



# Sentiment, Google queries and explosivity in the cryptocurrency market

Arianna Agosto<sup>\*,1</sup>, Paola Cerchiello<sup>1</sup>, Paolo Pagnottoni<sup>1</sup>

Department of Economics and Management, University of Pavia, Italy

## ARTICLE INFO

### Article history:

Received 6 May 2022

Received in revised form 29 July 2022

Available online 4 August 2022

### Keywords:

Bitcoin

Cryptocurrency market

Sentiment

Speculative bubbles

Google trends

Big data application

## 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 exploiting 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 using 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 statistics in concomitance with price surges, highlighting the ability of sentiment to predict speculative bubble occurrences. We also show how a polarised version of investors' sentiment plays an overall more determinant role, if compared to news volume and Google queries, in providing an early warning signal of market bubble episodes in cryptocurrencies.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

Since Bitcoin was first conceived Nakamoto et al. [1], research on cryptocurrencies 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. [2], de Souza et al. [3], Lee [4], Jaquart et al. [5]. On the other hand, methods pertaining to such innovative fields often lack a sound and robust state-of-the-art econometric and statistical framework, to develop models accurately, conduct appropriate inference, and, most importantly, improve model performance. It is against this background that we propose a novel sentiment-based testing procedure for cryptocurrency explosiveness, which explicitly takes into account any possible price predictors. We develop our empirical application upon a large set of financial news, to anticipate speculative bubble occurrences in cryptocurrency prices.

The 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. [6] 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

\* Corresponding author.

E-mail address: [arianna.agosto@unipv.it](mailto:arianna.agosto@unipv.it) (A. Agosto).

<sup>1</sup> All authors contributed equally to this work.

Pagnottoni [7], Giudici and Pagnottoni [8] explore the dynamic connectedness of Bitcoin exchanges and provide evidence of 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. [9], Pagnottoni and Dimpfl [10], Giudici and Polinesi [11], Dimpfl and Peter [12]. Bouri et al. [13], Resta et al. [14] 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 [15].

A flourishing 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. [16] analyse the presence of bubbles in Bitcoin prices from 18 July 2010 to 18 February 2014 by employing Phillips et al. [17] methodology. The 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. Fry and Cheah [18] 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 that is composed of a Wiener process and a jump process to control for a constant intrinsic rate of return and intrinsic level of risk. They show that explosive behaviour exists in the Bitcoin market, and the random walk hypothesis is rejected. In Corbet et al. [19] 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, 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 to 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. [20] for a systematic review on cryptocurrency bubbles.

More recently, Bouri et al. [21] have studied co-explosivity on data about Bitcoin, Ripple, Ethereum, Litecoin, Nem, Dash and Stellar, over the period from 7 August 2015 to 31 December 2017, 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, the market values of the 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 [22] employs price data on Bitcoin, Ethereum, Ripple, Litecoin and Stellar, over the period 1 January 2017–31 December 2018, to examine co-explosivity of cryptocurrency time series, covering the price surge period at the end of 2017. Partly contrasting with Bouri et al. [21], they find Ethereum particularly important in determining explosivity of other cryptocurrencies, and Bitcoin to be quite highly impacted by the dynamics of the other cryptocurrency prices. A recent contribution from Gronwald [23] 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. Moreover, 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, was developed by Cretarola and Figà-Talamanca [24,25]. 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 analyse the impact of news and investors' sentiment on the dynamics of financial markets. They mainly deal with the statistical and econometric analysis of nonconventional data; see, for instance, Bollen et al. [26], Bordino et al. [27], Choi and Varian [28], Feldman [29], Huang et al. [30], Cerchiello and Giudici [31], Cerchiello and Nicola [32], and Scaramozzino et al. [33], 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 [34] demonstrate how prices affect sentiment and viceversa, with noticeable differences in intensities and number of significant interactions.

Against those 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 large size distortions; see Caporale and Pittis [35].

We apply our methodology to detect explosivity in the price series of three major cryptocurrencies in terms of market capitalisation, 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 with the names of each cryptocurrency.

We find a clear divergence of the two test statistics with and without covariates, 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 that the sentiment indicator, which does take into account the positive, negative or neutral meaning of

cryptocurrency-related news, seems to act as a better early warning indicator of price surges, compared to term search volumes.

Our contribution to the literature is twofold. From a methodological perspective, we extend the Covariate-Augmented Dickey–Fuller setting for unit root testing Hansen [36] to the framework of bubble detection and explosivity testing Phillips et al. [17]. In this way, we provide recursive ADF testing procedures that 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 the 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. Bubble behaviour is defined in the literature as an extreme price acceleration that cannot be driven by the underlying fundamental economic variables Case and Shiller [37], Dreger and Zhang [38]. The end of this phase, often referred to as bubble burst, leads to drastic price drops, causing severe losses to investors.

Several recent works have provided empirical evidence of the presence of bubbles in cryptocurrency prices Fry and Cheah [18], Corbet and Yarovya [39]. From a methodological viewpoint, most of them resorted to the right-tailed unit root testing approach based on Augmented Dickey and Fuller [40] regression. Indeed, the extremely rapid price increase, to which the definition of a financial bubble refers, can be described by exponential growth, whose occurrence can be detected through right-sided unit root tests.

Specifically, Phillips et al. [17] proposed a univariate approach to test for end-of-sample bubbles through estimation of the ADF regression:

$$y_t = \mu + \phi y_{t-1} + \sum_{i=1}^p \psi_i \Delta y_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (1)$$

performed on the full sample of data, where  $\mu$  is the intercept and  $p$  is the number of lags of the differenced dependent variable  $\Delta y_t$ , chosen through some model selection criteria. A possible alternative specification includes a deterministic trend.

In the standard ADF test, Eq. (1) is used to test the null hypothesis  $\phi = 1$ , corresponding to the random walk process, against the alternative of stationarity ( $\phi < 1$ ). In the right-tailed unit root test for explosivity proposed by Phillips et al. [17], the null hypothesis is still the random walk hypothesis, but the alternative is  $\phi > 1$ , corresponding to 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 performing a sequential estimation of the ADF statistics by recursively applying the test to subsamples of the data. Specifically, denoting by  $ADF(p)_{r_1}^{r_2}$  the ADF test performed on the subsample  $t = [r_1 T], \dots, [r_2 T]$ , Phillips et al. [41] proposed the following test statistic:

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

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  $[r_0 T]$ . 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 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, to understand whether a particular observation belongs to a bubble phase in the overall trajectory.

Based on these motivations, Phillips et al. [41] 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  $r_2 - r_0$  (the sample fraction corresponding to the origination of the window):

$$BSADF_{r_2} := \sup_{r_1 \in [0, r_2 - r_0]} \{ADF(p)_{r_1}^{r_2}\} \quad (3)$$

Thus, at each window end time  $r_2$ , the BSADF test statistic is the supremum of the right-tailed ADF statistics computed on all subsamples ending at date  $t$  subject to a minimum sample size  $[r_0 T]$ .

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 the termination date ( $\hat{r}_e$ ) of the same bubble:

$$\hat{r}_s = \inf_{r_2 \in [0, 1-r_0]} \{r_2 : BSADF_{r_2}(r_0) > cv_{r_2}^\alpha\} \quad (4)$$

$$\hat{r}_e = \inf_{r_2 \in [r_s + \delta T, 1]} \{r_2 : BSADF_{r_2}(r_0) < cv_{r_2}^\alpha\} \quad (5)$$

where  $cv_{r_2}^\alpha$  is the critical value of the BSADF statistic at time  $r_2$  for the significance level  $\alpha$  and  $\delta$  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. [41].

## 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 [36] proposed including 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:

$$y_t = \mu + \phi y_{t-1} + \sum_{i=1}^p \psi_i \Delta y_{t-i} + \sum_{i=1}^q \xi_i \Delta x_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (6)$$

Similar to the conventional ADF test, the CADF test verifies the null hypothesis that a unit root is present, i.e.,  $H_0 : \phi = 1$ , against the left-sided alternative  $H_1 : \phi < 1$ . Hansen [36] 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  $v_t = \sum_{i=1}^q \xi_i \Delta x_{t-i} + \varepsilon_t$  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  $\Omega := \sum_{k=-\infty}^{\infty} E(\epsilon_t \epsilon'_{t-k})$ , with  $\epsilon_t := (v_t, \varepsilon_t)'$ , from which the long-run squared correlation between  $v_t$  and  $\varepsilon_t$ ,  $\rho^2$ , can be derived. When  $\Delta x_t$  has no explanatory power on the long-run movement of  $v_t$ , then  $\rho^2$  approaches 1. In contrast, when  $\Delta x_t$  explains most of the long-run variability of  $v_t$ ,  $\rho^2$  approaches 0. The asymptotic distribution, and, thus, power, of the test statistic depends on the nuisance parameter  $\rho^2$  but, provided that  $\rho^2$  is given, it can be simulated using standard techniques. Heteroskedasticity and autocorrelation consistent covariance estimators that can be used to obtain estimates of  $\rho^2$  have been proposed by, e.g., Andrews [42], Zeileis [43,44], Kleibers and Zeileis [45].

## 2.3. Bubble detection through covariate-augmented backward supremum ADF tests

In the above specified framework, we propose merging the right-tailed and covariate-augmented approaches described in Sections 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 crypto price patterns.

To this aim, we refer to the methodology by Korkos [46], 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):

$$BSCADF_{r_2} := \sup_{r_1 \in [0, r_2 - r_0]} \{CADF(p)_{r_1}^2\} \quad (7)$$

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 price dynamics, improving the power and/or the size of explosivity tests and, thus, providing a signal that can be effectively used by analysts and policy-makers 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.<sup>2</sup> The choice of those three cryptocurrencies is due to the need for sufficient news

<sup>2</sup> Data are available at <https://coinmarketcap.com/>.

**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.

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 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.

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 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.

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

coverage along the considered period for the reliable calculation of a daily sentiment indicator. Indeed, the latter is proposed and produced by Brain<sup>3</sup> a research company specialising 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 approximately 2000 financial media sources. 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.

As a counterpart of mood monitoring and for comparability purposes, we consider a news volume indicator, based on daily data from Google search queries for the terms “Bitcoin”, “Ethereum” and “Ripple” over the same period from Google Trends.<sup>4</sup> Specifically, we analyse the daily Google Search Index, which measures the relative volume of term searches performed, normalised with a value equal to 100 corresponding to the day with the highest volume. Notice that, in this case, sentiment and search term volume inputs 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 1–3, respectively. We note that the cryptocurrency returns and the Brain Sentiment indicator show similar patterns with average values close to 0 and totally comparable standard deviations of approximately 0.05–0.07. On the other hand, the completely different nature and measurement level of Google trends is clearly reported in Table 3. It is worth noting 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 normalised to 100 as for the first considered day, should be taken into account when looking at such data.

In Figs. 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. In contrast, we notice an increasing trend for Google search with one large hype for Ethereum during the spring 2021.

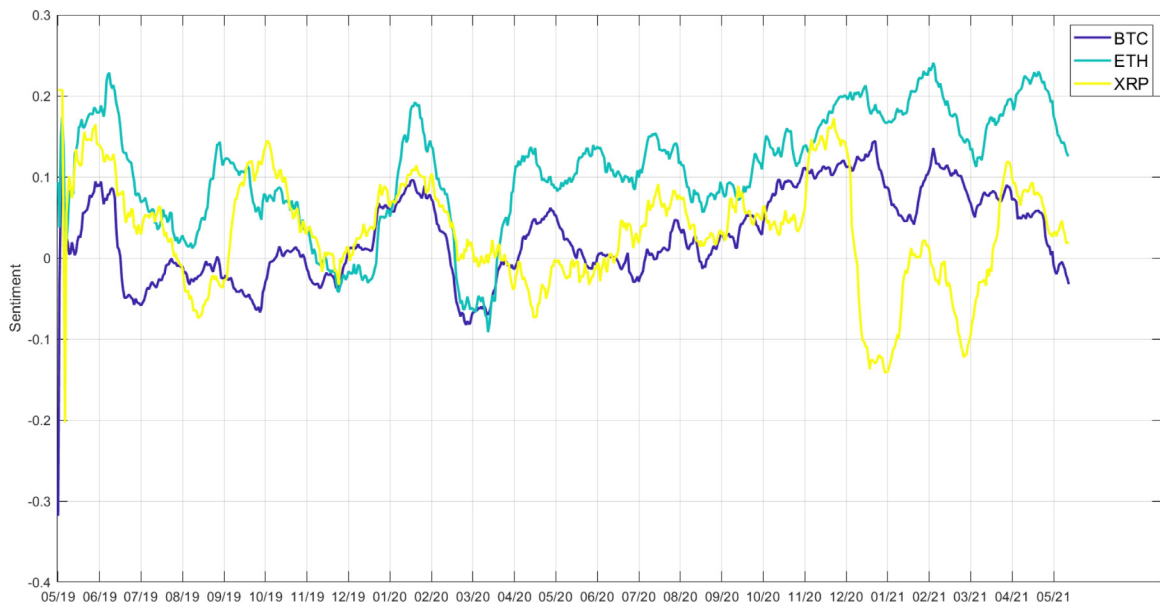
We report, additionally, the correlation matrix and a scatter and empirical distribution plot of our variables in Figs. A.8 and A.9 in Appendix.

### 3.2. Results

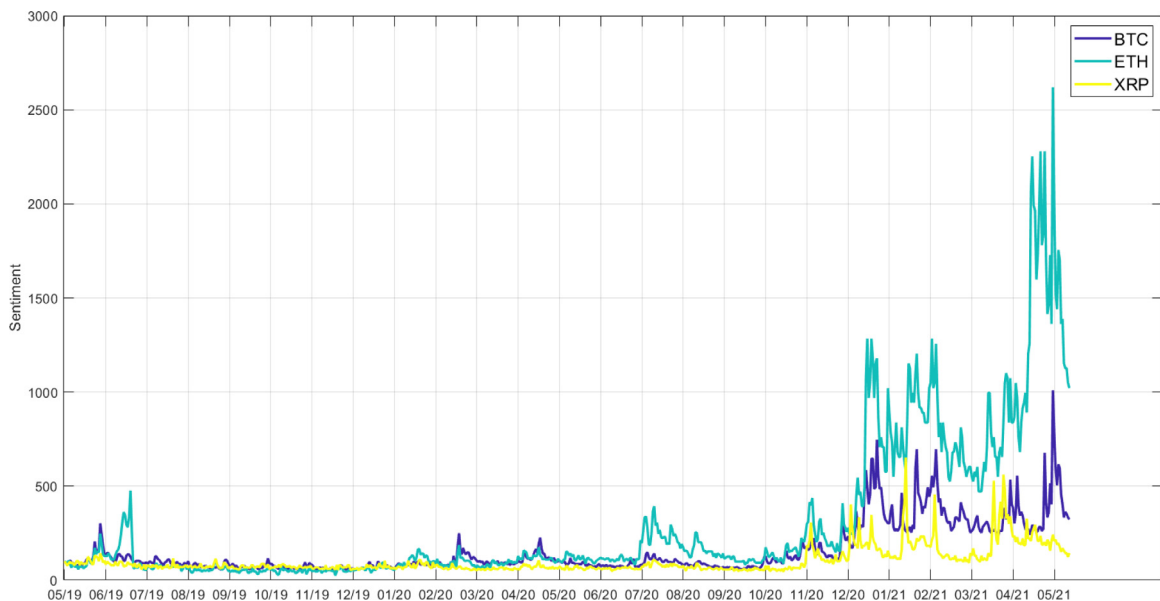
Before proceeding with our price bubble analysis, we verify the possible presence of nonstationarity in the analysed time series. We remark that nonstationarity is not an issue, but, rather, a prerequisite to conduct an explosivity analysis of a target variable (cryptocurrency prices). In contrast, we need to ensure that the considered covariates, which enter the model specification of the recursive ADF test, are stationary, to conduct proper inference on the autoregressive parameter

<sup>3</sup> <https://braincompany.co>

<sup>4</sup> Data are available at <https://trends.google.com/trends/>



**Fig. 1.** Brain Sentiment Indicator time series. The figure shows the dynamics of the daily sentiment index for the three selected cryptocurrencies over the period 31 May 2019–31 May 2021.

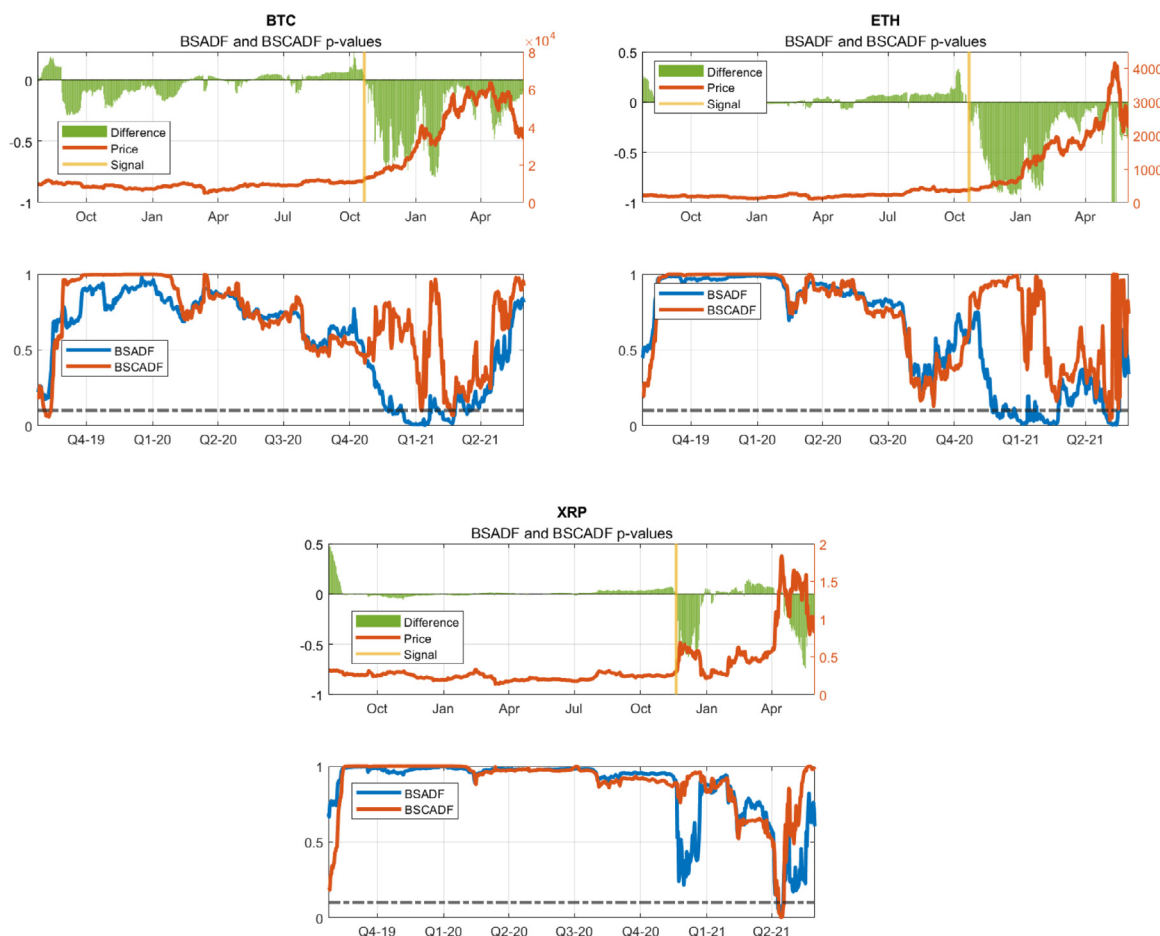


**Fig. 2.** Google Search Index time series. The figure shows the dynamics of the daily Google Search Index for the three selected cryptocurrencies over the period 31 May 2019–31 May 2021.

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 cryptocurrency prices are highly nonstationary, particularly Ethereum and Bitcoin prices. All other covariates, 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.

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. Fig. 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





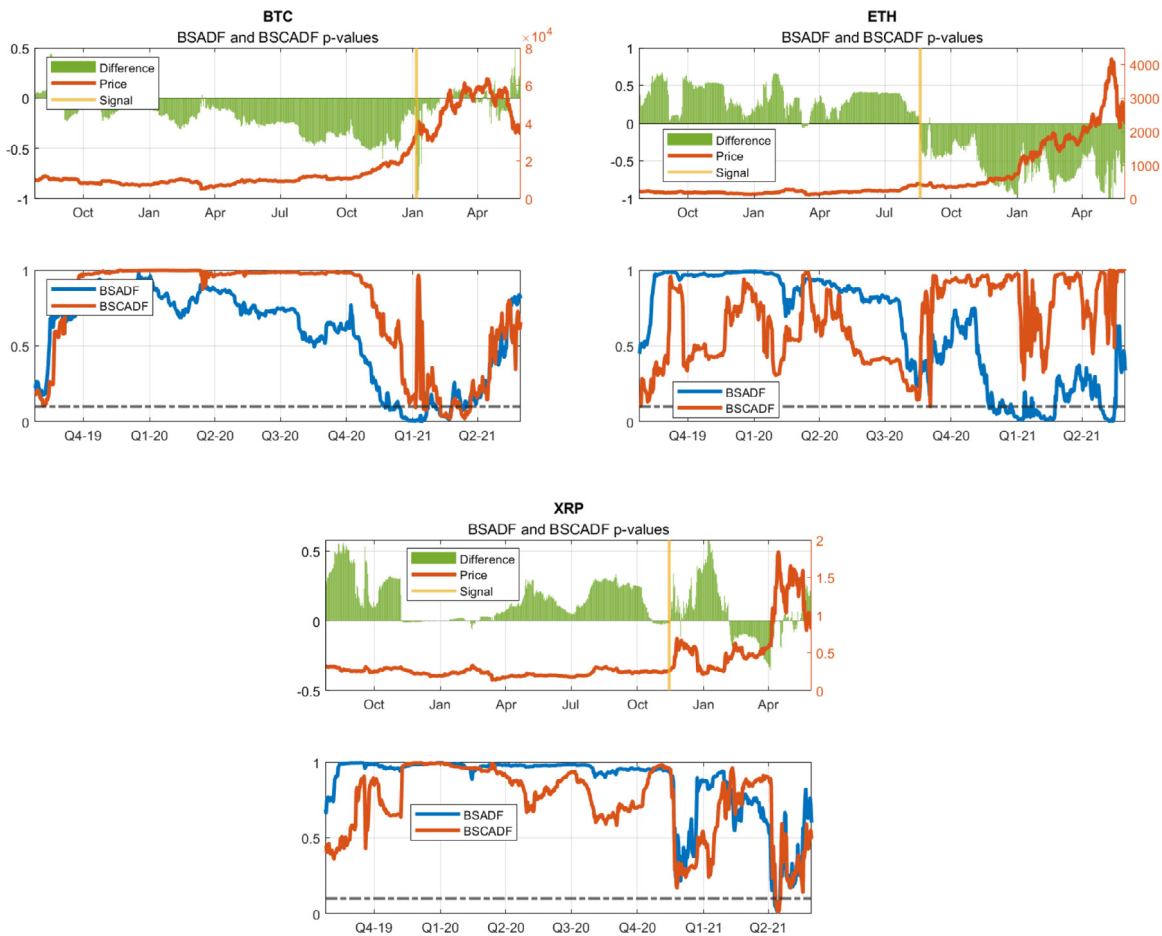
**Fig. 3.** BSADF and BSCADF Brain Sentiment Index test  $p$ -values. The figure shows the BSADF and BSCADF  $p$ -values (bottom panels), their difference and closing price dynamics (top panels) of the three cryptocurrencies, using the Brain Sentiment Index as a covariate, over the full sample period. Black dashed lines indicate the 10% significance level.

**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 specification 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).

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

that during the period of tranquil market dynamics the BSADF and BSCADF test statistics tend to comove strongly, giving rise 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



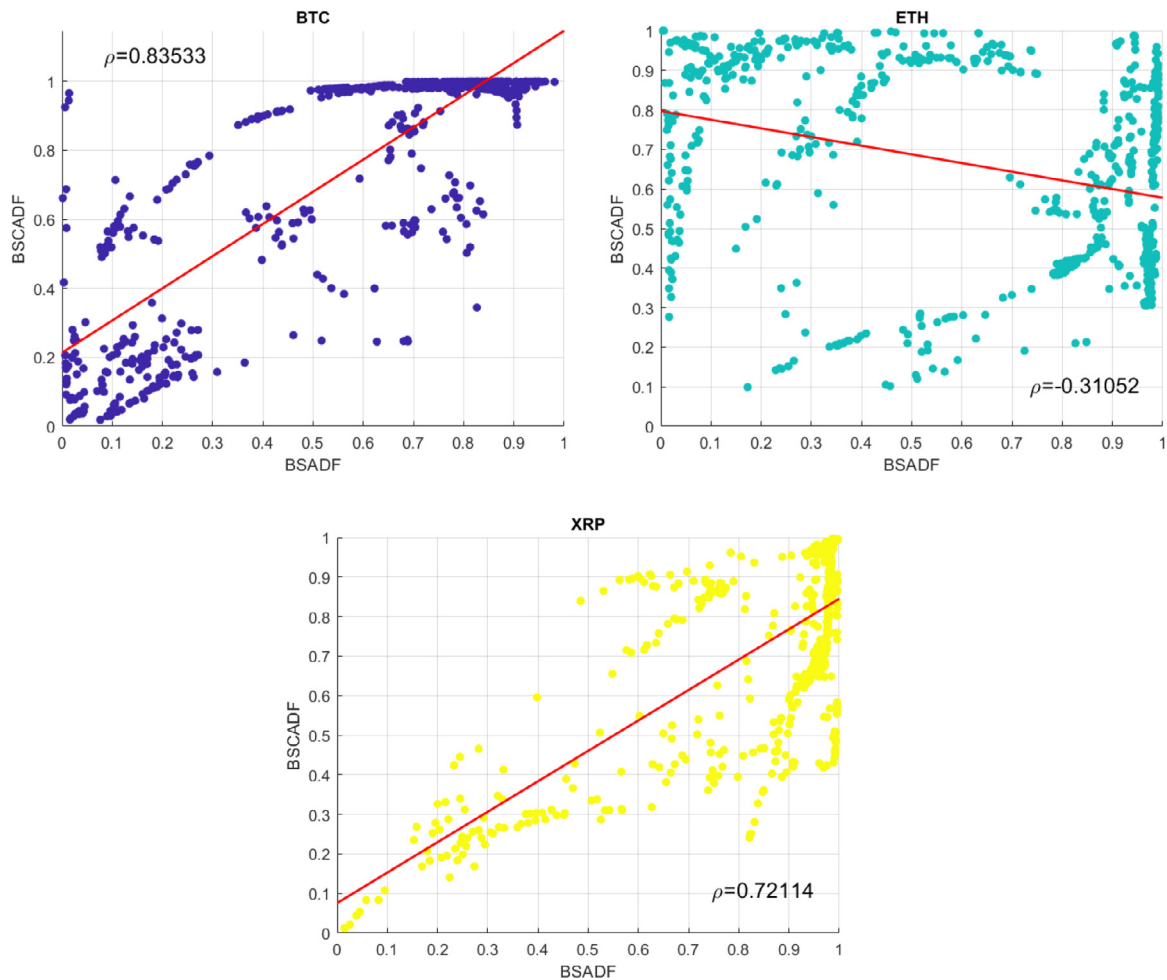
**Fig. 4.** BSADF and BSCADF Google Search Index test  $p$ -values. The figure shows the BSADF and BSCADF  $p$ -values (bottom panels), their difference and closing price dynamics (top panels) of the three cryptocurrencies, using the Google Search Index as a covariate, over the full sample period. Black dashed lines indicate the 10% significance level.

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 with a consequent surge in 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. In contrast, when 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 analysed 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 begin to be explosive, even conditional on market sentiment.

On the other hand, Fig. 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 not yet 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 the possible positive, negative and neutral polarity of news, might be more useful as an early warning indicator of time series explosiveness than search volumes. Indeed, while the latter does measure the level of interest of the community in 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.





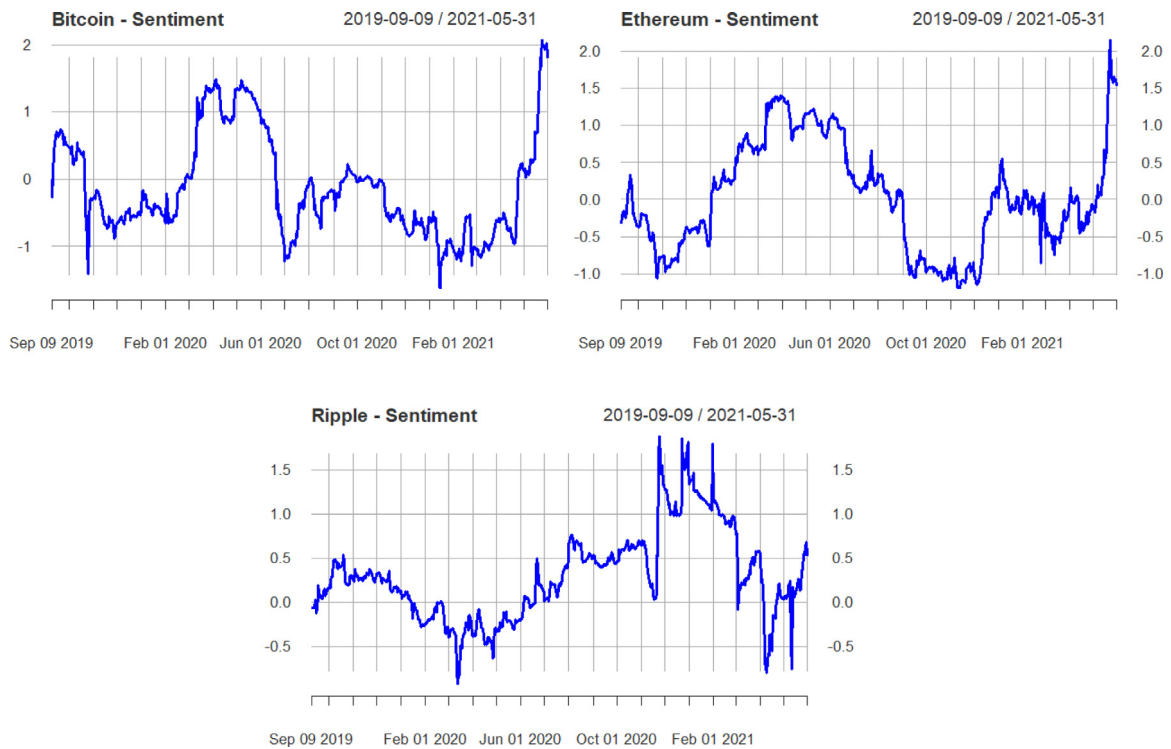
**Fig. 5.** Scatter plot of BSADF and Brain Sentiment Index BSCADF  $p$ -values. The figure shows the scatter plot of the BSADF and Brain Sentiment Index BSCADF  $p$ -values related to the three cryptocurrencies over the considered sample period. The value of  $\rho$  indicates the correlation between the two test statistics.

With the aim of more deeply comparing the differences in the BSADF and BSCADF test outcomes, Fig. 5 shows the scatter plot of the BSADF and Brain Sentiment Index BSCADF test  $p$ -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 Fig. A.10. From the figure we notice that there exists a relatively high degree of correlation 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 Fig. A.10 in Appendix.

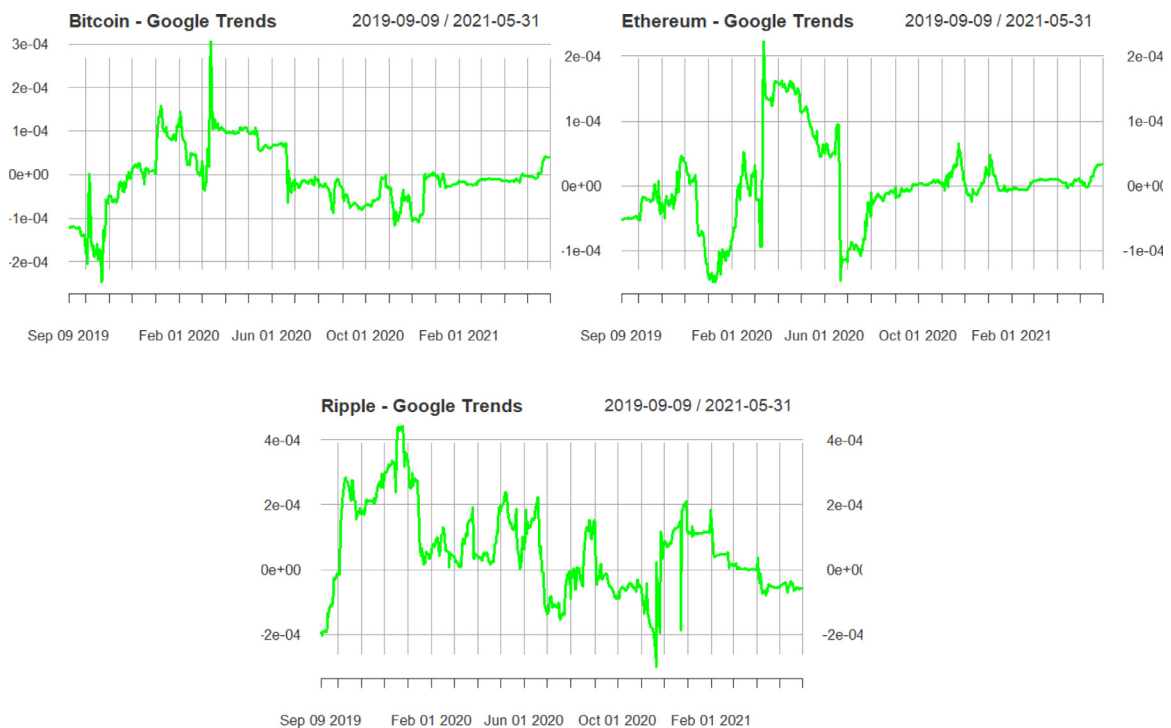
To gain further insight into the time-changing relationship between the sentiment indicator and the cryptocurrency price behaviour, we perform a rolling linear regression exercise, over 100-day windows, where the response variable is the cryptocurrency return and the regressor is the lagged sentiment indicator. Fig. 6 shows that, starting from September 2020, the estimated coefficients associated with the sentiment indicators grow sharply. It can be noticed from Fig. 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.

#### 4. Concluding remarks

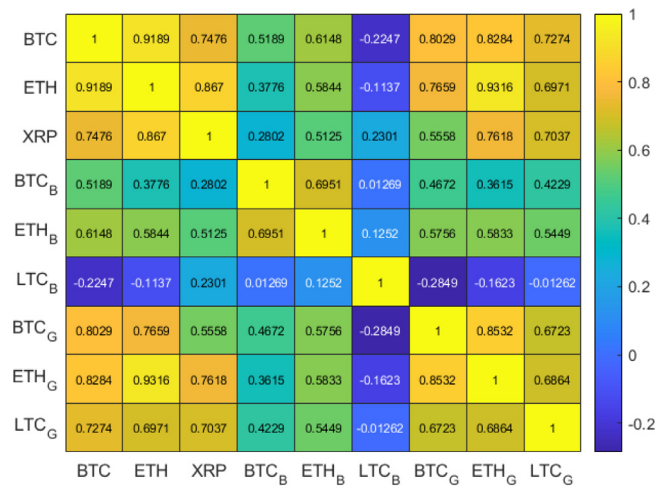
The recent advent of high-volatile, sentiment-based markets, such as cryptocurrency, calls for the design and implementation of innovative data science techniques, based upon sound econometric and statistical models, which can provide a timely signal of explosivity in the price behaviour of digital currency,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



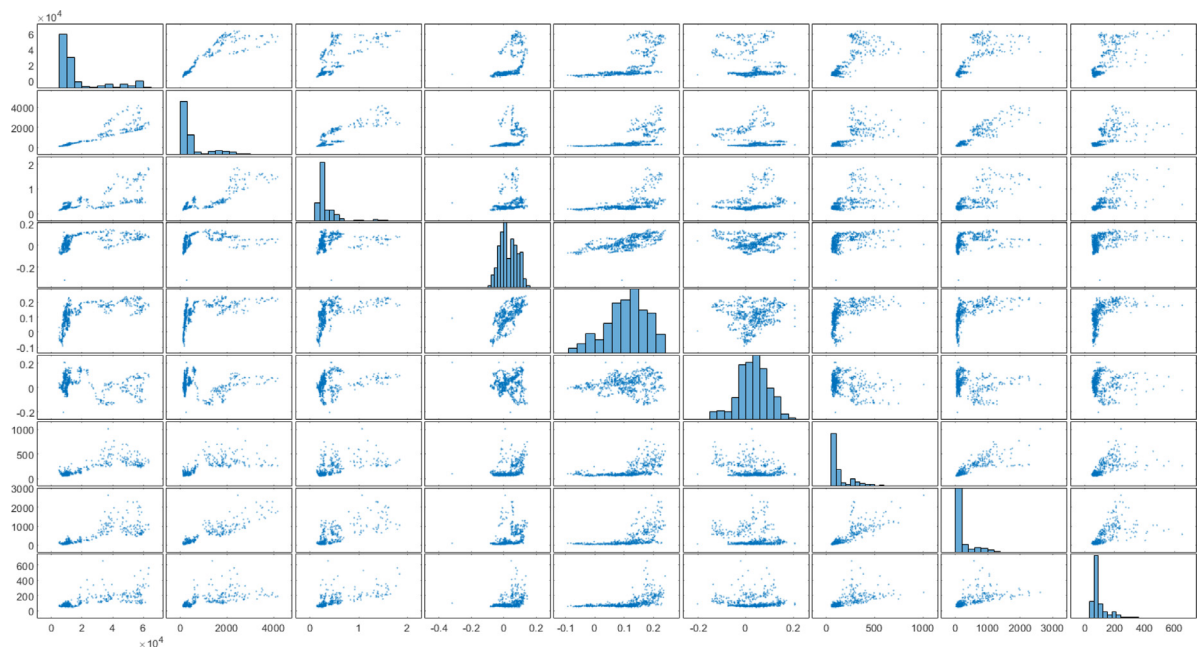
**Fig. 6.** Sentiment Index rolling regression coefficients. The figure shows the estimated coefficients associated with the lagged Brain Sentiment index for the selected cryptocurrencies in a rolling linear regression exercise, where the response variable is the cryptocurrency return. The rolling window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



**Fig. 7.** Google Search Index rolling regression coefficients. The figure shows the estimated coefficients associated with 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.



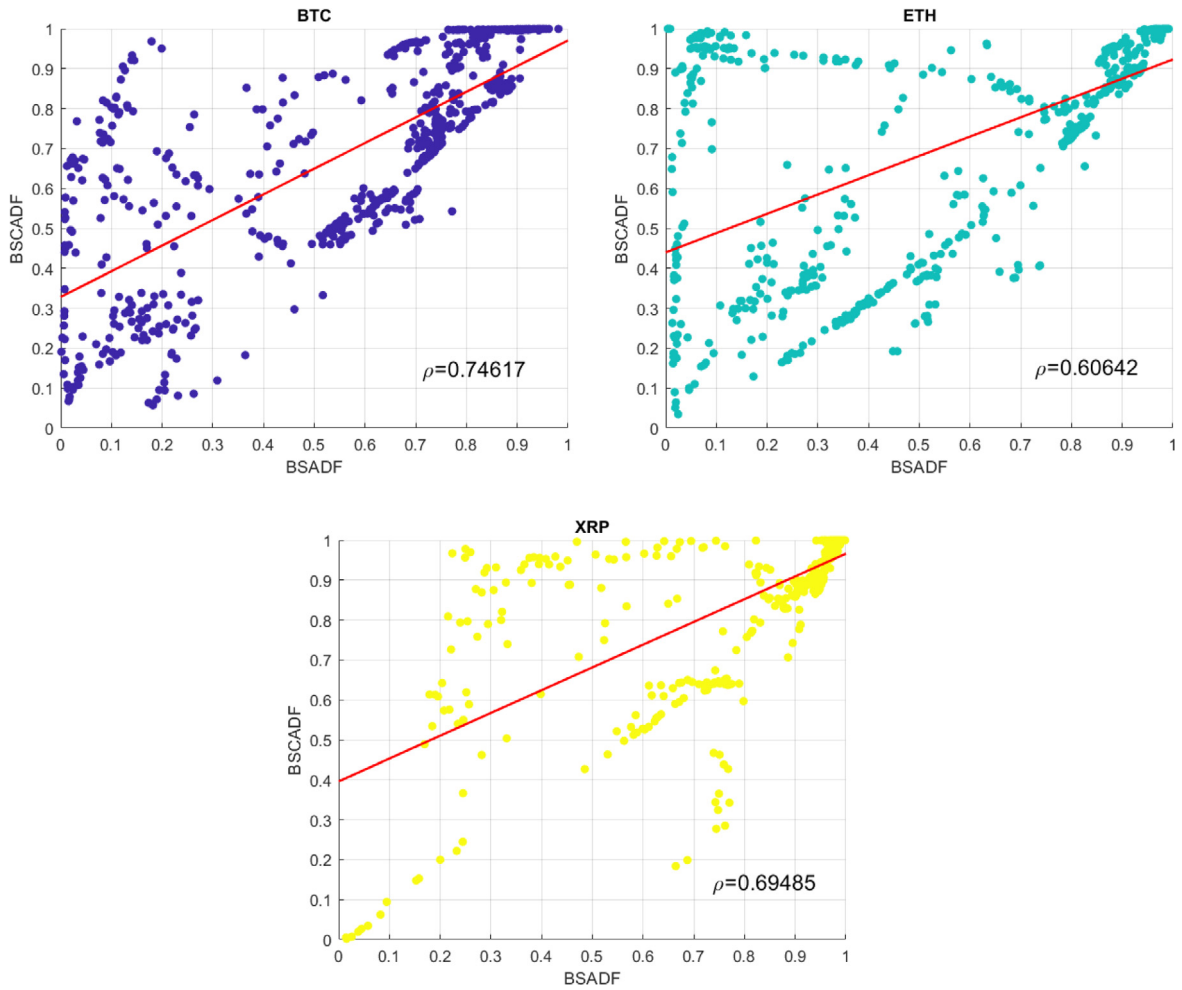
**Fig. A.8.** Data correlation matrix. The figure shows the correlation matrix of Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) prices, Brain Sentiment Indicator (B) and Google Search Indices (G) over the full sample period.



**Fig. A.9.** Scatter and empirical distribution function plots. The figure 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.

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 that takes into account the dependence of the price series on any relevant covariates, to detect explosive price behaviour. This approach allows us to improve bubble identification, providing a more effective signal of the exponential growth of cryptocurrency prices.

Our results show that the BSCADF with sentiment does significantly deviate 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 contrast, during explosive price phases, the two tests tend to converge in supporting the rejection of the random walk hypothesis in favour of explosivity. Additionally, our results provide evidence 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.



**Fig. A.10.** Scatter plot of BSADF and Google Search Index BSCADF  $p$ -values. The figure shows the scatter plot of the BSADF and Google Search Index BSCADF  $p$ -values related for the three cryptocurrencies over the considered sample period. The value of  $\rho$  indicates the correlation between the two test statistics.

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 is due 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 of taking into account multiple covariates when detecting time series explosiveness. It is worth mentioning that additional text-based data Cerchiello and Giudici [47] and cyber risk determinants Aldasoro et al. [48] could be incorporated into the model for bubble prediction purposes. Moreover, another natural extension of the present work, which would take into account a different class of models, is represented by regime-switching ones enriched with a Bayesian estimation process. In particular, we could refer to the Markov switching GARCH model, which can recognise structural breaks in the variance endogenously, as shown in Bauwens et al. [49]. Such an approach would be useful for comparative purposes, although our main interest in the present work is in detecting which external drivers would be able to improve the power or the size of the considered BSCADF test.

**Originality** The data employed in this paper and some preliminary results have been presented at SIS 2021 conference (50th edition of the Scientific Meeting of the Italian Statistical Society), but the article in its present form has not been published elsewhere.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## Acknowledgments

This research was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101016233.

## Appendix. Additional results

See Figs. A.8–A.10.

## References

- [1] S. Nakamoto, et al., Bitcoin: A peer-to-peer electronic cash system, 2008.
- [2] J.-Z. Huang, W. Huang, J. Ni, Predicting bitcoin returns using high-dimensional technical indicators, *J. Finance Data Sci.* 5 (3) (2019) 140–155.
- [3] M.J.S. de Souza, F.W. Almudhaf, B.M. Henrique, A.B.S. Negredo, D.G.F. Ramos, V.A. Sobreiro, H. Kimura, Can artificial intelligence enhance the bitcoin bonanza, *J. Finance Data Sci.* 5 (2) (2019) 83–98.
- [4] R.S. Lee, COSMOS trader-chaotic neuro-oscillatory multiagent financial prediction and trading system, *J. Finance Data Sci.* 5 (2) (2019) 61–82.
- [5] P. Jaquart, D. Dann, C. Weinhardt, Short-term bitcoin market prediction via machine learning, *J. Finance Data Sci.* 7 (2021) 45–66.
- [6] S. Corbet, A. Meegan, C. Larkin, B. Lucey, L. Yarovaya, Exploring the dynamic relationships between cryptocurrencies and other financial assets, *Econom. Lett.* 165 (2018) 28–34.
- [7] P. Giudici, P. Pagnottoni, High frequency price change spillovers in bitcoin markets, *Risks* 7 (4) (2019) 111.
- [8] P. Giudici, P. Pagnottoni, Vector error correction models to measure connectedness of bitcoin exchange markets, *Appl. Stoch. Models Bus. Ind.* 36 (1) (2020) 95–109.
- [9] M. Brandvold, P. Molnár, K. Vagstad, O.C.A. Valstad, Price discovery on bitcoin exchanges, *J. Int. Financial Mark. Inst. Money* 36 (2015) 18–35.
- [10] P. Pagnottoni, T. Dimpfl, Price discovery on bitcoin markets, *Digit. Finance* 1 (1) (2019) 139–161.
- [11] P. Giudici, G. Polinesi, Crypto price discovery through correlation networks, *Ann. Oper. Res.* 299 (1) (2021) 443–457.
- [12] T. Dimpfl, F.J. Peter, Nothing but noise? Price discovery across cryptocurrency exchanges, *J. Financial Mark.* (2020) 100584.
- [13] E. Bouri, C.K.M. Lau, B. Lucey, D. Roubaud, Trading volume and the predictability of return and volatility in the cryptocurrency market, *Finance Res. Lett.* 29 (2019) 340–346.
- [14] M. Resta, P. Pagnottoni, M.E. De Giuli, Technical analysis on the bitcoin market: Trading opportunities or investors' pitfall? *Risks* 8 (2) (2020) 44.
- [15] J.-Z. Huang, Z.J. Huang, Testing moving average trading strategies on ETFs, *J. Empir. Financ.* 57 (2020) 16–32.
- [16] A. Cheung, E. Roca, J.-J. Su, Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices, *Appl. Econ.* 47 (23) (2015) 2348–2358.
- [17] P. Phillips, Y. Wu, J. Yu, Explosive behavior in the 1990s NASDAQ: when did exuberance escalate asset values, *Internat. Econom. Rev.* 52 (1) (2011) 201–226.
- [18] J. Fry, E. Cheah, Negative bubbles and shocks in cryptocurrency markets, *Int. Rev. Financ. Anal.* 47 (2016) 343–352.
- [19] S. Corbet, B. Lucey, L. Yarovaya, Datedstamping the bitcoin and ethereum bubbles, *Finance Res. Lett.* 26 (2018) 81–88.
- [20] N. Kyriazis, S. Papadamou, S. Corbet, A systematic review of the bubble dynamics of cryptocurrency prices, *Res. Int. Bus. Finance* 54 (2020) 101254.
- [21] E. Bouri, R. Gupta, D. Roubaud, Herding behaviour in cryptocurrencies, *Finance Res. Lett.* 29 (2019) 216–221.
- [22] A. Agosto, A. Cafferata, Financial bubbles: a study of co-explosivity in the cryptocurrency market, *Risks* 8 (2) (2020) 34.
- [23] M. Gronwald, How explosive are cryptocurrency prices? *Finance Res. Lett.* 38 (2021) 101603.
- [24] A. Cretarola, G. Figà-Talamanca, Detecting bubbles in bitcoin price dynamics via market exuberance, *Ann. Oper. Res.* (2019) 1–21.
- [25] A. Cretarola, G. Figà-Talamanca, Bubble regime identification in an attention-based model for bitcoin and ethereum price dynamics, *Econom. Lett.* 191 (2020) 108831.
- [26] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, *J. Comput. Sci.* 2 (1) (2011) 1–8.
- [27] I. Bordino, S. Battiston, G. Caldarelli, M. Cristelli, A. Ukkonen, I. Weber, Web search queries can predict stock market volumes, *PLoS One* 7 (7) (2012) e40014.
- [28] H. Choi, H. Varian, Predicting the present with google trends, *Econ. Rec.* 88 (s1) (2012) 2–9.
- [29] R. Feldman, Techniques and applications for sentiment analysis, *Commun. ACM* 56 (4) (2013) 82–89.
- [30] J.-Z. Huang, M. Rossi, Y. Wang, Sentiment and corporate bond valuations before and after the onset of the credit crisis, *J. Fixed Income* 25 (1) (2015) 34–57.
- [31] P. Cerchiello, P. Giudici, How to measure the quality of financial tweets, *Qual. Quant.* 50 (4) (2016) 1695–1713.
- [32] P. Cerchiello, G. Nicola, Assessing news contagion in finance, *Econometrics* 6 (1) (2018).
- [33] R. Scaramozzino, P. Cerchiello, T. Aste, Information theoretic causality detection between financial and sentiment data, *Entropy* 23 (5) (2021).
- [34] T. Aste, Cryptocurrency market structure: Connecting emotions and economics, *Digital Finance* 1 (1–4) (2019) 5–21.
- [35] G.M. Caporale, N. Pittis, Unit root testing using covariates: some theory and evidence, *Oxf. Bull. Econ. Stat.* 61 (4) (1999) 583–595.
- [36] E.B. Hansen, Rethinking the univariate approach to unit root testing: using covariates to increase power, *Econom. Theory* 11 (1995) 1148–1171.
- [37] K. Case, R. Shiller, Is there a bubble in the housing market? *Brook. Pap. Econ. Act.* 2 (2003) 299–362.
- [38] C. Dreger, R. Zhang, Is there a bubble in the Chinese housing market? *Urban Policy Res.* 31 (1) (2013) 27–39.
- [39] B. Corbet, L. Yarovaya, Datedstamping the bitcoin and ethereum bubbles, *Finance Res. Lett.* 26 (2018) 81–88.
- [40] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, *J. Amer. Statist. Assoc.* 74 (1979) 427–431.
- [41] P. Phillips, Y. Wu, J. Yu, Testing for multiple bubbles : Historical episodes of exuberance and collapse in the SP500, *Internat. Econom. Rev.* 56 (4) (2015) 1043–1078.
- [42] D. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59 (3) (1991) 817–858.
- [43] A. Zeileis, Econometric computing with HC and HAC covariance matrix estimators, *J. Stat. Softw.* 11 (10) (2004) 1–17.
- [44] A. Zeileis, Object-oriented computation of sandwich estimators, *J. Stat. Softw.* 16 (9) (2006) 1–16.
- [45] C. Kleiber, A. Zeileis, *Applied Econometrics with R*, Springer-Verlag, New York, 2008.
- [46] I. Korkos, On Explosive Time Series (Ph.D. thesis), Essex Business School, 2020.
- [47] P. Cerchiello, P. Giudici, Non parametric statistical models for on-line text classification, *Adv. Data Anal. Classif.* 6 (4) (2012) 277–288.
- [48] I. Aldasoro, L. Gambacorta, P. Giudici, T. Leach, The drivers of cyber risk, *J. Financial Stab.* 60 (2022) 100989.
- [49] L. Bauwens, B. De Backer, A. Dufays, A Bayesian method of change-point estimation with recurrent regimes: Application to GARCH models, *J. Empir. Financ.* 29 (C) (2014) 207–229.