AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(1): 896–917.
DOI: 10.3934/math.2024045
Received: 19 September 2023
Revised: 22 November 2023
Accepted: 24 November 2023
Published: 04 December 2023

Research article

Mild explocivity, persistent homology and cryptocurrencies' bubbles: An empirical exercise

Stelios Arvanitis* **and Michalis Detsis**

Department of Economics of Athens University of Economics and Business, Athens, Greece

* **Correspondence:** Email: stelios@aueb.gr; Tel: +302108203437.

Abstract: An empirical investigation was held regarding whether topological properties associated with point clouds formed by cryptocurrencies' prices could contain information on (locally) explosive dynamics of the processes involved. Those dynamics are associated with financial bubbles. The Phillips, Shi and Yu [33, 34] (PSY) timestamping method as well as notions associated with the Topological Data Analysis (TDA) like persistent simplicial homology and landscapes were employed on a dataset consisting of the time series of daily closing prices of the Bitcoin, Ethereum, Ripple and Litecoin. The note provides some empirical evidence that TDA could be useful in detecting and timestamping financial bubbles. If robust, such an empirical conclusion opens some interesting paths of further research.

Keywords: financial bubbles; mild explocivity; PSY; bubble detection and timestamping; topological data analysis; persistent simplicial homology; persistent landscapes; EGARCH; cryptocurrencies
Mathematics Subject Classification: 55, 62, 91

1. Introduction

The present note provides an empirical investigation into whether topological properties associated with point clouds formed by cryptocurrencies' prices could contain information on (locally) explosive dynamics of the processes involved. Those dynamics are associated with financial bubbles. The interest lies on both the issue of the statistical inference on the existence and timestamping of bubbles, as well as on the empirical predictability of their formation and/or termination dates. The empirical analysis here utilizes tools from the econometrics of locally explosive auto-regressive processes as well as from the Topological Data Analysis-hereafter TDA-toolkits.

Cryptocurrencies are financial applications of the technology of the blockchain. They, among others, facilitate financial transactions without the presence of financial intermediaries. They may be regarded as new financial assets distinct from usual currencies or commodities since they form a brand new framework of decentralized transaction tools, with total market capitalization of approximately one trillion US dollars.

The literature on cryptocurrencies is becoming voluminous (see, for example, Fang et al. [20] and Cai et al. [9] for comprehensive reviews and surveys). The empirical cryptocurrencies' analysis now occupies a large strand of the empirical finance literature, given research questions about their potential diversification benefits, their relations to other asset classes, etc. (see Anyfantaki et al. [1], along with the references therein). Indicative papers related to issues of price (co-) explosive dynamics and complexity are Bouri et al. [5], Bonifazi et al. [4], Haykir et al. [25] and Shahzad et al. [37], and where sentiment analysis concerning tweets were applied, Kukacka and Kristoufek [27]. In a related issue, indicative papers adhering to the existence of co-jumps in the cryptocurrencies' returns dynamics are Bouri et al. [6], Xu et al. [40] and Zhang et al. [41]. Indicative papers that investigate new proof systems that result to less variability, construct complexity indices associated with price dynamics out of relevant articles that appear in important databases or investigate periodic anomalies associated with prices are, respectively, Bazzanella and Gangemi [2], Lucey et al. [28] and Tosunoglu et al. [39].

Bubbles are known to form in the price processes of financial assets due to speculative behavior (see Diba and Grossman [16]). The determination of whether they have already occurred in some historical samples could be of interest to theoretical and empirical finance and economics. Early empirical detection of the formation and, more importantly, of the collapse of a financial bubble could also be important to theoreticians and/or practitioners since speculative bubbles may be associated to financial crashes, with sometimes detrimental effects for the functioning of the financial markets and the real economy.

Bubbles are partially latent. One methodology for their empirical detection and timestamping, which is also used in this note, is proposed by Phillips, Shi and Yu [33,34], hereafter PSY. The method is based on the Philips and Magdalinos [32] argument that underlying bubble behavior is signaled via locally explosive behavior of asset prices. The PSY method relies on right-tailed Dickey-Fuller unit root tests via a recursive estimation over rolling windows of increasing sizes. It can detect the existence of more than one bubble within a sample, as opposed to the method of Phillips, Wu and Yu [35]. Also, PSY can consistently timestamp bubbles associated with mildly explosive autoregressive linear dynamics, i.e., in such probabilistic environments it can estimate consistently the origination and the termination date. It thus provides an empirical account for the existence, duration and timestamping of in-sample speculative bubbles. One question regarding PSY is whether it is robust to local explosivity that deviates from linear dynamics.

TDA constitutes a recent and fast growing branch of computational and applied mathematics relying on the field of algebraic topology (see, for example, Hatcher [24] and Munkres [29]). Its applications spread out to several fields with highly significant contributions, such as the case of detecting a new subgroup of breast cancers (see Nicolau, Levine and Carlsson [30]) or the study concerning the spread of coronavirus (see Chen and Volie [13]). It extracts robust topological information from complex and high dimensional datasets with noisy elements with computational convenience. It can also provide useful tools for data analysis as it employs topological and geometric techniques see Edelsbrunner, Letscher and Zomorodian [18] in order to observe how data can be analyzed in specific spaces, how their analysis can be quantified and how statistics and other computational methods can be used for investigating a plethora of questions and topics in different fields and subfields, extracting useful and robust conclusions. For financial time series analysis see, for example, Gidea [22] and Gidea and

Katz [23].

The results of Gidea and Katz [23] essentially motivate the present empirical exercise. They find that topological information associated with persistent homology can provide an empirical early warning for financial crashes. The research question here is whether there is empirical evidence on whether TDA could either provide tools that could help in the detection and timestamping of speculative bubbles and/or provide some early indicators for their initiation and/or burst. Given the non-parametric and topological nature of the analysis, such tools could remain robust to deviations from linear locally explosive processes, while simultaneously detecting explosive patterns in the price dynamics emerging from the dynamic behavior of high conditional moments. As manifested by the aforementioned strand of literature that investigates issues of co-explosivity for cryptocurrencies, the empirically documented locally explosive dynamics of those imply that the TDA tools could be particularly suitable for their analysis. The fact that those tools can simultaneously incorporate joint information from several cryptocurrencies, as well as from explanatory variables like the complexity price indices of Lucey et al. [28] or the analogous indices in Rudkin et al. [36] (who also uses variables associated with other financial markets and commodity returns), could imply efficiency gains compared to one-dimensional timestamping technologies like the PSY. Thus, given that TDA applications are not employed in the aforementioned literature, they could constitute complementary methodologies for addressing research questions about cryptocurrencies' dynamics like the above. Specifically, in the present note tools related to persistent homology are employed in order to investigate whether there is empirical topological information that signals the formation or the beginning or collapse of financial bubbles already empirically timestamped via the aforementioned PSY.

The PSY method and the TDA are employed on a dataset consisting of the time series of daily closing prices for the four largest cryptocurrencies by market capitalization, i.e. the Bitcoin, Ethereum, Ripple and Litecoin. Those empiricals show asymmetric risk profiles since their returns exhibit high volatility along with significant (and often negative) empirical skewness and kurtosis (see Table 1 of summary statistics). Their dynamic behavior is consistent with the existence of bubbles and mild explosivity (see again Anyfantaki et al. [1]), making them an ideal dynamic empirical environment for the current research question.

TDA tools have been already applied to cryptocurrencies; as mentioned above, Rudkin et al. [36] employed persistence norms to embed volatility dynamics and connectedness between coins and complemented the topological analysis with explanatory variables from complexity indices and other financial returns. They also demonstrated an empirical ability of forewarning crashes. Our methodology differs in that we employ a partial modification of the aforementioned tools in order to obtain evidence on the possibility of explosive dynamics.

The results do not seem to indicate that TDA could provide early warnings of crashes, as in Gidea and Katz [23] or Rudkin et al. [36]. They, however, provide some empirical evidence that TDA could be useful in detecting and timestamping financial bubbles. Given the nonlinear dynamic behavior posited for cryptocurrencies by papers like Kukacka and Kristoufek [27], if robust, such an empirical conclusion opens some interesting paths of further research.

The remaining note is organized as follow: In the following section the PSY and TDA methodologies used are presented. In section three the data set is described and the empirical analysis provided, while the final section concludes.

2. Methodology

The present methodology consists of three steps. Initially, the PSY algorithm is applied on the dataset consisting of time series of daily logarithmic prices of the aforementioned four cryptocurrencies. This intends to detect and timestamp in-sample mildly explosive behavior (see Phillips and Magdalinos [32]). Any such period of explosive dynamics is interpreted as a speculative bubble. As mentioned before, the algorithm provides consistent estimates regarding the existence and the location of multiple bubbles (see Phillips, Shi and Yu [34]). As such, it is considered a benchmark technology for the detection of strong empirical evidence of linear locally explosive dynamics for the cryptocurrencies' prices (see Enoksen et al. [19]). It is then used here in order to provide reliable estimates of the cryptocurrencies' bubble periods.

The PSY methodology relies on the augmented Dickey-Fuller (ADF) test, which tests whether the prices follow a random walk against the alternative of mild explocivity (see Phillips and Magdalinos [32]). Specifically in our context, for the sample size $T \in \mathbb{N}^{\star}$ of the observed time series, $\{0, \ldots, T\}$ is partitioned in K mild-explosivity periods B_k, $k = 1, \ldots, K$ and the remaining stationary periods $\cap_{k=1}^{K} B_k^c$. Furthermore, the logarithmic price process of the j^{th} asset, $j = 1, \ldots, 4$, is assumed to be initiated by the random variable $X_{j,0} = O_p(1)$, and then to satisfy the auto-regressive recursion:

$$X_{j,t} = \left(1 + \sum_{k=1}^{K} \frac{C_{j,k}}{M(T, j, k)} \mathbb{I}\left\{t \in B_{j,k}\right\}\right) X_{j,t-1} + \varepsilon_{j,t},\ t > 0.$$

Here, the noise sequence $\left(\varepsilon_{j,t}\right)_{t \in \mathbb{N}}$ is assumed to be a stationary and strong mixing process with a mixing coefficient sequence that converges to zero at an appropriate analysis rate. This is general enough to allow for a large variety of linear and/or conditionally heteroskedastic noise processes typically used for the analysis of stationary parts of logarithmic returns in empirical finance (see, for example, Drost and Nijman [17]). Therefore, the PSY methods allows noise, stationary, ergodic and geometrically mixing Autoregressive Moving Average (ARMA) and/or multivariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and stochastic volatility types of temporal dynamics with parameter restrictions that are empirically relevant, as well as with innovation's marginal distributions that have sufficiently smooth densities (see, for example, Boussama et al. [7] for a case of a multivariate GARCH-type model).

Moreover, $C_{j,k}$ is a positive explosivity coefficient at the k^{th} explosive period $B_{j,k}$. $M(T, j, k)$ is strictly positive and diverges to infinity as $T \to \infty$, representing the rate at which the k^{th} explosive behavior vanishes as a function of T-e.g. $M(T, j, k) = \frac{c_{j,k}}{T^{\eta_k}}$, $c_{j,k} > 0$, $\eta_k \in (0, 1)$. The structure of the auto-regressive parameter is compatible with the existence of K sub-periods of non-stationary bubbles in parts of the asset process. K is not assumed predetermined, and a fortiori can be the case that $K \to \infty$ as $T \to \infty$; hence, the analysis allows for the number of bubbles not to be asymptotically stabilized.

The null hypothesis for the j^{th} asset, $j = 1, \ldots, 4$, is that $C_{j,k} = 0$ for all k and all possible K, thus it posits that the logarithmic prices globally have unit root dynamics. The alternative hypothesis posits the existence of at least one sub-period with mild explosivity. The PSY test relies on a recursive estimation of right tailed Dickey-Fuller tests over rolling windows of increasing sizes, where r_0 is the smallest sample window width fraction (specified by the user to initialize computation) and one is the largest window fraction, i.e., the total sample size.

Given a time series sample realization, say $(X_{t,j})_{t=1,\ldots,T}$, of the price process, the Generalized Supremum ADF test statistic (GADF) of the PSY method is defined as the supremum of the ADF statistic sequence over all feasible values of r_1 and r_2. Hereafter, dependence on j is supressed for brevity:

$$\mathrm{GADF}\,(r_0) := \sup_{\substack{r_2\in[r_0,1]\\ r_1\in[0,r_2-r_0]}} \mathrm{ADF}_{r_1,r_2},$$

where the starting points r_1 are allowed to vary within the range $[0, r_2 - r_0]$ and where r_2 is the respective endpoint of the estimation window. If the null hypothesis is rejected, the bubble-test procedure can also be used under general regularity conditions as a date-stamping strategy for the estimation of the origination and termination of bubbles. Specifically, the strategy performs a supremum ADF test on a backward expanding sample sequence where the endpoint of each sample is r_2 and the start point varies within the range $[0, r_2 - r_0]$. The corresponding ADF statistic sequence is $\left\{\mathrm{ADF}_{r_1}^{r_2}\right\}_{r_1\in[0,r_2-r_0]}$ and the Backward ADF statistic $\mathrm{BSADF}_{r_2}\,(r_0)$ is the supremum value of the ADF statistic sequence over this interval.

The initiation of the bubble is estimated at

$$r_{\text{init}} := \inf_{r_2\in[r_0,1]} \left\{r_2 : \mathrm{BSADF}_{r_2}\,(r_0) > cv_{r_2}\,(\alpha_T)\right\}$$

and the burst is estimated at

$$r_{\text{burst}} := \inf_{r_2\in[r_{\text{init}},1]} \left\{r_2 : \mathrm{BSADF}_{r_2}\,(r_0) < cv_{r_2}\,(\alpha_T)\right\},$$

where $cv_{r_2}\,(\alpha_T)$ is the $100\,(1-\alpha_T)$ critical value of the PSY statistic based on $\lfloor T r_2 \rceil$ observations. To eliminate Type I error, there is a need for $\alpha_T \to 0$ as $T \to \infty$.

In the second step of the methodology, TDA tools are applied on the aforementioned time series in order to obtain some empirical indication on the existence of topological information that could be either used as an early warning for the formation or the burst of a bubble or for the construction of alternative methods of timestamping. The tools used are related to the concept of persistent homology and they are implemented on the logarithmic returns of the cryptocurrencies under consideration. We work either with pairs or with the ensemble of cryptocurrencies, thus constructing via the relevant time series, point clouds inside $\mathbb{R}^2$ or $\mathbb{R}^4$, respectively (see Carlsson [10, 11]). Specifically, selecting a sliding window $w \ll T$, where as mentioned above T is the size of the sample, we construct $T - w + 1$ point clouds, each of which has the form of the $w \times 2$ and $w \times 4$ matrix $Y_t := \left(\ln X_{t,j} - \ln X_{i-1,j}, \ln X_{i+1,j} - \ln X_{t,j}, \ldots, \ln X_{w+i-1,j} - \ln X_{w+i-2,j}\right)_{t,j}$, $t = 1, \cdots, T - w + 1$, $j = 1, \ldots, m$, $m = 2, 4$, where $X_{t,j}$ as before is the observed price at the time instance i in the sample of the j^{th} cryptocurrency included in the analysis.

Then, for arbitrary $\varepsilon > 0$, each point cloud is transformed into an abstract simplicial complex. Specifically, the Vietoris-Rips complex, hereafter VRC- $\mathcal{R}\,(Y_t, \varepsilon)$, of the point cloud is considered (see, for example, Ch. 2 of Ghrist [21]). There, a k-simplex is actually the set of $k+1$ points in the cloud, if any, defined by the property that the Euclidean distance between each pair of points in the simplex is less than or equal to ε. Allowing the radius ε to vary, for each $t = 1, \cdots, T - w + 1$, a filtration of VRCs $(\mathcal{R}\,(Y_t, \varepsilon))_{\varepsilon>0}$ is obtained, since $0 < \varepsilon_1 < \varepsilon_2$ implies that $\mathcal{R}\,(Y_t, \varepsilon_1) \subseteq \mathcal{R}\,(Y_t, \varepsilon_2)$.

For each VRC, its k-dimensional simplicial homology group $H_k\,(\mathcal{R}\,(Y_t, \varepsilon))$ is considered (see, Ch. 4 of Ghrist [21]). There, $H_0\,(\mathcal{R}\,(Y_t, \varepsilon))$ is the group generated by independent elements that correspond to connected components, $H_1(\mathcal{R}(Y_t, \varepsilon))$ is generated by independent elements that correspond to loops, $H_2(R(Y_t, \varepsilon))$ is the group generated by elements that correspond to voids and, generally, $H_k(R(Y_t, \varepsilon))$ is the group generated by independent elements that correspond to k-dimensional holes. The simplicial homology (with integer coefficients) of $\mathcal{R}(Y_t, \varepsilon)$ is then the sequence of homology groups$(H_k(R(Y_t, \varepsilon)))_{k\in\mathbb{N}}$. The filtration property of the VRCs directly implies an analogous property for each level of homology; $0 < \varepsilon_1 < \varepsilon_2$ implies that $H_k(R(Y_t, \varepsilon_1)) \subseteq H_k(R(Y_t, \varepsilon_2))$. This means that for any k, there exists canonical inclusion homomorhisms $H_k(R(Y_t, \varepsilon_1)) \hookrightarrow H_k(R(Y_t, \varepsilon_2))$, which, along with obvious arguments of total boundedness for the point clouds at hand, implies that for any $\varepsilon > 0$, and any homology class $\mathfrak{c} \in H_k(R(Y_t, \varepsilon))$, there exists $0 < \varepsilon_1 \leq \varepsilon < \varepsilon_2$ such that $\forall 0 < \delta < \varepsilon_1$, $\mathfrak{c} \notin H_k(R(Y_t, \varepsilon_1 - \delta))$, $\mathfrak{c} \hookrightarrow \mathfrak{c}_{\varepsilon^\star} \neq 0$, $\forall \varepsilon_1 \leq \varepsilon^\star < \varepsilon_2$ and $\mathfrak{c} \hookrightarrow 0$, $\forall \varepsilon^\star \geq \varepsilon_2$. In simple terms, $\mathfrak{c}$ is born at the time ε_1 and dies at the time ε_2. Hereafter, $b_\mathfrak{c} := \varepsilon_1$ and $d_\mathfrak{c} := \varepsilon_2$ denote the birth and the death of the topological features represented by the particular homology class and the interval $[b_\mathfrak{c}, d_\mathfrak{c}]$ denotes the lifespan that it persists. The accounting of the lifespan of the underlying homology classes is then termed persistent homology; this constitutes an algebraic method that stores information about the lifespans of the topological features that reside in the VRCs. If a topological feature 'lives' for a large time period, then it is considered as a significant feature. On the other hand, if its 'life' is small, then it is considered a noisy one. In the present note analysis, it is restricted to $H_1(R(Y_t, \varepsilon))$ for each VRC and ε as in Gidea and Katz [23]. Those are expected to convey information on temporally persistent large gaps between pairs of logarithmic returns.

A way to represent the persistence information of the generators of the order one homology group is via persistence diagrams of order one (see, for example, Carlsson et al. [12]). Those are two-dimensional. Their horizontal axis shows birth values while their vertical one shows the death values. The diagrams contain the birth and death values of each homology class at the group, along with information about its multiplicity. They also contain the diagonal of $\mathbb{R}^2$, the points of which are interpreted as trivial homology generators with zero life span and infinite multiplicity. Analysis of the persistence diagrams can be facilitated by endowing the set of all possible suchlike diagrams with the Wasserstein metric of degree $p > 1$ (see Gidea and Katz [23] and the references therein); this has the advantage of pertubation robustness, yet it does not enjoy useful analytical properties like completeness that could facilitate statistical analysis. A way to overcome this is by embedding the aforementioned space, to some complete function space (see Bubenik [8]). As in the present work, the TDA methodology of Gidea and Katz [23] is followed and we work with such an embedding producing the notion of persistence landscapes of order one: If P is a persistence diagram (of order one) and $(b, d) \in P$ off the diagonal, define the piecewise linear function:

$$f(x) := \begin{cases} x - b & , x \in [b, (b+d)/2] \\ d - x & , x \in [(b+d)/2, d] \\ 0 & , x \notin [b, d]. \end{cases}$$

Then the first persistence landscape function is obtained as a pointwise maximum w.r.t. the off-diagonal elements of the persistence diagram:

$$\lambda_1(x) := \max \{ f(x) | (b, d) \in P,\ (b, d) \text{ non diagonal} \}.$$

Whenever the maximum does not exist as a real number, $\lambda_1(x)$ is set equal to zero. This is not, however, relevant to our analysis, which concerns by construction finite point clouds.

Figures 1 and 2 present a simple example of the methodology up to now; two point clouds with eight (left) and 10 (right) points are constructed from different pairs of cryptocurrencies. They correspond to Bitcoin, Ethereum observations (left) from 28 February 2016 up to 5 March 2016, and Ripple, Litecoin observations (right) 20 up to 29 November 2015. On the bottom panels of Figure 1, the resulting Vietoris-Rips simplicial complexes are constructed for a sufficiently large radius. Figure 2 presents the respective persistence diagrams and their corresponding persistence landscapes for the aforementioned complexes. The large radius chosen for the example produces trivial topological features for the associated simplices, which are interpreted as noisy features.

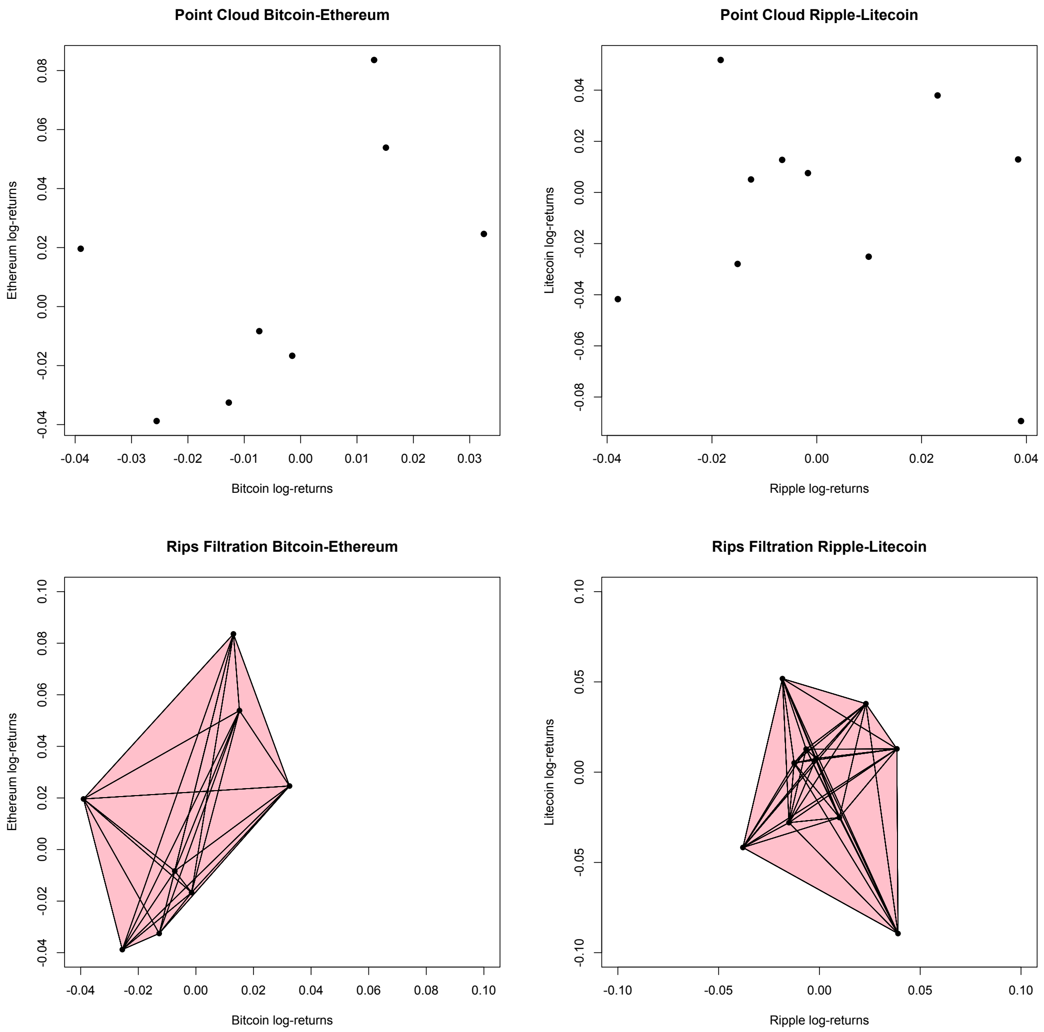

Figure 1. Two point clouds with eight (left) and ten (right) points are constructed from different pairs of cryptocurrencies. They correspond to Bitcoin, Ethereum observations (left) from 28 February 2016 up to 5 March 2016, and Ripple, Litecoin observations (right) 20 up to 29 November 2015. On the bottom panels, the resulting Vietoris-Rips simplicial complexes are constructed for a sufficiently large radius.

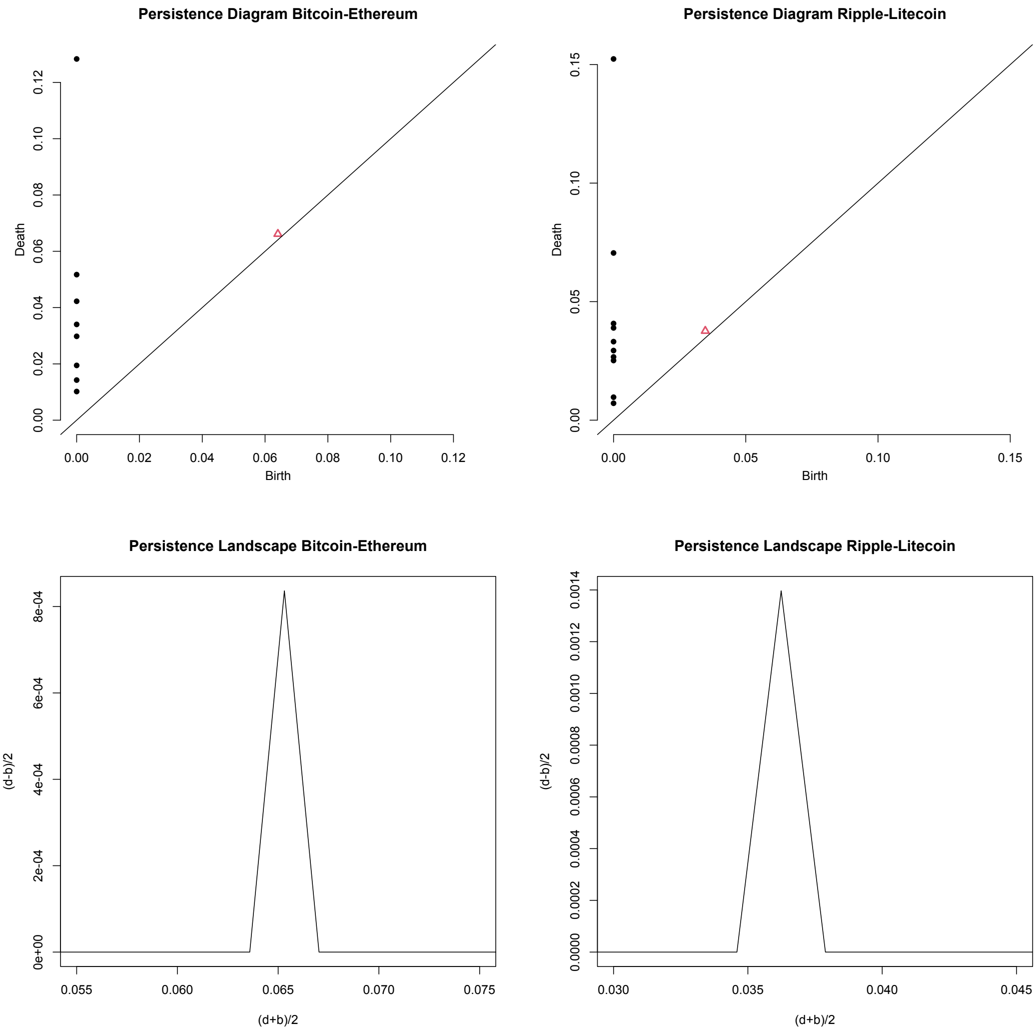

Figure 2. The respective persistence diagrams and their corresponding persistence landscapes are constructed for the complexes of the previous figure. The large radius chosen for the example produces trivial topological features for the associated simplices, which are interpreted as noisy features of the series.

Each persistent diagram of order one is represented via a bounded Lebesgue integrable real valued function. It lives on a function set that is completely metrized when endowed with the standard L^p-norm, $p \geq 1$, w.r.t. the Lebesgue measure, i.e., $\|\lambda_1\|_p^p := \int_{\mathbb{R}} |\lambda_1(x)|^p dx$. Thus, the topological information present in the persistent homology (of order one) of each point cloud in the analysis, is represented by a real number: The L^p norm of the associated persistent landscape of order one. When the above is performed at each point cloud defined in the rolling window, a time series of such norms is obtained $(\|\lambda_1\|_p)_{t=1,\cdots,n-w+1}$.

The analysis is then focused on properties of this series. Gidea and Katz [23] provide empirical evidence that the growth and the moving window variability of the L^p norms of particular financial time series seem to provide some early information of financial crashes. The present paper takes a somewhat different route. As in the previous work, the analysis is restricted to $p = 1, 2$. Instead of constructing explicit moving average variability filters for the times series of the associated norms, a volatility filter is provided via the maximization of the Gaussian (Quasi-) likelihood function of the Exponential Generalized Autoregressive Conditional Heteroskedasticity of order one (EGARCH(1,1)) model (see, for example, Straumann [38]). Specifically, the demeaned norms time series $(k_t)_{t=1,\cdots,T-w+1}$ is assumed to be approximated by a conditionally heteroskedastic process of the form: $k_t = z_t\sqrt{h_t}$, $h_t := \exp(\omega + \alpha z_{i-1} + \gamma|z_{i-1}| + \beta\ln(h_{i-1}))$, where z_t represents a martingale difference process, and h_t is a conditional volatility process obtained as a solution of the stochastic recurrence equation above. Hence, (k) is approximated by a martingale transform process, whereas both elements of the transform, along with the (pseudo-) true values of the associated parameter $\theta := (\omega, \alpha, \gamma, \beta)$, are to be optimally chosen. The particular conditionally heteroskedastic model is selected due to its versatility in embodying several stylized facts of conditional second moments of financial time series (see Straumann [38]). The methodology, however, allows for the consideration of other suchlike models.

Given the sample $(k_t)_{t=1,\cdots,T-w+1}$, and an initial condition $\hat{h}_1$ for the latent volatility process, the Gaussian Log-likelihood function $\ell(\omega, \alpha, \gamma, \beta; \hat{h}_0) := \sum_{t=1}^{T-w+1} \ln(\hat{h}_t(\theta)) + \frac{k_t^2}{\hat{h}_t(\theta)}$, $\ln(\hat{h}_t(\theta)) = \omega + \alpha\frac{k_{t-1}}{\sqrt{\hat{h}_t(\theta)}} + \gamma\frac{|k_{t-1}|}{\sqrt{\hat{h}_t(\theta)}} + \ln(\hat{h}_{t-1}(\theta))$, $i > 1$ is then maximized w.r.t. θ to obtain the Gaussian Quasi Maximum Likelihood Estimator (QMLE) $\hat{\theta}$ for the parameter, upon which the initial condition of the volatility filter $(\hat{h}_t(\hat{\theta}))_{t=1,T-w+1}$ is then constructed. Blasques et al. [3] provided sufficient conditions that ensure the strong approximation of the volatility filter by a process that conveys probabilistic properties of the associated time series as $T \to \infty$, even if the EGARCH model is misspecified, as expected. The resulting filter is the best maximum entropy approximation of the conditional second moment of the demeaned norms' time series; its marginal ergodic distribution minimizes the corresponding expected Kullback-Liebler divergence (see Cover and Thomas [15]) with the analogous distribution of the conditional second moment of k_0.

Finally, in the spirit of Gidea and Katz [23], the time-path of the filter is contrasted to the aforementioned reliable PSY timestamping of the bubbles. The purpose is to descriptively discern the relevance of the empirically extracted topological information with the bubbles' formations and bursts, indicated by the PSY method.

3. Empirical analysis

3.1. Data

The financial time series used in the analysis consists of the four cryptocurrencies' daily closing prices (in US dollars) that span the period between August 07, 2015 and August 31, 2021. The choice of the cryptocurrencies included on the analysis is based on their market cap and the data availability to the authors. The choice of the period of the analysis is also based on data availability in conjunction with the empirical observation that several episodes of price booms and bursts are recorded during this period for every cryptocurrency involved. The data was obtained from the Bitfinex exchange market through the CoinMarketCap. In total, the dataset involves 2.217 daily observations on each cryptocurrency involved. TDA is performed on the relevant daily logarithmic returns; this sample contains 2.216 observations for each cryptocurrency. Table 1 exhibits summary statistics for the latter.

Table 1. Summary statistics of the logarithmic returns of each cryptocurrency.

	Mean	S.D.	Skewness	Kurtosis
Bitcoin	0	0.04	-0.81	11.77
Ethereum	0	0.07	-3.23	67.17
Ripple	0	0.07	2.08	33.87
Litecoin	0	0.06	0.35	11.93

As mentioned in the introduction, those empirical moments suggest a risk profile that is characterized by high volatility (compared to the mean) and even higher levels of (absolute) skewness and kurtosis. It is noted though that the possibility of local non-stationarity for the returns could imply that those empirical moments may not be close to the analogous population moments (even if asymptotically stationary versions of the later are well defined).

3.2. Numerical analysis

The numerical aspects of the PSY and the TDA analyses are mostly performed inside the programming environment of R. Specifically, the package psymonitor is used for the PSY method and the package TDA is used for the extraction of persistence homology and landscapes. The optimization of the Gaussian log-likelihood function and the subsequent derivation of the filter was performed in Matlab, via the optimization routine fmincon.

3.3. Results

The PSY algorithm is applied to the daily logarithmic closing prices of each cryptocurrency. The minimum window size equals 106 implied by the algorithm's formula $T \cdot (0.01 + {}^{1.8}/\sqrt{T})$, where T is the length of logarithmic prices, i.e., 2.217. For each cryptocurrency, the sequence of BSADF test statistics is a vector of dimension 2.112. Its elements are compared to the analogous critical values obtained via bootstrap, and every exceedance is counted as a bubble date. Table 2 presents the bootstrap critical values for each cryptocurrency for the 90%, 95% and 99% significance levels.

Table 2. The bootstrap critical values for Bitcoin, Ethereum, Ripple and Litecoin.

	Bitcoin	Ethereum	Ripple	Litecoin
90%	0.26	0.28	0.27	0.34
95%	0.58	0.72	0.74	0.60
99%	1.42	1.35	1.36	1.39

The first day the BSADF test statistic lies above the corresponding 95% level critical values is counted as the origination day of a bubble. Given an origination, the consequent first day at which the statistic lies below the critical value counts as the termination date for the particular epoch of mild explosivity. Figure 3 depicts the resulting inference for the in-sample bubbles superimposed to the time series path of the logarithmic closing prices for the Bitcoin, Ethereum, Ripple and Litecoin.

Figure 3 presents the PSY method timestamping results. All cryptocurrencies seem to have several instances of explosive behavior for the time period at hand and at least one period of significant duration initiated during 2017 in all cases. Given our larger and more recent dataset, the current results refresh the analogous and similar results of Par. 3.1 of Anyfantaki, Arvanitis and Topaloglou [1].

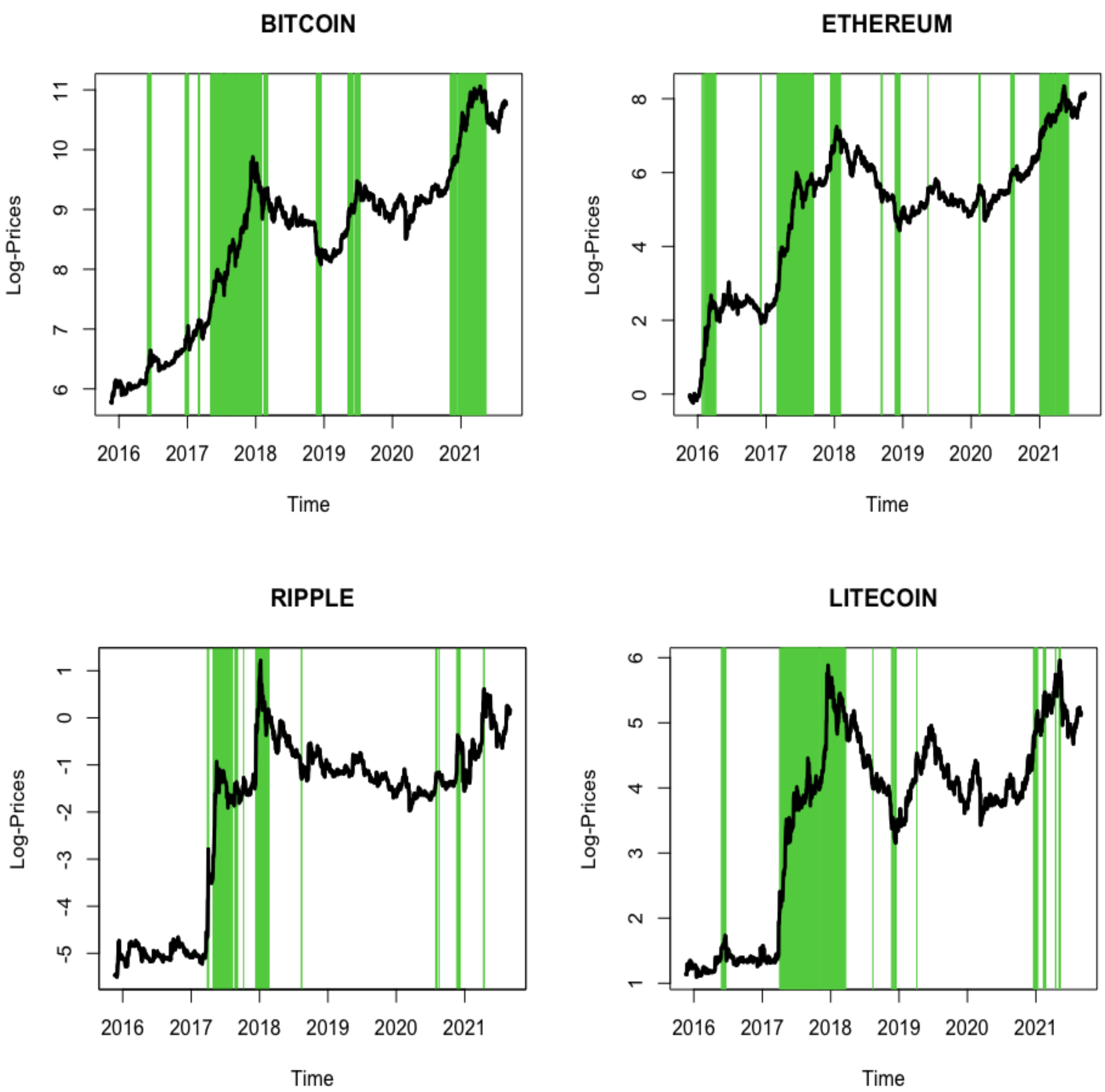

Figure 3. The timestamping of explosive periods generated by the PSY algorithm for each one of the cryptocurrencies separately. The green areas represent the periods of bubbles for Bitcoin (upper left), Ethereum (upper right), Ripple (lower left) and Litecoin (lower right), superimposed to the time series of the logarithmic closing prices of them.

The TDA methodology is then implemented to the time series of the logarithmic returns of the four cryptocurrencies each of 2.216 observations. The data is transformed to sequences of point clouds via a) the choice of the set of cryptocurrencies included in the analysis and b) the choice of the sliding window. As mentioned in the methodology section, the analysis is either performed on every possible pair of cryptocurrencies, hence the resulting clouds are subsets of $\mathbb{R}^2$, or it is performed in the totality of the assets, hence the resulting clouds are subsets of $\mathbb{R}^4$. The results of the analysis of the pairs are presented below in some indicative cases; the remaining cases are similar and available to the interested reader upon request. For b), two values for w are investigated. The first is relevant to the window used in the PSY procedure, i.e., $w = 105$. In this case, 2.112 point clouds are obtained. The second choice of sliding window covers the duration of the largest bubble period as estimated in-sample by PSY; $w' = 200$, then 2.017 point clouds are obtained. For each choice of the sliding window, the time series of the L^p-norms and normalized L^p-norms of the persistent landscapes of order one, for $p = 1, 2$, are then derived.

Figures 4 and 5 show the paths of the aforementioned norms for each choice of the sliding window. Superimposed to the PSY estimates of the bubble periods, it is noted that the trajectories for both the normalized norms for the point clouds are obtained from the totality of the cryptocurrencies. For both choices of the sliding window (see the fourth panel in Figures 4 and 5) they seem to have a neighborhood of their maxima at which the paths assume quite large and volatile values, and those neighborhoods seem to lie in close vicinity to the large duration bubbles filtered by the PSY analysis. Furthermore, normalization reduces the uniform distance between the p=1 and p=2 cases.

Figures 6–9 depict the analogous analyses for the respective filtered EGARCH volatility paths of the associated norms. Those are superimposed in each case to the PSY timestamps of the relevant bubbles.

Figures 6 and 7 show the results of analysis for the first choice of sliding windows w=105, whereas Figures 8 and 9 show the corresponding results for the second choice w=200. In the left panel of Figure 6, where the TDA is performed based on the pair Bitcoin-Ethereum, the results seem to suggest that the norms could be able to pick up at least the large bubble periods (as timestamped by the PSY method) for Bitcoin and Ethereum. This seems to be the case when all four cryptocurrencies are included in the TDA analysis in the right panel of Figure 6. The norm volatility filter appears to have smaller variation, but it seems to depict the main bubble episodes, even though it appears to have considerable variation in a period not associated by the PSY method with any major bubble episode for any of the four cryptocurrencies, namely, the mid to late 2016 period. The results in Figure 8 that correspond to a larger window, seem qualitatively worse, even though the large bubble period in 2017-18 seems to be partially depicted; the window chosen seems too large in order to pick up fine details and, therefore, the optimal choice of the window is-as expected-an issue for further investigation. It seems that an initial choice similar to the minimum windows size in the PSY method is functional.

Analogous remarks hold for Figures 7 and 9. The small window case seems to provide quite comparable results to the PSY timestamping for both the Ripple and Litecoin cryptocurrencies in both the right and the left panel analyses of Figure 7. The larger window results of Figure 9 seem inferior even though the large bubble periods seem partially depicted.

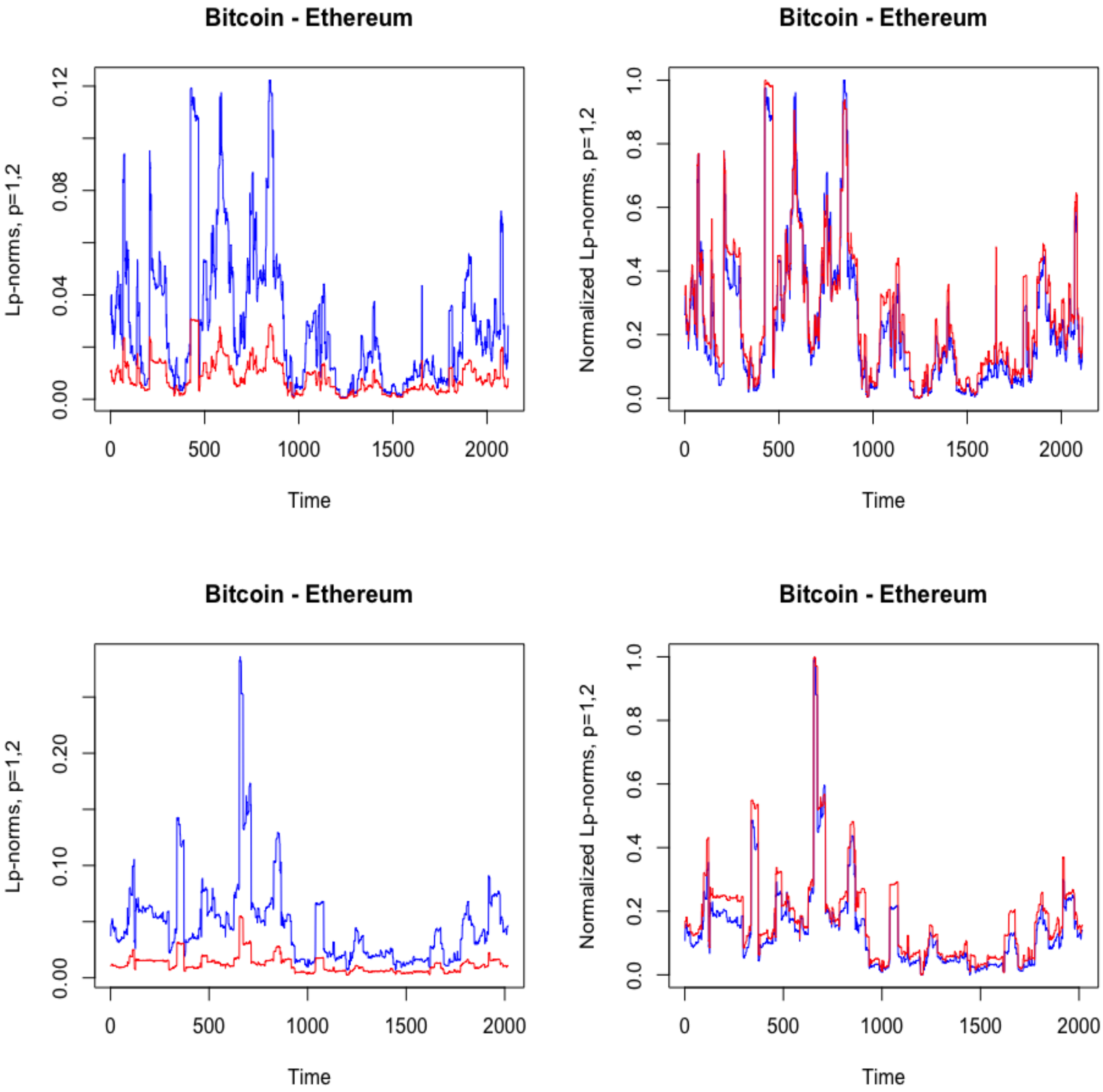

Figure 4. Time series of Lp-norms (left) and normalized L^p-norms (right), p = 1 (blue), p = 2 (red), for the pair Bitcoin - Ethereum, for w = 105 (up) and for w' = 200 (down). The larger window reduces variability in all cases. Significant variability in trajectories of the norms is in any case observed for both sliding windows approximately between the 500th and 1.000th observations. Normalization reduces the uniform distance between the p=1 and p=2 cases.

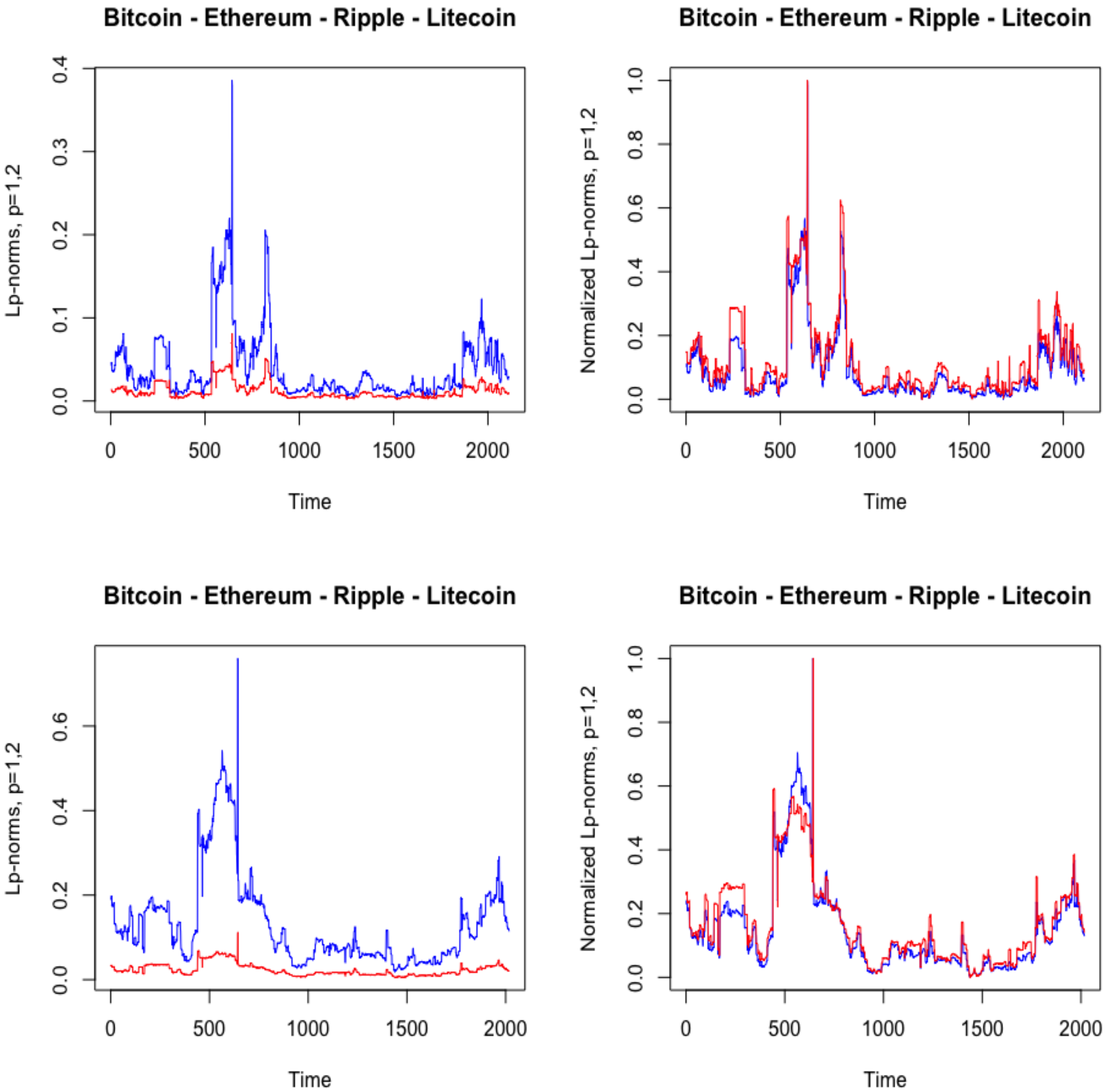

Figure 5. Time series of L^p-norms (left) and normalized L^p-norms (right), $p = 1$ (blue), $p = 2$ (red), for the ensemble of cryptocurrencies, for $w = 105$ (up) and for $w' = 200$ (down). Variability is not necessarily reduced for the larger window. Significant variability persists between the 500th and 1.000th observations for both the sliding windows. Normalization reduces the uniform distance between the p=1 and p=2 cases.

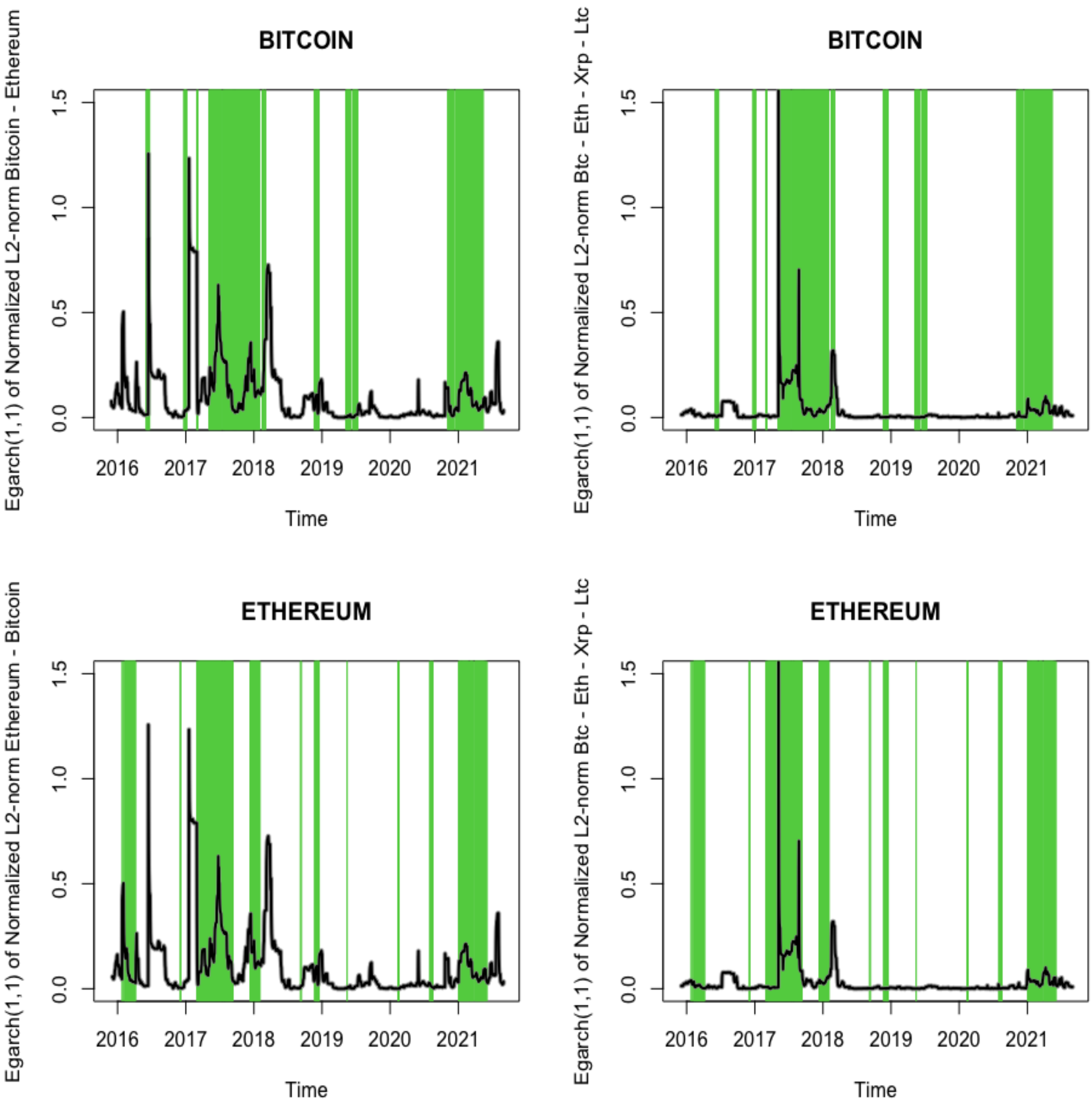

Figure 6. Time series of EGARCH(1,1) filtered volatility of the normalized L^2-norms for the pair Bitcoin - Ethereum (left panel) and the ensemble of cryptocurrencies (right panel), juxtaposed to the PSY timestamping of bubbles for the Bitcoin (upper panel) and Ethereum (lower panel) (w = 105). The norms seem be able to pick up at least the large bubble periods (as timestamped by the PSY method) for Bitcoin and Ethereum.

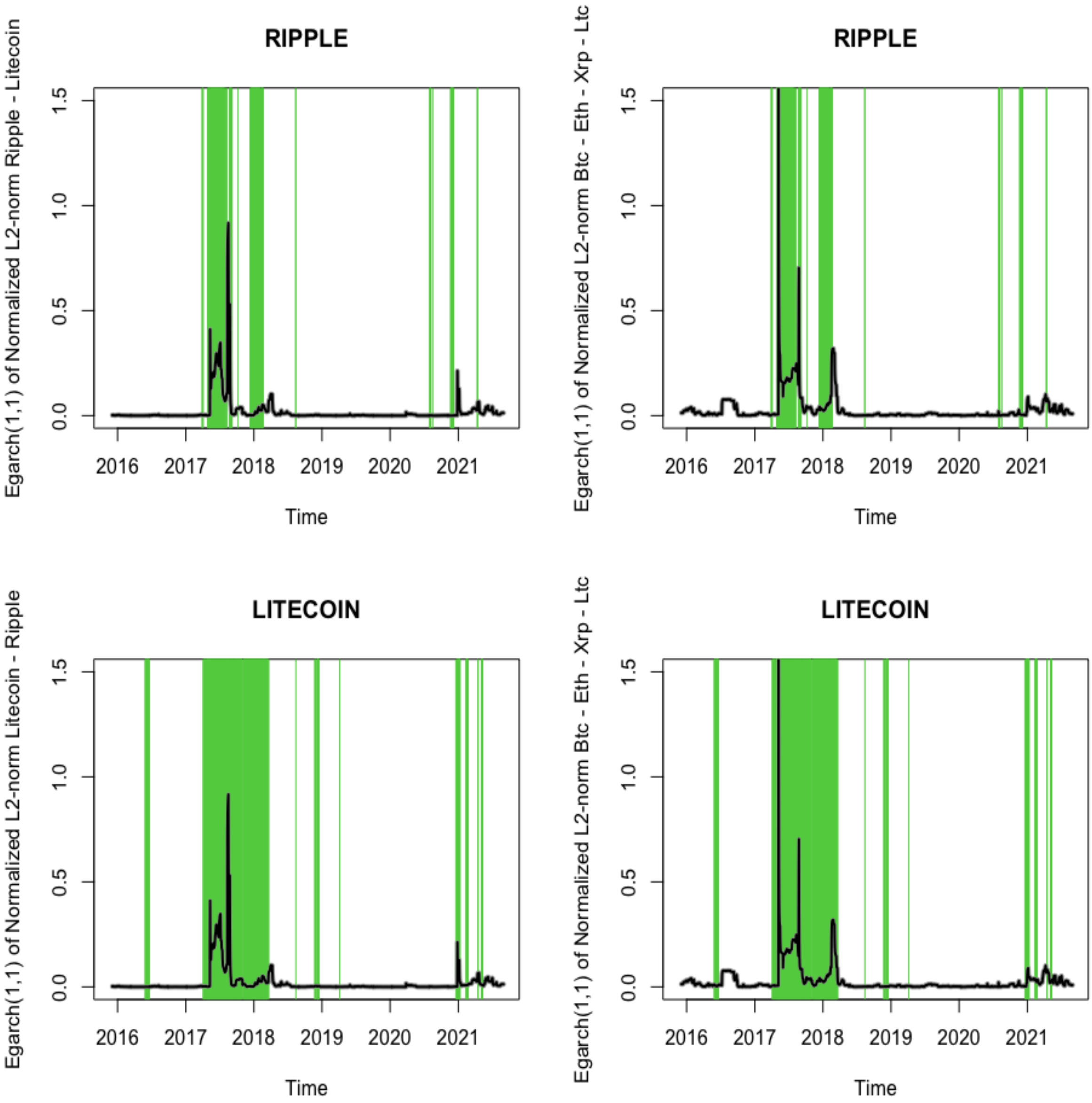

Figure 7. Time series of EGARCH(1,1) filtered volatility of the normalized L^2-norms for the pair Ripple - Litecoin, for the pair Bitcoin - Ethereum (left panel) and the ensemble of cryptocurrencies (right panel) juxtaposed to the PSY timestamping of bubbles for the periods of bubbles of Ripple (upper panel) and Litecoin (lower panel) ($w = 105$). As in the previous case, the norms seem be able to pick up at least the large bubble periods (as timestamped by the PSY method) for Ripple and Litecoin.

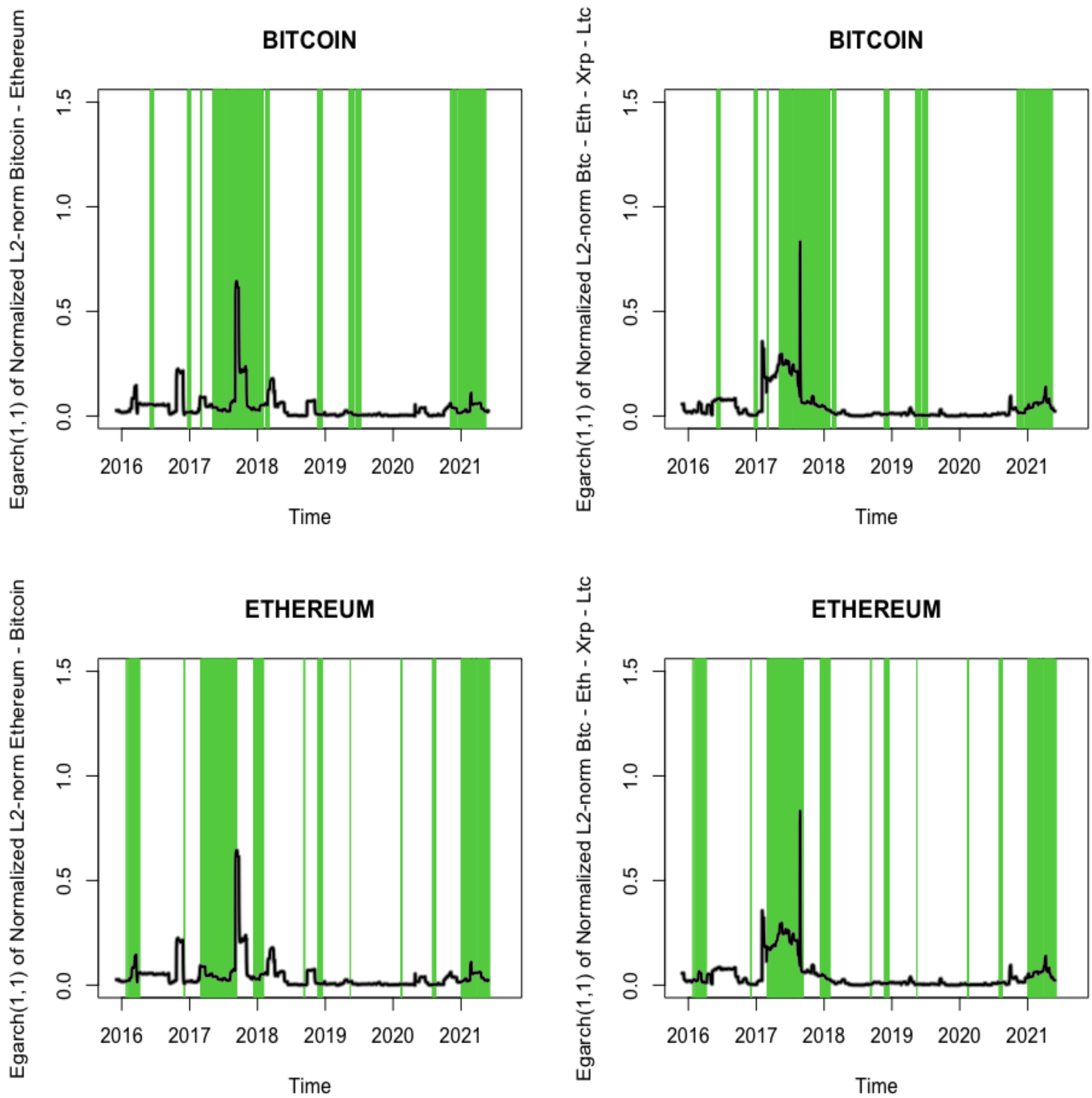

Figure 8. Time series of EGARCH(1,1) filtered volatility of the normalized L^2-norms for the pair Bitcoin - Ethereum (left panel) and the ensemble of cryptocurrencies (right panel), juxtaposed to the PSY timestamping of bubbles for the Bitcoin (upper panel) and Ethereum (lower panel) ($w' = 200$). The results seem qualitatively worse than the ones of Figure 6, even though the large bubble period in 2017-18 seems to be partially depicted.

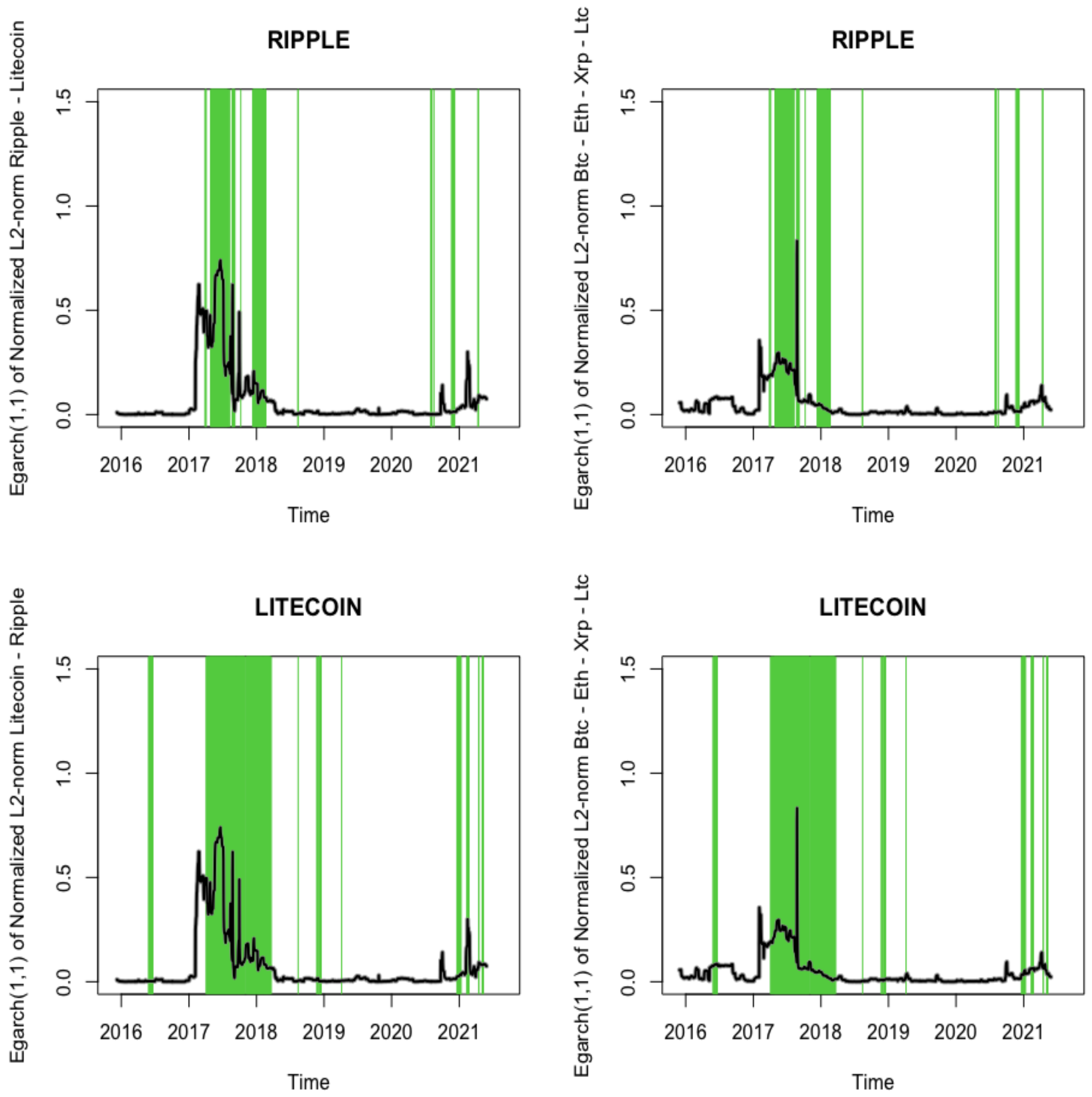

Figure 9. Time series of EGARCH(1,1) filtered volatility of the normalized L^2-norms for the pair Ripple - Litecoin(left panel) and the ensemble of cryptocurrencies (right panel) juxtaposed to the PSY timestamping of bubbles for the periods of bubbles of Ripple (upper panel) and Litecoin (lower panel) ($w' = 200$). The results seem qualitatively worse than the ones of Figure 7, even though the large bubble period in 2017-18 seems to be partially depicted.

4. Discussion and conclusions

The optimality or the robustness of the aforementioned results can be tested via further choices of the fudge parameters that appear in the analysis: The sliding windows that specify the sequence of available point clouds, the sets of assets that are included in the clouds, the orders of the homology groups utilized, the choice of norms and of the volatility models. Moreover, further models of conditional heteroskewness and/or heterokurtosis can be employed in order to assess the behavior of the aforementioned conditional higher moments of the norms of the persistent landscapes during bubbles.

The results seem somewhat dissimilar to the empirical results regarding early warnings of crashes of Gidea and Katz [23] or Rudkin et al. [36]. What is interpreted as an early warning there could be extreme variability of the associated topological information inside a bubble. The intensive activity of the trajectories of the norms and of their volatilities seem to be related to the PSY timestamps of bubbles of considerable duration, even though they do not seem to provide some early warning for the formation and/or the burst of speculative bubbles.

The above raise the following research question: Could, complementarily to methodologies like PSY, formal inferential procedures based on persistent homology and landscapes be designed for the detection and timestamping of time series' locally explosive behavior? This could be important for at least two reasons: First, the persistent homology approach can by construction accommodate joint information from multiple time series. As such it could potentially provide a timestamping tool for (say d-dimensional) vector locally explosive linear dynamics of the form:

$$\boldsymbol{X}_t = \left(\mathrm{Id} + \sum_{k=1}^{K} \frac{C_k}{M(T,k)} \mathbb{I}\{t \in B_k\}\right) \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t,\ t > 0,$$

where the explosive periods B_k are defined as in the previous section, C_k is a positive $d \times d$ explosivity coefficient matrix at the k^{th} explosive period, $M(T,k) > 0$ and diverges to infinity as $T \to \infty$ and $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{N}}$ is an $\mathbb{R}^d$-valued stationary and strong mixing noise process. Such vector dynamics provide more efficient characterizations of local explocivity, since at each bubble period k the structure of the explosivity coefficient matrix C_k is general enough to allow for intra-bubble dependence of currently explosive base assets on the dynamics of other currently and/or previously explosive assets, as well as on assets that are never explosive. Such structures are not accommodated by the one-dimensional PSY method. Several studies investigating issues of cryptocurrencies co-explocivity (see for example Bouri et al. [5]) use the PSY method to timestamp explosive behavior and tools like logistic regression in order to investigate inter-relatedness between bubbles in different cryptocurrencies. Procedures based on TDA, if developable, could help such investigations gain more efficiency.

Second, given its non-parametric topological nature, it could be quite robust to deviations from linear dynamics. For example, it could accommodate periods of explosivity emerging from processes of the form:

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} + \boldsymbol{z}_t \odot \boldsymbol{H}_t,\ t > 0,$$
$$\ln \boldsymbol{H}_t = \left(\mathrm{Id} + \sum_{k=1}^{K} \frac{C_k}{M(T,k)} \mathbb{I}\{t \in B_k\}\right) \ln \boldsymbol{H}_{t-1} + \boldsymbol{v}_{t-1},\ t > 0,$$

where $(\boldsymbol{z}_t)_{t\in\mathbb{N}}$ and $(\boldsymbol{v}_t)_{t\in\mathbb{N}}$ are $\mathbb{R}^d$-valued stationary martingale difference processes and $(\boldsymbol{H}_t)_{t\in\mathbb{N}}$ is a conditional stochastic volatility process. For the (pointwise) logarithm that obeys the mildly explosive linear dynamics for the empirical application of multivariate BEKK-GARCH-type models on cryptocurrencies, see Katsiampa et al. [26]. It could potentially timestamp bubble-like dynamic behavior as the one produced in the catastrophe-augmented diffusion model of Kukacka and Kristoufek [27]. To the best of our knowledge it is not known whether the PSY method remains robust to such deviations from linearity.

Moreover, it could be the case that the time series behavior of the L^p norms above is closely related to the uncertainty price indices of Lucey et al. [28] constructed from the prices' related articles that appear in relevant databases; the probit models results of Chowdhuryand and Damianov [14] reveal statistically significant relations between the PSY cryptocurrencies bubbles' timestamping and the aforementioned index. It could be of further interest to include such indices in the TDA analysis in order to enhance efficiency. Such time series (and/or indices that could be further constructed from textual/sentiment analyses on cryptocurrencies) could potentially assume the role of "topologically explanatory variables" and enhance the analysis by providing further sources of information besides observed prices and returns.

In any case, the specification of the probabilistic properties of the persistent landscapes that may carry the topological information of local explosivity could be of central importance to the aforementioned research question and is thereby left to future research. Such research could also be benefited from results related to the derivation of the limiting properties of random elements associated with persistent homology in the spirit of Owada [31], especially in cases where the point clouds employed are associated with non-stationary time series and heavy tailed marginals. This seems quite an exciting line of further research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

1. S. Anyfantaki, S. Arvanitis, N. Topaloglou, Diversification benefits in the cryptocurrency market under mild explosivity, *Eur. J. Oper. Res.*, **295** (2021), 378–393. https://doi.org/10.1016/j.ejor.2021.02.058
2. D. Bazzanella, A. Gangemi, Bitcoin: A new proof-of-work system with reduced variance, *Financ. Innov.*, **9** (2023), 91. https://doi.org/10.1186/s40854-023-00505-2
3. F. Blasques, P. Gorgi, S. J. Koopman, O. Wintenberger, Feasible invertibility conditions and maximum likelihood estimation for observation-driven models, *Electron. J. Statist.*, **12** (2018), 1019–1052. http://dx.doi.org/10.1214/18-EJS1416
4. G. Bonifazi, E. Corradini, D. Ursino, L. Virgili, A social network analysis based approach to investigate user behavior during a cryptocurrency speculative bubble, *J. Inform. Sci.*, **49** (2023), 1060–1085.
5. E. Bouri, D. Roubaudb, S. J. H. Shahzad, Co-explosivity in the cryptocurrency market, *Financ. Res. Lett.*, **29** (2019), 178–183. https://doi.org/10.1016/j.frl.2018.07.005
6. E. Bouri, D. Roubaudb, S. J. H. Shahzadb, Do Bitcoin and other large cryptocurrencies jump together? *Q. Rev. Econ. Financ.*, **76** (2020), 396–409. https://doi.org/10.1016/j.qref.2019.09.003

7. F. Boussama, F. Fuchs, R. Stelzer, Stationarity and geometric ergodicity of BEKK multivariate GARCH models, *Stoch. Proc. Appl.*, **121** (2011), 2331–2360. https://doi.org/10.1016/j.spa.2011.06.001

8. P. Bubenik, Statistical topological data analysis for persistence landscapes, *J. Mach. Learn. Res.*, **16** (2015), 77–102.

9. C. W. Cai, R. Xue, B. Zhou, Cryptocurrency puzzles: A comprehensive review and re-introduction, *J. Account. Lit.*, 2023. https://doi.org/10.1108/JAL-02-2023-0023

10. G. Carlsson, Topology and Data, *B. Am. Math. Soc.*, **46** (2009), 255–308. https://doi.org/10.1090/S0273-0979-09-01249-X

11. G. Carlsson, Topological pattern recognition for point cloud data, *Acta Numer.*, **23** (2014), 289–368. https://doi.org/10.1017/S0962492914000051

12. G. Carlsson, A. Zomorodian, A. Collins, L. Guibas, Persistence barcodes for shapes, *Int. Shape Model.*, **11** (2005), 149–188. https://doi.org/10.1145/1057432.1057449

13. Y. Chen, I. Volic, Topological data analysis model for the spread of the coronavirus, *PLoS One*, **16** (2021). https://doi.org/10.1371/journal.pone.0255584

14. M. S. R. Chowdhury, D. S. Damianov, Uncertainty and bubbles in cryptocurrencies: Evidence from newly developed uncertainty indices, *Int. Rev. Financ. Anal.*, **91** (2024), 1029–1049. https://doi.org/10.1016/j.irfa.2023.102949

15. T. Cover, J. Thomas, *Elements of information theory*, 2 Eds., Wiley, 2006.

16. B. T. Diba, H. I. Grossman, The theory of rational bubbles in stock prices, *Econ. J.*, **98** (1988), 746–754. https://doi.org/10.2307/2233912

17. F. C. Drost, T. E. Nijman, Temporal aggregation of GARCH processes, *Econometric*, **61** (1993), 909–927. https://doi.org/10.2307/2951767

18. H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification, *Discrete Comput. Geom.*, **28** (2002), 511–533. https://doi.org/10.1007/s00454-002-2885-2

19. F. A. Enoksen, C. J. Landsnes, K. Lucivjanska, P. Molnar, Understanding risk of bubbles in cryptocurrencies, *J. Econ. Behav. Organ.*, **176** (2020), 129–144. https://doi.org/10.1016/j.jebo.2020.05.005

20. F. Fang, C. Ventre, M. Basios, L. Kanthan, D. Martinez-Rego, F. Wu, et al., Cryptocurrency trading: A comprehensive survey, *Financ. Innov.*, **8** (2022), 13. https://doi.org/10.1186/s40854-021-00321-6

21. R. W. Ghrist, *Elementary applied topology*, Seattle: Createspace, 2014.

22. M. Gidea, *Topology data analysis of critical transitions in financial networks*, International Conference and School on Network Science, Berlin: Springer, 2017, 47–59. https://doi.org/10.48550/arXiv.1701.06081

23. M. Gidea, Y. Katz, Topological data analysis of financial time series: Landscapes of crashes, *Physica A*, **491** (2018), 820–834. https://doi.org/10.1016/j.physa.2017.09.028

24. A. Hatcher, *Algebraic topology*, Cambridge University Press, 2001.

25. O. Haykir, I. Yagli, Speculative bubbles and herding in cryptocurrencies, *Financ. Innov.*, **8** (2022), 78. https://doi.org/10.1186/s40854-022-00383-0

26. P. Katsiampa, S. Corbet, B. Lucey, Volatility spillover effects in leading cryptocurrencies: A BEKK-MGARCH analysis, *Financ. Res. Lett.*, **29** (2019), 68–74. https://doi.org/10.1016/j.frl.2019.03.009
27. J. Kukacka, L. Kristoufek, Fundamental and speculative components of the cryptocurrency pricing dynamics, *Financ. Innov.*, **9** (2023), 61. https://doi.org/10.1186/s40854-023-00465-7
28. B. Lucey, S. Vigne, L. Yarovaya, Y. Wang, The cryptocurrency uncertainty index, *Financ. Res. Lett.*, **45** (2022), 102147. https://doi.org/10.1016/j.frl.2021.102147
29. J. R. Munkres, *Elements of algebraic topology*, Boca Raton, Florida, USA: CRC Press, 2018.
30. M. Nicolau, A. J. Levine, G. Carlsson, Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival, *P. Natl. Acad. Sci. USA*, **108** (2011), 7265–7270. https://doi.org/10.1073/pnas.1102826108
31. T. Owada, Limit theorems for Betti numbers of extreme sample clouds with application to persistence barcodes, *Ann. Probab.*, **28** (2018), 2814–2854. https://doi.org/10.1214/17-AAP1375
32. P. Phillips, T. Magdalinos, Limit theory for moderate deviations from a unit root, *J. Econometrics*, **136** (2007), 115–130. https://doi.org/10.1016/j.jeconom.2005.08.002
33. P. Phillips, S. Shi, J. Yu, Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500, *Int. Econ. Rev.*, **56** (2015), 1043–1078. https://doi.org/10.1111/iere.12132
34. P. Phillips, S. Shi, J. Yu, Testing for multiple bubbles: Limit theory of real-time detectors, *Int. Econ. Rev.*, **56** (2015), 1079–1134. https://doi.org/10.1111/iere.12131
35. P. Phillips, Y. Wu, J. Yu, Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *Int. Econ. Rev.*, **52** (2011), 201–226. https://doi.org/10.1111/j.1468-2354.2010.00625.x
36. S. Rudkin, W. Rudkin, P. Dlotko, On the topology of cryptocurrency markets, *Int. Rev. Financ. Anal.*, **89** (2023), 102759. https://doi.org/10.1016/j.irfa.2023.102759
37. S. J. H. Shahzad, M. Anas, E. Bouri, Price explosiveness in cryptocurrencies and Elon Musk's tweets, *Financ. Res. Lett.*, **47** (2022), 102695. https://doi.org/10.1016/j.frl.2022.102695
38. D. Straumann, *Lecture notes in statistics: Estimation in conditionally heteroskedastic time series models*, Berlin: Springer, 2005.
39. N. Tosunoglu, H. Abaci, G. Ates, N. S. Akkaya, Artificial neural network analysis of the day of the week anomaly in cryptocurrencies, *Financ. Innov.*, **9** (2023), 88. https://doi.org/10.1186/s40854-023-00499-x
40. F. Xu, E. Bouri, O. Cepni, Blockchain and crypto-exposed US companies and major cryptocurrencies: The role of jumps and co-jumps, *Financ. Res. Lett.*, **50** (2022), 103201. https://doi.org/10.1016/j.frl.2022.103201
41. L. Zhang, E. Bouri, Y. Chen, Co-jump dynamicity in the cryptocurrency market: A network modelling perspective, *Financ. Res. Lett.*, **58** (2023), 104372. https://doi.org/10.1016/j.frl.2023.104372