

WHY TOPOLOGICAL DATA ANALYSIS DETECTS FINANCIAL BUBBLES?

SAMUEL W. AKINGBADE, MARIAN GIDEA, MATTEO MANZI,
AND VAHID NATEGHI

ABSTRACT. We present a heuristic argument for the propensity of Topological Data Analysis (TDA) to detect early warning signals of critical transitions in financial time series. Our argument is based on the Log-Periodic Power Law Singularity (LPPLS) model, which characterizes financial bubbles as super-exponential growth (or decay) of an asset price superimposed with oscillations increasing in frequency and decreasing in amplitude when approaching a critical transition (tipping point). We show that whenever the LPPLS model is fitting with the data, TDA generates early warning signals. As an application, we illustrate this approach on a sample of positive and negative bubbles in the Bitcoin historical price.

1. INTRODUCTION

Topological Data Analysis (TDA) has emerged in recent years as a powerful methodology in time-series analysis and signal processing. A prevalent approach relies on time-delay coordinate embedding, which is used to construct, from the time-series, a point-cloud in some high-dimensional Euclidean space [Tak81]. The dynamics on the point-cloud allows one to recover the evolution laws driving the time-series. The shape of the point-cloud can be characterized via Persistent Homology – a fundamental tool in TDA – which measures the number of holes of a geometric approximation of the point-cloud, as well as their ‘visibility’ at any given ‘resolution’ level. Applying a sliding window to a time series allows one to measure how the shape of the point-cloud associated to each window changes over time. The robustness of persistent homology of the point-cloud under small perturbations makes this approach particularly suitable for analyzing noisy time series. The method can be used, for example, for detection of periodic behavior, attractors, and bifurcations. More broadly, it can be used for detection of critical transitions (tipping points). These represent abrupt shifts of a system from one steady state to another, due to small changes in external conditions. Critical transitions are relevant to many disciplines, including finance, climate, ecology, and biomedicine; see, e.g., [SBB⁺09, Len11, TS11, DCB⁺12, BSP⁺21]. Some applications of TDA to detect critical transitions have been developed in [BGVJ14, BG14, SSGC⁺18, IYG20, DUC20].

An important class of applications of TDA concerns critical transitions in financial time series, particularly market crashes and financial bubbles. A positive bubble represents a rapid increase of the price of an asset – which does not reflect the asset’s intrinsic value – followed by a crash. A negative bubble represents a rapid decay followed by a quick rebound. In this context, the asset price peak in a positive bubble prior to a crash, and the trough in a negative bubble prior to a rebound, represent tipping points of the system. There is a practical interest in the detection of early warning signals of critical transitions, which would allow minimizing the losses from a bubble crash, or maximize the profit from a rebound.

It is important to distinguish between crashes of endogenous origins, caused by factors within the financial market, and crashes of exogenous origins, caused by external factors. On the one hand, endogenous crashes are associated with speculative unsustainable accelerating price growth, being fueled by self-reinforcements of cooperative herding and imitative behaviors of market agents. These determine specific, precursory signatures in the trajectory of the price [Bla79, BW82, Fer06, SH03, JS⁺10]. On the other hand, exogenous crashes may be caused by natural cataclysms, pandemics, fraud, or political events [SSZ22, SSZ21, ACAMNH21]. Exogenous crashes are less likely to yield early warning signals in comparison to the endogenous ones [Fry12].

Several mathematical models for endogenous bubbles have been proposed in recent years.

An important model – which we will focus on in this paper – is the Log-Periodic Power Law Singularity (LPPLS) model. This characterizes positive bubbles by super-exponential growth superimposed with oscillations of increasing frequency and decreasing amplitude when approaching the tipping point. A similar characterization holds for negative bubbles. The LPPLS model stems from the economic theory of rational expectation, behavioral finance of herding of traders, and the mathematical and statistical physics of bifurcations and phase transitions. See, e.g., [SJB96, JSW⁺96, FF96, JS99a, JS99b, JSL99, JLS00, YRWS12, LS13, SS16, Sor17].

Another significant model describes a bubble as a price process over a finite horizon, which is a strict local martingale (i.e., a local martingale that is not a true martingale) with respect to a risk neutral measure. The strict local martingale property is equivalent to the volatility function having a certain asymptotic behavior [Pro13]. One can test the specific asymptotic behavior of the volatility function, and assess the occurrence of bubbles, by using various estimators (parametric or non-parametric), or via Reproducing Kernel Hilbert Spaces (RKHS) theory [JKP11].

There are also models that characterize a bubble by three regimes that carry different processes: an initial normal regime, when the price follows a pure random walk, a second, market exuberance regime, when the price follows a mildly explosive process which ends in an abrupt collapse, and a final normal regime, when the price returns to a martingale process

[PWY11]. Other models have four regimes, as they insert a market recovery regime, between the collapse and the return to normal market behavior [PSY15a, PSY15b, HLS15]. Regression methods can be used to detect crises and estimate their associated turning points [PY11, HLST16, PS18].

In addition to model-based approaches, numerous empirical methods have also been developed to detect financial uncertainties and anomalies. TDA methods based on persistence homology have been used to study the 2000 dot-com bubble, the 2008 subprime mortgage crisis, and the 2018 cryptocurrency crash in [Gid17, GK18, GGK⁺20, KB21, RQD21]. Relatedly, Ball Mapper algorithms have been used to identify anomalies in the association between financial ratios and stock returns in [DR21].

Most of the evidence, so far, on the ability of TDA to detect financial bubbles has been empirical. A key question remains: what features of a time series yield to significant changes in the shape of the associated point cloud, more precisely, in terms of the holes that appear in the point-cloud? Some previous works attempted to explain this by drawing analogies between a market undergoing a crash and bifurcations in a complex dynamical system, or in terms of changes in drift and volatility of the strong non-stationary process described by a market near a crash [Gid17, GK18, GGK⁺20, AKV21, GK23].

In this paper, we propose a heuristic argument for why TDA is able to detect financial bubbles. As we are interested in early warning signals, we focus on the time period of a bubble before the tipping point. The ansatz is that the time series during such a period follows the LPPLS model. From a topological view-point, the time-delay embedding of the LPPLS oscillatory signal yields loops (surrounding holes) in the point-clouds. Changes in the frequency and/or amplitude of the oscillations yield changes in the structure of the holes in the point-cloud, which can be measured via persistent homology. When approaching the tipping point of a bubble, there is a significant change in the features of the oscillatory signal and consequently in the TDA output.

The upshot of our work is that whenever the LPPLS model for the detection of a positive (negative) bubble applies to some data set, the TDA method on that data set provides early warning signals of critical transition. As far as we know, this is the first time when the TDA method is justified in terms of a deterministic model for the expected dynamics of financial bubbles.

As an illustration, we apply both the LPPLS model and the TDA to Bitcoin – the largest cryptocurrency by market cap. Bitcoin has been experiencing numerous phases of extreme price growth and massive crashes, hence it provides a rich data set for exploring financial bubbles. Some related works on detection of bubbles in the cryptocurrency market include [FC16, GDS18, ADGK18, WWS18, RCPB19, WSH⁺19, GGK⁺20, ALGK20, SZ20b, SNC20].

Our experiments show a good agreement between the TDA method and the LPPLS model. Moreover, even when the LPPLS model fits poorly with some of the data, the TDA method can show relatively strong signal prior to the tipping point. While the TDA method is robust, the LPPLS appears to be sensitive to noise [BCP13].

The structure of the paper is as follows. In Section 2 we provide some background: in Section 2.1 we recall the LPPLS model; in Section 2.2 we briefly review the TDA approach based on persistence homology; in Section 2.3 we describe how to apply this method to time-series. In Section 3 we consider some examples of periodic, quasi-periodic, and frequency-changing oscillatory time-series, and observe the dependence of the TDA output on the time-series parameters as well as on the TDA parameters. The main contribution is Section 4, in which we explain how TDA method extracts topological features from time-series generated by the LPPLS model. Next, we illustrate the application of the TDA method on some concrete financial time-series. In Section 5 we review some methods to segment a given financial time series into up- and down-trends. In Section 6 we take the Bitcoin price data between October 1st, 2021 and July 1st, 2022, we segment it into up- and down-trends, and then we apply both the LPPLS model and the TDA method to the segments. We show the results in Appendix A. Conclusions and future directions are delineated in Section 7.

2. BACKGROUND

2.1. Log-Periodic Power Law Singularity (LPPLS) Model. Below we summarize the LPPLS model following [JSW⁺96, JLS00, YRWS12, SS16, Sor17, GDS18]. This is a model for both positive bubbles as well as for negative bubbles. In either case the market reaches a point when it no longer reflects the real underlying value of the asset (or it violates the market efficiency with respect to its time-scale [SS17]), which triggers a critical transition.

The LPPLS model can be derived from a network description of the market, which assumes that the trading of the asset is driven by two types of agents: a group consisting of traders with rational expectations, and another group formed by noise traders. Noise traders are likely to deviate from fundamental valuation in an accelerating mode, due to positive feedback and herding behaviour as a group. The collective behavior of the agents eventually leads to market instability by pushing the market away from an equilibrium between supply and demand. See [JLS00, SH05, SS16, Sor17].

It is important to keep in mind that the LPPLS model is concerned only with bubble generated endogenously by the dynamics of the market, and does not apply to bubbles of exogenous origins produced by external shocks.

According to the model, the asset price $p(t)$ during a (positive or negative) bubble evolves according to the stochastic differential equation

$$(2.1) \quad \frac{dp}{p} = B'(t_c - t)^{m-1} + C'(t_c - t)^{m-1} \cos(\omega \ln(t_c - t) - \phi') + \sigma(t)dW(t)$$

where t_c represents the critical time, $B', C', \phi' \in \mathbb{R}$, $0 < m < 1$ are parameters, $\sigma(t)$ is the volatility, and $dW(t)$ is the increment of a standard Wiener process with zero mean and unit variance. Hence the expectation of the logarithm of the asset price satisfies

$$(2.2) \quad \begin{aligned} E[\ln p(t)] &= A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi) \\ &= A + B(t_c - t)^m + C_1(t_c - t)^m \cos(\omega \ln(t_c - t)) \\ &\quad + C_2(t_c - t)^m \sin(\omega \ln(t_c - t)). \end{aligned}$$

Here $A > 0$, $0 < m < 1$, B, C and ϕ are parameters, whose role is explained below. Also, $C_1 = C \cos(\phi)$, $C_2 = -C \sin(\phi)$, and $C = \sqrt{C_1^2 + C_2^2}$.

The function in (2.2) has a singularity at $t = t_c$. Since

$$\lim_{t \rightarrow t_c} (t_c - t)^m \cos(\omega \ln(t_c - t)) = 0,$$

the function can be extended by continuity at $t = t_c$ by $E[\ln p(t_c)] = A$. The critical time t_c marks the termination of the bubble regime and the transition time to a different regime, which could be a major crash or rebound. The parameter A represents the expected value of the log price $\ln p(t)$ at the critical time t_c .

The condition $m < 1$ ensures that there is a singularity at the critical time t_c , while $m > 0$ ensures that the expected price remains finite at t_c . The parameter m is associated to the non-linearity of the trend, that is, $m \approx 0$ corresponds to a strong super-linear trend of $\ln p(t)$, while $m \approx 1$ corresponds to a nearly linear trend. The parameter m is also affecting, jointly with the parameters C_1 and C_2 , the changing amplitude of the oscillations.

The case when $B < 0$ corresponds to a super-exponential growth of the price $p(t)$ as the time t approaches t_c , while $B > 0$ corresponds to a super-exponential decay.

Thus, the main features of the model (2.1) are the following

- the price of the asset oscillates about a trend-line that is characterized by super-exponential growth or decay;
- the amplitude of the oscillations decreases when approaching the critical time t_c ;
- the frequency of the oscillations increases when approaching the critical time t_c ;
- the oscillations are superimposed with Gaussian noise.

The LPPLS model can be fitted to data and allows for predictions of the time of the crash [YRWS12, GDS18]. However, we note that the LPPLS model is fairly sensitive to the system parameters, to where to start and end

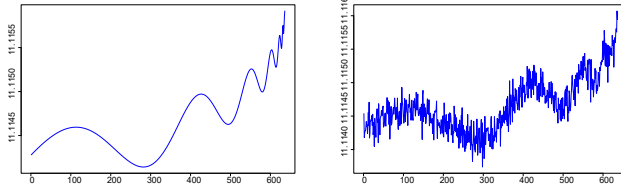


FIGURE 1. Synthetic LPPLS signal: without noise (left), with noise (right).

the fit of a bubble, and to the length of the data window (see [BCP13] and the references listed there).

2.2. Persistent homology. In this section we summarize the main topological tools that we will use: persistent homology and persistence landscapes. See [EH22, B⁺15, BD17]. Persistent homology converts a given point cloud into a geometric shape and characterizes it in terms of its homology generators. Persistence landscapes convert the information derived from persistent homology into elements of a functional Banach space. Through this process, a point cloud can be characterized by a single numerical value, which is the norm of the corresponding persistence landscape.

2.2.1. Point-cloud. The input of persistent homology is a point-cloud $Z = \{z_0, \dots, z_{m-1}\}$ embedded in some Euclidean space $\mathbb{R}^N$. The computation of persistent homology involves a sequence of steps.

2.2.2. Vietoris-Rips simplicial complex. We associate to the point cloud Z a topological space, thought of as a ‘shape’, depending on some parameter $\varepsilon > 0$, where ε represents the resolution at which the data is analyzed. The Vietoris-Rips simplicial complex $R(Z, \varepsilon)$, or, simply Rips complex, is defined as follows:

- For each $k = 0, 1, \dots$, a k -simplex of vertices $\{z_{i_0}, \dots, z_{i_k}\}$ is part of the simplicial complex $R(Z, \varepsilon)$ if and only if the mutual distance between any pair of its vertices is less than ε , that is

$$d(z_{i_j}, z_{i_{j'}}) < \varepsilon, \text{ for all } z_{i_j}, z_{i_{j'}} \in \{z_{i_0}, \dots, z_{i_k}\}.$$

In other words, a k -simplex is included in $R(Z, \varepsilon)$ whenever the vertices of that simplex are ‘indistinguishable from one another’ at a resolution level ε .

2.2.3. Filtration of simplicial complexes. The Rips simplicial complexes $R(Z, \varepsilon)$ form a filtration with respect to the resolution parameter ε , that is,

$$R(Z, \varepsilon) \subseteq R(Z, \varepsilon') \text{ provided } \varepsilon < \varepsilon'.$$

2.2.4. *Homology groups of simplicial complexes.* At each resolution level ε , the corresponding shape can be characterized in terms of its homology.

- For each complex $R(Z, \varepsilon)$, we compute its n -dimensional homology $H_n(R(Z, \varepsilon))$ with coefficients in some field, e.g., $\mathbb{Z}_2$.

The generators of the 0-dimensional homology module $H_0(R(Z, \varepsilon))$ correspond to the connected components of $R(Z, \varepsilon)$, the generators of the 1-dimensional homology module $H_1(R(Z, \varepsilon))$ correspond to ‘holes’ in $R(Z, \varepsilon)$, the generators of the 2-dimensional homology module $H_2(R(Z, \varepsilon))$ correspond to ‘voids’ in $R(Z, \varepsilon)$, etc. For the rest of the paper we will use only 1-dimensional homology.

2.2.5. *Filtration of homology modules.* The filtration property of the Rips complexes induces a filtration on the corresponding homologies, that is

$$H_n(R(Z, \varepsilon)) \subseteq H_n(R(Z, \varepsilon')) \text{ provided } \varepsilon < \varepsilon' \text{ for each } n.$$

These inclusions determine canonical homomorphisms

$$H_n(R(Z, \varepsilon)) \hookrightarrow H_n(R(Z, \varepsilon')), \text{ for } \varepsilon < \varepsilon'.$$

2.2.6. *Persistent homology.* For each non-zero n -dimensional homology generator α there exists a pair of values $\varepsilon_1 < \varepsilon_2$, such that:

- α is in $H_n(R(Z, \varepsilon_1))$ but is not in the image of any element in $H_n(R(Z, \varepsilon_1 - \delta))$ under the corresponding homomorphism, for $\delta > 0$;
- the image of α in $H_n(R(Z, \varepsilon'))$ is non-zero for all $\varepsilon_1 < \varepsilon' < \varepsilon_2$, but the image of α in $H_n(R(Z, \varepsilon_2))$ is zero.

In this case, one says that the generator α is ‘born’ at the parameter value $b_\alpha := \varepsilon_1$, and ‘dies’ at the parameter value $d_\alpha := \varepsilon_2$. The pair (b_α, d_α) represents the ‘birth’ and ‘death’ indices of α . The multiplicity $\mu_\alpha(b_\alpha, d_\alpha)$ of the point (b_α, d_α) equals the number of classes α that are born at b_α and die at d_α . This multiplicity is finite since the simplicial complex is finite.

2.2.7. *Persistence diagram.* The information on the n -dimensional homology generators at all resolution levels can be encoded in a *persistence diagram* P_n . Such a diagram consists of:

- for each n -dimensional homology class α one assigns a point $p_\alpha = p_\alpha(b_\alpha, d_\alpha) \in \mathbb{R}^2$ together with its multiplicity $\mu_\alpha = \mu_\alpha(b_\alpha, d_\alpha)$;
- in addition, P_n contains all points on the positive diagonal of $\mathbb{R}^2$; these points represent all trivial homology generators that are born and instantly die at every level; each point on the diagonal has infinite multiplicity.

The axes of a persistence diagram are birth values on the horizontal axis and death values on the vertical axis.

2.2.8. Persistence landscapes. The space of persistence diagrams can be embedded into a Banach space. One such embedding is based on *persistence landscapes*, consisting of sequences of functions in the Banach space $L^p(\mathbb{N} \times \mathbb{R})$, with $p \geq 1$.

- For each birth-death point $(b_\alpha, d_\alpha) \in P_n$, we first define a piecewise linear function

$$(2.3) \quad f_{(b_\alpha, d_\alpha)}(x) = \begin{cases} x - b_\alpha, & \text{if } x \in (b_\alpha, \frac{b_\alpha + d_\alpha}{2}]; \\ -x + d_\alpha, & \text{if } x \in (\frac{b_\alpha + d_\alpha}{2}, d_\alpha); \\ 0, & \text{if } x \notin (b_\alpha, d_\alpha). \end{cases}$$

- To a persistence diagram P_n consisting of a finite number of off-diagonal points, we associate a sequence of functions $\lambda_n = (\lambda_i)_{i \in \mathbb{N}}$, where $\lambda_n(i) : \mathbb{R} \rightarrow [0, +\infty]$ is given by

$$(2.4) \quad \lambda_n(i)(x) = i\text{-max}\{f_{(b_\alpha, d_\alpha)}(x) \mid (b_\alpha, d_\alpha) \in P_n\}$$

where $i\text{-max}$ denotes the i -th largest value of a function. We set $\lambda_n(i)(x) = 0$ if the i -th largest value does not exist. Thus, the persistence landscapes form a subset of the Banach space $L^p(\mathbb{N} \times \mathbb{R})$.

2.2.9. Norm of a persistence landscape. For each persistence landscape we can compute its norm, and so we can compare persistence landscapes by computing the mutual distances, using the metric derived from the norm.

- The norm of $\lambda_n \in L^p(\mathbb{N} \times \mathbb{R})$ is given by

$$(2.5) \quad \|\lambda_n\|_p = \left(\sum_{i=1}^{\infty} \|\lambda_n(i)\|_p^p \right)^{1/p}.$$

Above, $\|\cdot\|_p$ denotes the L^p -norm, $p \geq 1$, i.e., $\|f\|_p = (\int_{\mathbb{R}} |f|^p)^{1/p}$, where the integration is with respect to the Lebesgue measure on $\mathbb{R}$.

We recall from [B⁺15, BD17] that the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ can be computed via direct formulas

$$(2.6) \quad \begin{aligned} \|\lambda_n\|_1 &= \frac{1}{4} \sum_{\alpha} (d_\alpha - b_\alpha)^2, \\ \|\lambda_n\|_\infty &= \frac{1}{2} \sup_{\alpha} (d_\alpha - b_\alpha). \end{aligned}$$

Throughout the paper we will only refer to the L^1 -norms of persistence diagrams.

2.2.10. Stability. Persistence diagrams and persistence landscapes are ‘stable’ under small perturbations of a point-cloud [CDSO14, B⁺15].

2.3. Topological data analysis of time-series. As input, consider a time series of numerical values of length L :

$$(2.7) \quad X = \{x_0, x_1, \dots, x_{L-1}\}.$$

2.3.1. *Time-series of time-delay coordinate vectors.* We choose and fix an embedded dimension N and a time-delay $d \geq 1$.

- We transform the time series X into a sequence of N -dimensional time-delay coordinate vectors

$$\begin{aligned}
 z_0 &= (x_0, x_d, \dots, x_{(N-1)d}), \\
 z_1 &= (x_1, x_{1+d}, \dots, x_{1+(N-1)d}), \\
 &\dots \\
 z_t &= (x_t, x_{t+d}, \dots, x_{t+(N-1)d}), \\
 &\dots \\
 z_{L-1-(N-1)d} &= (x_{L-1-(N-1)d}, x_{L-1-(N-2)d}, \dots, x_{L-1}).
 \end{aligned}
 \tag{2.8}$$

In our applications below, we will only be interested in detecting $D = 1$ -dimensional objects (loops), in which case it will be sufficient to choose $N \geq 2D + 1 = 3$. The theoretical considerations behind the choice of the embedding dimension N rely on Takens' embedding theorem [Tak81] and its generalization by Sauer, Yorke and Casdagli [TJM91]. Also, in our applications we choose the delay d empirically. If we choose d too small, then the coordinates x_{t+id} and $x_{t+(i+1)d}$ can be too close to each other, and so they do not represent two 'independent' coordinates in a statistical sense. Similarly, if d is too large, then x_{t+id} and $x_{t+(i+1)d}$ are completely independent of each other, and so geometric information gets lost. Some empirical rules on choosing a suitable delay d are discussed in [ABST93].

2.3.2. *Sliding windows.* To detect qualitative changes along a time series, we apply a sliding window and assess how its features change along the sliding window. In our case:

- We apply a sliding window of size w to the sequence (2.8), for w sufficiently large with $N \ll w \ll L$, obtaining a time-varying point-cloud embedded in $\mathbb{R}^N$:

$$Z^t = \{z_t, z_{t+1}, \dots, z_{t+w-1}\}, \text{ for } t \in \{0, \dots, K-1\},
 \tag{2.9}$$

where we denote $K-1 = L-1-(N-1)d-(w-1)$.

2.3.3. *Time series of norms of persistent landscapes.* Fix $n \geq 1$ and $p \geq 1$.

- For each point-cloud Z^t we compute the norm of the persistent landscape $\|\lambda_n^t\|_p$, thus obtaining a time-series of norms

$$\|\lambda_n^t\|_p, \text{ for } t \in \{0, \dots, K-1\}.$$

3. TDA APPLIED TO OSCILLATORY TIME-SERIES

The main question is to understand what features of a time-series yield 1-dimensional 'holes' in the persistence diagram and result in 'peaks' of the norms of the persistence landscapes.

Time-series generated by periodic and quasi-periodic signals are some basic type of signals that result in ‘holes’ in the persistence diagram. We will also consider oscillatory time series whose frequency changes over time.

3.1. Periodic time-series. First, consider time-series generated by a periodic signal of the form

$$X(t) = A_1 \sin(\omega_1 t + \phi_1) + \sigma G(t)$$

where A_1, ω_1, ϕ_1 are the parameters of the deterministic signal, σ is the noise intensity, and $G(t)$ is Gaussian noise with zero mean and unit variance. When $\sigma = 0$, the embedding of $X(t)$ in $\mathbb{R}^N$ with $N \geq 2$ fills densely a topological circle, which is a homeomorphic copy of $\mathbb{T}^1$. It has a single hole, which is 1-dimensional, thus the 1-dimensional homology is $H_1(\mathbb{T}^1) = \mathbb{Z}_2$ whose generator is a 1-dimensional loop. When we add noise with intensity σ small, we obtain a ‘noisy’ circle.

To apply the TDA procedure, we first discretize the time t , i.e., we let $t_{i+1} = t_i + \Delta t$, with a step size of $\Delta t > 0$ small. Then we choose the window size w and the delay d . It is important that the size of the sliding window w is larger than the period $T_1 = 2\pi/\omega_1$ of the signal, and that the delay d is less than $2\pi/(\omega_1 \Delta t)$ but not too small. Otherwise, if $T_1 > w$, then the embedding of a window does not yield a closed curve. If d is too small, the embedding of the window may yield a point cloud with a narrow ‘hole’. Indeed, if Δt is small and d is small then the components of $z_t = (x_t, x_{t+d}, \dots, x_{t+(N-1)d})$ are very close to one another, and thus the points z_t are all located within a small neighborhood of the diagonal set $\{\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N : x_1 = \dots = x_N\}$. In such a case, the persistence of the homology generator for the topological circle may be relatively small, and hence the corresponding norm of the persistence landscape may be dominated by the effect of the noise. Some precise formulations on how to choose the window size w and the delay d can be found in [PH15]. Related works on applying TDA to periodic signals can be found in [KM18, DQR19].

An example of a periodic time-series without noise is shown in Fig. 2. We note that delay $d = 5$ yields a stronger signal than $d = 1$ in both the persistence diagram and the norm of the persistence landscape.

In either case, the norms of the persistence landscapes stay nearly constant, since the reconstructed set in $\mathbb{R}^N$ – a topological circle – is the same for each sliding window. The deviations from a constant value are due to the discretization of the signal.

An example with a periodic time-series with noise is shown in Fig. 3; in this case we add noise with $\sigma = 0.1$. We note that delay $d = 5$ still yields a stronger signal than $d = 1$. We also notice that adding the noise weakens the TDA signal from the no-noise case. This is because the noise narrows the ‘hole’ in the point cloud.

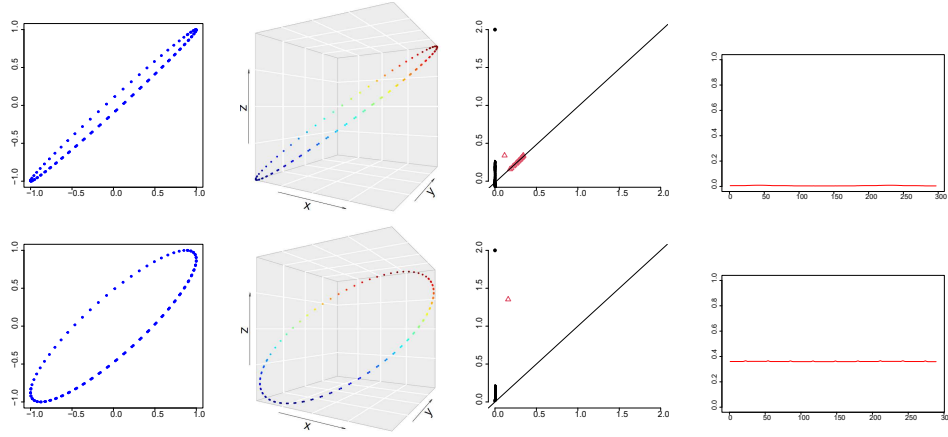


FIGURE 2. Periodic signal without noise reconstructed in 2D and 3D, persistence diagram and norms of persistence landscapes: $d = 1$ (top), and $d = 5$ (bottom)

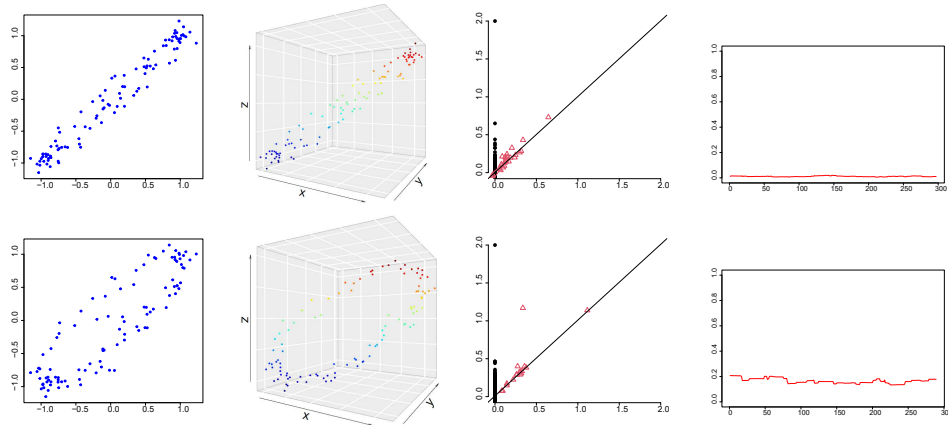


FIGURE 3. Periodic signal with noise reconstructed in 2D and 3D, persistence diagram and norms of persistence landscapes: $d = 1$ (top), and $d = 5$ (bottom)

3.2. Quasi-periodic time-series. Second, consider time-series generated by a quasi-periodic signal with two incommensurate frequencies

$$X(t) = A_1 \sin(\omega_1 t + \phi_1) + A_2 \sin(\omega_2 t + \phi_2) + \sigma G(t)$$

where $A_1, \omega_1, \phi_1, A_2, \omega_2, \phi_2$ are the parameters of the deterministic signal, and $\sigma, G(t)$ are as before; we assume that ω_1/ω_2 is an irrational number. When $\sigma = 0$, for suitable window size w and delay d , the embedding of $X(t)$ in $\mathbb{R}^N$ with $N \geq 3$ yields a topological 2-dimensional torus $\mathbb{T}^2$. This has two

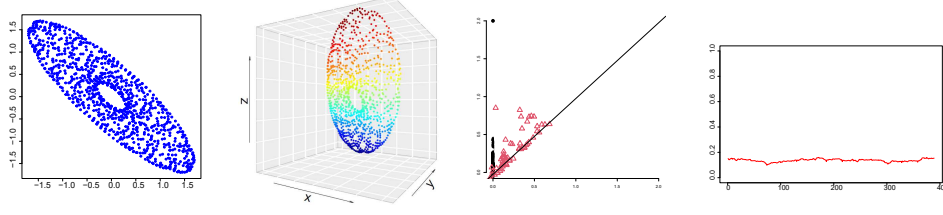


FIGURE 4. Quasi-periodic signal without noise reconstructed in 2D and 3D, persistence diagram and norms of persistence landscapes: $d = 1$ (top), and $d = 5$ (bottom)

1-dimensional homology generators, since $H_1(\mathbb{T}^2) = \mathbb{Z}_2^2$. In addition, there is a single 2-dimensional homology generator, i.e., $H_2(\mathbb{T}^2) = \mathbb{Z}$. See Fig. 4.

The norms of the persistence landscapes stay nearly constant, since the reconstructed set in $\mathbb{R}^N$ – a topological torus – is the same for each sliding window. The deviations from a constant value are due to the discretization of the signal.

Similarly, we can consider general quasi-periodic signals with $m \geq 1$ incommensurate frequencies

$$X(t) = \sum_{i=1}^m A_i \sin(\omega_i t + \phi_i) + \sigma G(t).$$

There are specific conditions on how to choose the windows size w , the delay d and the embedding dimension N ; see [PH15, Per16]. The reconstructed set is an m -dimensional torus $\mathbb{T}^m$, whose homology is given by $H_k(\mathbb{T}^m) = \mathbb{Z}_2^{\binom{m}{k}}$.

3.3. Oscillatory time-series with changing frequency. Another model that produces ‘holes’ in point-cloud is an oscillatory signal of the form

$$X(t) = A_1 \sin(\omega_1 \ln(t_c - t) + \phi_1) + \sigma G(t),$$

where t_c represents a critical time and $t \in [0, t_c)$. Note that when $t \rightarrow t_c$ the frequency of the oscillations $\omega_1 \ln(t_c - t) \rightarrow \infty$ at a logarithmic rate, while the amplitude of the oscillations stays constant. This type of signal is related to the LPPLS model (2.2).

There is a strong dependence of the TDA signal on the range $[0, t_c)$ of the time-series, on the parameter ω_1 , and on size of the sliding window w . Note that at any moment of time t the approximate period of the signal is $T = \frac{2\pi}{\omega_1 \ln(t_c - t)}$. Each full oscillation in the time-series produces a hole in point-cloud. As t increases, more holes appear in point-cloud for the corresponding sliding window. When w is larger than the initial period $T = \frac{2\pi}{\omega_1 \ln(t_c)}$, then the TDA signal will capture the first oscillation in the time-series as well as the subsequent ones. However, when $w < \frac{2\pi}{\omega_1 \ln(t_c)}$, the TDA signal will only capture oscillations in the time-series whose period $\frac{2\pi}{\omega_1 \ln(t_c - t)} < w$.

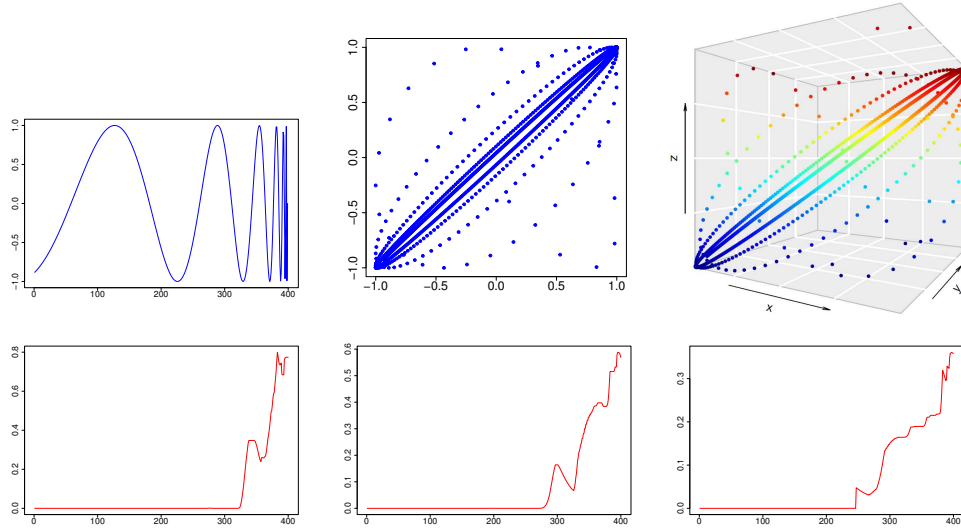


FIGURE 5. Oscillatory time-series with changing frequency and its reconstructions in 2D and 3D (top), norms of persistence landscapes for $w = 100, 150, 230$ (bottom).

In Fig. 5 we show an oscillatory signal as above with no noise, and its TDA signal for various sizes of the sliding window.

4. TDA APPLIED TO THE LPPLS MODEL

We recall that for the LPPLS model, the price of the asset oscillates with increasing frequency and decreasing amplitude about a super-exponential trend line. When we apply the TDA procedure with a fixed window size we observe the following:

- At the beginning of the time-series, the behavior is nearly periodic, and the TDA output is similar to that of a periodic signal, producing a single, persistent 1-loop dimensional loop in the point cloud;
- As the time increases, the time-series exhibits more than one frequency, and TDA output produces a few persistent 1-dimensional loops in the point cloud;
- As the time approaches the critical time, the time-series exhibits oscillations of growing frequencies, and the TDA output yields more and more persistent 1-dimensional loops, which decay in size. In addition, the super-exponential trend makes the points in the point-cloud more spread out, so the growing-in-size ‘gaps’ in the point-cloud contribute with additional 1-dimensional loops. Consequently, the norms of the persistent landscapes spike in the proximity of the critical time.

To capture such behavior, the size of the sliding window w should be suitably chosen; this depends on the length of the data set as well as on the parameters of the LPPLS model. For instance, using a larger window size w will result in the larger loops in the earlier part of the time series to carry more weight towards the TDA signal, which will result in a spike earlier in the range as opposed to the end of the range. On the other hand, a smaller window size w will not capture the larger loops but only the smaller ones in the latter part of the time series, which will result in a spike towards the end of the range.

In practice, one needs to explore a range of window size w and observe the presence of a spike in the TDA signal that appears earlier in the range for large w , and shifts towards the end of the range for smaller w .

As an example, we consider a noisy LPPLS model for a positive bubble, with the following parameters

$$(4.1) \quad \begin{aligned} t_c &= 637, m = 3.003e - 01, \omega = 6.889e + 00, \\ A &= 1.111e + 01, B = -2.937e - 04 \\ C &= 5.515e - 05, C_1 = 4.372e - 05, C_2 = -3.362e - 05. \end{aligned}$$

These parameters corresponds to fitting the LPPLS model to the Bitcoin price between the dates 2021/03/22/–2021/04/16, sampled hourly.

We use the following TDA sliding-window size w , delay d , and embedding dimension N :

$$(4.2) \quad w = 72, d = 5, N = 4.$$

The TDA output for various levels of added noise is shown in Fig. 6. We observe a peak in the TDA signal prior to the critical time, and note that the growing trend towards the peak appears earlier when the noise intensity is larger.

4.1. Dependence of the TDA signal on parameters. We consider synthetic time series generated by (2.2), of 200 data points and with $t_c = 200$. We start with the following model parameters $m = 0.3$, $\omega = 6.7$, $A = 11$, $B = -0.0003$, $C_1 = 0.000044$, $C_2 = -0.000034$, and with a sliding widow of size $w = 48$. These parameters are similar to the ones in (4.1), except for the length of the time series, which is chosen shorter for computational efficiency. In the experiments below we explore the changes in the TDA signal when we change the parameters $w, \omega, m, B, C_1, C_2$, one at a time. A more detailed explorations would be needed to study the interdependence among these parameters, which appears to be quite intricate.

4.1.1. Dependence on sliding window size w . The experiments displayed in Fig. 7 show that as the window size is decreased, the peak of the TDA signal gets shifted towards the end of the range of the time series. This is because windows of larger size w capture more of the larger loops in the point-cloud, while windows of smaller size capture more of the smaller loops in the point-cloud. This is consistent with the remarks in Section 3.

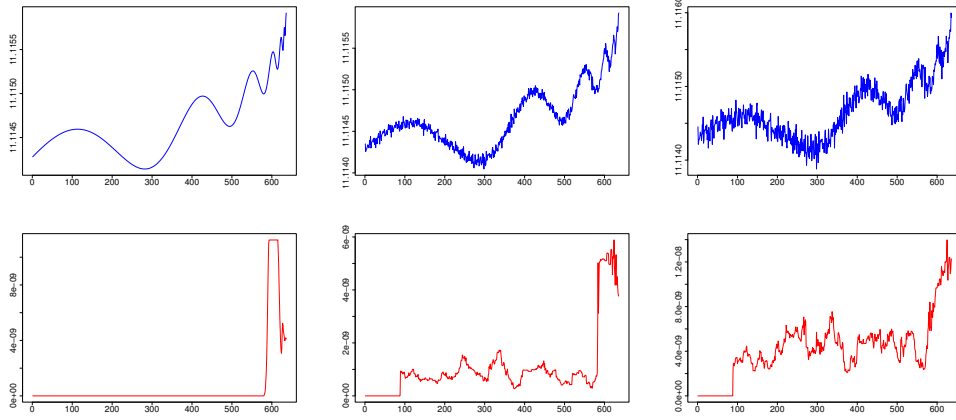


FIGURE 6. LPPLS model (top) and norm of persistence landscapes (bottom) for a positive bubble and increasing noise intensity

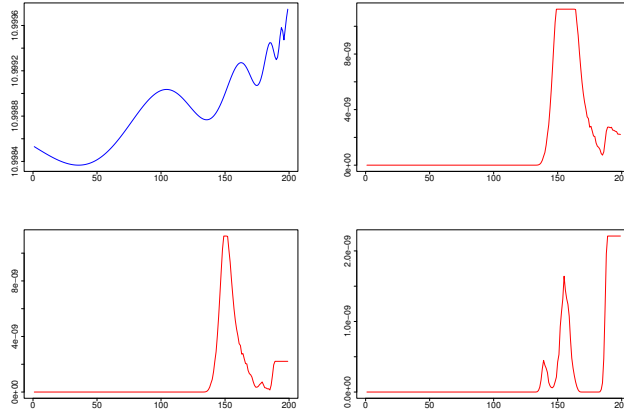


FIGURE 7. Synthetic LPPLS signal and corresponding TDA signal for window sizes $w = 72, 60, 48$.

4.1.2. *Dependence on frequency ω .* The experiments displayed in Fig. 8 show that as the frequency ω is increased, the TDA signal increases in strength.

4.1.3. *Dependence on non-linearity parameter m .* We note that when $m \approx 1$, the main trend line of the log price is nearly linear. The experiments displayed in Fig. 9 show that as the frequency m is decreased, the TDA signal gets shifted towards the end of the range.

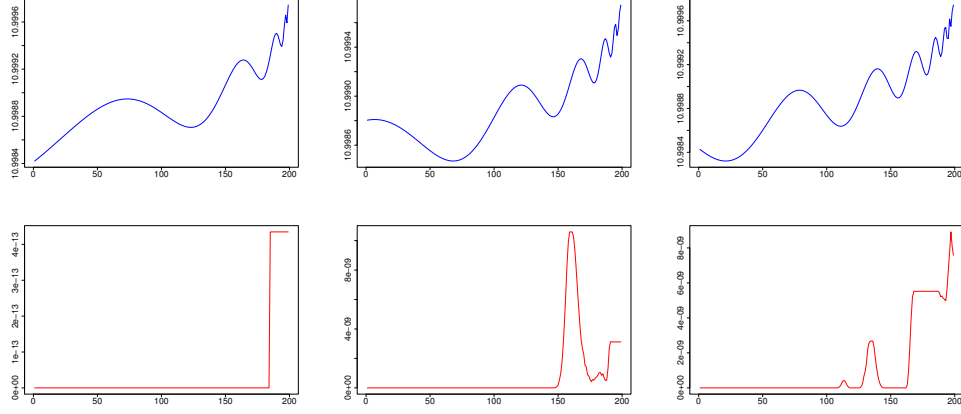


FIGURE 8. Synthetic LPPLS signal (top) and corresponding TDA signal (bottom) for $\omega = 5, 7, 9$.

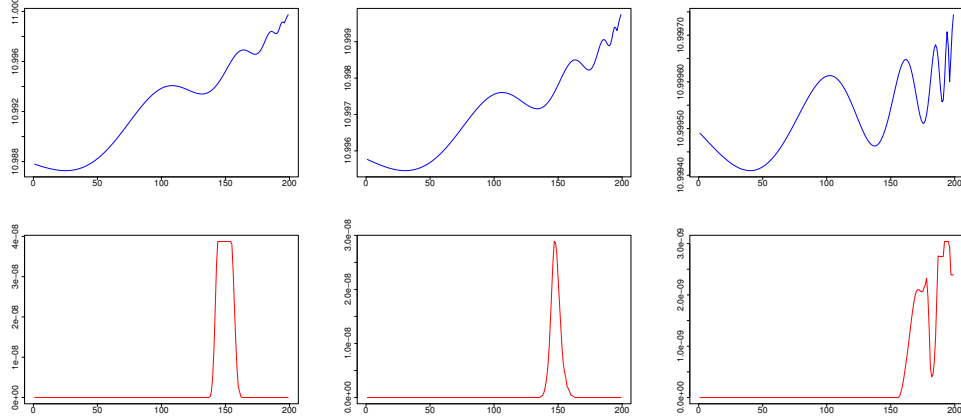


FIGURE 9. Synthetic LPPLS signal (top) and corresponding TDA signal (bottom) for $m = 0.7, 0.5, 0.1$

4.1.4. *Dependence of positive/negative bubble parameter B .* The sign of the parameter b is associated with the bubble-type: $B < 0$ corresponds to positive bubble and $b > 0$ corresponds to negative bubble. In Figure 10 we show a synthetic LPPLS model for a negative bubble. It is remarkable that for both positive and negative bubbles the TDA signal spikes when approaching the critical time.

4.1.5. *Dependence on amplitude parameters C_1, C_2 .* We explore the dependence of the TDA signal on the parameters C_1, C_2 which are responsible of the amplitude of the oscillations (here we use $w = 60$). When the amplitude

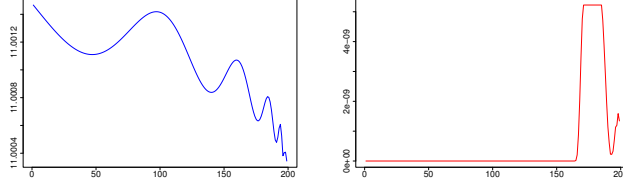


FIGURE 10. Synthetic LPPLS model and norm of persistence landscapes for a negative bubble $B > 0$.

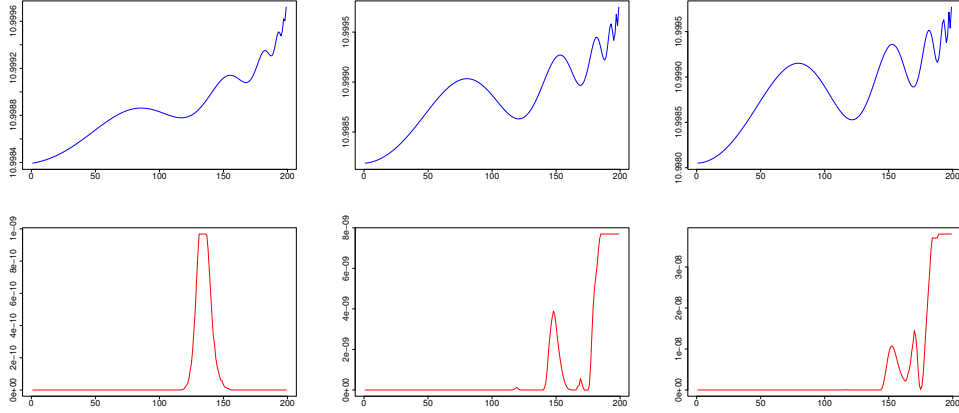


FIGURE 11. Synthetic LPPLS signal (top) and corresponding TDA signal (bottom) for $C_1 = C_2 = 0.00002, 0.00005, 0.00007$.

of the signal is increased, the TDA signal grows in strength and becomes more prominent towards the end of the range. See Fig. 11.

5. DATA SEGMENTATION INTO POSITIVE AND NEGATIVE BUBBLES

In this section we describe a method to segment a given financial time series into positive and negative bubbles. For a positive bubble the main idea is there will be a succession of positive returns, which may be interrupted by negative returns no larger in amplitude than some pre-specified tolerance level. An analogous description holds for negative bubbles. Our segmentation algorithm is a simplified version of the “epsilon-drawdown” method developed in [JS02, JS⁺10, FS15], and of some of its variants in [SZ20a, SZ20b, YRdSTT20].

Let $x(t_i)$ be the time-series of the price of an asset, where $t_i = t_{i-1} + \Delta t$. The log-return of the asset at time t_i is

$$p_i = \ln x(t_i) - \ln x(t_{i-1}).$$

Assume that some initial time t_{i_0} is the beginning of a up- (down-) trend. The cumulative return at time i is

$$p_{i_0,i} = \sum_{i=i_0+1}^i (\ln x(t_i) - \ln x(t_{i-1})) = \ln x(t_i) - \ln x(t_{i_0}).$$

The largest deviation $\delta_{i_0,i}$ of the price trajectory from a previous minimum (maximum) is given by

$$(5.1) \quad \delta_{i_0,i} = \begin{cases} \max_{i_0 \leq k \leq i} p_{i_0,k} - p_{i_0,i}, & \text{for an up-trend,} \\ p_{i_0,i} - \min_{i_0 \leq k \leq i} p_{i_0,k}, & \text{for a down-trend.} \end{cases}$$

To find the end of the up-trend, we first compute i_{cross_1} the first time $i > i_0$ when $\delta_{i_0,i}$ crosses a certain tolerance level ε_i ,

$$(5.2) \quad \delta_{i_0,i} - \varepsilon_i > 0$$

and then define the end of the up- (down-) trend as the time

$$(5.3) \quad t_{i_1} = \begin{cases} \arg \max_{i_0 \leq k \leq i_{\text{cross}_1}} p_{i_0,k}, & \text{for an up-trend,} \\ \arg \min_{i_0 \leq k \leq i_{\text{cross}_1}} p_{i_0,k}, & \text{for a down-trend.} \end{cases}$$

The time t_{i_1} is considered as the beginning of the next down- (up-) trend. Through this procedure, the data is successively segmented in up and down trends, which alternate one after the other. The procedure is carried out for the whole time series and yields a sequence of peaks

$$\{t_{u,1}, t_{u,2}, \dots\},$$

and a sequence of troughs

$$\{t_{d,1}, t_{d,2}, \dots\},$$

such that peak is always followed by a trough and vice-versa.

At the end of this procedure, each up-trend corresponds to a positive bubble, and each down-trend corresponds to a negative bubble.

The tolerance level ε_i can be chosen in multiple ways. One option is to choose

$$\varepsilon_i = \varepsilon,$$

where $\varepsilon > 0$ is some preset constant. Another option is to choose a time-dependent tolerance level ε_i proportional to the volatility (standard deviation) $\sigma_i = \sigma_i(w_0)$ at time i estimated over a preceding time window of some fixed size w_0 , i.e.,

$$(5.4) \quad \varepsilon_i = \varepsilon_0 \sigma_i,$$

where ε_0 is some scaling parameter. Thus the tolerance is more permissive when the market presents high volatility and is stricter during calmer periods. There are other ways to choose the tolerance level (see, e.g., [YRdSTT20]).

The segmentation procedure in [GDS18] does not fix a single scaling parameter ε_0 and a single window-size w_0 for the computation of the standard

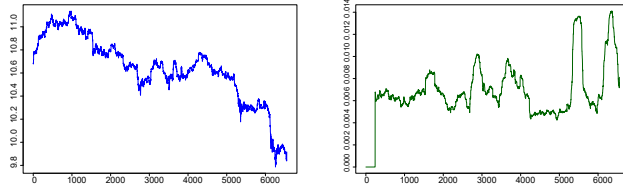


FIGURE 12. Log-price and standard deviation of Bitcoin over 10/01/2021-07/01/2022

deviation, but rather considers some range of scaling parameters $\varepsilon_0 \in [\varepsilon_1, \varepsilon_2]$ and of window-sizes $w_0 \in [w_1, w_2]$, each sampled with some frequency, and computes a collection of peaks and a collection of troughs for each ε_0 and w_0 . Finally, it selects the most frequent peaks and troughs that appear over the whole range of ε_0, w_0 .

In Fig. 12 we show the log-price of Bitcoin over the period 10/01/2021-07/01/2022 sampled at 1 hour intervals (resulting in 6552 data points), and the standard deviation measured over a preceding time window of size $w_0 = 240$.

In the examples in Section 6, as we are interesting in ‘large’ bubbles, we will fix the scaling parameter at some large value of ε_0 .

6. APPLICATION TO DETECTION OF BITCOIN BUBBLES

In this section we present a series of numerical experiments on hourly Bitcoin price during the period 2021-10-01 and 2022-07-01. In Section 6.1 we segment the data into positive bubbles and negative bubbles following the methodology described in Section 5. In Section 6.2 we fit the LPPLS model to each of the segments. In Section 6 we apply the TDA procedure to each segment, following the methodology described in Sections 2.2 and 2.3.

6.1. Data segmentation. The hourly Bitcoin price data set between 2021-10-01 and 2022-07-01 consists of 6552 data points. We segment the data following the procedure described in Section 5. To compute the standard deviation (volatility) σ_i we use a fixed window size of $w_0 = 240$ data points. We also use a fixed scaling parameter $\varepsilon_0 = 15.0$. Thus, the tolerance level is taken as $\varepsilon_i = \varepsilon_0 \sigma_i$, where σ_i is computed over a sliding window of length w_0 .

In the table below we identify the peak-time (P) and the crossing-time (C) for each up-trend.

Table 2. Up-trend peak and crossing times:

P	471	977	2105	3184	3662	3831	4290	4861	5447	5824	6443
C	538	1104	2130	3355	3718	3934	4370	4887	5729	5894	6536

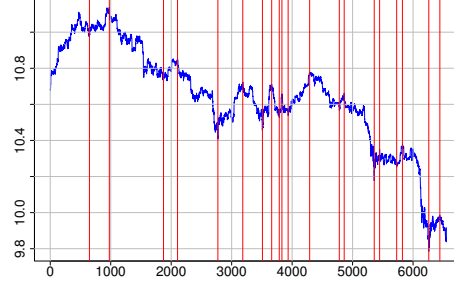


FIGURE 13. Log price with peaks and troughs marked by vertical lines

In the table below we identify the trough-time (T) and the crossing-time (C) for each down-trend.

Table 3. Down-trend troughs and crossing times:

T	648	1873	2772	3509	3787	3934	4780	5358	5729	6260
C	783	2010	2821	3531	3831	4075	4835	5601	5804	6415

Both peaks and troughs from Tables 1 and 2 are shown in Fig. 13.

Then we define up-trend (U) and down-trend (D) segments to be fitted with LPPLS models as well as analyzed with TDA.

First, each segment is determined by the following criterion. For an up- (down-) trend, the end-point is a peak (trough) from the list, and the starting-point of the trend is the nearest crossing time in the past from the previous down- (up-) trend. If the nearest crossing time in the past occurs before the previous trough (peak), then the starting-point is chosen to be that particular trough (peak).

The reason for choosing the starting-point of a trend at the previous crossing-time is that, right after a trough (peak), the price may grow (drop) dramatically, which yields a strong TDA signal that may eclipse the TDA signal preceding the next bubble burst. We want the starting-point to be after the recovery period for the previous bubble burst. Therefore, we empirically choose the crossing time as a mark for the end of the recovery period. As mentioned before, in some exceptional cases, the previous crossing time may occur before the previous trough/peak, in which case we choose the starting-point at that trough/peak. Making these choices results in up- and down-trends that have no overlap, which is convenient for the data analysis.

Table 4. Up-trend and down-trend segments

U	783-977	2010-2105	2821-3184	3531-3662	3783-3831	4075-4290
U	4835-4861	5358-5447	5804-5824	6415-6443		
D	538-648	1104-1873	2130-2772	3355-3509	3718-3787	3831-3934
D	4370-4780	4887-5358	5447-5729	5894-6260		

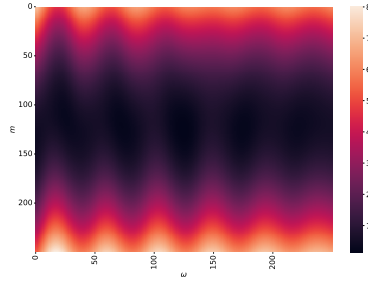


FIGURE 14. Fitness function in the domain $[m, \omega]$, for constant values of the remaining parameters. The objective function displays strong non-convexity.

Second, we perform some additional adjustments. We examine each of the resulting segments and if we observe significant peaks or troughs inside a segment we do an ad-hoc adjustment of the segment: we cut the portion from the beginning of the segment to the last observed a peak/through. This adjustment is occasionally needed since using the segmentation procedure with fixed w_0 and fixed ε_0 can miss some peaks/troughs.

Third, from the remaining segments we dismiss those that are very short, since we cannot apply the LPPLS fitting and the TDA procedure to them. Thus, the segments 3783 – 3831, 4835 – 4861, 5804 – 5824 and 6415 – 6443, all corresponding to up-trends, were deemed as too short to perform the analysis. We also disregard the first segment 1 – 471 as it was used to initialize the process.

We retain the following:

Table 5. Adjusted up-trend and down-trend segments

U	783-977	2010-2105	3020-3184	3531-3662	4175-4290
U	5358-5447				
D	538-648	1104-1873	2500-2772	3355-3509	3718-3787
D	3831-3934	4370-4780	5150-5358	5447-5729	5894-6260

6.2. Fitting the LPPLS model to Bitcoin time-series. Once a specific segment is defined, we address the optimization problem, in which we aim at best fitting the LPPLS model to the time series. More formally, using (6.1) to obtain a vector of estimates of the expected log-price $\hat{y} = LPPLS(y)$ from the segment x , the goal is to solve the unconstrained, non-convex optimization problem:

$$(6.1) \quad \min_{tc, m, \omega, A, B, C_1, C_2} \|\hat{y} - y\|_2^2$$

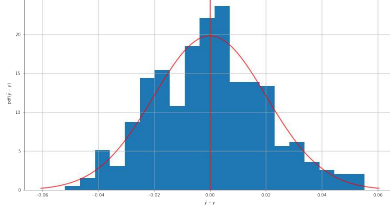


FIGURE 15. Distribution of residuals for a fitted model.

Figure 14 represents the fitness function in terms of two variables m and ω . The figure shows that the problem is highly non-convex and we cannot easily use any local minimizers to find an optimum solution.

Given the geometry of the problem, a global optimization scheme based on differential evolution [SP97] has been used. This method, as opposed to gradient-based local optimizers, is based on performing function evaluation on a set of sample points drawn from a large candidate space followed by a local optimizer at the end. While similar works propose the use of grid search in the space of initial conditions, each followed by local optimization, also [SZ20b] makes use of evolutionary algorithms (specifically, covariance matrix adaptation evolution strategy).

In order to speed up convergence, the problem is reformulated as

$$(6.2) \quad \min_{p=[tc, m, \omega]} \|\hat{y}(p, A(p), B(p), C_1(p), C_2(p)) - y\|_2^2$$

In which $q = [A(p), B(p), C_1(p), C_2(p)]$ is determined, for each p , solving a linear least square problem:

$$(6.3) \quad q = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{p}.$$

For the optimization problem (6.1), we have assumed that the noise (the residuals in (6.1)) follows a Gaussian distribution with zero mean and homoscedasticity, that is, time-independent variance, indicating that the variance of samples taken from the time-series remains constant over time. Our assumption is justified by the distribution of residuals depicted in Figure 15. Under this assumption, solving for the fitness function in (6.1) is equivalent to a maximum likelihood estimation.

We note that, when the time series exhibit heteroscedasticity, where the variance of the time-series samples changes across time, a weighting scheme can be introduced as a modification to the definition of the fitness function. Such a weighting scheme assigns smaller weights to residuals with higher variances, so that outliers have less impact on the estimates on the fitness function (see, e.g., [HGBF21, NM23]).

The software developed in this context is built on top of the open-source package [Bou]. The results of the LPPLS fitting are depicted in Appendix A.

6.3. TDA applied to Bitcoin time-series and comparison to fitted LPPLS models. We apply TDA to each adjusted up-trend and down-trend segment from Table 5. For each segments we have chosen a suitable set of TDA parameters from the following range:

$$(6.4) \quad w = 48, 60, 72, d = 3, 5, N = 4.$$

As noted in Section 4.1, adjustments of the TDA parameter are needed because the segments have different lengths, and the trends of the corresponding time-series (reflected by the fitted LPPLS models) are different.

In Appendix A we show, side-by-side, the fitted LPPLS model and TDA signal for each up-/down-trend, in chronological order (i.e., alternating up- and down-trends).

We observe that most of the segments show a primary peak in the TDA signal towards the end of the range. Some of the segments show some secondary peaks in the TDA signal (e.g., Appendix A – Figs.16b, 16c, 16e, 16h, 16p), which seem to capture some intermediate price jumps or drops.

In a few instances (e.g., Appendix A – Fig. 16l, 16n)) the primary peak in the TDA signal occurs earlier than the end of the range. We note that in these instance that an apparent tipping point seems to occur not at the end of the range but closer to the middle. The fact that the segmentation procedure used in Section 5 does not identify this apparent tipping point is likely due to the dependence of the threshold (5.4) on volatility; if the preceding time period is highly volatile, the threshold for the crossing time (5.2) is higher, and if the volatility becomes lower at a later time, then the threshold will also become lower, and hence the crossing time will be registered at a later time.

We also note that some of the fitted LPPLS models do not display the desired characteristics, that is, super-exponential growth (or decay) superimposed with oscillations increasing in frequency and decreasing in amplitude. For instance, Appendix A – Figs. 16d, 16j, 16l) show very few oscillations, or weak nonlinearity.

The poor fit of some of the LPPLS model means that the LPPLS regime does not hold for those segments. In fact, only LPPLS models that pass certain filters should be selected as viable solutions. See [GDS18] and the references listed there.

Nevertheless, the TDA signal is clearly distinguishable even in cases when the LPPLS regime does not hold.

7. CONCLUSIONS

We provide an empirical argument on why TDA is able to provide early warning signals for critical transitions in financial time series, for both positive and negative financial bubbles. The premise of our investigation is that

financial bubbles are reflected by the LPPLS model, asserting that the price of an asset exhibits super-exponential growth (or decay) superimposed with oscillations that increase in frequency and decrease in amplitude when approaching a critical transition. When the time series is transformed into a point cloud via time-delay coordinate embedding, these oscillations give rise to persistent 1-dimensional homology generators (loops), which can be quantified via persistence landscapes. By applying a sliding window to the time series, one can capture the changes in the oscillations through the growth in the norms of the corresponding persistence landscapes.

Our experiments show that, when the TDA procedure is applied to synthetic LPPLS time-series, the norms of the persistence landscapes grow as predicted. They also show an intricate dependence of the TDA signal on the parameters of the LPPLS model and of the TDA procedure.

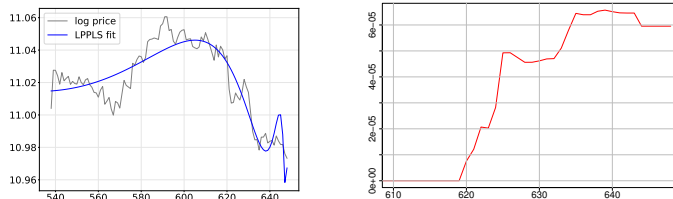
As a practical application, we run the TDA procedure on a collection of positive and negative bubbles of the Bitcoin price between October 1st 2021 and July 1st 2022. In most cases, the TDA gives early warnings for tipping points, even when the LPPLS regime does not seem to hold.

In terms of future work, we plan to undertake a more rigorous study of the dependence of the TDA parameters on the underlying time-series, which would enable automatic parameter selection. Additionally, it would be pertinent to explore further quantitative measures for TDA-based early warning signals, encompassing projections regarding the timing, magnitude, and likelihood of a crash. We also plan to validate the TDA-based approach to financial bubble detection through other models, including those described in Section 1.

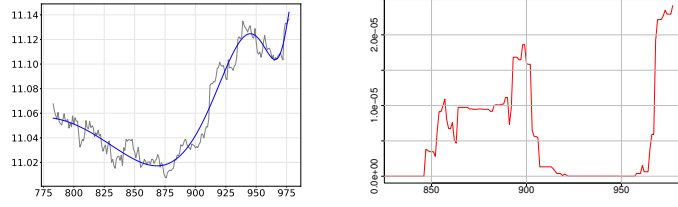
ACKNOWLEDGEMENT

We would like to thank the referees, as well as Didier Sornette, for useful comments and suggestions. Samuel W. Akingbade and Marian Gidea were partially supported by National Science Foundation (NSF) grant DMS-2307718

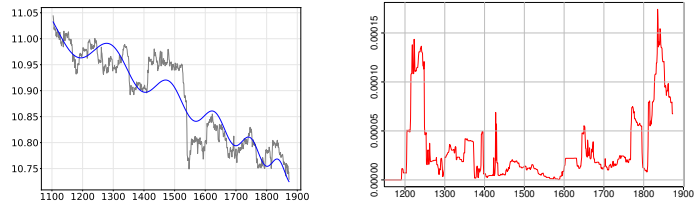
APPENDIX A. LPPLS AND TDA APPLIED TO BITCOIN SEGMENTS



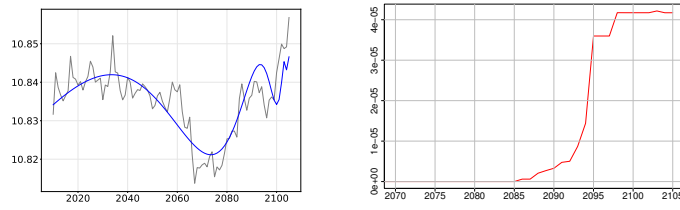
(A) Data set 538-648.



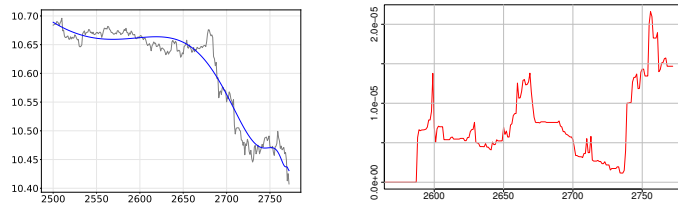
(B) Data set 783-977.



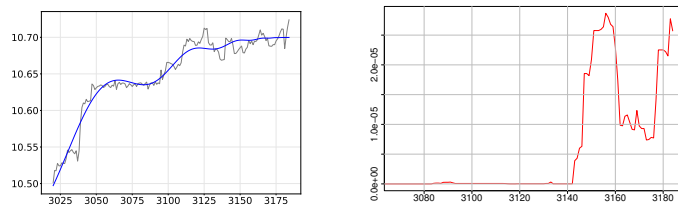
(C) Data set 1104-1873.



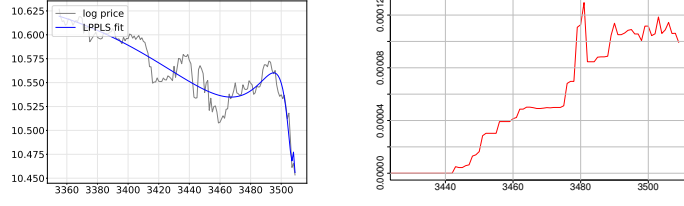
(D) Data set 2010-2105.



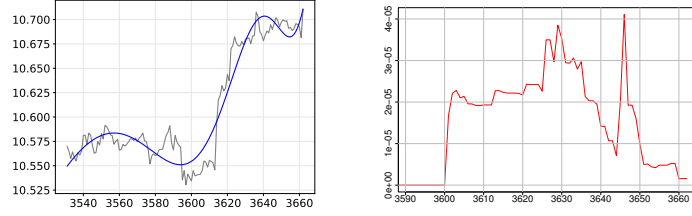
(E) Data set 2500-2772.



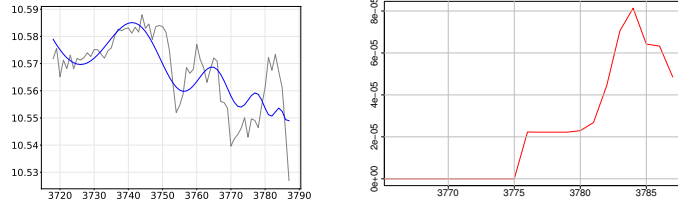
(F) Data set 3020-3184.



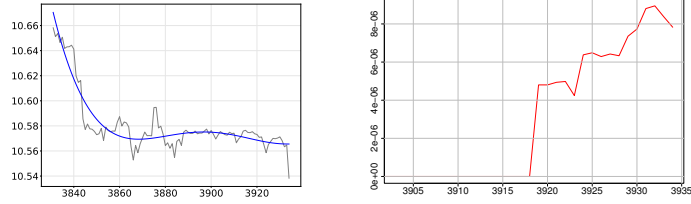
(G) Data set 3355-3509.



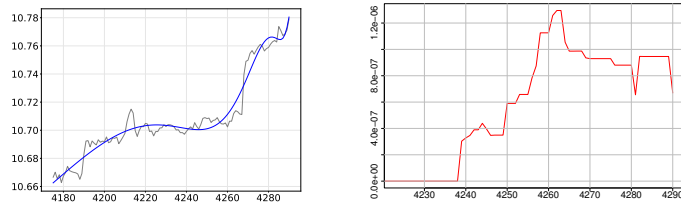
(H) Data set 3531-3662.



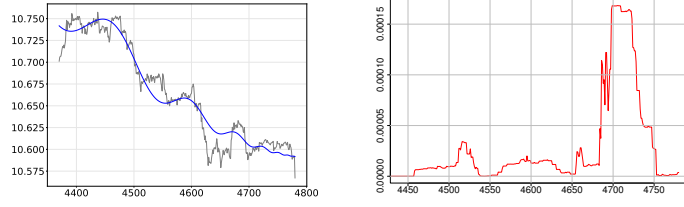
(I) Data set 3718-3787.



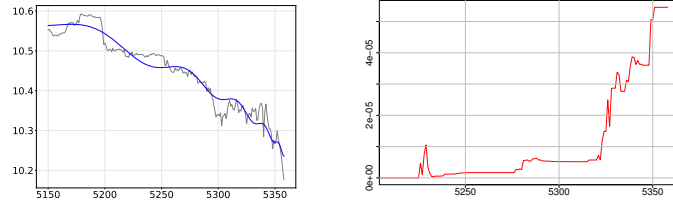
(J) Data set 3831-3934.



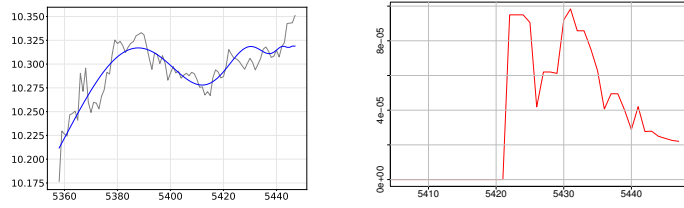
(K) Data set 4175-4290.



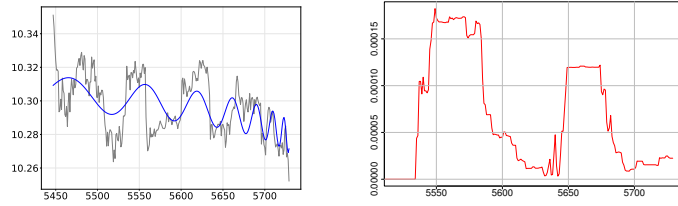
(L) Data set 4370-4780.



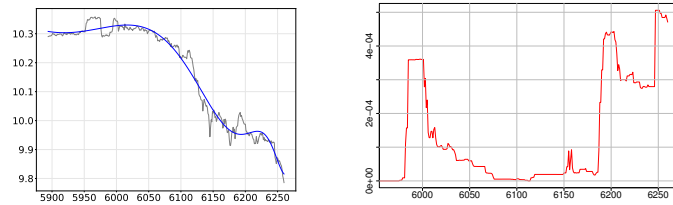
(M) Data set 5150-5358.



(N) Data set 5358-5447.



(O) Data set 5447-5729.



(P) Data set 5894-6260.

FIGURE 16. Fitted LPPLS models and norms of persistence landscapes for the segments from Table 5. The log-price and the LPPLS fit is shown on the left, and the TDA signal is shown on the right.

REFERENCES

- [ABST93] Henry DI Abarbanel, Reggie Brown, John J Sidorowich, and Lev Sh Tsimring. The analysis of observed chaotic data in physical systems. *Reviews of modern physics*, 65(4):1331, 1993.
- [ACAMNH21] Karina Arias-Calluari, Fernando Alonso-Marroquin, Morteza N. Najafi, and Michael Harré. Methods for forecasting the effect of exogenous risks on stock markets. *Physica A: Statistical Mechanics and its Applications*, 568:125587, 2021.
- [ADGK18] Cuneyt Gurcan Akcora, Matthew F Dixon, Yulia R Gel, and Murat Kantarcioglu. Bitcoin risk modeling with blockchain graphs. *Economics Letters*, 173:138–142, 2018.
- [AKV21] Lloyd L Aromi, Yuri A Katz, and Josep Vives. Topological features of multivariate distributions: Dependency on the covariance matrix. *Communications in Nonlinear Science and Numerical Simulation*, 103:105996, 2021.
- [ALGK20] Cuneyt G Akcora, Yitao Li, Yulia R Gel, and Murat Kantarcioglu. Bitcoin-heist: Topological data analysis for ransomware prediction on the bitcoin blockchain. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020.
- [B⁺15] Peter Bubenik et al. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, 2015.
- [BCP13] David S Brée, Damien Challet, and Pier Paolo Peirano. Prediction accuracy and sloppiness of log-periodic functions. *Quantitative Finance*, 13(2):275–280, 2013.
- [BD17] Peter Bubenik and Paweł Dłotko. A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78:91–114, 2017.
- [BG14] Jesse Berwald and Marian Gidea. Critical transitions in a model of a genetic regulatory system. *Mathematical Biosciences and Engineering*, 11(4):723–740, 2014.
- [BGVJ14] JJ Berwald, M Gidea, and Mikael Vejdemo-Johansson. Automatic recognition and tagging of topologically different regimes in dynamical systems. *Discontinuity, Nonlinearity, and Complexity*, 3(4):413–426, 2014.
- [Bla79] Olivier Jean Blanchard. Speculative bubbles, crashes and rational expectations. *Economics letters*, 3(4):387–389, 1979.
- [Bou] Boulder-Investment-Technologies, howpublished = <https://github.com/boulder-investment-technologies/lppls>, note = Accessed: 2023-03-15.
- [BSP⁺21] Thomas M Bury, RI Sujith, Induja Pavithran, Marten Scheffer, Timothy M Lenton, Madhur Anand, and Chris T Bauch. Deep learning for early warning signals of tipping points. *Proceedings of the National Academy of Sciences*, 118(39):e2106140118, 2021.
- [BW82] Olivier J Blanchard and Mark W Watson. Bubbles, rational expectations and financial markets, 1982.
- [CDSO14] Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.
- [DCB⁺12] Vasilis Dakos, Stephen R Carpenter, William A Brock, Aaron M Ellison, Vishweshha Guttal, Anthony R Ives, Sonia Kéfi, Valerie Livina, David A Seekell, Egbert H van Nes, et al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PloS one*, 7(7):e41010, 2012.
- [DQR19] Paweł Dłotko, Wanling Qiu, and Simon Rudkin. Cyclicity, periodicity and the topology of time series. *arXiv preprint arXiv:1905.12118*, 2019.

- [DR21] P Dłotko and ST Rudkin. Financial ratios and stock returns reappraised through a topological data analysis lens. *The European Journal of Finance*, pages 1–25, 2021.
- [DUC20] Meryll Dindin, Yuhei Umeda, and Frederic Chazal. Topological data analysis for arrhythmia detection through modular neural networks. In *Advances in Artificial Intelligence: 33rd Canadian Conference on Artificial Intelligence, Canadian AI 2020, Ottawa, ON, Canada, May 13–15, 2020, Proceedings 33*, pages 177–188. Springer, 2020.
- [EH22] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.
- [FC16] John Fry and Eng-Tuck Cheah. Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis*, 47:343–352, 2016.
- [Fer06] Marcelo Fernandes. Financial crashes as endogenous jumps: estimation, testing and forecasting. *Journal of Economic Dynamics and Control*, 30(1):111–141, 2006.
- [FF96] James A. Feigenbaum and Peter G.O. Freund. Discrete scale invariance in stock markets before crashes. *International Journal of Modern Physics B*, 10(27):3737–3745, 1996.
- [Fry12] JM Fry. Exogenous and endogenous market crashes as phase transitions in complex financial systems. *The European Physical Journal B*, 85:1–6, 2012.
- [FS15] Vladimir Filimonov and Didier Sornette. Power law scaling and “Dragon-Kings” in distributions of intraday financial drawdowns. *Chaos, Solitons & Fractals*, 74:27–45, 2015.
- [GDS18] Jan-Christian Gerlach, Guilherme Demos, and Didier Sornette. Dissection of Bitcoin’s multiscale bubble history. *Swiss Finance Institute Research Paper*, (18-30), 2018.
- [GGK⁺20] Marian Gidea, Daniel Goldsmith, Yuri Katz, Pablo Roldan, and Yonah Shmalo. Topological recognition of critical transitions in time series of cryptocurrencies. *Physica A: Statistical Mechanics and its Applications*, 548:123843, 2020.
- [Gid17] Marian Gidea. Topological data analysis of critical transitions in financial networks. In *International conference and school on network science*, pages 47–59. Springer, 2017.
- [GK18] Marian Gidea and Yuri Katz. Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834, 2018.
- [GK23] Victor E Gluzberg and Yuri A Katz. Topological data analysis of noise: Uniform unimodal distributions. *Communications in Nonlinear Science and Numerical Simulation*, page 107216, 2023.
- [HGBF21] David Hong, Kyle Gilman, Laura Balzano, and Jeffrey A Fessler. Heppcat: Probabilistic PCA for data with heteroscedastic noise. *IEEE Transactions on Signal Processing*, 69:4819–4834, 2021.
- [HLS15] David I Harvey, Stephen J Leybourne, and Robert Sollis. Recursive right-tailed unit root tests for an explosive asset price bubble. *Journal of Financial Econometrics*, 13(1):166–187, 2015.
- [HLST16] David I Harvey, Stephen J Leybourne, Robert Sollis, and AM Robert Taylor. Tests for explosive financial bubbles in the presence of non-stationary volatility. *Journal of Empirical Finance*, 38:548–574, 2016.
- [IYG20] Umar Islambekov, Monisha Yuvaraj, and Yulia R Gel. Harnessing the power of topological data analysis to detect change points. *Environmetrics*, 31(1):e2612, 2020.

- [JKP11] Robert Jarrow, Younes Kchia, and Philip Protter. How to detect an asset bubble. *SIAM Journal on Financial Mathematics*, 2(1):839–865, 2011.
- [JLS00] Anders Johansen, Olivier Ledoit, and Didier Sornette. Crashes as critical points. *International Journal of Theoretical and Applied Finance*, 3(02):219–255, 2000.
- [JS02] Anders Johansen and Didier Sornette. Large stock market price drawdowns are outliers. *Journal of Risk*, 4(92):69–110, Winter 2001/02.
- [JS99a] Anders Johansen and Didier Sornette. Critical crashes. *arXiv preprint cond-mat/9901035*, 1999.
- [JS99b] Anders Johansen and Didier Sornette. Modeling the stock market prior to large crashes. *The European Physical Journal B-Condensed Matter and Complex Systems*, 9:167–174, 1999.
- [JS⁺10] Anders Johansen, Didier Sornette, et al. Shocks, crashes and bubbles in financial markets. *Brussels Economic Review*, 53(2):201–253, 2010.
- [JSL99] Anders Johansen, Didier Sornette, and Olivier Ledoit. Predicting financial crashes using discrete scale invariance. *Journal of Risk*, 1(4):5–32, 1999.
- [JSW⁺96] Anders Johansen, Didier Sornette, Hiroshi Wakita, Urumu Tsunogai, William I Newman, and Hubert Saleur. Discrete scaling in earthquake precursory phenomena: evidence in the Kobe earthquake, Japan. *Journal de Physique I*, 6(10):1391–1402, 1996.
- [KB21] Yuri A Katz and Alain Biem. Time-resolved topological data analysis of market instabilities. *Physica A: Statistical Mechanics and its Applications*, 571:125816, 2021.
- [KM18] Firas A Khasawneh and Elizabeth Munch. Topological data analysis for true step detection in periodic piecewise constant signals. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2218):20180027, 2018.
- [Len11] Timothy M Lenton. Early warning of climate tipping points. *Nature climate change*, 1(4):201–209, 2011.
- [LS13] Li Lin and Didier Sornette. Diagnostics of rational expectation financial bubbles with stochastic mean-reverting termination times. *The European Journal of Finance*, 19(5):344–365, 2013.
- [NM23] Vahid Nateghi and Matteo Manzi. Machine learning methods for nonlinear dimensionality reduction of the thermospheric density field. *Advances in Space Research*, 2023.
- [Per16] Jose A Perea. Persistent homology of toroidal sliding window embeddings. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6435–6439. IEEE, 2016.
- [PH15] Jose A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.
- [Pro13] Philip Protter. *A Mathematical Theory of Financial Bubbles*, pages 1–108. Springer International Publishing, Cham, 2013.
- [PS18] Peter CB Phillips and Shu-Ping Shi. Financial bubble implosion and reverse regression. *Econometric Theory*, 34(4):705–753, 2018.
- [PSY15a] Peter CB Phillips, Shuping Shi, and Jun Yu. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International economic review*, 56(4):1043–1078, 2015.
- [PSY15b] Peter CB Phillips, Shuping Shi, and Jun Yu. Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, 56(4):1079–1134, 2015.

- [PWY11] Peter CB Phillips, Yangru Wu, and Jun Yu. Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1):201–226, 2011.
- [PY11] Peter CB Phillips and Jun Yu. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491, 2011.
- [RCPB19] Rodrigo Rivera-Castro, Polina Pilyugina, and Evgeny Burnaev. Topological data analysis for portfolio management of cryptocurrencies. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pages 238–243. IEEE, 2019.
- [RQD21] Simon Rudkin, Wanling Qiu, and Pawel Dlotko. Uncertainty, volatility and the persistence norms of financial time series. *arXiv preprint arXiv:2110.00098*, 2021.
- [SBB⁺09] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.
- [SH03] D Sornette and A Helmstetter. Endogenous versus exogenous shocks in systems with memory. *Physica A: Statistical Mechanics and its Applications*, 318(3):577–591, 2003.
- [SH05] K Suchecki and JA Holyst. Log-periodic oscillations in degree distributions of hierarchical scale-free networks. *Acta Physica Polonica. Series B*, 35, 2005.
- [SJB96] Didier Sornette, Anders Johansen, and Jean-Philippe Bouchaud. Stock market crashes, precursors and replicas. *Journal de Physique I*, 6(1):167–175, 1996.
- [SNC20] P. Saengduan, S. Noisagool, and F. Chamchod. Topological data analysis for identifying critical transitions in cryptocurrency time series. In *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 933–938, 2020.
- [Sor17] Didier Sornette. Why stock markets crash. In *Why Stock Markets Crash*. Princeton university press, 2017.
- [SP97] Rainer Storn and Kenneth Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341, 1997.
- [SS16] Maximilian Seyrich and Didier Sornette. Micro-foundation using percolation theory of the finite time singular behavior of the crash hazard rate in a class of rational expectation bubbles. *International Journal of Modern Physics C*, 27(10):1650113, 2016.
- [SS17] H Sohn and Didier Sornette. Bubbles as violations of efficient time-scales. *Institute for New Economic Thinking Working Paper Series*, (65), 2017.
- [SSGC⁺18] Manish Saggat, Olaf Sporns, Javier Gonzalez-Castillo, Peter A Bandettini, Gunnar Carlsson, Gary Glover, and Allan L Reiss. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, 9(1):1399, 2018.
- [SSZ21] Min Shu, Ruiqiang Song, and Wei Zhu. The ‘COVID’ crash of the 2020 US Stock market. *The North American Journal of Economics and Finance*, 58:101497, 2021.
- [SSZ22] Ruiqiang Song, Min Shu, and Wei Zhu. The 2020 global stock market crash: Endogenous or exogenous? *Physica A: Statistical Mechanics and Its Applications*, 585:126425, 2022.
- [SZ20a] Min Shu and Wei Zhu. Detection of Chinese stock market bubbles with LPPLS confidence indicator. *Physica A: Statistical Mechanics and Its Applications*, 557:124892, 2020.

- [SZ20b] Min Shu and Wei Zhu. Real-time prediction of Bitcoin bubble crashes. *Physica A: Statistical Mechanics and its Applications*, 548:124477, 2020.
- [Tak81] Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young, editors, *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.
- [TJM91] Sauer Tim, A Yorke James, and Casdagli Martin. Embedology. *Journal of Statistical Physics*, 65(3-4):579–616, 1991.
- [TS11] J Michael T Thompson and Jan Sieber. Climate tipping as a noisy bifurcation: a predictive technique. *IMA Journal of Applied Mathematics*, 76(1):27–46, 2011.
- [WSH⁺19] Spencer Wheatley, Didier Sornette, Tobias Huber, Max Reppen, and Robert N Gantner. Are bitcoin bubbles predictable? combining a generalized Metcalfe’s law and the log-periodic power law singularity model. *Royal Society open science*, 6(6):180538, 2019.
- [WWS18] Ke Wu, Spencer Wheatley, and Didier Sornette. Classification of cryptocurrency coins and tokens by the dynamics of their market capitalizations. *Royal Society open science*, 5(9):180381, 2018.
- [YRdSTT20] Arthur Matsuo Yamashita Rios de Sousa, Hideki Takayasu, and Misako Takayasu. Segmentation of time series in up-and down-trends using the epsilon-tau procedure with application to USD/JPY foreign exchange market data. *PloS one*, 15(9):e0239494, 2020.
- [YRWS12] Wanfeng Yan, Reda Rebib, Ryan Woodard, and Didier Sornette. Detection of crashes and rebounds in major equity markets. *International Journal of Portfolio Analysis and Management*, 1(1):59–79, 2012.

DEPARTMENT OF MATHEMATICAL SCIENCES, YESHIVA UNIVERSITY, USA
Email address: `sakingba@mail.yu.edu`

DEPARTMENT OF MATHEMATICAL SCIENCES, YESHIVA UNIVERSITY, USA
Email address: `marian.gidea@yu.edu`

CRUNCHDAO, FRANCE
Email address: `matteo.manzi@crunchdao.com`

AEROSPACE SCIENCES AND TECHNOLOGY DEPARTMENT, POLITECNICO DI MILANO,
 ITALY
Email address: `vahid.nateghi@mail.polimi.it`