?

AUTHOR(S)		SUPERVISOR: Peter Molnar
Candidate number:	Name:	
4007 …………………	Christian J. Landsnes …………………………………….	
4024 …………………	Fredrik A. Enoksen …………………………………….	

Acknowledgements

This thesis concludes our studies at the Univerisity of Stavanger (UiS). Our objective is to expand the current literature on cryptocurrency bubble detection and investigate the possible predictors of such bubbles. We would like to thank our supervisor, Professor Peter Molnar at the UiS Business School for invaluable guidance. Through countless hours of supervision, Peter has given us encouragement and constructive suggestions which we are grateful for.

University of Stavanger

Stavanger, June 2019

Christian J. Landsnes

Fredrik A. Enoksen

Abstract

In this paper we study variables that can predict bubbles in cryptocurrency prices. Bubble periods are detected by employing a recursive augmented Dickey-Fuller algorithm called the PSY test, developed by Phillips et al. (2015a,b). Through probit and linear regression models we study the possible predictors of the bubble periods. We utilize both detected days and the underlying test statistics produced by the algorithm as dependent variables in the analysis. Compared to other studies, we emphasize uncertainty measures as predictors and include an extended selection of cryptocurrencies. We apply panel regressions to investigate predictors across cryptocurrencies and time series regressions to study predictors for specific cryptocurrencies. We detect multiple bubble periods in all cryptocurrencies, particularly in 2017 and early 2018. The predictive ability of the variables appear to be dependent on the cryptocurrency studied. Though in general, we find that higher volatility and trading volume is positively associated with the presence of bubbles across cryptocurrencies. When it comes to uncertainty variables, the VIX-index consistently demonstrates negative relationships with bubble behavior. Furthermore, transactions and the EPU-index mostly exhibit positive associations with bubbles, but the effects are dependent on the cryptocurrency examined. In terms of bubble prediction, the probit models perform better than the linear models.

Keywords – Cryptocurrency bubbles; detection; prediction; PSY test; probit regression model; linear regression model

Contents

List of Figures

List of Tables

1 Introduction

The emergence of digital currencies has been one of the most remarkable financial innovations of the last decade. Their futuristic properties and extreme price behavior have earned excessive media coverage, as well as attention from regulators and researchers. Most cryptocurrency prices are known to be volatile, and has experienced dramatic increases and collapses in the recent years. This has triggered a conversation about bubbles and whether the price levels can be justified by a fundamental value. Bubbles are commonly interpreted as deviations from intrinsic value. In this paper, we try to detect bubbles by analyzing the statistical properties of the cryptocurrencies prices.

Bitcoin has been the most prominent of the cryptocurrencies and has experienced severe price fluctuations over the recent years, with the price reaching a peak in late 2017. Bitcoin was originally intended to be digital money. It supposedly contributes to a more reliable and trustworthy transaction system with lower costs (Grinberg, 2012). Bitcoin and other cryptocurrencies have 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 made it subject to criticism (Grinberg, 2012). The question of classification as either a speculative asset or a means of exchange have also been discussed. 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 these assets are naturally a topic of research interest. Multiple studies have detected bubbles in cryptocurrencies, mainly in the Bitcoin price (Cheah and Fry, 2015; Corbet et al., 2018; Su et al., 2018).

A novel method for detecting bubbles was developed by Phillips et al. (2015a). This framework is commonly referred to as the PSY framework and is based on the preceding PWY framework (Phillips et al., 2011). The PSY procedure has been shown effective for bubble detection and is employed in asset monitoring by central banks, particularly for several real estate markets (Phillips and Shi, 2018). By utilizing a recursive augmented Dickey-Fuller test algorithm it can detect periods were the price behavior deviates from its assumed normal behavior. It tests whether the observed asset prices follow a martingale process with a mild drift, which is assumed to be normal behavior. The alternative behavior is that the price is in an expansionary *bubble* phase were it follows a mildly explosive process, or that the price is in a depreciative *crisis* phase defined as a martingale process with a random drift term (Phillips and Shi, 2018). The PSY procedure is presented in detail in the methodology section.

This study uses the PSY framework to locate bubbles and generate corresponding test statistics for the cryptocurrencies Bitcoin, Ethereum, Ripple, Litecoin, Monero, Dash coin, Nem coin and Dogecoin. We consider the ability to predict bubbles as an important contribution in both normative analysis, such as utilization in market monitoring, and in positive analysis for understanding the price dynamics of the cryptocurrency markets. The relationship between the bubbles/test statistics is studied by estimating different regression models with relevant explanatory variables. Compared to other papers, we attempt to look at an extended selection of cryptocurrencies, with an emphasis on some uncertainly measures. We study the predictive ability of Google search queries, volatility, transactions, trading volume, EPU-index, VIX-index and the TED-spread. We find that volatility and trading volume consistently exhibit a positive relationship with bubble behavior, while the VIX-index demonstrates a negative association. Other variables exhibit significant effects, but they appear to be more dependent on the estimated model and cryptocurrency studied.

The remainder of the paper is structured in the following way: Section 2 provides an overview of the background literature. Section 3 describes the data utilized in the paper. Section 4 briefly explains the methodology used. An analysis of the results are conducted in section 5. Finally, a conclusion is provided in section 6.

2 Literature Review

This section provides a review of literature related to our paper. First, we review the literature about the drivers of cryptocurrency prices. Second, we give an overview of the development in the research of empirical bubble detection strategies. Third, we look at the application of bubble detection strategies in cryptocurrency markets.

2.1 Cryptocurrency Price Determinants

Many papers have studied cryptocurrency price determinants. Examined variables can be categorized into two groups: intrinsic variables and extrinsic variables. Intrinsic variables can be referred to as variables that are direct properties of the cryptocurrency phenomenon, such as trading volume, volatility, search interest etc. Extrinsic variables are not directly connected to cryptocurrencies and can be stock market returns, gold price, interest rates etc. This research lays the foundation for the conducted analysis and has contributed to the choice of relevant variables studied in the regression models.

An asset's volatility and returns together with it's transactions and trading volume are common research variables. These variables has also been studied for cryptocurrencies. Ciaian et al. (2016) shows that the price formation of Bitcoin, to a large extent, can be explained by traditional economic models. The aggregated number of unique Bitcoin transactions, which is considered a demand side variable, has greater influence than the total supply of Bitcoins. The research of Balcilar et al. (2017) shows that trading volume has predictive ability of future returns in Bitcoin. Though, the relationship is subject to non-linearity and structural breaks. Contrarily, Aalborg et al. (2018) do not find predictive ability of trading volume, but rather unique addresses and transactions. They also show that price volatility appears to be strongly positively related to the previously observed levels of volatility. Blau (2017) studies the level of speculative trading in Bitcoin. Interestingly, the research do not demonstrate an association between the level of speculative trading and the level of volatility or extreme returns.

Studying the relationship between proxies of interest and cryptocurrency prices is common in cryptocurrency research. Web search queries, such as Google Trends and Wikipedia, are common proxies. Kristoufek (2013) concludes that there is a strong correlation between the Bitcoin price and the frequency of "Bitcoin" search queries on Google Trends and Wikipedia. The correlation is bidirectional, meaning that the search frequency has an impact on the Bitcoin price and vice versa. Panagiotidis et al. (2018) find that Google searches exhibit a positive relationship with Bitcoin returns when it is above the 7-day trend of Google search queries and a negative relationship when below trend. Aalborg et al. (2018) report that Google search frequency is negatively associated with future trading volume. All referenced papers note that web search queries can serve as a proxy for public interest and that the variable has demonstrated relationships with cryptocurrency prices. An interesting paper by Kristoufek (2015) states that the Bitcoin price is influenced by investors' interest in the asset. Particularly, greater investor interest generates price rises during explosive periods in the Bitcoin price, and vice verca.

Other studies, like Panagiotidis et al. (2018) and Demir et al. (2018), have examined different uncertainty variables' relationship to the Bitcoin price. Panagiotidis et al. (2018) examine multiple determinants for Bitcoin returns in their research. The most influential variables of those examined were Google search frequency, gold returns and economic policy uncertainty indices. Uncertainty indices such as the VIX-index and different economic policy indices were mostly negatively associated with returns. The research of Demir et al. (2018) find that the EPU-index has predictive potential of Bitcoin returns and that the cryptocurrency demonstrates hedging capabilities. The results show that a higher EPU indicates lower returns, but not for the the higher and lower quantiles of returns where the relationship is positive.

2.2 Empirical Bubble Detection Strategies

Empirical bubble detection has been a topic of study for decades. One of the most common methods to identify bubbles empirically is to investigate the time series properties of the underlying asset's price. Asset pricing theory proposes that a bubble component in an asset price exists if the prices demonstrate explosive behavior. This lays the foundation for establishing econometric tests of the price's stochastic properties, which is targeted at detecting episodes of explosiveness in time series data (Caspi, 2017).

There have been numerous attempts to develop statistical procedures for bubble identification. Diba and Grossman (1988) were early to apply a unit root test in order to detect explosive behavior in asset prices. Phillips et al. (2011) and Phillips et al. (2015a,b) continued the research of Diba and Grossman (1988), by expanding on their methodology. The PWY and PSY strategies apply different forms of the augmented Dickey-Fuller test to identify and date-stamp bubbles. Phillips et al. (2015a,b) show through different tests that the PSY method outperforms the PWY method when it comes to detecting multiple bubbles.

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

2.3 Explosive Behavior in Cryptocurrency Prices

Previous researchers have utilized different types of bubble detection strategies to detect explosiveness in cryptocurrencies. The most common method used for detecting bubbles in cryptocurrency prices, and especially Bitcoin, has been PSY framework.

Cheung et al. (2015) and Su et al. (2018) apply different versions of the PSY test to detect bubbles in the Bitcoin price. Cheung et al. (2015) employs the PSY framework presented in Phillips et al. (2013), which is a early version of the PSY method, to detect bubbles in the Bitcoin market during the period 2010-2014. They identify three extensive periods of bubble behavior in the time span ranging from 2011 to 2013. The bubble periods lasted from 66 days to 106 days, and seems to be influenced by major events that appeared in the Bitcoin market. Su et al. (2018) uses the PSY procedure presented in Phillips et al. (2015a,b) to identify and date-stamp bubbles in the Bitcoin market. In total, the researchers detects 4 bubble periods during the period 2011-2017. They identify two bubbles in 2013 and two bubbles in 2017. Similar to the research of Cheung et al. (2015), Su et al. (2018) also find the bubble periods to 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). They also investigate the fundamental drivers of the prices, contrary to the papers mentioned above. Their conclusion is that there is no clear relationship between the fundamental variables and bubble development in both Bitcoin and Ethereum. The paper proposes that there are short periods were fundamental variables (hashrate, block size, volatility and liquidity) affect the price in both currencies, but these influences disappear. Bouri et al. (2018) identify bubbles in Bitcoin, Ripple, Ethereum, Litecoin, NEM, Dash and Stellar. They also apply a logistic regression to study the co-explosivity between the cryptocurrencies. The paper shows that there were numerous bubble periods in all cryptocurrencies, particularly in 2017. The results from the logistic regression shows that the likelihood of bubble periods in one cryptocurrency typically is contingent on the existence of bubbles in other cryptocurrencies, implying a high-degree of co-explosivity.

There are papers which do not use the PSY framework. Cheah and Fry (2015) use economic and econometric models to examine the fundamental value of Bitcoin and if there exists speculative bubbles. They find that the Bitcoin price is prone to speculative bubbles similarly to other assets. Furthermore, the paper proposes that Bitcoin appear to behave more like an asset than a currency. Fry and Cheah (2016) employ econophysics models to identify bubbles in Bitcoin and Ripple. In the analyzed period from 2011 to 2015, the researchers detects negative bubbles in both cryptocurrencies from 2014 and onwards. The paper further note that there is a spillover from Ripple to Bitcoin that intensifies price decreases in the latter.

3 Data

The data used in this paper cover the time period December 27, 2013 to February 15, 2019. Starting dates vary depending on the availability of data for the individual cryptocurrencies studied (see table 3.1). The considered cryptocurrencies are primarily chosen based on the length of their data sets, their respective popularity and total market value. The VIX-index, which is used in our analysis, is not reported on weekends and on certain holidays etc. These days have been omitted from our analysis and the gaps have been dealt with by using the last observed value when necessary.

The daily price data and trading volume of the cryptocurrencies are collected from CoinMarketCap through an API in R Studio. Transaction volume has been collected from Coinmetrics. Though it is possible to get earlier data from other sources, we chose to use these data sets due to their apparent reliability compared to other available sources. Economic policy uncertainty index (EPU) is collected from the Economic Policy Uncertainty web page. Data on the TED-spread and VIX-index are collected from the FRED database, the 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 parenthesis in table 3.1.

Table 3.1: Time Period Employed for Each Cryptocurrency

The table presents the start and end dates of the price dataset for 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

3.1 Variables

Volatility

To measure the volatility of the cryptocurrencies we use the method originally proposed by Garman and Klass (1980), later applied in Kim et al. (2018). This estimator utilizes the trading price range during a day. The method is considered an improvement in accuracy compared to the common method of measuring volatility by standard deviation of returns (Molnár, 2012). Daily volatility is calculated as follows:

$$\text{Volatility}_t = \sqrt{\frac{1}{2}(h_t - l_t)^2 - (2\log 2 - 1)c_t^2}, \tag{3.1}$$

where

$c_t = \log(\text{close}_t) - \log(\text{open}_t)$,

$l_t = \log(\text{low}_t) - \log(\text{open}_t)$,

$h_t = \log(\text{high}_t) - \log(\text{open}_t)$.

In order to deal with possible weekly seasonality, we convert the preceding daily values into a 7-day arithmetic average by the following equation:

$$\overline{\text{Volatility}_t} = \frac{1}{7}\sum_{t=-7}^{-1} \text{Volatility}_t. \tag{3.2}$$

Transactions

Transfers of cryptocurrencies can either be done over an exchange or directly between users within the blockchain network. In general, direct transfers is assumed to be more regularly used as a means of exchange, as opposed to transfers over an exchange. For this reason it is useful to differentiate between these forms of transfers. In our paper, transactions is classified as direct transfers of a cryptocurrency between users. Transactions (TV) is standardized the same way as in Aalborg et al. (2018). It is standardized by estimating the deviation from the average volume over the last year, and is divided by the standard deviation in the same period:

$$\text{Transactions}_t = \frac{\text{TV}_t - \overline{\text{TV}}}{\sigma(\text{TV})}. \tag{3.3}$$

Volume

In our paper, trading volume for a cryptocurrency is classified as transfers over an exchange, which do not include direct transfers between users. The time series for trading volume of Bitcoin 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.

Google Trends: Adjusted Search Volume Index

Search volume from Google trends is applied in the analysis because it measures public interest in the specific cryptocurrencies analyzed. The variable is constructed as the relative level of web searches provided by Google, and have previously demonstrated to have predictive potential, as Choi and Varian (2009, 2012); Molnár and Bašta (2017); Bijl et al. (2016) have reported. The 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 of 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. (2018). 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. The following equation has been used:

$$\text{Google}_t = \log[\text{SVI}_t] - \log\left[\text{Median}\left(\text{SVI}_{t-7}, \text{SVI}_{t-14}, \ldots, \text{SVI}_{t-56}\right)\right]. \tag{3.4}$$

EPU-index

The EPU-index can be considered a proxy for the level of 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 contains certain key words and has a theme of economic 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:

$$\text{EPU}_t = \log\left[\frac{1}{7}\sum_{t=-7}^{-1} \text{EPU}_t.\right] \tag{3.5}$$

VIX-index

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 SP 500, with an expiration date of approximately one month (CBOE, 2019). Historically the VIX-index has exhibited a negative relationship with Bitcoin returns, which might be due to a "safe haven" property of the cryptocurrencies, as examined by Bouri et al. (2017). The variable has been converted by logarithmic transformation in our analysis.

TED-spread

The TED-spread is used as a proxy for the level of credit risk in the economy. It is constructed as the difference between the US inter-bank rate and the risk free US treasury rate. The intuition behind the metric is that the spread between the inter-bank 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 more volatile and at a higher level (Boudt et al., 2017). The variable has been converted by logarithmic transformation in our analysis.

3.2 Summary Statistics

Table 3.2 provides a short summary of the variables used in the analysis of bubble predictors.

Table 3.2: Variable Summary

Variable	Definition	Data Source
Google	Google search frequency for a particular cryptocurrency	API Google Trends
Volatility	Range-based volatility of cryptocurrency prices	API CoinMarketCap
Transactions	Transfers of cryptocurrency, directly between users	Coinmetric
Volume	Transfers of cryptocurrency, over exchanges	API CoinMarketCap
EPU-index	US policy uncertainty proxy	Economic Policy Uncertainty
VIX-index	Stock market uncertainty proxy	FRED database
TED-spread	Credit risk proxy	FRED database

Table 3.3 provides the descriptive statistics of the specific variables included in the analysis. The term *specific* reflects that the variables are distinct for each cryptocurrency.

Table 3.3: Descriptive Statistics for Specific Variables

	BTC		ETH		XRP		LTC	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Google	0.026	0.309	0.045	0.481	0.017	0.317	0.018	0.331
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

	XMR		DASH		XEM		DOGE	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Google	0.007	0.303	-0.688	0.805	-0.129	1.038	0.018	0.308
Volatility	0.055	0.026	0.049	0.027	0.062	0.034	0.048	0.031
Transactions	0.45	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

Table 3.4 provides the descriptive statistics of the non-specific variables included in

the analysis.[1] The term *non-specific* express that the variables are independent of the cryptocurrencies. All the non-specific variables can be considered uncertainty variables.

Table 3.4: Descriptive Statistics for Non-Specific Variables

Variables	N	Mean	St. Dev.	Min	Max	Skew	Kurtosis
EPU	1259	4.37	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

The correlations between the variables are presented in table 3.5. It is notable that the correlation between volume and volatility, as well as volume and transactions are relatively high with 47% and 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 does not seem to be a problem and that the variables seemingly capture different aspects or forms of uncertainty.

Table 3.5: Correlation Matrix

The table illustrates the correlations between the independent variables used in the analysis. We apply the same methodology as Da et al. (2011) when estimating the correlations in table 3.5. 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

[1]For the remainder of this paper (regression tables and equations), EPU, VIX and TED are respectively abbreviations for the EPU-index, VIX-index and TED-spread.

4 PSY Methodology

In the following paragraphs we present the PSY procedure. 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.

4.1 Identification of Price Explosiveness

Phillips and Magdalinos (2007) propose that explosive behavior in asset price series can be regarded as a warning signal of market explosiveness in the expansionary phase of a bubble period. It is this assumption that lays the foundation for econometric testing of time series market data by applying recursive right-tailed unit root test procedures. Although the PWY, the sequential PWY and the PSY date-stamping strategies uses distinctive recursive algorithms for each strategy, 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 appear over long time series. By adding an asymptotically negligible drift to the martingale null they incorporate this effect. The null hypothesis (H_0) of the date stamping strategies assumes normal market behavior and has the following form:

$$y_t = dT^{-\eta} + \theta y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right), \quad \theta = 1 \tag{4.1}$$

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 consequently is asymptotically negligible. The localizing coefficient η is a parameter that regulates the impact of the intercept and drift as the sample size goes to infinity $T \to \infty$.

Solving equation 4.1 gives $y_t = d\frac{t}{T^\eta} + \sum_{j=1}^{\infty} \epsilon_j + y_0$. 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 $\eta > \frac{1}{2}$. The standardized output $T^{-\frac{1}{2}}y_t$ also 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.

4.2 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 contains 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. A general rolling window ADF (RADF) test procedure is illustrated by figure 4.1 below.

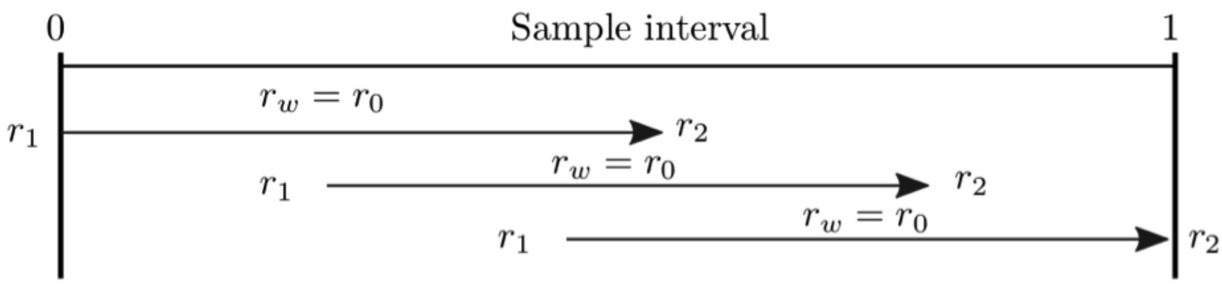


Figure 4.1: Illustration of RADF Procedure (Caspi, 2017)

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

$$\Delta y_t = \hat{\alpha}_{r_1,r_2} + \hat{\beta}_{r_1,r_2} y_{t-1} + \sum_{i=1}^{k} \hat{\psi}^i_{r_1,r_2} \Delta y_{t-i} + \hat{\epsilon}_t, \quad \epsilon_t \overset{iid}{\sim} (0, \sigma^2) \tag{4.2}$$

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 represents the starting and ending point in the regression window of the total sample (T). The observation quantity in the regression is denoted by $T_w = \lfloor Tr_w \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. The ADF statistic (t-ratio) from the regression, denoted by $ADF_{r_1}^{r_2}$, 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

$$\begin{aligned} H_0 &: \hat{\beta}_{r_1,r_2} = 1, \\ H_1 &: \hat{\beta}_{r_1,r_2} > 1. \end{aligned} \tag{4.3}$$

The null and alternative hypothesis is 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 4.2.3 *Comparison of Bubble Identification Tests.*

4.2.1 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

$$SADF\,(r_0) = \sup_{r_2 \in [r_0,1]} \{ADF_0^{r_2}\}. \tag{4.4}$$

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. The recursion of the PWY test is illustrated below in figure 4.2.

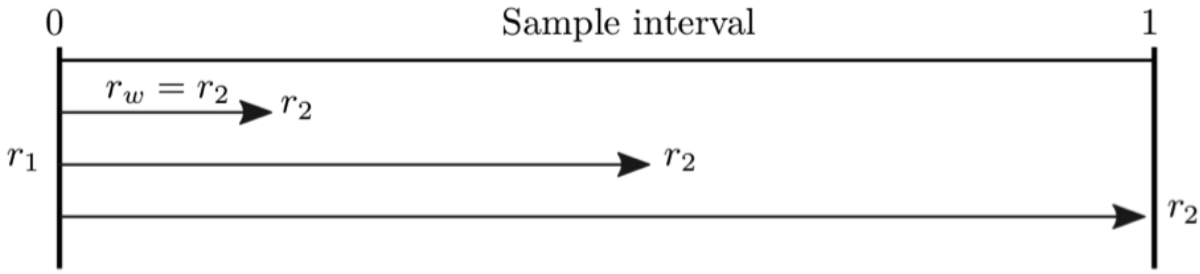


Figure 4.2: Illlustration of SADF Procedure (Phillips et al., 2015a)

4.2.2 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 4.3 *Date-stamping Bubbles*. 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 allows to adjust both the starting and ending point over a reasonable range of flexible windows. The PSY test allows the starting point in the ADF regression model 4.2 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 of the PWY test. The power of the GSADF statistic is hence larger compared to the SADF statistic. The recursion of the PSY test is illustrated in figure 4.3 below. Formally the GSADF statistic is defined as

$$GSADF\,(r_0) = \sup_{r_2 \in [r_0,1], r_1 \in [0, r_2 - r_0]} \{ADF_{r_1}^{r_2}\}. \tag{4.5}$$

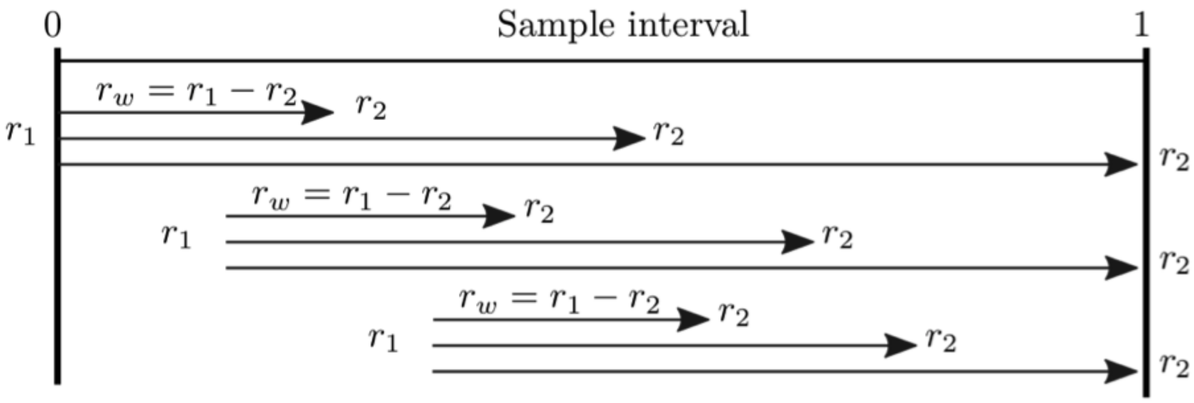


Figure 4.3: Illlustration of GSADF Procedure (Phillips et al., 2015a)

4.2.3 Comparison of Bubble Identification Tests

In Phillips et al. (2015a) it is shown that the PSY method outperforms the PWY approach, a modified sequential PWY algorithm developed in the same paper, as well as a procedure called the CUSUM approach. The main reasons for the outperformance is that the PSY approach covers more subsamples and have superior 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. The inconsistencies becomes even more evident when using long time series or swiftly fluctuating market data where more than one bubble period is expected.

The high degree of volatility in cryptocurrency prices makes the PWY method unsuitable to employ in 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 to use when identifying bubbles in cryptocurrencies because of its rapidly changing price behavior. We therefore use the PSY approach further in this paper.

4.3 Date-stamping Bubbles

The PSY test allows for date stamping the origination and termination points of a bubble. 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 performing the bubble date-stamping. The PSY statistic is defined as the supremum of the ADF statistic sequence, i.e.,

$$PSY_{r^\dagger}(r_0) = \sup_{r_1 \in [0, r^\dagger - r_0], r_2 = r^\dagger} \{ADF_{r_1}^{r_2}\}. \tag{4.6}$$

The PSY framework then suggest 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 to detect explosive behaviour at time $T_{r^\dagger}$. The first chronological observation where the ADF statistics 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 statistics goes below the critical value from above. The origination and termination of the explosiveness is respectively stated according to the following crossing time fractions:

$$\hat{r_e} = \inf_{r^\dagger \in [r_0, 1]} \left\{ r^\dagger : PSY_{r^\dagger}(r_0) > cv_{r^\dagger}(\beta_T) \right\}, \tag{4.7}$$

$$\hat{r_f} = \inf_{r^\dagger \in [\hat{r_e}, 1]} \left\{ r^\dagger : PSY_{r^\dagger}(r_0) < cv_{r^\dagger}(\beta_T) \right\}, \tag{4.8}$$

where $cv_{r^\dagger}(\beta_T)$ is the $100(1 - \beta_T)$ critical values of the $PSY_{r^\dagger}(r_0)$ statistic and β_T is the test size.

4.4 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 setting of bubble identification, the alternative hypothesis is a mildly explosive process (described in subsection 4.4.1). When it comes to detecting crisis, the alternative hypothesis is a random-drift martingale process (explained in subsection 4.4.2).

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 crisis). We do not distinguish between bubbles and crisis since the PSY algorithm doesn't separate the two of them.[2] In the following two subsections we present the rationale associated with the PSY test for bubble and crisis identification, respectively. Table 4.1 summarizes the null and alternative hypotheses for bubble and crisis identification.

Table 4.1: The PSY Test for Bubble and Crisis Identification

Identification	Null Hypothesis (Normal Market Conditions)	Alternative Hypothesis (Bubble/Crisis)
Bubble Identification	Martingale process with mild drift	Bubble: Mildly explosive process
Crisis Identification	Martingale process with mild drift	Crisis: Random-drift martingale process

4.4.1 The PSY Test for Bubble Identification

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

$$\log P_t = \delta_T \log P_{t-1} + u_t, \tag{4.9}$$

[2] 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).

in which $\delta_T = 1 + cT^{-\eta}$ is a autoregressive coefficient which mildly exceeds unity (with $c > 0$ and $\eta \in (0, 1)$).

As presented in table 4.1, bubble identication is achieved by testing the null hypothesis of normal market conditions (martingale process with a drift) against the alternative of bubble (mildly explosive process). When it comes to bubble identification, the null and alternative hypotheses of the empirical regression model equation 4.2 can be stated as

$$\begin{aligned} &H_0 : \mu = g_T \text{ and } \rho = 0 \\ &H_{1,bubble} : \mu = 0 \text{ and } \rho > 0. \end{aligned} \tag{4.10}$$

4.4.2 The PSY Test for Crisis Identification

Phillips (2017) modeled the dynamics of asset prices during market collapses as a random drift martingale process. The logarithmic price change ($\log P_t - \log P_{t-1}$) is affected by a random sequence term ($-L_t$) and the martingale difference innovations u_t, expressed by the following equation

$$\log P_t - \log P_{t-1} = -L_t + u_t. \tag{4.11}$$

u_t are the superposition of martingale differences with variance σ^2. 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 type of crises, and is given by

$$L_t = Lb_t, \quad b_t \overset{iid}{\sim} U\left[-\epsilon, 1\right], \quad 0 < \epsilon < 1, \tag{4.12}$$

where L is a positive scale quantity which represents the shock intensity and b_t is uniform on the interval from $-\epsilon$ to 1.

As summarized in table 4.1, crisis identification is done 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 equation 4.2 can then be written as

$$\begin{aligned} &H_0 : \mu = dT^{-\eta} \text{ and } \rho = 0 \\ &H_{1,crash} : \mu = K \text{ and } \rho = 0, \end{aligned} \tag{4.13}$$

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 equation 4.1.

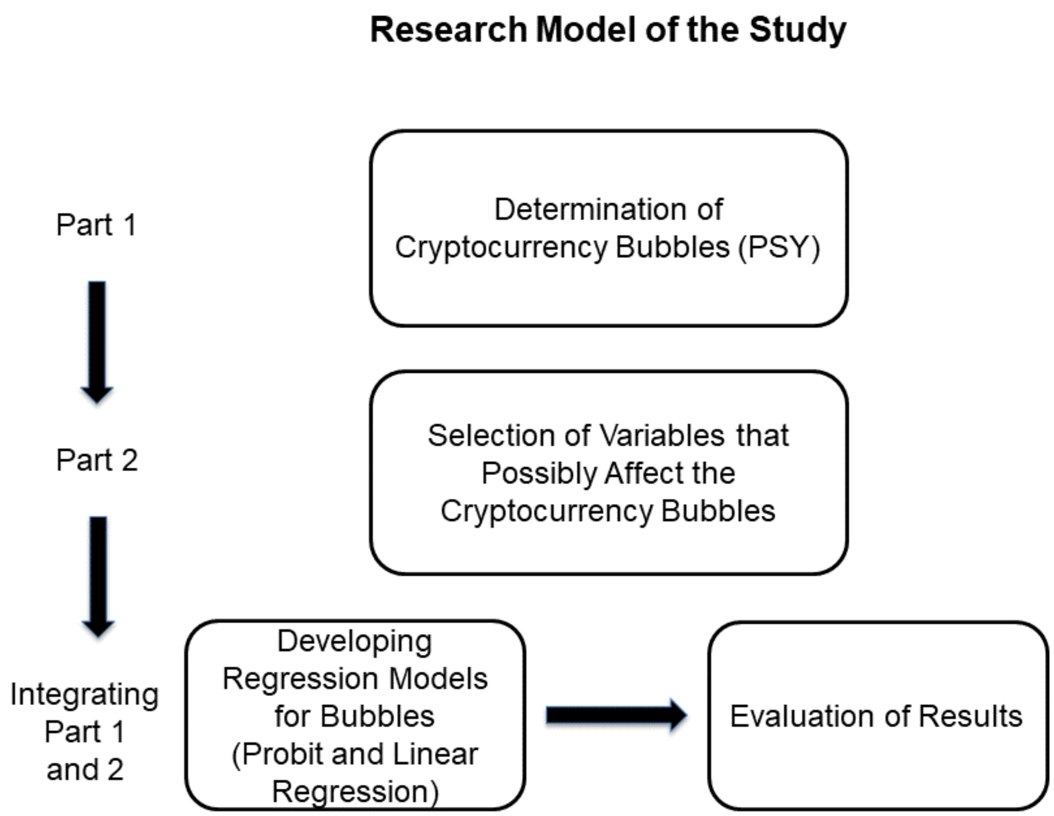
5 Analysis & Results

This section presents the results of our analysis. First, we describe the research model of the study. In subsection 5.2 *Bubble Detection - PSY Test*, we display the results from running the PSY algorithm, and provide some general statistics and graphics of the bubble periods. In subsection 5.3 *Bubble Predictors - Regression Models*, we study bubble predictors through regression models.

5.1 Research Model of the Study

The analysis in this paper consists of two parts that are integrated to evaluate the main issue of this paper, to detect and predict bubbles in cryptocurrencies. The framework for the paper is illustrated in figure 5.1 below.

Figure 5.1: Illustration of the Framework for the Paper

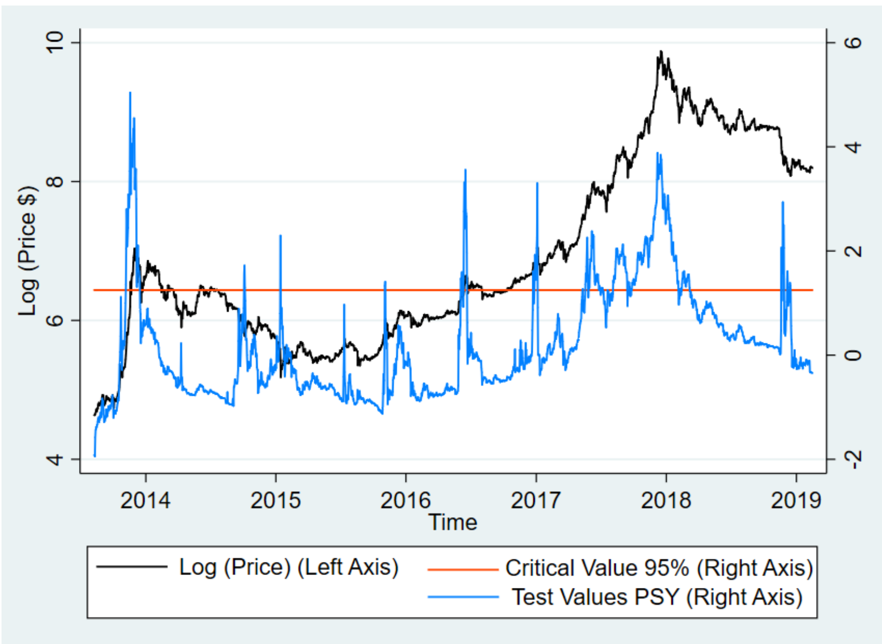


First, we employ the PSY test to identify and date-stamp bubble periods in each cryptocurrency separately. Then we investigate variables that can possibly predict explosive periods in the cryptocurrency prices. Thereafter we develop regression models to study the relationships between the chosen predictors and the cryptocurrency bubbles. In the probit models we use a dummy variable as dependent variable. The variable is generated by giving the value 1 to the bubble dates and the value 0 to the dates where no explosive behavior is observed. In the linear regression models we use the PSY statistic[3] as dependent variable. Finally, we evaluate the results.

5.2 Bubble Detection - PSY Test

The results from application of the PSY algorithm show that there have been several bubbles in each of the cryptocurrencies investigated. Figure 5.2 illustrates the PSY test, when applied to the logarithm of the Bitcoin price (represented by the 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 occur when the PSY test values, illustrated by the blue line, exceeds the critical value. Evidently, there have been numerous bubble periods in Bitcoin in the observed sample.

Figure 5.2: PSY Test of Bitcoin Bubbles



[3]As defined in the PSY methodology section, the PSY values are the suprema of the ADF statistics (generated by the algorithm) for each observation.

The PSY procedure has also been performed for the seven other cryptocurrencies too. Table 5.1 provides descriptive statistics of the generated PSY statistics and the bootstrapped 95% critical values for each of the cryptocurrencies.

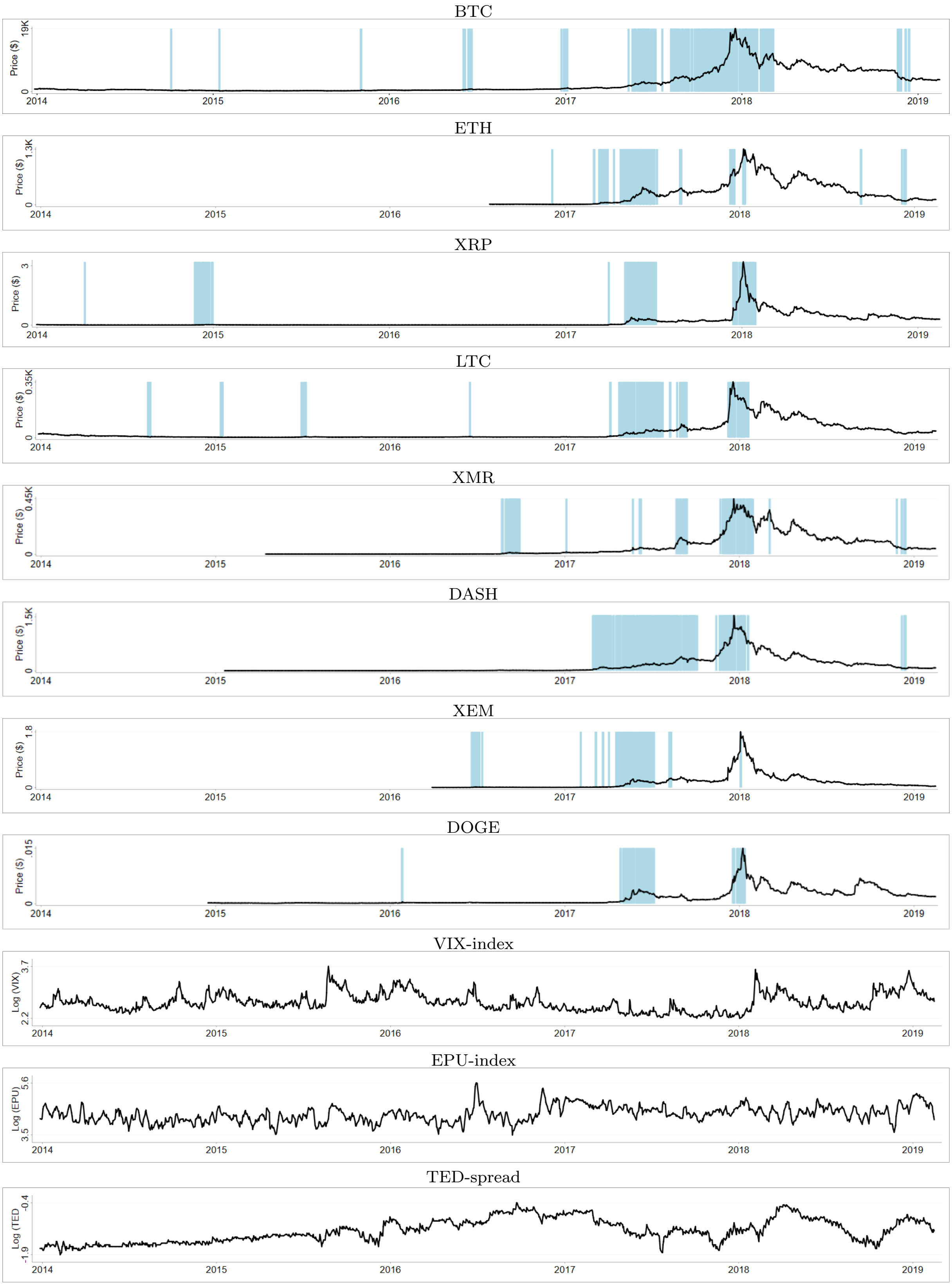
Table 5.1: Descriptive Statistics of the PSY Values

	Mean	St. Dev.	Min	Max	Skew	Kurtosis	95% CV
BTC	0.185	1.049	-1.945	5.044	1.250	1.461	1.251
ETH	0.314	0.961	-2.823	4.572	0.781	1.148	1.167
XRP	-0.152	1.090	-1.885	8.005	1.759	5.568	1.330
LTC	-0.233	1.146	-3.049	5.478	0.828	1.021	1.338
XMR	0.022	0.898	-2.300	3.297	-0.055	-0.408	1.171
DASH	0.015	1.085	-1.631	4.770	0.958	0.680	1.368
XEM	0.021	1.155	-2.753	4.262	0.185	1.776	1.244
DOGE	-0.505	1.027	-1.949	6.070	2.046	6.885	1.315

Figure 5.3 illustrates the time-stamped bubble periods of the PSY test and development of the variables measuring uncertainty (VIX-index, EPU-index and TED-spread) employed in the regression models. For all of the studied cryptocurrencies, we detect 925 days of explosiveness in total. Most of the explosive periods last only for a few days, with the exception of some extensive long-lived bubbles. The short-lived bubbles occur at different time periods for the individual cryptocurrencies. The long-lived bubbles coincide to a greater extent compared to the short-lived bubbles.

Figure 5.3: Bubble Periods in Cryptocurrencies and Uncertainty Variables

Coloured 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 line starts where the dataset of prices begins for the individual cryptocurrency and ends at February 15, 2019. The black lines for the uncertainty variables VIX-index, EPU-index and TED-spread display their historical development.



The prices for all cryptocurrencies studied in this paper increased dramatically during 2017. As can be seen from figure 5.3, the PSY algorithm reveals that there were bubbles in most of the cryptocurrencies in large parts of 2017. Especially Bitcoin exhibit long-lived bubble periods in 2017 and 2018. The date-stamped bubble periods for each cryptocurrency ended some time after the price collapse in January 2018. Notably, the price decline seems to coincide with a substantial increase in the VIX-index. By February 15, 2019, the analyzed cryptocurrencies declined on average 90.4% from their peak in December 2017/January 2018 (see table 5.2).

Table 5.2: Price Decline from Peak for Each Cryptocurrency

The table provides the price decline from all-time high to February 15, 2019, for the eight cryptocurrencies.

Cryptocurrency	**Pric decline from peak**
BTC	81.4 %
ETH	91.3 %
XRP	91.1 %
LTC	88.1 %
XMR	89.9 %
DASH	94.9 %
XEM	97.7 %
DOGE	88.8 %
Average	90.4 %

An overview of bubble periods is provided in table 5.3. Panel A specifies the number of bubble days, where BTC and DASH display the highest number of total bubble days with 193 days and 188 days, respectively. Most bubble days occurs in 2017. DASH had the highest frequency of bubble days in 2017 (174 days). Panel B indicates that the percentage of days with explosiveness is higher in 2017 compared with other years. The explosive periods occured more in DASH (10.3% of days with explosiveness in the time period 2015-2019) and less in DOGE (2.9% of days with explosiveness in the time period 2014-2019) compared to the other cryptocurrencies.

Table 5.3: Statistics of Bubble Periods

Panel A specifies the number of bubble days for the individual cryptocurrencies. Panel B provides the % of days with explosiveness. Total % is the average % of days with explosiveness over the 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 %

5.3 Bubble Predictors - Regression Models

Having applied the PSY framework, we generate the PSY statistics for each of the cryptocurrencies. We then analyze the results by performing various regressions in order to evaluate which variables can predict cryptocurrency bubbles. We estimate both probit models and regular linear regression models. First, we estimate panel models with all cryptocurrencies in the same sample. Second, we estimate models for each cryptocurrency separately.

The two dependent variables (bubble dates dummy and PSY test statistic) applied in the regressions do to some extent measure the same property, as they both are derived from the PSY statistics. As described in subsection 2.1, previous studies have shown that there are correlations between the 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 predictor vector $x_{i,t-1}$ for the panel regression models has been established as:

$$\begin{aligned} x_{i,t-1} = [&\mathrm{Google}_{i,t-1}, \mathrm{Volatility}_{i,t-1}, \mathrm{Transactions}_{i,t-1}, \mathrm{Volume}_{i,t-1}, \mathrm{EPU}_{i,t-1}, \mathrm{VIX}_{i,t-1}, \\ &\mathrm{TED}_{i,t-1}], \end{aligned} \tag{5.1}$$

where $\mathrm{Google}_{i,t-1}$, $\mathrm{Volatility}_{i,t-1}$, $\mathrm{Transactions}_{i,t-1}$, and $\mathrm{Volume}_{i,t-1}$ are cryptocurrency specific variables, and $\mathrm{EPU}_{i,t-1}$, $\mathrm{VIX}_{i,t-1}$ and $\mathrm{TED}_{i,t-1}$ can be categorized as uncertainty variables for individual cryptocurrencies, $i = \mathrm{BTC}, \mathrm{ETH}, \ldots, \mathrm{DOGE}$.

The predictor vector for the time series regression models can be stated as:

$$\begin{aligned} x_{t-1} = [&\mathrm{Google}_{t-1}, \mathrm{Volatility}_{t-1}, \mathrm{Transactions}_{t-1}, \mathrm{Volume}_{t-1}, \mathrm{EPU}_{t-1}, \mathrm{VIX}_{t-1}, \\ &\mathrm{TED}_{t-1}], \end{aligned} \tag{5.2}$$

where the variables have the same interpretation as in predictor vector 5.1, only that we examine one cryptocurrency at a time.

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 the values 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 cryptocurrency (bubble phase) and zero when below (no bubble):

$$\mathrm{BUB}_{i,t} = \begin{cases} 1, & \text{if } PSY_{i,t}(r_0) > cv_{i,t}(\beta_T) \\ 0, & \text{if } PSY_{i,t}(r_0) < cv_{i,t}(\beta_T) \end{cases} \tag{5.3}$$

$$\mathrm{BUB}_{t} = \begin{cases} 1, & \text{if } PSY_{t}(r_0) > cv_{t}(\beta_T) \\ 0, & \text{if } PSY_{t}(r_0) < cv_{t}(\beta_T) \end{cases} \tag{5.4}$$

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

$$P(\mathrm{BUB}_{i,t} = 1) = \Phi(\beta x_{i,t-1} + \nu_i), \tag{5.5}$$

$$P(\mathrm{BUB}_t = 1) = \Phi(\beta x_{t-1}), \tag{5.6}$$

where $\Phi(\cdot)$ is the cumulative distribution function. In the panel probit models, $x_{i,t-1}$ is the vector of lagged predictors (equation 5.1) in cryptocurrency $i = \mathrm{BTC}, \mathrm{ETH}, \ldots, \mathrm{DOGE}$ and $\nu_i \overset{iid}{\sim} N(0, \sigma_\nu^2)$. x_{t-1} is a vector of lagged predictors (equation 5.2) in the time series probit models.

The linear regression models use the generated PSY statistic as dependent variable. The PSY statistic is the supremum of the estimated ADF statistics for the respective observation, generated by the algorithm (as defined in equation 4.6). The estimated models try to predict what variables affects this statistic, independent of the generated critical value.

The linear panel regression model is specified as follows:

$$\begin{aligned} PSY_{i,t}(r_0) = \beta_0 &+ \beta_1 \mathrm{Google}_{i,t-1} + \beta_2 \mathrm{Volatility}_{i,t-1} + \beta_3 \mathrm{Transactions}_{i,t-1} + \beta_4 \mathrm{Volume}_{i,t-1} \\ &+ \beta_5 \mathrm{EPU}_{i,t-1} + \beta_6 \mathrm{VIX}_{i,t-1} + \beta_7 \mathrm{TED}_{i,t-1} + \epsilon_{i,t}, \end{aligned} \tag{5.7}$$

while the linear time series regression model is specified as follows:

$$\begin{aligned} PSY_t(r_0) = \beta_0 &+ \beta_1 \mathrm{Google}_{t-1} + \beta_2 \mathrm{Volatility}_{t-1} + \beta_3 \mathrm{Transactions}_{t-1} + \beta_4 \mathrm{Volume}_{t-1} \\ &+ \beta_5 \mathrm{EPU}_{t-1} + \beta_6 \mathrm{VIX}_{t-1} + \beta_7 \mathrm{TED}_{t-1} + \epsilon_t. \end{aligned} \tag{5.8}$$

The following subsections present the regression results of our models. An overview of the models used are presented in table 5.4. The models includes samples from either all or individual cryptocurrencies. Due to some autocorrelation and heteroscedasticity, we apply models more suitable to deal with this issue.

The panel probit models is estimated with random effects and cluster robust standard errors, 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 time series models are 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).[4] All variables are stationary.

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, 1974).

Table 5.4: Summary of Regression 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
Individual	Bubble dummy	Probit with optimal lag Newey-West standard errors
Individual	PSY statistics	OLS with optimal lag Newey-West standard errors

5.3.1 Panel Regressions: All Cryptocurrencies

The regression results from the probit panel regressions and the PSY statistic panel regressions is provided in table 5.5 and table 5.6, respectively. We use panel regressions to analyze the variables' predictive effects across cryptocurrencies. We estimate both single variable models, termed *univariate models*, and models which include all variables studied, termed *complete models*. The univariate models investigate one explanatory variable at a time, for each cryptocurrency.

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

For the probit models, positive coefficients indicate a higher predicted probability. An increase in the variable is thus associated with a higher likelihood of bubbles. A negative coefficent would similarly decrease the likelihood of bubbles. In the linear models, an increase in a variable with a positive coefficient indicates a higher predicted PSY statistic. Given the definition of the PSY statistics, this can imply that there is a tendency of changes in price to be affected by the previous observed price level. A negative coefficient implies a lower PSY statistics, indicating an opposite effect.

The probit models and the linear models essentially identify the same predictors of bubbles, with a few differences. In the estimated univariate models, almost all the independent variables exhibit highly significant associations, except the EPU-index and TED-spread, in table 5.6 (linear panel regression).

Volatility exhibits positive effects in all panel models. Thus, an increase in this variable raises the likelihood of bubble states. The research of Bekiros et al. (2017) states that herding behavior[5] is usually more prevalent in periods of excessive volatility, which might make volatility a natural property of bubbles. Volume exhibits a positive relationship with the dependent variables in all models. Thus, increases in volume corresponds with a higher likelihood of bubbles. This can possibly be explained by theories such as rational bubbles[6] 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 findings of Blau (2017), which does not find a connection between speculative trading and extreme market behavior. Google searches and transactions have positive effects on bubble behavior in all panel models. Though, these effects are not significant when other variables are controlled for in the complete probit panel model. We suspect that both these variables are closely connected to the trading volume, which can explain why the effects disappear in the complete probit panel model, when trading volume is included. To an extent the three variables; volume, Google searches and

[5]In behavorial finance, herding behavior is when an investor's decisions is based on the trend of past trades (Avery and Zemsky, 1998).

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

transactions are similar, as they are related to the market demand for cryptocurrencies. The fact that they demonstrate the same direction of effects can support this suspicion.

When it comes to uncertainty variables, the VIX-index is significant in both the probit models and the linear models. An increase in the VIX-index demonstrates negative relationships with bubbles in all panel models. An increase in the VIX-index usually implies higher volatility in the stock market. The negative relationship might be related to the "safe haven" property, which is discussed by Bouri et al. (2017). The referenced paper indicates a negative correlation between the volatility of Bitcoin and the VIX-index. The EPU-index is positive and significant for the probit models, but not for the linear models. This implies that the probability of bubbles is higher when US political uncertainty increases. As stated in our literature review, Demir et al. (2018) identifies that the EPU-index is generally negatively associated with future returns, but exhibits a positive relationship with high and low quantiles of returns. This can possibly explain why the EPU-index is only positively associated with bubbles in the probit model, which only captures extreme PSY values. The TED-spread is significant and negatively associated with bubbles in the univariate probit model, but the effect disappears when other variables are included in the complete model.

Considering the measures of fit metrics of the panel probit models, the McFadden R-squared shows that the different models display varying ability to predict bubbles. The VIX-index marginally displays the highest value and the TED-spread displays the lowest value. For the linear panel models, the R-squared is generally very low for the univariate models, which indicates that the estimated models predict only a small share of the variance in the PSY statistics. In the complete model, R-squared is considerably higher, but still relatively low. It is able to predict 8.3% of the total variance in the PSY statistics.

Table 5.5: Probit Regression Results - Panel Regression

The dependent binary variable $\text{BUB}_{i,t}$ only takes the values 1 (explosive dates) and 0 (non-explosive dates). Independent variables are described in the data section. The sample includes *all cryptocurrencies* (see table 3.1 for the individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. All the reported estimates are coefficients with corresponding cluster-robust standard errors, by cryptocurrency. The McFadden R-squared for this table has been calculated manually.

	Dependent variable: $\text{BUB}_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Google}_{i,t-1}$	0.894***							0.180
	(0.251)							(0.276)
$\text{Volatility}_{i,t-1}$		19.94***						7.241***
		(1.671)						(1.997)
$\text{Transactions}_{i,t-1}$			0.368***					0.129
			(0.103)					(0.0791)
$\text{Volume}_{i,t-1}$				0.920***				0.651***
				(0.0891)				(0.124)
$\text{EPU}_{i,t-1}$					0.618***			0.432**
					(0.115)			(0.218)
$\text{VIX}_{i,t-1}$						-2.724***		-1.441***
						(0.381)		(0.359)
$\text{TED}_{i,t-1}$							-0.557***	0.152
							(0.156)	(0.221)
Intercept	-1.347***	-2.309***	-1.626***	-13.05***	-3.971***	5.764***	-1.855***	-8.190***
	(0.0866)	(0.129)	(0.167)	(1.868)	(0.510)	(0.975)	(0.179)	(1.464)
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

Table 5.6: Linear Regression Results - Panel Regression

The dependent variable is the PSY statistic. Independent variables are described in the data section. The sample includes *all cryptocurrencies* (see table 3.1 for the individual time spans). *, ** and *** represents significance at the 10%, 5% and 1% level, respectively. The coefficients are are estimated by Prais-Winsten regression. 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

5.3.2 Time Series Regressions: Individual Cryptocurrencies

The results from the estimated probit regressions and linear regressions for individual cryptocurrencies are shown in table 5.7 and table 5.8, respectively.[7] We study the cryptocurrencies separately to examine whether the predictive effects seem to be cryptocurrency-dependent or consistent across cryptocurrencies.

Similar to the results of the panel regressions, volatility and volume exhibit positive associations with bubbles for most cryptocurrencies. This implicates that increases in volatility or volume corresponds with a higher likelihood of bubbles, as demonstrated

[7]We have also estimated univariate regressions for the individual cryptocurrencies, which are included in the appendix (table A0.1 and A0.2).

in the panel regression models. Google searches shows varying direction of effects and predictive ability depending on the cryptocurrency studied. It displays positive effects for BTC and ETH in both the linear models and the probit models. On the other hand, Google searches are negatively associated with bubbles for DASH in both time series regression models and XMR in the linear regression models. The variation in effects can possibly be explained by the differences in total market value for the cryptocurrencies. Transactions generally demonstrates a positive relationships with bubbles. The exceptions are transactions for BTC in both time series models and XRP in the linear regression model, where the variables display a negative direction of effect. A possible explanation for this exception is that BTC and XRP are among the cryptocurrencies which ranks highest in terms of total market value. An increase in transactions for these cryptocurrencies can imply a higher degree of use as means of exchange. This can 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 in the panel regression models.

The examined uncertainty variables; EPU-index, VIX-index and TED-spread shows varying relationships with bubble states when it comes to direction of effects. As the panel models indicate, the EPU-index is positively associated with bubbles. Though, this relationship seems to be very dependent on the cryptocurrency studied, as many models fails to demonstrate an significant effect. The VIX-index are in general negatively associated with bubbles across cryptocurrencies, similarly to the panel regression models. In the probit models, BTC, DASH and DOGE is negatively associated with bubbles, which might explain why the panel models exhibit the same effects. In the linear models, all the effects are negative, but not significant for LTC and DOGE. ETH is an exception in terms of the predictive effect for the VIX-index. It is positively associated with bubble states in both time series regression models. The TED-spread shows a positive relationship with bubbles for BTC in both the time series models, as opposed to the panel models, where there are weak indications of an effect. In the probit models, the TED-spread shows no significant effects for other cryptocurrencies. In the linear models, the TED-spread shows differing effects in terms of direction and significance.

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

Table 5.7: Probit Regression Results - Time Series Regressions

The dependent binary variable BUB_t only takes the values 1 (explosive dates) and 0 (non-explosive dates). Independent variables are described in the data section. The sample includes *all* dates for the *respective cryptocurrency* (see table 3.1 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_t							
	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE
$Google_{t-1}$	1.406***	1.464***	0.0818	0.438	0.798*	-0.509**	-0.00304	0.239
	(0.392)	(0.336)	(0.273)	(0.322)	(0.458)	(0.251)	(0.0860)	(0.333)
$Volatility_{t-1}$	21.31***	8.548*	4.692	5.828*	12.93***	10.21*	6.168*	-1.956
	(5.462)	(4.657)	(2.907)	(3.214)	(4.083)	(5.514)	(3.632)	(4.435)
$Transactions_{t-1}$	-0.271**	0.448***	0.0083	0.277**	0.154**	0.178	0.312***	-0.135
	(0.106)	(0.165)	(0.123)	(0.112)	(0.0730)	(0.161)	(0.120)	(0.167)
$Volume_{t-1}$	0.903***	0.147	0.952***	0.532***	0.543***	0.861***	0.280***	1.394***
	(0.211)	(0.123)	(0.162)	(0.140)	(0.179)	(0.315)	(0.0849)	(0.253)
EPU_{t-1}	-0.566	0.445	0.351	-0.036	0.243	1.592***	0.601	1.240***
	(0.414)	(0.401)	(0.570)	(0.461)	(0.441)	(0.414)	(0.366)	(0.428)
VIX_{t-1}	-1.566**	1.060**	-1.062	0.539	-0.43	-2.996***	-0.822	-2.793**
	(0.641)	(0.447)	(0.708)	(0.645)	(0.770)	(1.155)	(0.616)	(1.324)
TED_{t-1}	0.879**	-0.327	-0.876	0.305	0.593	-0.0515	-0.471	0.790
	(0.345)	(0.459)	(0.574)	(0.416)	(0.512)	(0.600)	(0.391)	(0.510)
Intercept	-9.335**	-9.919***	-14.06***	-8.659***	-6.432**	-9.919***	-6.093**	-16.34***
	(3.754)	(3.123)	(3.266)	(3.230)	(3.004)	(3.521)	(2.457)	(3.704)
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

Table 5.8: Linear Regression Results - Time Series Regressions

The dependent variable is the PSY statistic. Independent variables are described in the data section. The sample includes *all* dates for the *respective cryptocurrency* (see table 3.1 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

5.3.3 Summary of Regressions Results

In general, the cryptocurrency-specific variables volatility and trading volume demonstrate similar and consistent results for 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 demonstrates varying effects depending on the cryptocurrency 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 show 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 demonstrates 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.

5.3.4 Predictive Ability of Models

Table 5.9 presents a comparison between the time series models' ability to predict the estimated bubble dates from the PSY framework. The models utilized to test the predictive ability are the complete models displayed in table 5.7 and table 5.8. 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[8] (generated by the PSY framework) for the respective cryptocurrency.

Table 5.9: Predictive Ability of Models

% True Bubble Days Predicted is the share of bubble days detected by the PSY framework which the respective model is able to predict. *% Correct Predictions* is the share of predicted bubble days, which are true PSY bubble days.

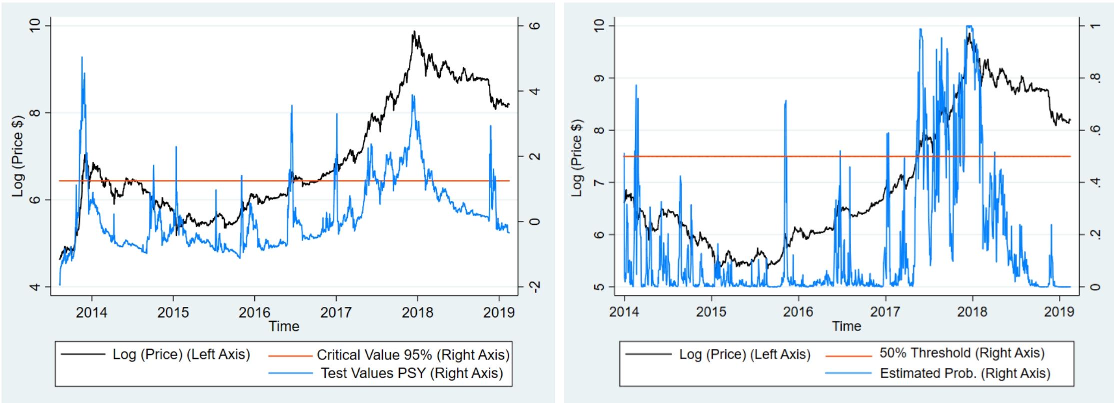
	BTC	ETH	XRP	LTC	XMR	DASH	XEM	DOGE	Average
PSY Detected Bubbles Days	193	91	100	118	92	188	79	64	
Panel A: Probit regression									
Predicted Bubble Days	144	76	69	94	53	169	46	54	
% True Bubble Days Predicted	58.6 %	65.9 %	53.00 %	62.7 %	46.7 %	75.5 %	40.51 %	70.3 %	59.2 %
% Correct Predictions	78.5 %	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 %

[8] The critical values can be found in table 5.1.

The results in table 5.9 indicate that the probit models are generally superior to the linear regression models, except when it comes to *% True Bubble Days Predicted* for XRP and XEM. These results contradict our a priori suspicion that the linear models would perform better than the probit models. We suspected that by trying to predict the underlying PSY values, the consequence would be improved predictive accuracy.

Figure 5.4: Comparison of Linear Model and Probit Model on BTC

The figure to left represents the linear model, while the figure to the right represents the probit model.



We speculate that the reason the probit models are superior to the linear models are 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, we suspect that the extreme values are better fitted in 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.

An illustration of the two different regression models applied on Bitcoin is given in figure 5.4. In this figure we see that the blue lines for the models (*Test Values PSY* for the linear model and *Estimated Probability* for the probit model) are different. The probit model estimates more extreme values than the linear model, which can be an indicator of why the probit model is superior to the linear model. If the objective is to predict bubbles, the results indicates that the probit approach is preferred. If the objective is to analyze the tendency of the price to be affected by the previous observed prices, the PSY approach might be better.

6 Conclusion

In this paper, we examine variables that have predictive ability of bubbles in cryptocurrency prices. We utilize the novel PSY framework and explore variables' predictive effects. The ability to predict bubbles can be an important contribution to market monitoring and in the understanding of price dynamics for cryptocurrencies. To our knowledge, this is the first study to examine predictors of PSY detected bubbles in cryptocurrencies. We use the price determinants research, presented in subsection 2.1, to identify which variables have an impact on the cryptocurrency prices and use this as a basis for our selection of predictors in the regression models.

Similar to the research presented in subsection 2.3, our results from running the PSY test reveals multiple bubble periods in all cryptocurrencies. Recently published papers, like Corbet et al. (2018) and especially Bouri et al. (2018), identifies long-lived bubble periods in multiple cryptocurrencies during 2017 and 2018. These findings coincide to a large extent with our results, which also detects extensive cryptocurrency bubbles in the same periods. Furthermore, Bouri et al. (2018) finds that particularly Bitcoin demonstrates extensive price explosivity, which is also in line with our findings.

The conclusion of our paper is that several variables demonstrate predictive ability of cryptocurrency bubbles. Of cryptocurrency-specific variables, volatility and volume are distinctly associated with bubble behavior across cryptocurrencies. Google trends and transactions mostly demonstrates positive relationships with bubbles, but the effects are dependent on the cryptocurrency studied and type of regression model. For the uncertainty variables, the VIX-index generally exhibits a negative association with bubbles. The EPU-index demonstrates positive relationships with bubbles, but the effects are dependent on the cryptocurrency investigated and type of regression model. The TED-spread exhibits a more ambiguous relationship with bubbles. We find that the probit models demonstrates better predictive ability, compared to the linear models. In summary, many variables exhibit predictive potential of bubbles, where trading volume, volatility and the VIX-index appear to be particularly prevalent.

References

Aalborg, H. A., Molnár, P., and de Vries, J. E. (2018). What can explain the price, volatility and trading volume of Bitcoin? *Finance Research Letters*.

Avery, C. and Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. *American economic review*, pages 724–748.

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

Bekiros, S., Jlassi, M., Lucey, B., Naoui, K., and Uddin, G. S. (2017). Herding behavior, market sentiment and volatility: Will the bubble resume? *The North American journal of economics and finance*, 42:107–131.

Bijl, L., Kringhaug, G., Molnár, P., and Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, 45:150–156.

Blanchard, O. and Watson, M. (1982). Bubbles, Rational Expectations and Financial Markets. NBER Working Papers 0945, National Bureau of Economic Research, Inc.

Blau, B. M. (2017). Price dynamics and speculative trading in Bitcoin. *Research in International Business and Finance*, 41:493–499.

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

Boudt, K., Paulus, E. C., and Rosenthal, D. W. (2017). Funding liquidity, market liquidity and TED spread: A two-regime model. *Journal of Empirical Finance*, 43:143–158.

Bouri, E., Azzi, G., and Dyhrberg, A. H. (2017). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics: The Open-Access, Open-Assessment E-Journal*, 11(2017-2):1–16.

Bouri, E., Shahzad, S. J. H., and Roubaud, D. (2018). Co-explosivity in the cryptocurrency market. *Finance Research Letters*.

Caspi, I. (2017). Rtadf: Testing for Bubbles with EViews. *Journal of Statistical Software, Code Snippets*, 81(1):1–16.

CBOE (2019). White Paper Cboe Volatility Index. Retrieved May 29, 2019, from https://www.cboe.com/micro/vix/vixwhite.pdf.

Cheah, E.-T. and Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130:32–36.

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

Choi, H. and Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, pages 1–5.

Choi, H. and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88:2–9.

Ciaian, P., Rajcaniova, M., and Kancs, d. (2016). The economics of BitCoin price formation. *Applied Economics*, 48(19):1799–1815.

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

Da, Z., Engelberg, J., and Gao, P. (2011). In search of Attention. *The Journal of Finance*, 66(5):1461–1499.

Datta, D. and Du, W. (2012). Nonparametric HAC estimation for time series data with missing observations. *FRB International Finance Discussion Paper*, (1060).

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

Diba, B. T. and Grossman, H. I. (1988). Explosive rational bubbles in stock prices? *The American Economic Review*, 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.

Fry, J. and Cheah, E.-T. (2016). Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis*, 47:343–352.

Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of business*, pages 67–78.

Gebka, B. and Wohar, M. E. (2013). Causality between trading volume and returns: Evidence from quantile regressions. *International Review of Economics & Finance*, 27:144–159.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M., and Siering, M. (2014). Bitcoin - Asset or currency? Revealing users' hidden intentions. *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*.

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. *The stata journal*, 7(3):281–312.

Kim, N., Lučivjanská, K., Molnár, P., and Villa, R. (2018). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*.

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific reports*, 3:3415.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*, 10(4):e0123923.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*, pages 105–142.

Molnár, P. (2012). Properties of range-based volatility estimators. *International Review of Financial Analysis*, 23:20–29.

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

Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787.

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

Phillips, P. C. (2017). Detecting Financial Collapse and Ballooning Sovereign Risk. Technical report, Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.

Phillips, P. C. and Shi, S. (2018). Real Time Monitoring of Asset Markets: Bubbles and Crises. Technical report, Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C., Shi, S., and Yu, J. (2012). Testing for multiple bubbles. Cowles Foundation Discussion Paper No. 1843. Yale University.

Phillips, P. C., Shi, S., and Yu, J. (2013). Testing for Multiple Bubbles 1: Historical Episodes of Exuberance and Collapse in the SP 500. *Research Collection School Of Economics*, pages 1–47.

Phillips, P. C., Shi, S., and Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4):1043–1078.

Phillips, P. C., Shi, S., and Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, 56(4):1079–1134.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1):201–226.

Su, C.-W., Li, Z.-Z., Tao, R., and Si, D.-K. (2018). Testing for multiple bubbles in bitcoin markets: A generalized sup ADF test. *Japan and the World Economy*, 46:56–63.

Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In *Handbook of digital currency*, pages 31–43. Elsevier.

Appendix

The regression results from the probit univariate regressions and the PSY statistic univariate regressions is provided in table A0.1 and A0.2, 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 A0.1: 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). Independent variables are described in the data section. The sample includes *all* dates for the *respective cryptocurrency* (see table 3.1 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 A0.2: Linear Regression Results - Univariate Time Series Regressions

The dependent variable is the PSY-statistic. Independent variables are described in the data section. The sample includes *all* bubble dates for the *respective cryptocurrency* i (see table 3.1 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: $PSY_{r^\dagger}(r_0)_{i,t}$							
	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)
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