Highlights

Sentiment, Google Queries and Explosivity in the Cryptocurrency Market

- We propose a Backward Superior Covariate-Augmented Dickey-Fuller (BSCADF) test including sentiment as a covariate for detecting explosive financial time series behaviour;
- We exploit the information derived from a large set of news and Google Search Indices to detect the presence of speculative bubbles in cryptocurrency prices;
- The BSCADF test statistics tends to significantly diverge from its univariate counterpart during price surges;
- Evidence shows how investors' sentiment plays a more determinant role if compared to Google queries in providing an early warning bubble signal.

Sentiment, Google Queries and Explosivity in the Cryptocurrency Market

Abstract

The lack of fundamental values in the cryptocurrency market paves the way for the rise of unprecedented speculative bubble phenomena, which are often associated with alternating phases of investors' fear and greed. We propose to exploit the information derived from a large set of cryptocurrency news and Google Search Indices to detect and, possibly, anticipate the presence of speculative bubbles in cryptocurrency prices. This is done by means of a Backward Superior Covariate-Augmented Dickey-Fuller (BSCADF) test, which allows us to explicitly account for market sentiment when testing the presence of an explosive root in cryptocurrency prices. Our results show that the covariate test statistics does significantly diverge from the traditional one in concomitance with price surges, highlighting the ability of sentiment to foresee speculative bubble occurrences. We also show how a polarised version of investors' sentiment plays overall a more determinant role, if compared to news volume and Google queries, in providing an early warning signal of market bubble episodes in cryptocurrencies.

Keywords: Bitcoin, Cryptocurrency Market, Sentiment, Speculative Bubbles, Google Trends, Big Data Application

1. Introduction

Since cryptocurrencies were conceived in first place under the advent of Bitcoin (Nakamoto et al., 2008), research on this topic has been prospering in a highly multidisciplinary way. The emergence of fields such as machine learning (ML), deep learning (DL), Big Data Analytics (BDA), eXplainable artificial intelligence (XAI) and Automated Trading Systems (ATS) have, at the same time, brought new challenges to both academics and practitioners calling for rapid knowledge advances in several disciplines, particularly computer science, from a wide variety of perspectives - see for instance Huang et al. (2019); de Souza et al. (2019); Lee (2019); Jaquart et al. (2021); Yi et al. (2021); Corea (2016). On the other hand, methods pertaining to such innovative fields often lack a sound and robust state-of-the-art econometric and statistical framework, so to develop models accurately, conduct appropriate inference, and - most importantly - improve model performances. It is against this background that we propose a novel sentiment-based testing procedure for cryptocurrency explosiveness, which explicitly takes into account for any possible price predictors. We develop our empirical application upon a large set of financial news, in order to anticipate speculative bubble occurrences in cryptocurrency prices.

Statistical properties of cryptocurrencies, price connectedness and their usage for investment purposes have been largely analysed in applied research, due to the appealing and unique market features of these novel financial instruments. In Corbet et al. (2018b) the authors investigate the dynamic spillovers of cryptocurrencies with other financial assets, demonstrating that the two categories of financial instruments are not significantly linked. Similarly, Giudici and Pagnottoni (2019, 2020) explore the dynamic connectedness of Bitcoin exchanges and provide evidence on their relative importance in transmitting information on the dynamics of the fundamental (unobserved) Bitcoin price. Their findings are linked, from a lead-lag relationship perspective, to those reached in the field of price discovery on Bitcoin (and cryptocurrency) exchanges – see Brandvold et al. (2015), Pagnottoni and Dimpfl (2019) and Dimpfl and Peter (2020). Bouri et al. (2019b) and Resta et al. (2020) examine the performances of a number of different technical trading rules in cryptocurrency markets and provide evidence on the fact that moving-average based strategies are, across all alternatives analysed, the best performing trading strategies. The latter analyses could be further extended by exploiting frameworks such as that developed by Huang and

Huang (2020).

A fluorishing literature has started investigating the cryptocurrency bubble phenomena, mostly focusing on the analysis of cryptocurrency price dynamics from a univariate time series perspective. Cheung et al. (2015) analyse the presence of bubbles in Bitcoin prices from 18 July 2010 to 18 February 2014 by employing the Phillips et al. (2011) methodology. Evidence from the generalised Supremum Augmented Dickey Fuller (GSADF) statistics supports the existence of three large bubbles, most of which do not last for more than a few days period. Fry and Cheah (2016) investigate speculative bubble occurrences on the Bitcoin Coindesk Index over the period from 18 July 2010 to 17 July 2014 by means of a price model which is composed of a Wiener process and a jump process so to control for constant intrinsic rate of return and intrinsic level of risk. They show that an explosive behaviour exists in the Bitcoin market, and the random walk hypothesis is rejected. In Corbet et al. (2018a) the authors analyse Bitcoin and Ethereum data from 9 January 2009 to 9 November 2017 and from 7 August 2015 until 9 November 2017, respectively, so to identify intrinsic bubbles, herding behaviour and time-varying fundamentals of the cryptocurrency market. They employ a dynamic econometric approach using the Supremum (SADF), the Generalised Supremum (GSADF) and the Backward Supremum (BSADF) Augmented Dickey-Fuller specifications. They find evidence of a Bitcoin bubble at the turn of the year from 2013 and 2014 and, in line with the extant literature, that the occurrence of speculative bubbles in the digital currencies under examination did not last for a long time. We refer the reader to Kyriazis et al. (2020) for a systematic review on cryptocurrency bubbles.

More recently, Bouri et al. (2019a) have studied co-explosivity on data about Bitcoin, Ripple, Ethereum, Litecoin, Nem, Dash and Stellar that over the period from 7 August 2015 to 31 December 2017 in order to study co-explosivity in their markets. They find, among others, that: Bitcoin's explosivity tends to lower Ripple's explosivity; as price increases in Ethereum, Litecoin, Nem and Stellar, market values of Ripple market tend to consequently adjust; also digital currencies with small market capitalisation are in some cases influential in transmitting price shocks to others. Similarly, the study from Agosto and Cafferata (2020) employs price data on Bitcoin, Ethereum, Ripple, Litecoin and Stellar, over the period 1 January 2017 – 31 December 2018, to examine co-explositivy of cryptocurrency time series, covering the price surge period at the end of 2017. Partly contrasting with Bouri et al. (2019a), they find Ethereum particularly important in determining explosivity of other cryptocurrencies, and Bitcoin to be quite highly impacted by the dynamics of the other cryptocurrencies' prices. A recent contribution from Gronwald (2021) analyses the prices of Bitcoin, Ripple, Ethereum, and Litecoin expressed in US Dollars and, for the price of the latter three, additionally expressed in Bitcoin. Additionally, alternative approaches based on continuous time stochastic models for Bitcoin dynamics, which depend upon a market attention factor, i.e. an index based on Google search queries, have been developed by Cretarola and Figà-Talamanca (2019, 2020). Their findings indicate that, when the correlation between the market attention factor and Bitcoin returns is above a threshold, creating a vicious loop between the time series, a speculative bubble tends to emerge.

At the same time, numerous studies analyze the impact of news and investors' sentiment on the dynamics of financial markets. They mainly deal with the statistical and econometric analysis of non conventional data - see, for instance, Bollen et al. (2011), Bordino et al. (2012), Choi and Varian (2012), Feldman (2013), Huang et al. (2015),Cerchiello and Giudici (2016), Cerchiello and Nicola (2018), Scaramozzino et al. (2021) - who all show the added value of textual data and sentiment in economics and finance. In the specific context of cryptocurrency dynamics, prominent examples such as that of Aste (2019) demonstrate how prices affect sentiment and vice-versa, with noticeable differences in intensities and number of significant interactions.

Against this premises, we propose to exploit the information derived from a large set of cryptocurrency news to detect and foresee the presence of speculative bubble phenomena occurring in cryptocurrency prices. In other words, by means of a Backward Superior Covariate Augmented Dickey-Fuller (BSCADF) test, we are able to explicitly model the market sentiment in the context of testing for an explosive autoregressive coefficient in the cryptocurrency price series. The main advantage of the inclusion of the covariate in the model specification is that the power of the test, if compared to its univariate counterpart, can be increased - meaning that the price bubble can be detected more efficiently - without incurring in large size distortions - see Caporale and Pittis (1999).

We apply our methodology to detect explosivity in the price series of three major cryptocurrencies in terms of market capitalization, i.e. Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP), sampled at a daily frequency, over the period ranging from 31 May 2019 to 31 May 2021. We augment the standard recursive ADF procedures with two different categories of covariates: a) sentiment covariates, which consist of synthetic polarised (positive or negative) scores, assigned to a large set of cryptocurrency news, gathered from a variety of financial media sources; b) news volume covariates, namely the relative amount of Google queries for the search terms associated to the names of each cryptocurrency.

We find a clear divergence of the two test statistics with and without covariate, which arises well before price accelerations. As soon as explosivity of the time series comes into place - instead - both tests tend to indicate a rejection of the unit root hypothesis in favour of the alternative of an explosive price autoregressive coefficient. We further provide evidence on the fact that the sentiment indicator, which does take into account for the positive, negative or neutral meaning of cryptocurrency news, seems to act as a better early warning indicator of price surges, if compared to term search volumes.

Our contribution to the literature is twofold. From a methodological viewpoint, we extend the Covariate-Augmented Dickey-Fuller setting for unit root testing (Hansen, 1995) to the framework of bubble detection and explosivity testing (Phillips et al., 2011). In this way, we provide recursive ADF testing procedures which exploit the information of any relevant covariates to detect the explosive behaviour of a target time series. From an empirical viewpoint, to the best of our knowledge, this is the first empirical application of such a testing strategy in the literature on cryptocurrency bubbles. This setup is able to unveil the nexus between cryptocurrency prices, search volumes and investors' sentiment, from a robust econometric viewpoint, providing an early warning signal of speculative bubble phenomena in the market.

The remainder of this paper proceeds as follows. Section 2 outlines our proposed methodology for sentiment-based cryptocurrency explosive behaviour detection. Section 3 illustrates empirical application and outcomes. Section 4 offers some concluding remarks.

2. Methodology

2.1. Testing for price explosivity

From an econometric point of view, one of the main research questions related to cryptocurrencies concerns the possible presence of bubbles in their price. An asset bubble is defined in literature as an extreme price acceleration that cannot be driven by the underlying fundamental economic variables (Case and Shiller (2003); Dreger and Zhang (2013)). The end of this phase,

often referred to as bubble burst, leads to drastic price drops, causing severe losses to investors.
Several recent works provided empirical evidence of the presence of bubbles in the cryptocurrency prices (Fry and Cheah (2016); Corbet and Yarovya (2018)). From a methodological viewpoint, most of them resorted to the right-tailed unit root testing approach based on Augmented Dickey Fuller (Dickey and Fuller (1979)) regression. Indeed, the extremely rapid price increase, to which the definition of financial bubble refers to, can be described by an exponential growth, whose occurrence can be detected through right-sided unit root tests.

Specifically, Phillips et al. (2011) proposed a univariate approach to test for end-of-sample bubbles through estimation of the ADF regression:

$$y_t = \quad + \quad y_{t \quad 1} + \sum_{i=1} \quad {}_i \Delta y_{t \quad i} + \varepsilon_t; \qquad \varepsilon_t \quad N(0; \quad ^2) \tag{1}$$

performed on the full sample of data, where is the intercept and p is the number of lags of the differenced dependent variable Δy_t, chosen through some model selection criteria. A possible alternative specification includes a deterministic trend.
In the standard ADF test, Equation (1) is used to test the null hypothesis $= 1$, corresponding to the random walk process, against the alternative of stationarity (< 1). In the right-tailed unit root test for explosivity proposed by Phillips et al. (2011), the null hypothesis is still the random walk one, but the alternative is > 1, corresponding to an explosive behaviour. The ADF test applied to the full sample can be denoted as $ADF_0^1(p)$.
However, full sample tests for explosivity can be shown to have poor power to detect short duration bubble episodes. To overcome this, Phillips et al. (2015) proposed to perform a sequential estimation of the ADF statistics by recursively applying the test to subsamples of the data.
Specifically, denoted by $ADF(p)_{r_1}^{r_2}$ an ADF test performed on the subsample $t = \lfloor r_1 T \rfloor, \ldots, \lfloor r_2 T \rfloor$, Phillips et al. (2015) proposed the following test statistic:

$$SADF := \sup_{r_2 \in [r_0, 1]} \{ADF(p)_0^{r_2}\} \tag{2}$$

The SADF test is thus the supremum (S) of right-tailed ADF statistics performed on a window of observations starting at $t = 1$ subject to a

minimum sample size br_0Tc. This recursive regression technique was shown to have high power in detecting periodically collapsing bubbles, which frequently occur in empirical market prices.
The same authors highlighted that the SADF test, being based on an expanding window, can lack of power in detecting end-of-sample bubble episodes, due to the relatively higher influence of early observations. Furthermore, financial analysts and authorities are mainly interested in tools for real-time detection of bubbles, so to understand whether a particular observation belongs to a bubble phase in the overall trajectory.

Based on these motivations, Phillips et al. (2015) proposed the following test statistic, based on repeated estimation of the SADF test on a backward expanding sample, where the endpoint of each sample is fixed at r_2 (the sample fraction corresponding to the endpoint of the window) and the start point varies from 0 to $r_2 \quad r_0$ (the sample fraction corresponding to the origination of the window):

$$BSADF_{r_2} := \sup_{r_1 2[0;r_2 \quad r_0]} ADF(p)_{r_1}^{r_2} \tag{3}$$

Thus, at each window end time r_2, the BSADF test statistic is the supremum of the right-tailed ADF statistics computed on all sub-samples ending at date t subject to a minimum sample size br_0Tc.
As it provides a value of the test statistic for each time $t = r_2$, the BSADF test can also be effectively used for date stamping past explosive episodes. In particular, the origination date of a bubble ($\hat{r}_s$) is calculated as the first time in which the calculated BSADF statistic exceeds the critical value, while the first observation following ($\hat{r}_s$) whose BSADF statistic goes back below the critical value is considered as the termination date ($\hat{r}_e$) of the same bubble:

$$\hat{r}_s = \inf_{r_2 2[0;1 \quad r_0]} \quad r_2 : BSADF_{r_2}(r_0) > cv_{r_2} \tag{4}$$

$$\hat{r}_e = \inf_{r_2 2[r_s + \quad T;1]} \quad r_2 : BSADF_{r_2}(r_0) < cv_{r_2} \tag{5}$$

where cv_{r_2} is the critical value of the BSADF statistic at time r_2 for the significance level and is the fraction of the total sample - chosen by the analyst - determining the minimum duration for an explosive pattern to be identified as a bubble.
Critical values for the BSADF test can be obtained through simulation experiments based on the limit theory in Phillips et al. (2015).

2.2. *Covariate-augmented unit root tests*

The literature on unit root testing was mainly developed in a univariate framework. However, it is quite simplistic to consider the financial and economic environment as univariate, and taking into account the multivariate nature of economic time series can in principle lead to better testing procedures.
Hansen (1995) proposed to include stationary covariates in the standard ADF framework, showing that exploiting the information embodied in related time series can increase the power of stationarity tests.
Formally, Hansen (1995) considered the following regression model:

$$y_t = \mu + \delta y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{i=1}^{q} \beta_i \Delta x_{t-i} + \varepsilon_t; \qquad \varepsilon_t \sim N(0; \sigma^2) \tag{6}$$

Similarly to the conventional ADF test, the CADF test verifies the null hypothesis that a unit root is present, i.e. $H_0 : \delta = 1$, against the left-sided alternative $H_1 : \delta < 1$. Hansen (1995) refers to the test statistic as the $CADF(p, q)$ statistic. As in the conventional ADF regression, a linear trend could also be included in the specification.
Denoted as $v_t = \sum_{i=1}^{q} \beta_i \Delta x_{t-i} + \varepsilon_t$ the stochastic component of the response dynamics driven by the covariate and the error process, the limit distribution of the CADF test statistic - and thus its asymptotic power - is determined by the long-run covariance matrix $\Omega := \sum_{k=-\infty}^{\infty} \mathrm{E}\left(\eta_t \eta_{t-k}'\right)$, with $\eta_t := (v_t; \varepsilon_t)'$, from which the long-run squared correlation between v_t and ε_t, ρ^2, can be derived. When Δx_t has no explanatory power on the long-run movement of v_t, then ρ^2 approaches 1. On the contrary, when Δx_t explains most of the long-run variability of v_t, then ρ^2 approaches 0. The asymptotic distribution - and thus power - of the test statistic depends on the nuisance parameter ρ^2 but, provided that ρ^2 is given, it can be simulated using standard techniques. Heteroskedasticity and autocorrelation consistent covariance estimators that can be used to obtain estimates of ρ^2 have been proposed by, e.g., Andrews (1991), Zeileis (2004), Zeileis (2006) and Kleiber and Zeleis (2008).

2.3. *Bubble detection through covariate-augmented Backward Supremum ADF tests*

In the above specified framework, we propose to merge the right-tailed and the covariate-augmented approaches described in Section 2.1 and 2.2 to

identify potential bubbles in the cryptocurrency price time series, also considering the role of sentiment and news volume in anticipating and explaining the crypto price patterns.
To this aim, we refer to the methodology by Korkos (2020), who proposed a covariate-augmented BSADF test, named BSCADF, based on repeated estimation of the CADF test statistic using the forward and backward rolling-windows technique in (3):

$$BSCADF_{r_2} := \sup_{r_1 \in [0, r_2 - r_0]} CADF(p)_{r_1}^{r_2} \tag{7}$$

The BSCADF test can also be used for date stamping the detected bubble occurrences, such as the BSADF.
The idea is that considering the sentiment and the volume of market information can potentially affect the estimates of the autoregressive component of the price dynamics, improving the power and/or the size of explosivity tests and, thus, providing a signal that can be effectively used for bubble detection.

3. Data and Empirical Findings

3.1. Data description

To test our proposal, we collect daily closing prices of Bitcoin, Ethereum and Ripple over the period 31 May 2019 - 31 May 2021 from Coinmarketcap[1]. The choice of those three cryptocurrencies is due to the need of having a sufficient news coverage along the considered period for the reliable calculation of a daily sentiment indicator. Indeed, the latter is proposed and produced by Brain[2], a research company specialized in the production of alternative datasets and in the development of proprietary algorithms for investment strategies on financial markets, which monitors public financial news on cryptocurrencies from about 2000 financial media sources. The sentiment scoring technology is based on a combination of various natural language processing techniques. An initial step applies model based sentiment analysis which is then further refined, in case of errors, through manual inspections. The sentiment score assigned to each cryptocurrency is a value ranging from -1 (most negative) to +1 (most positive), updated on a daily basis.

[1] Data is available at https://coinmarketcap.com/
[2] https://braincompany.co

As a counterpart of the mood monitoring, we consider a news volume indicator, based on daily data of Google search queries for the terms "Bitcoin", "Ethereum" and "Ripple" over the same period from Google Trends[3]. Specifically, the exogenous covariate we analyse is the daily Google Search Index, which measures the relative volume of term searches performed, normalized with a value equal to 100 corresponding to the day with the highest volume. Notice that, in this case, both sentiment and search term volume indicators can be obtained in a near-real time manner, therefore contributing to the setup of a practical toolbox for sentiment-based explosivity monitoring.

We present relevant summary statistics of cryptocurrency prices, Brain Sentiment Index and Google Search Index in Tables 1, 2 and 3, respectively. We can notice that the cryptocurrency returns and the Brain Sentiment indicator show similar patterns with average values close to 0 and totally comparable standard deviations around 0.05-0.07. On the other hand, the complete different nature and measurement level of Google trends is clearly reported in Table 3. It is worth noticing that Ethereum is by far the most searched crypto in the considered period, doubling the values of Bitcoin despite its global popularity. Indeed, the nature of Google trends data which cannot be gathered in absolute values but rather normalized to 100 as for the first considered day, should be taken into account when looking at such data.

Returns	Min	Max	Avg	Std	Skew	Kurt
BTC	-0.3717	0.1875	0.0028	0.0402	-0.8118	12.1680
ETH	-0.4235	0.2595	0.0046	0.0517	-0.6662	9.5027
XRP	-0.4233	0.5601	0.0034	0.0676	1.5363	16.7793

Table 1: Summary statistics of cryptocurrency returns. The table presents relevant summary statistics of cryptocurrency returns over the period 31 May 2019 - 31 May 2021.

In Figure 1 and 2, we show the time series dynamics of our covariates, namely the Brain Sentiment Index and the Google Search Index, respectively. The Sentiment indicator appears to be rather stable although some relevant drops appear clearly around March 2020 and first months of 2021. At the opposite we notice an increasing trend for Google search with one big hype for Ethereum during the spring 2021.

[3] Data is available at https://trends.google.com/trends/

Brain Sentiment	Min	Max	Avg	Std	Skew	Kurt
BTC	-0.3182	0.1439	0.0283	0.0536	-0.2580	1.2452
ETH	-0.0916	0.2408	0.1083	0.0712	-0.4925	-0.2632
XRP	-0.2031	0.2072	0.0318	0.0655	-0.2622	0.1932

Table 2: Summary statistics of Brain Sentiment Indicators. The table presents relevant summary statistics of the Brain Sentiment Indicators over the period 31 May 2019 - 31 May 2021.

Google Search	Min	Max	Avg	Std	Skew	Kurt
BTC	52.2727	1008.8601	154.9287	131.3087	2.1707	5.4673
ETH	26.0870	2619.3434	296.0966	405.3924	2.4256	6.3983
XRP	44.3860	651.8219	101.2452	69.8259	3.1347	13.5140

Table 3: Summary statistics of Google Search Indices. The table presents relevant summary statistics of the Google Search Indices over the period 31 May 2019 - 31 May 2021.

We report, additionally, the correlation matrix and a scatter and empirical distribution plot of our variables in Figures A.8 and A.9 in Appendix.

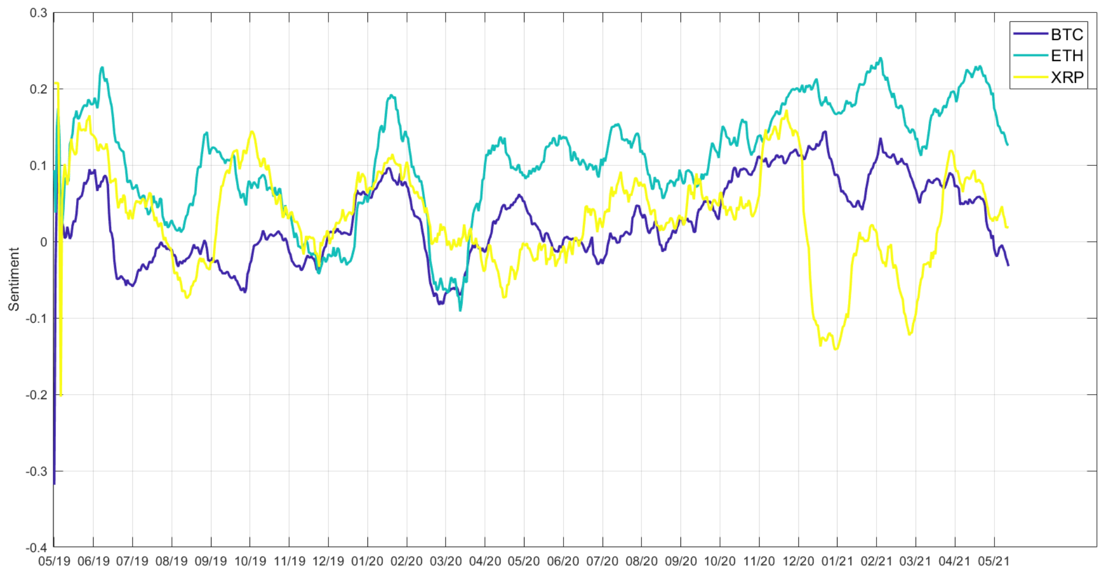

Figure 1: Brain Sentiment Indicator time series. The gure shows the dynamics of the daily sentiment index for the three selected cryptocurrencies over the period 31 May 2019 - 31 May 2021.

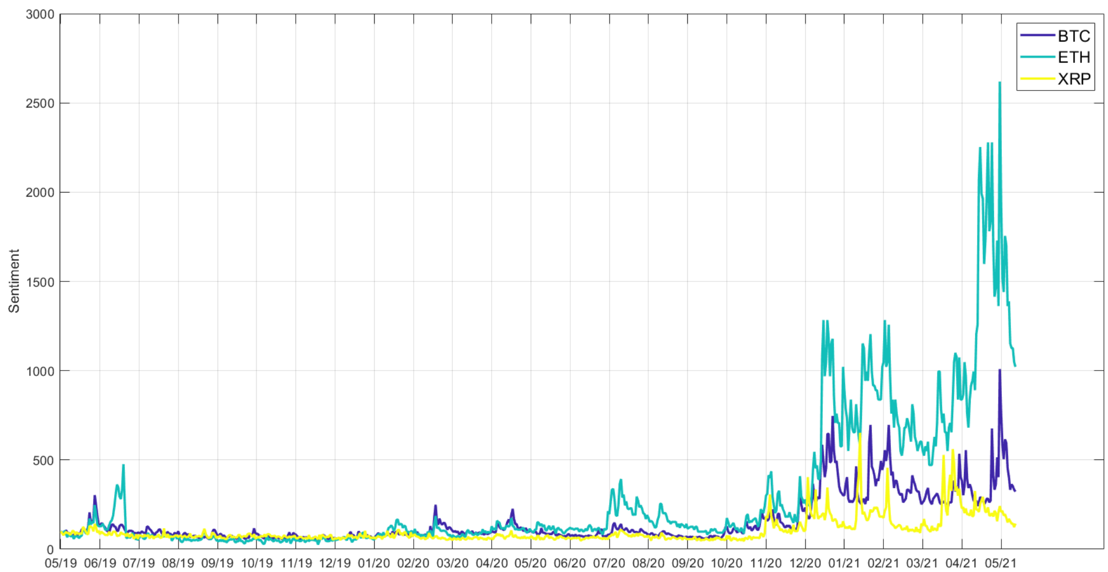

Figure 2: Google Search Index time series. The figure shows the dynamics of the daily Google Search Index for the three selected cryptocurrencies over the period 31 May 2019 - 31 May 2021.

3.2. Results

Before proceeding with our price bubble analysis, we verify the possible presence of non-stationarity in the analysed time series. We remark that non-stationarity is not an issue, but rather a pre-requisite to conduct explosivity analysis of a target variable (cryptocurrency prices). On the contrary, we need to ensure that the considered covariates, which enter the model specification of the recursive ADF test, are stationary, so to conduct proper inference on the autoregressive parameter of interest. To this aim, we present the full-sample results of the standard left-sided ADF unit root test performed on the three time series of price, sentiment and Google search queries related to Bitcoin, Ethereum and Ripple in Table 4. As expected, evidence shows that cyptocurrency prices are highly non-stationary, in particular Ethereum and Bitcoin. All other covariates instead, except for the Google Search Index related to Ethereum, are stationary in levels. We therefore proceed with our analysis by first differencing the latter variable and by applying our testing procedure to the rest of the variables in levels.

Time series	p-value
BTC Price	0.7526
ETH Price	0.8266
XRP Price	0.3481
BTC Sentiment	0.0986
ETH Sentiment	0.0423
ETH Sentiment	0.0198
BTC Search	0.0235
ETH Search	0.1840
XRP Search	0.0000

Table 4: Augmented Dickey-Fuller (ADF) tests. The table reports the results of the ADF unit root test performed on the three time series of price, sentiment and Google Search Indices related to Bitcoin, Ethereum and Ripple over the full sample period 31 May 2019 - 31 May 2021. The model speci cation under the null hypothesis is a random walk, with constant but no time trend. The test is performed considering the optimal number of lags of the dependent variable according to the Bayes-Schwarz Information Criterion (BIC).

Within our framework, we perform BSADF and BSCADF tests for the selected cryptocurrencies using, at each iteration, the optimal lag order of the exogenous covariate determined through the Bayes-Schwarz (BIC) information criterion. Figure 3 compares the dynamics of p-values for the BSADF and the BSCADF test - the latter considering the Brain Sentiment Index - along with their difference and the price dynamics of the considered cryptocurrencies. On the one hand, evidence shows that during period of tranquil market dynamics the BSADF and BSCADF test statistics tend to co-move strongly, giving raise to low-magnitude deviations between the two. On the other hand, from the beginning of October 2020 onwards, we observe that the two test statistics start diverging and, in most cases, the p-value of the BSADF test statistics is smaller than the p-value of the covariate-augmented test, meaning that the inclusion of sentiment as an explanatory variable reduces the estimated acceleration of the price pattern. Furthermore, notice that the rising difference between the two is associated to a consequent surge in the cryptocurrency prices.

The latter result is likely due to the influence of the lagged sentiment indicator, which is able to explain a large portion of the cryptocurrency price variations, and thus lowers the value of the test statistics inducing different outcomes with respect to those obtained by the test without covariates. In fact, this difference arises, in general, well before the cryptocurrency price surge, indicating that the misalignment between the two test statistics can be informative and act as an early-warning indicator for bubble detection purposes. On the contrary, when the price growth takes strength, the two tests tend to converge on low p-values, corresponding to rejection of the random walk hypothesis in favour of explosivity. This is evident for all the analyzed cryptocurrencies, for which the BSCADF p-value drops below the significance level - signalling the start of a bubble period - right before the beginning of an exponential price growth. When the price starts growing exponentially, the estimated autoregressive dynamics begins indeed to be explosive, even conditionally on market sentiment.

On the other hand, Figure 4 shows the p-value dynamics of the BSADF and the BSCADF test including the Google Search Index, along with their difference and the price dynamics of the considered cryptocurrencies. Although the divergence between the two test statistics' p-values can still provide a signal, this is yet not clear as in the sentiment case. On the one hand, we can notice a clear inversion of the sizes of the two p-values before the price surge in Ethereum, as it was for sentiment. On the other hand, the signal is not as pronounced for Bitcoin or Ripple. This suggests that a polarised indicator, which takes into account for the possible positive, negative and neutral polarity of news, might be more useful as an early warning indicator of time series explosiveness if compared to search volumes. Indeed, while the

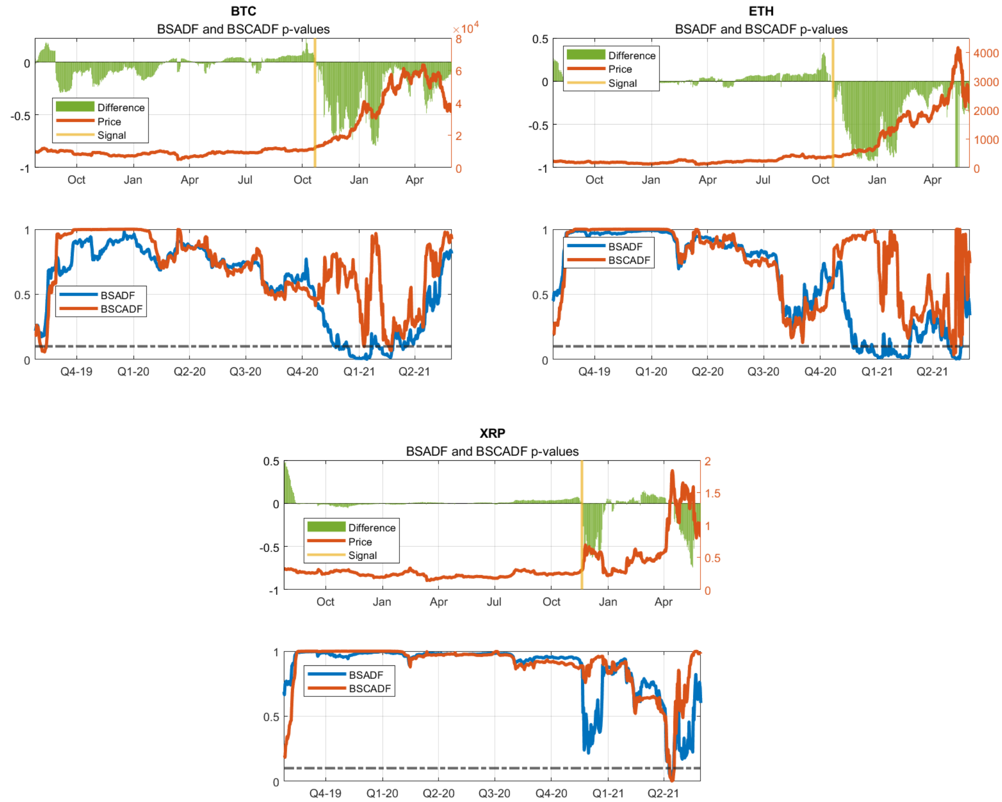

Figure 3: BSADF and BSCADF Brain Sentiment Index test p-values. The gure shows the BSADF and BSCADF p-values (bottom panels), their di erence and closing price dynamics (top panels) of the three cryptocurrencies, using the Brain Sentiment Index as a covariate, over the full sample period. Black dashed lines indicate the 10% signi cance level.

latter do measure the level of interest of the community into cryptocurrency matters, it does not, however, distinguish between positivity, negativity and neutrality of news, therefore constituting allegedly a more suitable candidate predictor of volatility, rather than of price direction.

With the aim of comparing more in depth the differences in the BSADF and BSCADF test outcomes, Figure 5 shows the scatter plot of the BSADF and Brain Sentiment Index BSCADF tests' p-values for the three analysed cryptocurrencies over the considered sample period. We report a similar comparison plot for the Google Search Index p-values in Figure A.10. From the figure we notice that there exists a relatively high degree of correlation

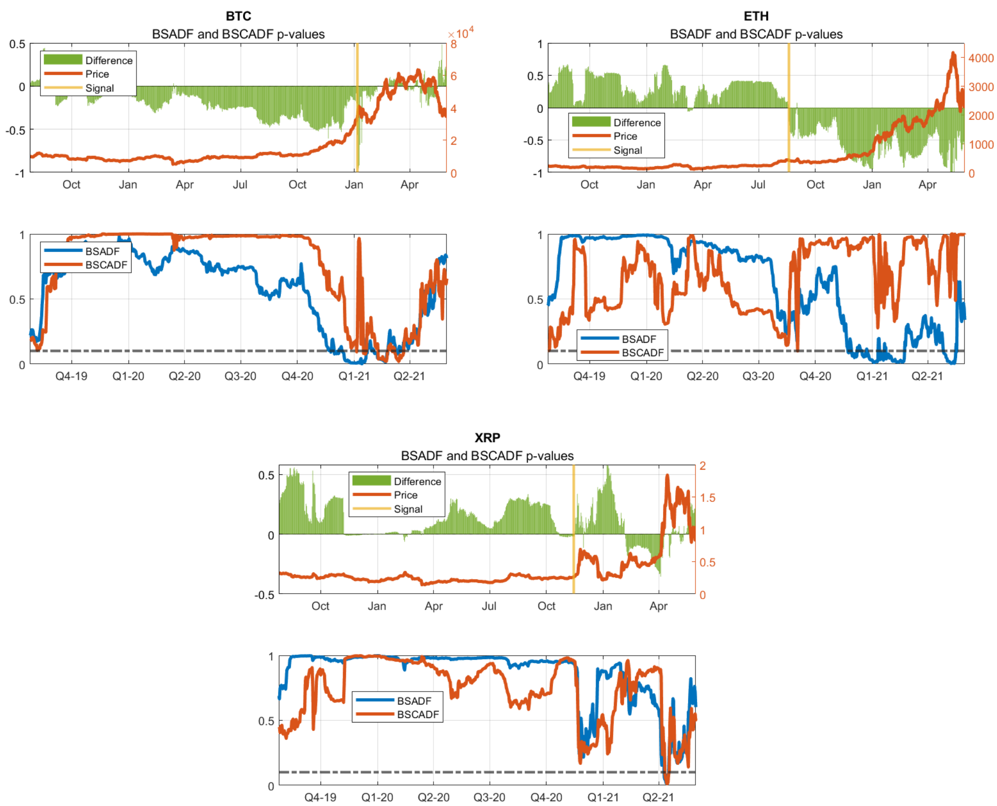

Figure 4: BSADF and BSCADF Google Search Index test p-values. The gure shows the BSADF and BSCADF p-values (bottom panels), their di erence and closing price dynamics (top panels) of the three cryptocurrencies, using the Google Sentiment Index as a covariate, over the full sample period. Black dashed lines indicate the 10% signi cance level.

between the two p-values of the test statistics in the case of Bitcoin (correlation: 0.8353) and Ripple (correlation: 0.7211). Interestingly, the correlation of the two p-values is consistently lower for Ethereum. This is also in line with the results obtained considering the Google Search Index as a covariate in the test specification - see Figure A.10 in Appendix.

To get a further insight into the time-changing relationship between the sentiment indicator and the cryptocurrency price behaviour, we perform a rolling linear regression exercise, over 100-day windows, where the response variable is the cryptocurrency return and the regressor is the lagged sentiment indicator. Figures 6 shows that, starting from September 2020, the

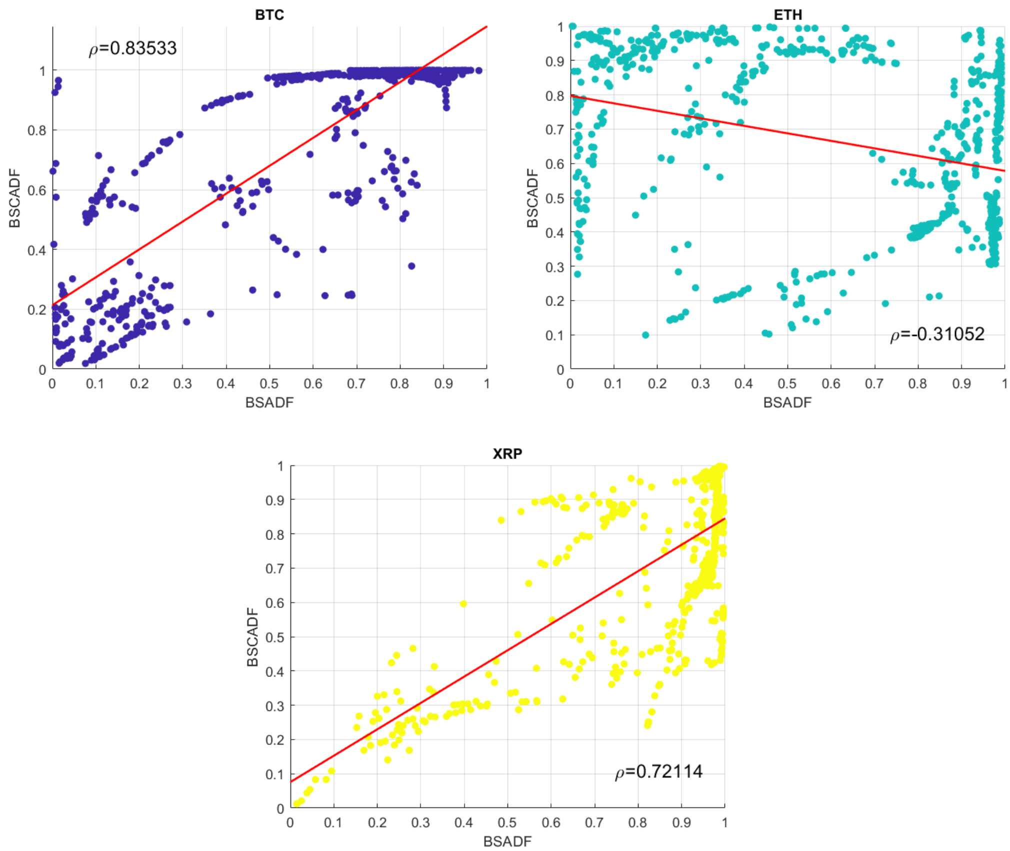

Figure 5: Scatter plot of BSADF and Brain Sentiment Index BSCADF p-values. The gure shows the scatter plot of the BSADF and Brain Sentiment Index BSCADF p-values related for the three cryptocurrencies over the considered sample period. The value of indicates the correlation between the two test statistics.

estimated coefficients associated to the sentiment indicators grow sharply. It can be noticed from Figure 7 that, repeating the same exercise using lagged Google Search Indices as a regressor, the estimated coefficients are relatively flatter. Therefore, in the most recent period, the sentiment indicators allegedly turn out to be more informative than the news volume in predicting explosive behaviours in the cryptocurrency prices.

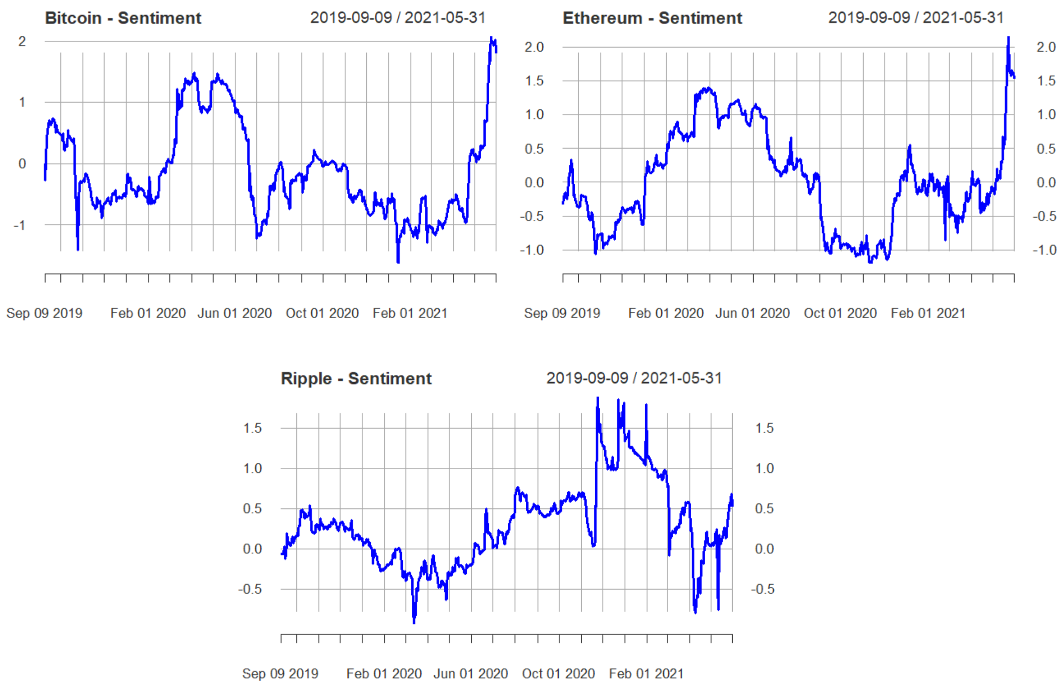

Figure 6: Sentiment Index rolling regression coefficients. The figure shows the estimated coefficients associated to the lagged Brain Sentiment index for the selected cryptocurrencies in a rolling linear regression exercise, where the response variable is the cryptocurrency return. The rolling window is set to 100 observations.

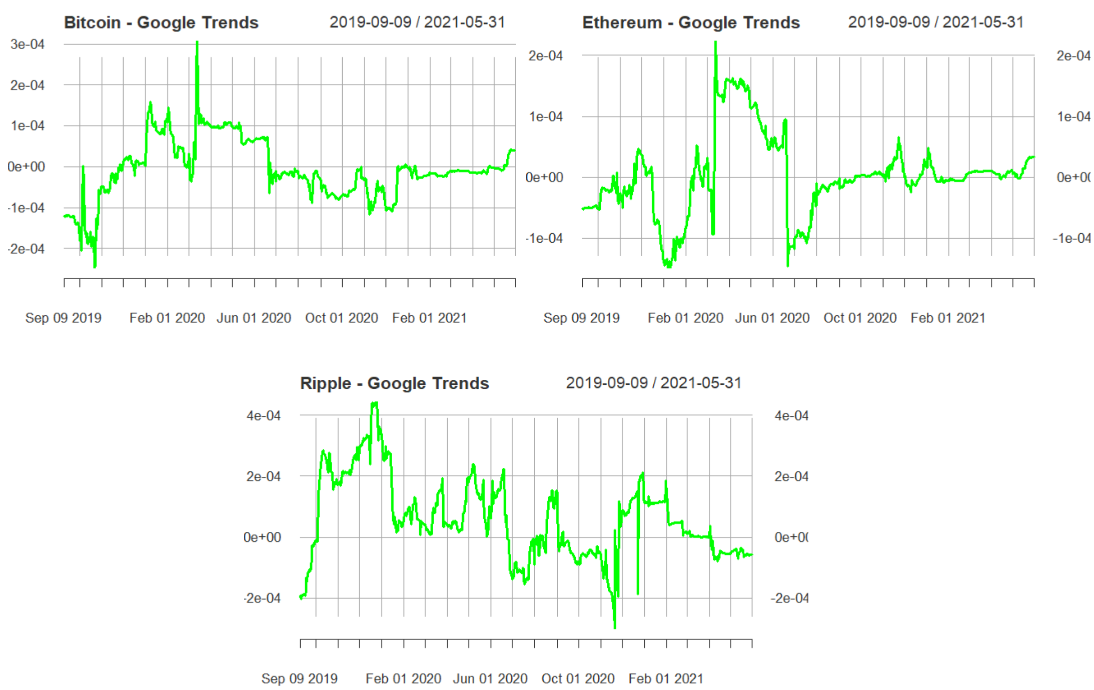

Figure 7: Google Search Index rolling regression coe cients. The gure shows the estimated coe cients associated to the lagged Google Search Index for the selected cryptocurrencies in a rolling linear regression exercise, where the response variable is the cryptocurrency return. The rolling window is set to 100 observations.

4. Concluding remarks

The recent affirmation of high-volatile, sentiment-based market such as the cryptocurrency one calls for the design and implementation of innovative data science techniques, based upon sound econometric and statistical models, that can provide a timely signal of explosivity in the price behaviour of digital currency, taking explicitly into account for investors' sentiment.

To this aim, we propose a recursive covariate Augmented Dickey-Fuller testing procedure, which leverages a high-dimensional set of cryptocurrency news and Google search volumes to detect and anticipate bubble phenomena in the price of three major cryptocurrencies, i.e. Bitcoin, Ethereum and Ripple, over the period 31 May 2019 - 31 May 2021. Specifically, we apply a recursive ADF testing procedure which takes into account dependence of the price series on any relevant covariates, so to detect explosive price behaviour. This approach allows us to improve bubble identification, providing a more effective signal of the exponential growth of the cryptocurrency prices.

Our results show that the BSCADF with sentiment does significantly diverge from the standard BSADF test well before price surges, thereby the usefulness of the misalignment between the two test statistics in providing an early-warning indicator for explosive time series behaviour. In concomitance with the explosive dynamics of prices, instead, the two tests tend to converge in supporting the rejection of the random walk hypothesis in favour of explosivity. Additionally, our results provide evidence on the fact that polarised sentiment indicators, rather than term search volumes, might be better at anticipating explosivity, given their capability to distinguish among negative, positive or neutral news volumes.

We remark that our methodology can be applied, without loss of generality, to any financial or other kind of time series to detect, with an improved power of the test, possible explosive trends in time series. This thanks to the explanatory gain given by adding any relevant predictors of the dependent variable to the specification of the standard univariate recursive ADF tests. Further research might generalise the econometric properties of the test to a multivariate framework, with the aim of designing an econometric testing procedure capable to take into account for multiple covariates when detecting time series explosiveness.

References

Agosto, A. and Cafferata, A. (2020). Financial bubbles: a study of co-explosivity in the cryptocurrency market. *Risks*, 8(2):34.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Aste, T. (2019). Cryptocurrency Market Structure: Connecting Emotions and Economics. *Digital Finance*, 1(1-4):5–21.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., and Weber, I. (2012). Web Search Queries Can Predict Stock Market Volumes. *PLoS One*, 7(7):e40014.

Bouri, E., Gupta, R., and Roubaud, D. (2019a). Herding behaviour in cryptocurrencies. *Finance Research Letters*, 29:216–221.

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

Brandvold, M., Molnár, P., Vagstad, K., and Valstad, O. C. A. (2015). Price discovery on bitcoin exchanges. *Journal of International Financial Markets, Institutions and Money*, 36:18–35.

Caporale, G. M. and Pittis, N. (1999). Unit root testing using covariates: some theory and evidence. *Oxford Bulletin of Economics and Statistics*, 61(4):583–595.

Case, K. and Shiller, R. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, 2:299–362.

Cerchiello, P. and Giudici, P. (2016). How to measure the quality of financial tweets. *Quality & Quantity*, 50(4):1695–1713.

Cerchiello, P. and Nicola, G. (2018). Assessing news contagion in finance. *Econometrics*, 6(1).

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. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(s1):2–9.

Corbet, S., L. B. and Yarovya, L. (2018). Datestamping the bitcoin and ethereum bubbles. *Finance Research Letters*, 26:81–88.

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

Corbet, S., Meegan, A., Larkin, C., Lucey, B., and Yarovaya, L. (2018b). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165:28–34.

Corea, F. (2016). Can twitter proxy the investors' sentiment? the case for the technology sector. *Big Data Research*, 4:70–74.

Cretarola, A. and Figà-Talamanca, G. (2019). Detecting bubbles in bitcoin price dynamics via market exuberance. *Annals of Operations Research*, pages 1–21.

Cretarola, A. and Figà-Talamanca, G. (2020). Bubble regime identification in an attention-based model for bitcoin and ethereum price dynamics. *Economics Letters*, 191:108831.

de Souza, M. J. S., Almudhaf, F. W., Henrique, B. M., Negredo, A. B. S., Ramos, D. G. F., Sobreiro, V. A., and Kimura, H. (2019). Can artificial intelligence enhance the bitcoin bonanza. *The Journal of Finance and Data Science*, 5(2):83–98.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Dimpfl, T. and Peter, F. J. (2020). Nothing but noise? price discovery across cryptocurrency exchanges. *Journal of Financial Markets*, page 100584.

Dreger, C. and Zhang, R. (2013). Is there a bubble in the chinese housing market? *Urban Policy and Research*, 31(1):27–39.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.

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

Giudici, P. and Pagnottoni, P. (2019). High frequency price change spillovers in bitcoin markets. *Risks*, 7(4):111.

Giudici, P. and Pagnottoni, P. (2020). Vector error correction models to measure connectedness of bitcoin exchange markets. *Applied Stochastic Models in Business and Industry*, 36(1):95–109.

Gronwald, M. (2021). How explosive are cryptocurrency prices? *Finance Research Letters*, 38:101603.

Hansen, E. B. (1995). Rethinking the univariate approach to unit root testing; using covariates to increase power. *Econometric Theory*, 11:1148–1171.

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

Huang, J.-Z. and Huang, Z. J. (2020). Testing moving average trading strategies on etfs. *Journal of Empirical Finance*, 57:16–32.

Huang, J.-Z., Rossi, M., and Wang, Y. (2015). Sentiment and corporate bond valuations before and after the onset of the credit crisis. *The Journal of Fixed Income*, 25(1):34–57.

Jaquart, P., Dann, D., and Weinhardt, C. (2021). Short-term bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*, 7:45–66.

Kleiber, C. and Zeleis, A. (2008). Applied econometrics with r. *Springer-Verlag, New York*.

Korkos, I. (2020). On explosive time series. *Essex Business School PhD thesis*.

Kyriazis, N., Papadamou, S., and Corbet, S. (2020). A systematic review of the bubble dynamics of cryptocurrency prices. *Research in International Business and Finance*, 54:101254.

Lee, R. S. (2019). Cosmos trader–chaotic neuro-oscillatory multiagent financial prediction and trading system. *The Journal of Finance and Data Science*, 5(2):61–82.

Nakamoto, S. et al. (2008). Bitcoin: A peer-to-peer electronic cash system.

Pagnottoni, P. and Dimpfl, T. (2019). Price discovery on bitcoin markets. *Digital Finance*, 1(1):139–161.

Phillips, P., 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.

Phillips, P., Wu, Y., and Yu, J. (2015). Testing for multiple bubbles : Historical episodes of exuberance and collapse in the sp500. *International Economic Review*, 56(4):1043–1078.

Resta, M., Pagnottoni, P., and De Giuli, M. E. (2020). Technical analysis on the bitcoin market: Trading opportunities or investors' pitfall? *Risks*, 8(2):44.

Scaramozzino, R., Cerchiello, P., and Aste, T. (2021). Information theoretic causality detection between financial and sentiment data. *Entropy*, 23(5).

Yi, H., Li, Y., Wang, M., Yan, Z., and Nie, Z. (2021). An efficient blockchain consensus algorithm based on post-quantum threshold signature. *Big Data Research*, 26:100268.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16(9):1–16.

Appendix A. Additional results

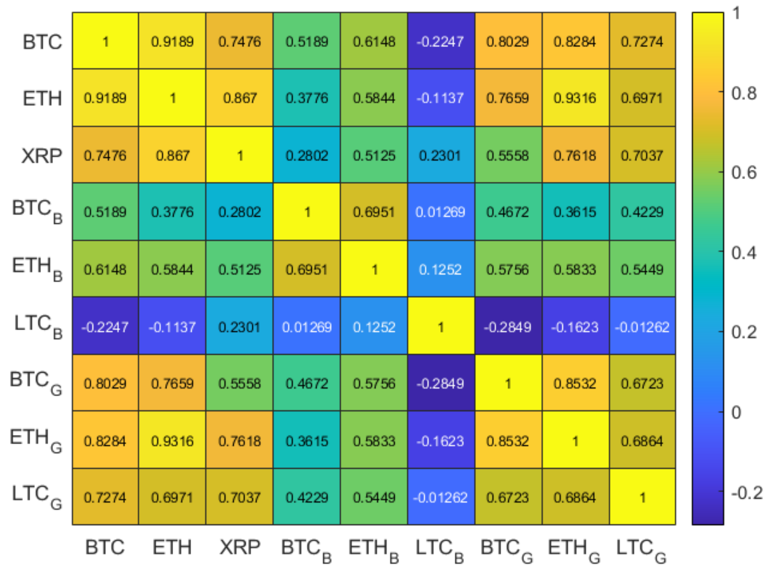

Figure A.8: Data correlation matrix. The gure shows the correlation matrix of Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) prices, Brain Sentiment Indicator (B) and Google Search Indices (G) over the full sample period.

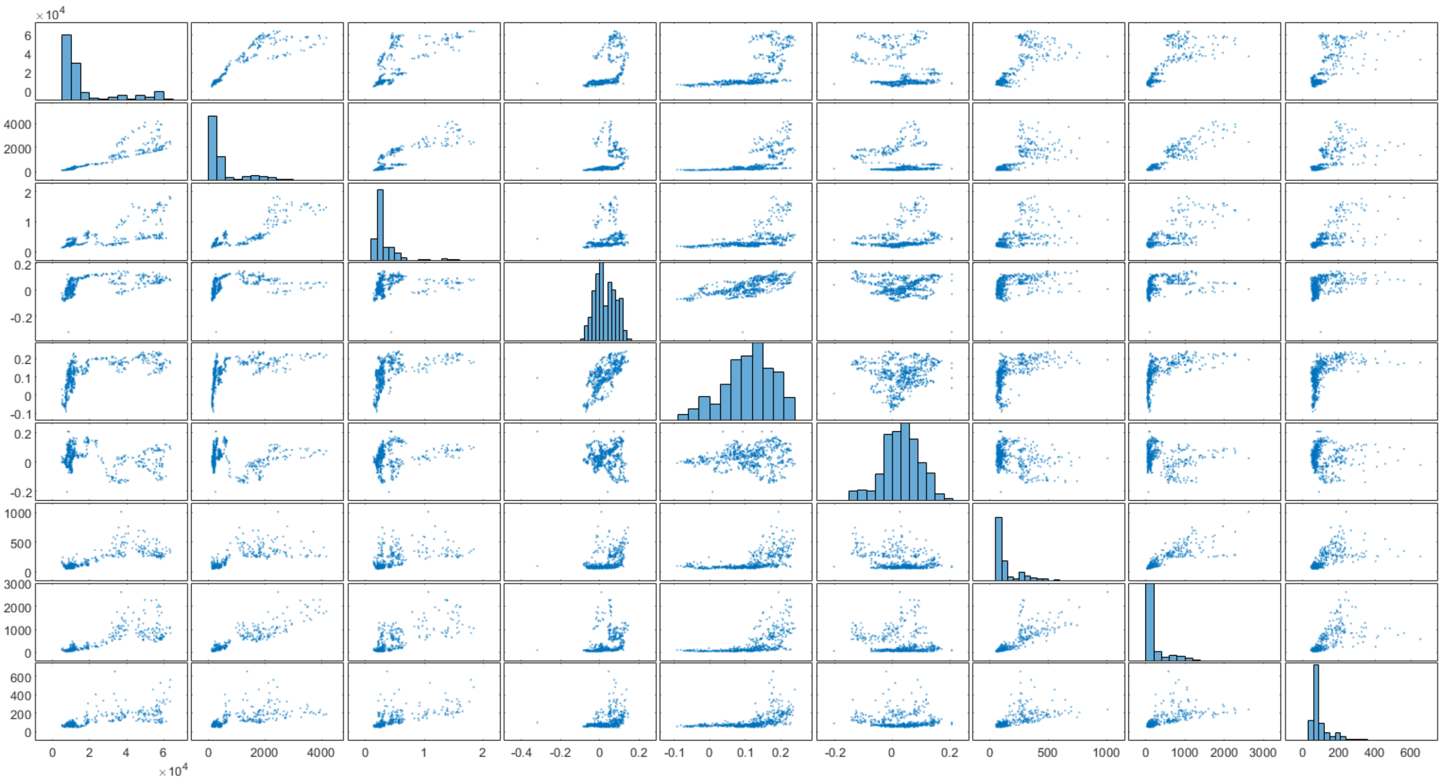

Figure A.9: Scatter and empirical distribution function plots. The gure shows the scatter and empirical distribution function plot of Bitcoin, Ethereum, Litecoin prices, Brain Sentiment Indicator and Google Search Indices over the full sample period.