Topological data analysis of financial time series: Landscapes of crashes

Marian Gidea [a,b,*], Yuri Katz [c]

[a] School of Civil Engineering and Architecture, Xiamen University of Technology, Fujian, China
[b] Department of Mathematical Sciences, Yeshiva University, New York, NY 10016, USA
[c] S&P Global Market Intelligence, 55 Water Str., New York, NY 10040, USA

HIGHLIGHTS

- We introduce a new method, based on topological data analysis (TDA), to analyze financial time series, and detect possible early signs prior to financial crashes.
- We analyze the time-series of daily log-returns of four major US stock market indices: S&P 500, DJIA, NASDAQ, and Russell 2000.
- We use persistence homology to detect and quantify topological patterns that appear in the multidimensional time series.
- We find that, in the vicinity of financial meltdowns, the L^p-norms of persistence landscapes exhibit strong growth prior to the primary peak, which ascends during a crash.
- Our method is very general and can be applied to any asset-types and mixtures of time series.

ARTICLE INFO

ABSTRACT

We explore the evolution of daily returns of four major US stock market indices during the technology crash of 2000, and the financial crisis of 2007–2009. Our methodology is based on topological data analysis (TDA). We use persistence homology to detect and quantify topological patterns that appear in multidimensional time series. Using a sliding window, we extract time-dependent point cloud data sets, to which we associate a topological space. We detect transient loops that appear in this space, and we measure their persistence. This is encoded in real-valued functions referred to as a 'persistence landscapes'. We quantify the temporal changes in persistence landscapes via their L^p-norms. We test this procedure on multidimensional time series generated by various non-linear and non-equilibrium models. We find that, in the vicinity of financial meltdowns, the L^p-norms exhibit strong growth prior to the primary peak, which ascends during a crash. Remarkably, the average spectral density at low frequencies of the time series of L^p-norms of the persistence landscapes demonstrates a strong rising trend for 250 trading days prior to either dotcom crash on 03/10/2000, or to the Lehman bankruptcy on 09/15/2008. Our study suggests that TDA provides a new type of econometric analysis, which complements the standard statistical measures. The method can be used to detect early warning signals of imminent market crashes. We believe that this approach can be used beyond the analysis of financial time series presented here.

© 2017 Elsevier B.V. All rights reserved.

---

* Corresponding author at: Department of Mathematical Sciences, Yeshiva University, New York, NY 10016, USA.
  E-mail addresses: Marian.Gidea@yu.edu (M. Gidea), yuri.katz@spglobal.com (Y. Katz).

1. Introduction

Topological Data Analysis (TDA) [1,2] refers to a combination of statistical, computational, and topological methods allowing to find shape-like structures in data. The TDA has proven to be a powerful exploratory approach for complex multi-dimensional and noisy datasets. For TDA to be applied, a data set is encoded as a finite set of points in some metric space. The general and intuitive principle underlying TDA is based on persistence of k-dimensional holes, e.g., connected components ($k = 0$), loops ($k = 1$), etc., in a topological space that is inferred from random samples for a wide range of scales (resolutions) at which data is looked at. Accordingly, persistent homology is the key topological property under consideration [3,4].

The procedure to compute persistent homology associated to a point cloud data set involves the construction of a filtration of simplicial complexes, ordered with respect to some resolution (scaling) parameter. As the resolution parameter changes, some topological features appear in the corresponding simplicial complex while others disappear. Thus, each topological feature is assigned a 'birth' and a 'death' value, and the difference between the two values represents the persistence of that feature. A topological feature that persists for a bigger range of scales can be viewed as a significant one, while a feature that persists for a smaller range can be viewed as a less significant, or a noisy feature. An important quality of the persistence homology method is that it does not require an artificial cutoff between 'signal' and 'noise'; all topological features that emerge from the data are kept, and assigned 'weights' according to their persistence. The output of the filtration procedure is captured in a concise form by a persistence diagram. The two coordinates of each point in the diagram represent the birth value and the death value of a k-dimensional hole. An alternative instrument to summarize the information contained in a persistence diagram is a persistence landscape [5,6]. The latter consists of a sequence of continuous, piecewise linear functions defined in the rescaled birth–death coordinates, which are derived from the persistence diagram. Persistence diagrams have a natural metric space structure whereas persistence landscapes are naturally embedded in a Banach space. Thus, one can study the statistical properties of persistence landscapes, e.g., compute expectations and variances, among other properties [5,7].

A remarkable property of persistence homology is that both persistence diagrams and persistence landscapes are robust under perturbations of the underlying data. That is, if the data set changes only little, the persistence diagrams/persistence landscapes move only by a small distance. This feature is a key ingredient for mathematically well-founded statistical developments using persistence homology.

Exploration of stable topological structures (or 'shapes') in nosy multidimensional datasets has led to new insights, including the discovery of a subgroup of breast cancers [8], is actively used in image processing [9], in signal and time-series analysis [10–15]. The latter has primarily been applied to detect and quantify periodic patterns in data [16,17], to understand the nature of chaotic attractors in the phase space of complex dynamical systems [18], to analyze turbulent flows [19], and stock correlation networks [20].

Motivated by these studies, in this paper, we investigate whether application of TDA to financial time series could help to detect a growing systemic risk in financial markets. Despite the obvious practical interest for policymakers and market participants, prediction of catastrophic market meltdowns is notoriously difficult due to complexity and non-stationarity of the financial system. During last decades, there has been a growing body of empirical and theoretical studies, inspired by analysis of abrupt transitions in complex natural systems, devising early warning signals (EWS) in financial markets, see, e.g., [21] and references therein. Observations on different markets show that financial crashes are preceded by a period of increasing variance of stock market indices, shifting of spectral density of time series towards low frequencies as well as growing cross-correlations. Yet, there is no consensus regarding the mechanism of financial crises. Moreover, even a relatively short-term forecasting of approaching financial disaster remains one of the open challenges.

We analyze the time-series of daily log-returns of four major US stock market indices: S&P 500, DJIA, NASDAQ, and Russell 2000. Collectively, these noisy 1D signals form a multi-dimensional time series in 4D-space. We apply a sliding window of certain length w along these time series, thereby obtaining a 4D-point cloud for each instance of the window. The sliding step is set to one day. Then we compute the L^p-norms ($p = 1$ and $p = 2$) of the persistence landscape of the loops (1D persistent homology) in each of the 4D-point clouds. The resulting time series of L^p-norms allows to track temporal changes in the state of the equity market. We find that the time series of L^p-norms exhibits strong growth prior to the primary peak, which ascends during a crash. Remarkably, the average spectral density of the time series of the norms of the persistence landscapes demonstrates a strong rising trend at low frequencies for 250 trading days prior to either the dotcom crash on 03/10/2000 or to the Lehman bankruptcy on 09/15/2008. Our study suggests that TDA provides a new type of econometric analysis that can be used to detect EWS of imminent market crashes. The method is very general and can be applied to any asset-types and mixtures of time series.

In Section 2 we provide a concise and informal review of the TDA background and key concepts employed in this paper. In Section 3 we compare our method outlined above with the traditional time-delay coordinate embedding procedure. We test the proposed method on several multidimensional time series, generated by various non-linear and non-equilibrium models. These tests help to assess how the characteristics of the underlying process affect the persistence landscapes. In each case, we let the parameters of the underlying process to change in specific ways, and observe which features of the signal makes the L^p-norms of the persistence landscapes grow. Section 4 presents our findings on financial data, which demonstrate that the time series of the norms of the persistence landscapes and its variability could be used as new EWS of approaching market crashes. Section 5 concludes the paper.

2. Background

TDA uses topology to extract information from noisy data sets. An important step is to assign topological invariants that remain robust under small changes of the scale at which the data set is being considered. This is particularly relevant for analyzing complex systems, when there is no information available on the underlying stochastic process. The method of persistence homology, which we use in this paper, allows computing the topological features at all scales, and ranking these features according to the range of scales at which these features can be observed.

Here we provide a brief account of persistent homology, with an emphasis on the tools that we use directly. For a rigorous exposition, we refer the interested reader to the textbook [2].

The input data that we analyze is a point cloud data set consisting, of a family of points $X = \{x_1, \ldots, x_n\}$ in an Euclidean space $\mathbb{R}^d$. We associate a topological space to such data set as follows. For each distance $\varepsilon > 0$ we define the so called Vietoris–Rips simplicial complex $R(X, \varepsilon)$, or, simply Rips complex, obtained as follows:

- for each $k = 0, 1, 2, \ldots$, a k-simplex of vertices $\{x_{i_1}, \ldots, x_{i_k}\}$ is part of $R(X, \varepsilon)$ if and only if the mutual distance between any pair of its vertices is less than ε, that is

  $$d(x_{i_j}, x_{i_l}) < \varepsilon, \text{ for all } x_{i_j}, x_{i_l} \in \{x_{i_1}, \ldots, x_{i_k}\}.$$

If we think of ε as the resolution at which that data is looked at, then a k-simplex is included in $R(X, \varepsilon)$ for every set of k data points that are indistinguishable from one another at resolution level ε.

The Rips simplicial complexes $R(X, \varepsilon)$ form a filtration, that is, $R(X, \varepsilon) \subseteq R(X, \varepsilon')$ whenever $\varepsilon < \varepsilon'$. For each such complex, we can compute its k-dimensional homology $H_k(R(X, \varepsilon))$ with coefficients in some field. Informally, the generators of the 0-dimensional homology group $H_0(R(X, \varepsilon))$ correspond to the connected components of $R(X, \varepsilon)$, the generators of the 1-dimensional homology group $H_1(R(X, \varepsilon))$ correspond to the 'independent loops' in $R(X, \varepsilon)$, the generators of the 2-dimensional homology group $H_0(R(X, \varepsilon))$ correspond to 'independent cavities' in $R(X, \varepsilon)$, etc.

In the sequel, we will use only the 1-dimensional homology. As an example of a loop at level ε, consider a set of 4 points $\{x_{i_0}, x_{i_1}, x_{i_2}, x_{i_3}\}$ in X such that the distance $d(x_{i_j}, x_{i_{j+1}}) < \varepsilon$, for $j = 0, \ldots, 3$, with j computed mod 4, but no other pair of points in this set is within a distance of ε. We note that, by the definition of the Rips complex, we cannot have a loop consisting of only 3-points.

The filtration property of the Rips complexes induces a filtration on the corresponding homologies, that is $H_k(R(X, \varepsilon)) \subseteq H_k(R(X, \varepsilon'))$ whenever $\varepsilon < \varepsilon'$, for each k. These inclusions determine canonical homomorphisms $H_k(R(X, \varepsilon)) \hookrightarrow H_k(R(x, \varepsilon'))$, for $\varepsilon < \varepsilon'$. Due to this family of induced mappings, for each non-zero k-dimensional homology class α there exists a pair of values $\varepsilon_1 < \varepsilon_2$, such that $\alpha \in H_k(R(X, \varepsilon_1))$ but α is not in the image of any $H_k(R(X, \varepsilon_1 - \delta))$, for $\delta > 0$, and the image of α in $H_k(R(X, \varepsilon'))$ is non-zero for all $\varepsilon_1 < \varepsilon' < \varepsilon_2$, but the image of α in $H_k(R(X, \varepsilon_2))$ is zero. In this case, one says that α is 'born' at the parameter value $b_\alpha := \varepsilon_1$, and 'dies' at the parameter value $d_\alpha = \varepsilon_2$, or that the pair (b_α, d_α) represent 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.

The information on the k-dimensional homology generators at all scales can be encoded in a so called persistence diagram P_k. Such a diagram consists of:

- For each k-dimensional homology class α one assigns a point $z_\alpha = (b_\alpha, d_\alpha) \in \mathbb{R}^2$ together with its multiplicity $\mu_\alpha(b_\alpha, d_\alpha)$;
- In addition, P_k contains all points in 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 indices on the horizontal axis and death indices on the vertical axis. For an example, see Fig. 1.

The space of persistence diagrams can be endowed with a metric space structure. A standard metric that can be used is the *degree p Wasserstein distance*, with $p \geq 1$, defined by

$$W_p(P_k^1, P_k^2) = \inf_\phi \left[\sum_{q \in P_k^1} \|x - \phi(x)\|_\infty^p \right]^{1/p},$$

where the summation is over all bijections $\phi : P_k^1 \to P_k^2$, and $\|\cdot\|_\infty$ denotes the sup norm. Since persistence diagrams include the diagonal set, the above summation includes pairing of points $(x, \phi(x))$ between off-diagonal points and diagonal points in P_k^1, P_k^2. When $p = \infty$ the Wasserstein distance W_∞ is known as the 'bottleneck' distance.

A remarkable property that makes persistence homology suitable to analyze noisy data is its *robustness* under small perturbations. Informally, this property says that if the underlying cloud point data set changes only 'little', then the corresponding persistence diagrams moves only a 'small' Wasserstein distance from the persistence diagram corresponding to the unperturbed data; see [22].
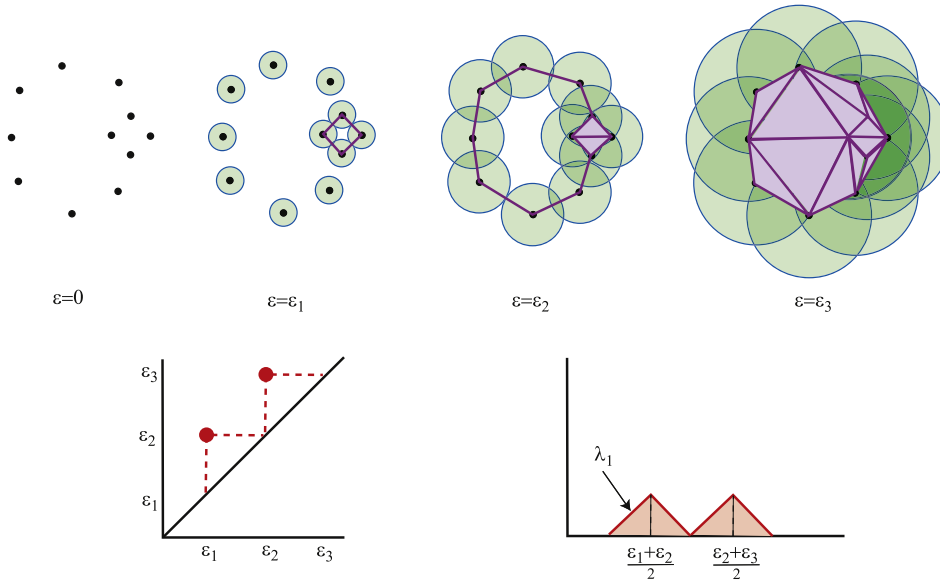
Fig. 1. Rips filtration of simplicial complexes illustrating the birth and death of loops; the 1-dimensional persistence diagram and the corresponding persistence landscape are shown below.

The space of persistence diagrams $\mathscr{P}$ endowed with the Wasserstein distance forms a metric space $(\mathscr{P}, W_p)$. However, this metric space is not complete, hence not appropriate for statistical treatment. This is an inconvenience for our purposes, since we want to study time series of persistence diagrams with statistical tools.

To address this issue, several works have been devoted to either modify the structure of the space of persistence diagrams endowed so that it has a more amenable structure (e.g., of a Polish space, or geodesic space), or to embed the space of persistence diagrams into function spaces. One such an embedding is based on persistence landscapes, consisting of sequences of functions in the Banach space $L^p(\mathbb{N} \times \mathbb{R})$; see [5,6]. We now define persistence landscapes. For each birth–death point $(b_\alpha, d_\alpha) \in P_k$, we first define a piecewise linear function

$$f_{(b_\alpha, d_\alpha)} = \begin{cases} x - b_\alpha, & \text{if } x \in \left(b_\alpha, \dfrac{b_\alpha + d_\alpha}{2}\right]; \\ -x + d_\alpha, & \text{if } x \in \left(\dfrac{b_\alpha + d_\alpha}{2}, d_\alpha\right); \\ 0, & \text{if } x \notin (b_\alpha, d_\alpha). \end{cases} \tag{2.1}$$

To avoid confusion, we point out that in Eq. (2.1), by (b_α, d_α) we mean the open interval with endpoints b_α and d_α, while earlier we also denoted by (b_α, d_α) a point in P_k of coordinates b_α and d_α.

To a persistence diagram P_k consisting of a finite number of off-diagonal points, we associate a sequence of functions $\lambda = (\lambda_k)_{k\in\mathbb{N}}$, where $\lambda_k : \mathbb{R} \to [0; 1]$ is given by

$$\lambda_k(x) = k\text{-}\max\{f_{(b_\alpha, d_\alpha)}(x) \mid (b_\alpha, d_\alpha) \in P_k\} \tag{2.2}$$

where the k-max denotes the kth largest value of a function. We set $\lambda_k(x) = 0$ if the kth largest value does not exist. For an example of a persistence landscape, see Fig. 1.

Thus, persistence landscapes form a subset of the Banach space $L^p(\mathbb{N} \times \mathbb{R})$. This consists of sequences of functions $\eta = (\eta_k)_{k\geq 0}$, where $\eta_k : \mathbb{R} \to \mathbb{R}$ for $k \geq 0$. This set has an obvious vector space structure, given by $(\eta^1 + \eta^2)_k(x) = \eta_k^1(x) + \eta_k^2(x)$, and $(c \cdot \eta)_k(x) = c \cdot \eta_k(x)$, for all $x \in \mathbb{R}$, and $k = 1, 2, \dots$ It becomes a Banach space when endowed with the norm

$$\|\eta\|_p = \left(\sum_{k=1}^{\infty} \|\eta_k\|_p^p\right)^{1/p}, \tag{2.3}$$

where $\| \cdot \|_p$ denotes the L^p-norm, i.e., $\|f\|_p = \left(\int_{\mathbb{R}} |f|^p\right)^{1/p}$, where the integration is with respect with the Lebesgue measure on $\mathbb{R}$. In the sequel, we will use only the L^1 and L^2 norms. Generally, it is not possible to go back and forth between persistence diagrams and landscapes. There exist landscapes that do not correspond to persistence diagrams, e.g., if λ^1, λ^2 are two persistence landscapes derived, via Eq. (2.1), from corresponding persistence diagrams $P^1, P^2 \in \mathscr{P}$, respectively, the mean $\bar{\lambda} = (\lambda^1 + \lambda^2)/2$ may not correspond, via Eq. (2.1), to any persistence diagram $\bar{P} \in \mathscr{P}$; see [5]. However, the Banach space structure makes persistence landscapes suitable for treatment via *statistical methods*.

3. Description of the method and testing on synthetic time series

In Section 4 we will analyze financial time series using persistence landscapes and their norms ($p = 1$ and $p = 2$). We will see that these norms exhibit substantial growth prior to financial crashes, while they exhibit a tame behavior when the market is stable. In this section, we will explain our method of analysis and test it on several synthetic multidimensional time series generated by various non-linear and non-equilibrium models. The objective is to assess how various features of the signal affect the growth of the L^p-norms of the persistence landscapes.

3.1. Method

Consider d time series $\{x_n^k\}_n$, where $k = 1, \ldots, d$, and a sliding window of size w. For each instance t_n, we have a point $x(t_n) = (x_n^1, \ldots, x_n^d) \in \mathbb{R}^d$. Then for each time-window of size w we have a point cloud data set consisting of w points in $\mathbb{R}^d$, namely $X_n = (x(t_n), x(t_{n+1}), \ldots, x(t_{n+w-1}))$. Algorithmically, the point cloud is formed by a $d \times w$ matrix, where d is the number of columns corresponding to the number of 1D time series under study and w is the size of the window, which determines the length of a column. Then TDA is applied on top of the time-ordered sequence of point clouds to study the time-varying topological properties of the multidimensional time series, from *window* to *window*. For each point cloud we compute the persistence diagram of the Rips filtration, the corresponding persistence landscape, and its L^p-norms. We use efficient computational algorithms and the relevant TDA libraries allowing computation of topological features from noisy observations [23]. In our implementation we employ the R-package "TDA" [24], which greatly simplifies the application programming interface.

In the following subsection we illustrate our method to analyze a time series generated by a stochastic version of the Hénon map.

Remark 3.1. We now compare our approach with the well-known method based on time-delay coordinate embedding [25], which is broadly used in the conventional analysis and TDA studies of 1D signals.

First, we describe the case when the underlying signal is stationary and represents the time series of a chaotic attractor (hidden in a 'black box'). A sliding window of size m is applied to this time series, with each instance of a window giving rise to a point in $\mathbb{R}^m$. The theory ensures the resulting set of points in $\mathbb{R}^m$ represents (approximately) an embedding of the attractor in the m-dimensional space, provided the dimension m is chosen large enough: at least twice the fractal dimension of the attractor [26]. There are several methods to estimate, directly from the time series, the minimal dimension m necessary to achieve the embedding [27,28]. We should stress that this approach becomes increasingly difficult to apply when the dimension of the attractor is large, or in the presence of noise [29]; see also [30].

The second case that we describe is when the 1D time series is associated to a non-stationary process; we follow [15]. Given a lengthy 1D time-series, one takes a partition of it into multiple segments of a predefined size w. Within each segment, a sliding window of size m is applied to the time series, where each instance of the window is consists of a sequence of m consecutive data-points from the time series, where $m \ll w$. Again, m has to be chosen suitably large. The time-ordered sequence of sliding windows for each segment forms a single point cloud of size $(w - m + 1)$ in $\mathbb{R}^m$. Thereby, one obtains a representation of each segment of length w of the 1D data set in the m-dimensional space. Then one applies TDA on top of the embedded point clouds to study the time-varying topological properties of the time series, *from segment to segment*.

We stress that our method is quite different from the time-delay embedding method. As we do not aim to obtain an embedding of the time series in some *high dimensional space* — as such embedding might be practically impossible to realize, due to the a priori lack of an attractor, and to the intrinsic stochasticity of the time series — we choose to represent the time series in a low dimensional space, namely in $\mathbb{R}^d$, where $d \ll w$, as we are dealing with a handful (that is, d) separate time series. Our approach has only one parameter, the size w of the window sliding along the time series, as in contrast with the time-delay method, which has two parameters, the size w of the segments, and the size m of a sliding window along each segment.

We stress that both the length of the segment w in [15] and the length of the sliding window w in our method play the same role, namely to determine the size of the point cloud to which we apply TDA, from one segment to another in the former, and from one window to another in the latter. The dimension m of the point cloud is derived from the time-delay embedding procedure in the former, and is naturally given by the number d of time series under consideration in the latter.

3.2. Chaotic time series with noise

First, we recall the Hénon map [31], which can be viewed as a 2-dimensional analogue of the well known logistic map. The latter has been used as a paradigm for economic cycles (see, e.g., [32,33]). The evolution of stable states, bifurcations, periodic and chaotic regimes – displayed by both the logistic map and the Hénon map – provides a useful analogy for the dynamics of economic oscillations and financial crashes (see, e.g., [34]). A model for financial crashes based on bifurcations in dynamical systems appears in [35].

The equations of the Hénon map are:

$$\begin{aligned} x_{n+1} &= 1 - a x_n^2 + b y_n, \\ y_{n+1} &= x_{n+1}. \end{aligned} \tag{3.1}$$
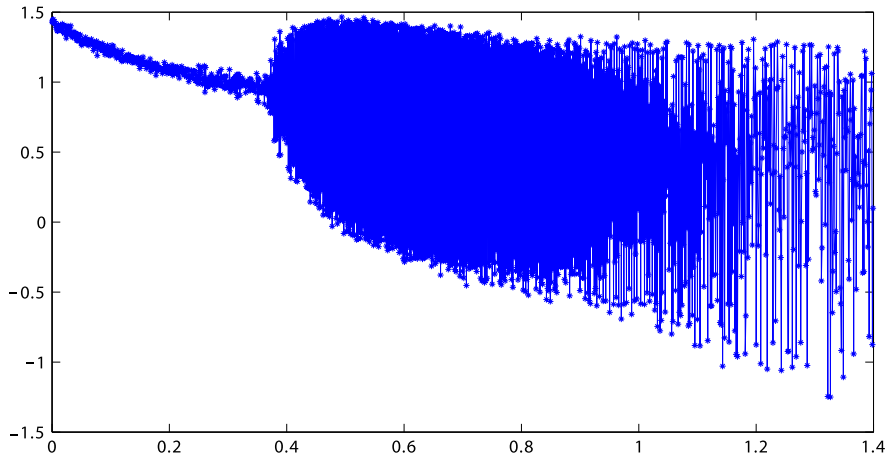
Fig. 2. An *x*-time series corresponding to the system Eq. (3.2). The horizontal axis corresponds to the values of the parameter *a*.

where *a*, *b* are real parameters. It has been shown that, for some range of values of *a*, *b*, every initial condition (x_0, y_0) in some appropriate region of the plane — the 'basin of attraction' –, the sequence (x_n, y_n) approaches the same invariant set, known as the Hénon attractor. For some values of *a* and *b*, the attractor is *chaotic*. Let us fix the parameter $b = 0.3$ and let $0 \leq a \leq 1.4$. Numerical experiments show that for fixed values of parameter the parameter *a* with $0 < a < 1.06$, there exists an attractor that undergoes period-doubling bifurcations, whereas at $a \approx 1.06$ a chaotic attractor emerges; for *a* between 1.06 and 1.4, there exist intervals of values of *a* for which there is a chaotic attractor, which are interspersed with intervals of values of *a* for which there is an attractive periodic orbit.

We modify the equations of the Hénon attractor by making one of the parameters change slowly in time (analogous to an external forcing), and we also dress it with noise. For fixed values of the parameter *b* ($b = 0.27$, $b = 0.28$, $b = 0.29$, and $b = 0.3$), we let the parameter *a* grow slowly and linearly in time, from $a = 0$ to $a = 1.4$, and also add the Gaussian noise. That is, we consider the system:

$$
\begin{aligned}
x_{n+1} &= 1 - a_n x_n^2 + b y_n + \sigma W_n \sqrt{\Delta t}, \\
y_{n+1} &= x_{n+1} + \sigma W_n \sqrt{\Delta t}, \\
a_{n+1} &= a_n + \Delta t,
\end{aligned}
\tag{3.2}
$$

where W_n is a normal random variable, $\Delta t > 0$ is a small step size, and $\sigma > 0$ is the noise intensity. The stochastic terms in Eq. (3.2) correspond to a diffusion process. The time changes according to $t_{n+1} = t_n + \Delta t$, hence a_n can be viewed as equivalent to the time variable. An realization of the *x*-time series for this system is shown in Fig. 2.

First, we generate $d = 4$ realizations of the *x*-time series $\{x_n^k\}_n$, where $k = 1, \ldots, d$, one for each value of the parameter $b = 0.27$, $b = 0.28$, $b = 0.29$, $b = 0.3$. Thus, for each time t_n, we have a point $x_n = (x_n^1, \ldots, x_n^d) \in \mathbb{R}^d$. For a sliding windows $w = 50$, we generate a sequence of point cloud data sets $X_n = (x_n, x_{n+1}, \ldots, x_{n+w-1})$. Hence each cloud consists of w points in $\mathbb{R}^d$. In Fig. 3, we show two-dimensional projections of point clouds corresponding to various instants of time; as time progresses, the value of the parameter *a* increases. Around each point, we draw a blue circle of a small radius (left column), and of a bigger radius (right column), to illustrate how loops are born and die. The panels at the top correspond to low values of the parameter *a*, when the deterministic Hénon map Eq. (3.1) has a stable fixed point or a stable periodic orbit of low period; the corresponding noisy Hénon map Eq. (3.2) shows a cluster of points, or a few clusters apart from one another. The panels at the bottom correspond to larger values of the parameter *a*, when the deterministic Hénon map Eq. (3.1) undergoes a period doubling bifurcation that leads to a chaotic attractor; the corresponding noisy Hénon map Eq. (3.2) shows a complex structure.

Next, for each cloud point X_n we associate the corresponding Rips filtration $R(X_n, \varepsilon)$, $\varepsilon > 0$, and we compute the 1D persistence diagram $P_1(X_n)$, the corresponding persistence landscape $\lambda(X_n)$, and the L^1 and L^2-norms: $\|\lambda(X_n)\|_1$ and $\|\lambda(X_n)\|_2$, respectively. In Fig. 4 we show the normalized time series of $\|\lambda(X_n)\|_1$ (in blue) and $\|\lambda(X_n)\|_2$ (in red). Both graphs show sharp increase around the parameter value *a* that marks the onset of chaotic behavior. We also point out that, when the *x*-time series move towards the chaotic regime, the cross-correlation among them decays; in the chaotic regime, where there is sensitive dependence on initial conditions, the cross-correlation typically gets below $\pm 10\%$. At the same time, the noise dressing added to the signal only plays a minor role.

The conclusion of this numerical experiment is that the L^p-norms of persistence landscapes have the ability to detect transitions from regular to chaotic regimes, in systems with slowly evolving parameters. Passing from regular to chaotic dynamics determines significant changes in the topology of the attractor, which are picked up very clearly by the time series of L^p-norms. We stress that we only use this example to test out the proposed method. In the next subsection we consider a model that is more relevant to empirical financial time series.
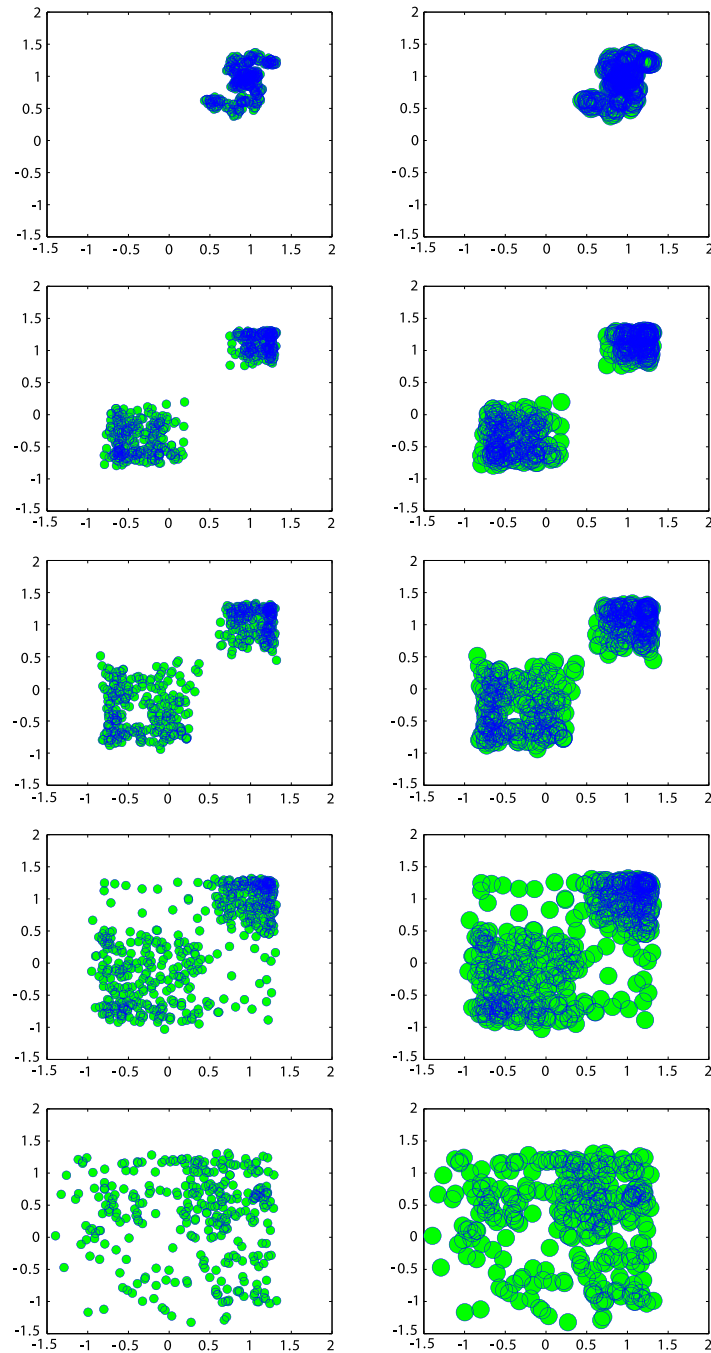
Fig. 3. Two-dimensional projections of point clouds corresponding to various instants of time; as time progresses, the value of the parameter *a* increases (from top to bottom). Around each point, we draw a blue circle of a small radius (left column), and of a bigger radius (right column), to illustrate the birth and death of various loops. Color online.

3.3. White noises with a growing variance

The classic assumption made in theoretical quantitative finance is that equity prices follow a geometric Brownian motion with constant drift and diffusion coefficients. This model leads to normally distributed stock returns, defined as the difference in the logarithm of prices separated by a given time lag. In the following we use our method and perform Monte Carlo simulations of the L^p-norms of persistence landscapes of a finite number $d = 4$ white noises. Specifically, we wish to test whether a growing variance of the underlying white noises leads to increase in values of L^p-norms.
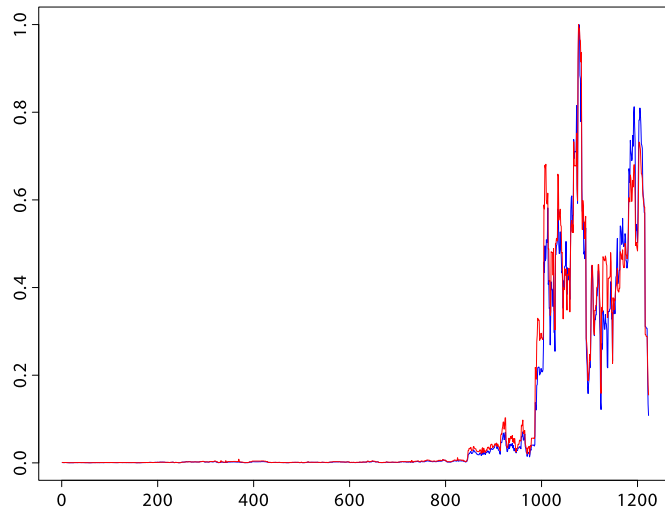
Fig. 4. Plots of the normalized L^1-norm (blue) and L^2-norm (red) of the persistence landscapes for a chaotic time series with noise. The horizontal axis corresponds to the steps in time, and the vertical axis to the norms (see text for details). Color online.
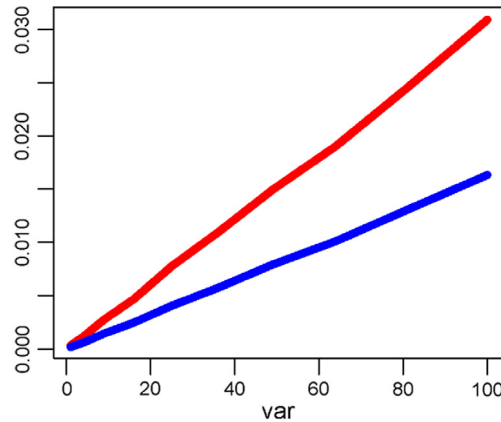
Fig. 5. Monte Carlo simulation, 100 realizations, of the dependency of the L^1 (red line) and L^2 (blue line) norms of persistence landscapes on the growing variance of white noises, marked on the horizontal axis (see text for details). Color online.

We design our numerical experiment as follows. First, we generate four data sets, each with 100 data points selected from Normal distributions $N(0, (\sigma + \delta_i)^2)$ with standard deviations $\sigma + \delta_i$, where $\sigma = 1$ and δ_i, $i = 1, \ldots, 4$, has been chosen from the uniform distribution $[-0.1, 0.1]$. Thereby, we construct a point cloud in the 4D space. We repeat this procedure multiple times. For each realization of the generated set X_n we associate the corresponding Rips filtration $R(X_n, \varepsilon)$, $\varepsilon > 0$, and compute the persistence diagram $P_1(X_n)$, the corresponding persistent landscape $\lambda(X_n)$, and the L^p-norms: $\|\lambda(X_n)\|_1$ and $\|\lambda(X_n)\|_2$. We collect the values of the L^p-norms obtained at each realization, and calculate their mean-values at the end of the first simulation. We run this simulation 10 times, with σ that has been sequentially increased from 1 to 10, and find that the average values of the L^1-and L^2-norms of persistence landscapes rapidly converge with the number of iterations towards *linearly increasing functions* of the average variance of the generated white noises with small positive cross-correlation (~ 0.1), see Fig. 5. This result is not surprising. When the average standard deviation of white noises is increased by some scaling factor F, the effect on the persistence diagram is that each point (b, d) becomes $(F \cdot b, F \cdot d)$, and the area under the graph of each function λ_k (see (2.2)) is changed by a factor of F^2. Hence, in this case, the values of the L^p-norms are proportional to the variance σ^2.

3.4. White noises with Gamma-distributed inverse variance

Since early 1960's it has been known that PDFs of equity returns have "fat tails" [36,37]. The non-Gaussian distributions of relative changes in value of different asset types and on multiple time scales is an established stylized fact, see, e.g., [38,39] and the references therein. There is strong evidence that the time evolution of the volatility of stock returns is determined

by its own stochastic process [40]. There is a growing interest in modeling financial time series within the framework of non-extensive statistics [32,41–45] and 'superstatistics' [46–52]. The superstatistical approach [53] significantly simplifies the description of the temporal evolution of the logarithm of asset returns, x. Briefly, the model is based on the assumption that the typical relaxation time τ of the local dynamics is much shorter than the time scale T_γ of externally driven fluctuations of its realized variance, $\sigma^2 = 1/\gamma$. One assumes that a given discrete time series of x containing n data points can be divided into $N = n/T_\gamma$ time slices with locally equilibrium Normal distribution of x and a distinct value of the realized local variance. Consequently, for $N \gg 1$, the long-term, $t \gg T_\gamma$, the unconditional probability distribution of x is approximated by the superposition of two distributions (therefore, superstatistics): locally normal distribution of x and the probability distribution $f(\gamma)$ of the inverse variance

$$p(x) = \int_0^\infty \sqrt{\gamma/2\pi} \exp(-\gamma x^2/2) f(\gamma) d\gamma. \tag{3.3}$$

Marginalization over γ expressed by this formula has been originally proposed by Praetz [54] who has shown that an infinite mixture of Gaussian PDFs of stock log-returns with the Gamma distributed precision γ:

$$f(\gamma; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^{\alpha-1} \exp(-\beta\gamma), \quad \gamma \geq 0, \alpha, \beta > 0, \tag{3.4}$$

yields the scaled Student's t-distribution, which provides a much better fit to empirical data than the normal distribution. Here $\Gamma(\alpha)$ is the Gamma function, α and β are the shape and the rate parameters of the Gamma distribution, respectively. For large α the Gamma distribution converges to the Gaussian distribution with mean $\mu = \alpha/\beta$ and variance $\sigma^2 = \alpha/\beta^2$. In this limit, Eq. (3.3) recovers the Normal distribution of log-returns, which follows from the classic geometric Brownian motion model of equity prices. Yet, for small values of the shape parameter corresponding to a high weight of realizations with high variance, see Fig. 6(a), Eq. (3.3) leads to fat-tailed distributions of x.

As Praetz pointed out, by analogy with Brownian motion, the variance of price changes over a unit time interval (the diffusion coefficient) is proportional to a slowly changing "…'temperature' of the share market …, which represents the degree of activity or energy of the markets". His insight and formulation of the problem are still incredibly valuable. Recent studies of the daily time series of DJIA and S&P 500 indices show that empirical distributions of γ are close to the Gamma distribution [47,55]. Since the main objective of this paper is to explore the utility of TDA in forecasting of an approaching financial crash, which is conventionally associated with a growing 'market temperature', in the following we use our method and perform Monte Carlo simulations of the L^p-norms of persistence landscapes of $d = 4$ white noises with Gamma-distributed inverse variance.

We generate 100 point clouds in the $4D$ space, each with 100 data points selected from four Normal distributions $N(0, \gamma^{-1/2})$, with γ that has been randomly taken from the set of 100 Gamma-distributed variables. To imitate the transition from the 'cold' (low variance) to the 'hot' (high variance) state of a market, for the first 75 point clouds we keep the scale parameter relatively high $\alpha = 8$. Then we decrease this parameter by 0.25 for each of the last 25 steps. We keep the rate parameter fixed, $\beta = 1$. For each generated point cloud X_n we associate the corresponding Rips filtration $R(X_n, \varepsilon), \varepsilon > 0$, and compute the 1D persistence diagram $P_1(X_n)$, the corresponding persistence landscape $\lambda(X_n)$, and the norms: $\|\lambda(X_n)\|_1$ and $\|\lambda(X_n)\|_2$, respectively. We repeat this procedure multiple times, collect the values of L^p-norms obtained at each realization, and calculate the relevant mean-values. In Fig. 6(b) we show the normalized time series of the L^1-norm and the average variance of underlining noises, averaged over 100 realizations; the behavior of the L^2-norm is almost indistinguishable from that of the L^1-norm and is not shown here. The plot shows a sharp increase of the L^1-norm of persistence landscapes with the decrease of the shape parameter, which corresponds to a growing variance of the distributions of x. This numerical experiment demonstrates that the L^p-norms of persistence landscapes are very sensitive to transitions in the state of a system from a regular to a 'heated' regime. Note that the cross-correlation between the generated noises at each realization is rather low: it never exceeds $\pm 20\%$. On the other hand, we do not observe any persistent loops for identical noises. In this special case, the L^p-norms are always equal to zero.

Once again, we stress that the goal of these simulations is to check the applicability of the proposed method to noisy multidimensional time series that are typical for complex systems. We certainly do not claim that the particular models considered here are representative for how crises occur in financial markets.

4. Empirical analysis of financial data

We analyze the daily time series of four major US stock market indices: S&P 500, DJIA, NASDAQ, and Russell 2000 between December 23, 1987 and December 08, 2016 (7301 trading days). The times series data were downloaded from Yahoo Finance. For each index and for each trading day we calculate log-returns defined as the forward daily changes in the logarithm of the ratio $r_{ij} = \ln(P_{i,j}/P_{i-1,j})$, where $P_{i,j}$ represents the adjusted closing value of the index j at the day i. We wish to explore the topological properties of this multi-dimensional time series. In our approach, the point cloud is formed by w points in $\mathbb{R}^d$, where the coordinates of each point in $\mathbb{R}^d$ represent the daily log-returns. That is, each point cloud is given by a $w \times d$ matrix, where the number of columns $d = 4$ is the number of 1D time series involved in our analysis, and the number of rows w is the size of the sliding window. We consider two sizes of the window: $w = 50$ and $w = 100$ trading days; see Fig. 7 for an example of 2D point clouds. The sliding step is set to one day, which in the case under consideration yields a $(7300 - w)$ time-ordered set of point clouds.
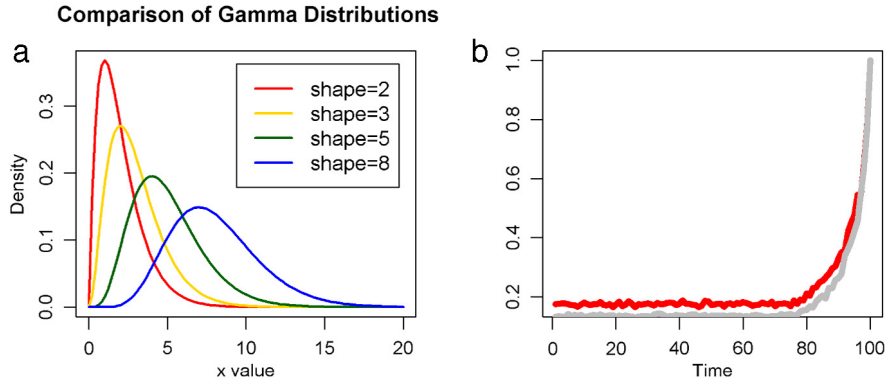
Fig. 6. (a) Gamma-distributions with different values of the shape parameter; (b) Monte Carlo simulation of the time series of the normalized L^1-norm (red line) of persistence landscapes derived for the process with the Gamma distributed inverse variance of the white noise. The gray line represents the time series of the normalized average variance of four underlining noises. For details see text. Color online.
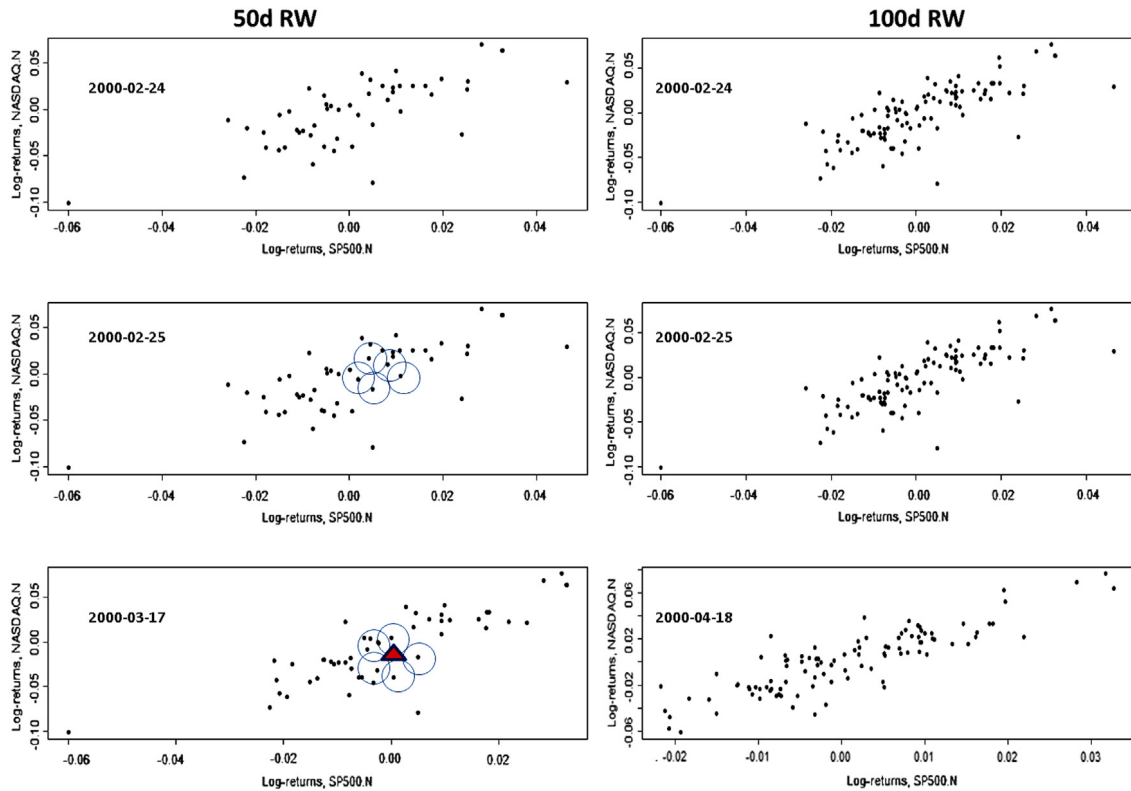
Fig. 7. Scattered plots of the time series of the normalized daily log-returns of S&P 500 and NASDAQ within 50 and 100 days intervals ending at selected dates. The left column schematically demonstrates birth of the $1D$ homology (loops) at certain scale (radius) in the $2D$ point cloud of 50 data points.

Fig. 8 shows the Rips persistence diagrams and the corresponding 1-st landscape function λ_1, defined by Eq. (2.2), calculated with the sliding window of 50 trading days at certain dates. The x and y coordinate of Rips persistence diagrams correspond to the times of birth (b) and death (d) of a topological feature, respectively. The x and y coordinate of persistence landscapes correspond to the rescaled axes: $x = (d+b)/2$, and $y = (d-b)/2$. In plain words, the x-coordinate of a persistence landscape is the average parameter value over which the feature exists, and the y-coordinate is the half-life of the feature. Fig. 8 demonstrates a visually clear separation of topological signals from a topological noise per point cloud. Two of selected dates correspond to the date of the technology crash on March 10, 2000 (8(a), right column) and the date of the Lehman bankruptcy on 09/15/2008 (Fig. 8(b), right column). It is clearly seen from these plots that as the stock market becomes more volatile loops in the relevant point clouds become much more pronounced.
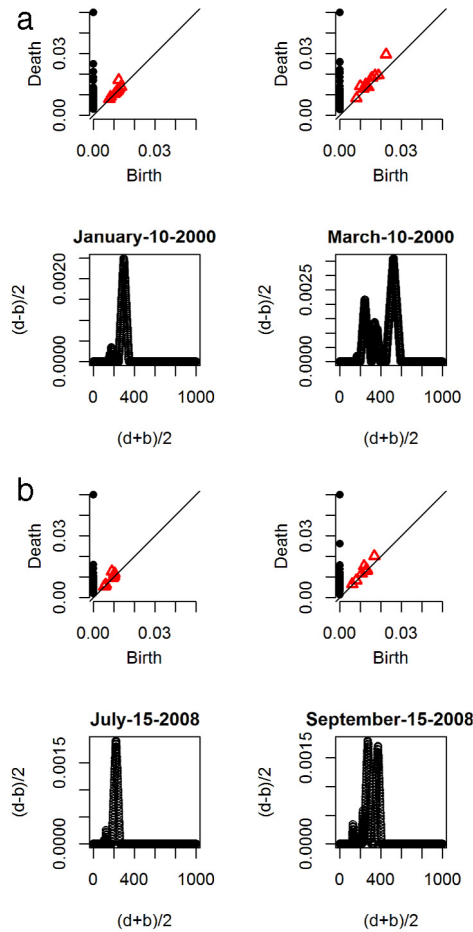
Fig. 8. The Rips persistence diagrams and the corresponding persistence landscapes calculated with the sliding window of 50 trading days on selected dates. The solid black dots represent connected components, red triangles represent loops. (a) Technology crash, 2000; (b) Financial Crisis, 2008. Color online.

To quantify this behavior, as a convenient summary of changes in topological features throughout filtration, we calculate the L^p norms ($p = 1, 2$) of the persistence landscapes λ for each rolling window. The set of these values forms the daily time series of these measures; see Fig. 9 for an example.

Fig. 10 shows the normalized time series of the L^1 norms of persistence landscapes calculated with the sliding window of $w = 50$ trading days around the date of the technology crash and the date of the Lehman bankruptcy. For comparison, we also show the corresponding segments of the normalized time series of S&P 500 and the volatility index VIX. The VIX measures the implied volatility in the S&P 500 index over the next 30 days and is often used for forecasting purposes. It is qualitatively clear from these plots that the time series of the L^1-norm demonstrates strong fore-shocks prior to the primary peak, which mounts during a crisis. The change of the size of the rolling window from 50 to 100 days as well as the temporal behavior of the L^2-norm closely resemble this picture (not shown here). The only notable differences in the outcome related to the change of the window size are an overall flattening of spikes in the time series of the L^p-norms and a slight shift of the primary maximum of these time series towards the later dates.

Having the time series of L^p norms of persistence landscapes in place, we apply the standard technique and analyze the leading statistical indicators of these time series. In that way, we further quantify temporal variations in the persistence of loops in the time-ordered set of point clouds with a growing market stress. We follow Guttal et al. [21], and employ the 500 days rolling window with the sliding step of one day to calculate the variance, the average spectral density at low frequencies, and the first lag of the auto-correlation function (ACF) for 250 trading days prior to the date of a market crash. This window should not be confused with the sliding window of $w = 50$ days that has been used for encapsulation of point clouds throughout the computation of the time series of L^p-norms of persistence landscapes. Fig. 11 shows the obtained results. It is clearly seen from the plots on this figure that the variance and the average spectral density at low frequencies of the time series of L^p-norms are substantially growing prior to any of the crashes. We quantify the observed trend with the Mann–Kendall test, which statistically assess a monotonic upward or downward movement of the variable over time [56,57].
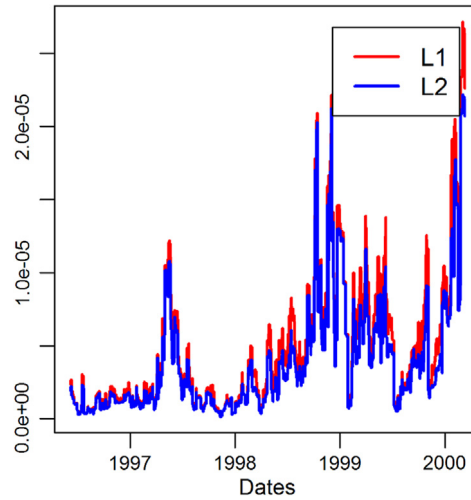
Fig. 9. The time series of normalized L^1 (blue line) and L^2 (red line) norms of persistence landscapes calculated with the sliding window of 50 days. Color online.
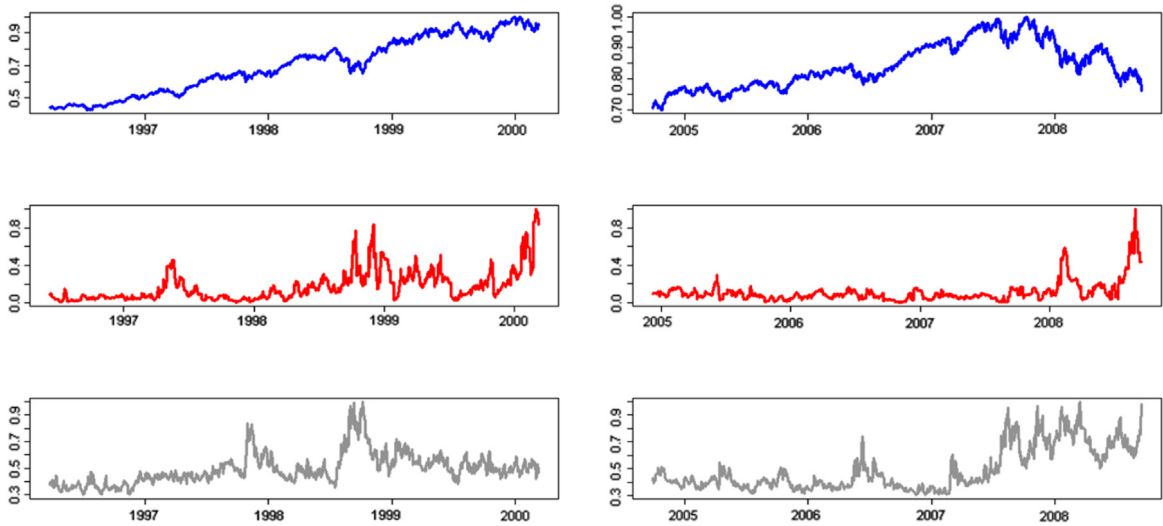

Fig. 10. 1000 trading-days prior to a crash. Top-down, the daily time series of the normalized S&P 500, the normalized L^1-norm of topological landscapes calculated with the sliding window of 50 days, and VIX. Left column: proximity of the 'dotcom crash' on 03/10/2000. Right column: proximity of the Lehman bankruptcy on 09/15/2008.

The strong trend of rising average spectral density at low frequencies for 250 trading days prior to either dotcom crash on 03/10/2000 or to the Lehman Bankruptcy on 09/15/2008 is especially pronounced: Kendall-tau rank correlation coefficients are 0.89 and 1.00, respectively. On the other hand, the ACF at lag-1 shows no trend prior to any of these events. We note that doubling the size of the point cloud to 100 data points does not significantly change this picture (not shown here). For comparison, we also present on Fig. 11 the temporal behavior of the same statistical indicators obtained with the same rolling window of 500 days from the daily time series of the volatility index. Interestingly, prior to the 2000 crash the time series of VIX shows a declining variability. On the other hand, similarly to the norms of persistence landscapes, the average spectral density at low frequencies of the time series of VIX demonstrates strong rising trend for 250 trading days prior to the Lehman Bankruptcy (Kendall-tau = 0.89).

5. Conclusions

We apply TDA for exploration of the temporal behavior of topological features in the well-known financial data as the state of the market is changing with time. The novelty of our approach to multi-source time series is twofold: (i) we use the
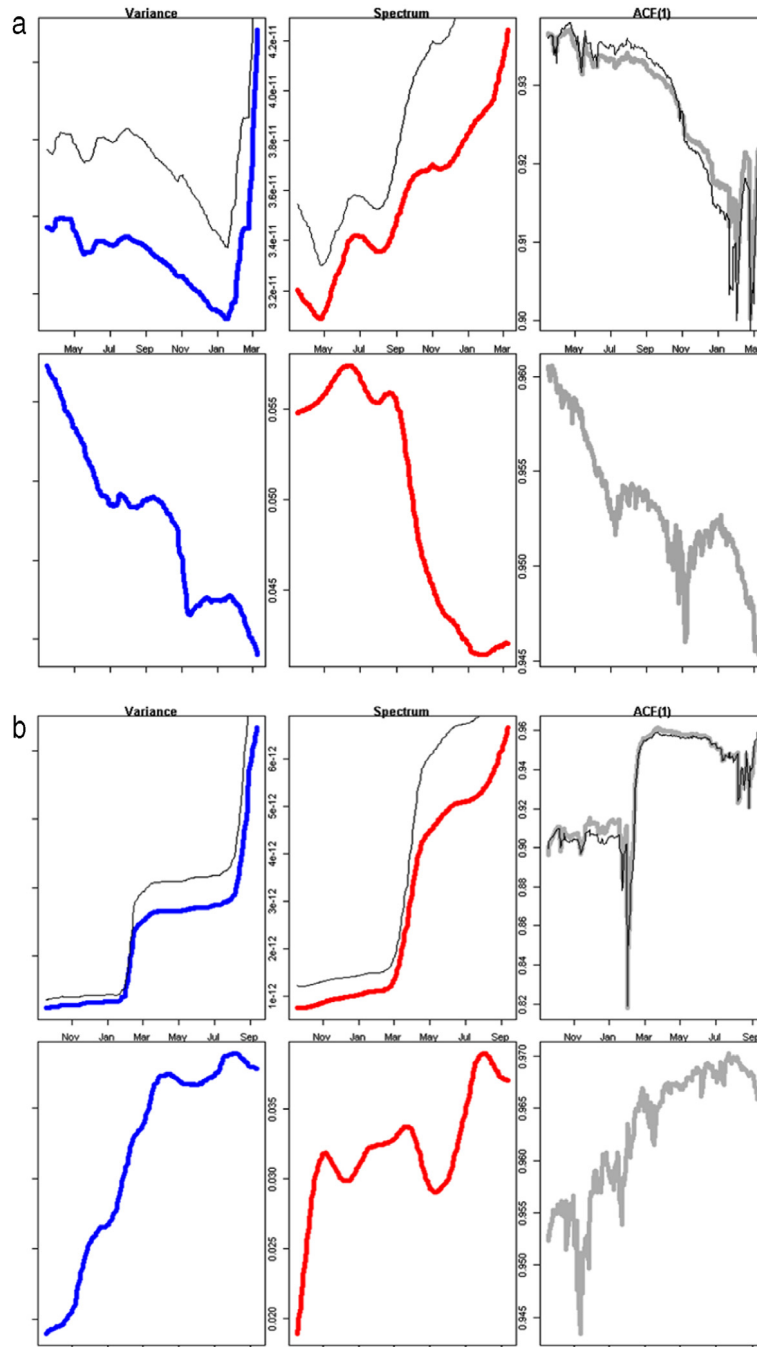
Fig. 11. The variance, the average spectral density at low frequencies, and the first lag of the ACF of the time series of L^p norms of persistence landscapes, and VIX for 250 trading-days prior to (a) the technology crash on 03/10/2000; (b) the Lehman bankruptcy on 09/15/2008 (see text for details). The thin black line corresponds to L^1.

single sliding window to enclose the point cloud in the d-dimensional space determined by the number of time series under consideration; (ii) we employ the L^p-norm of persistence landscapes as a quantifier of stability of topological features. We believe that this approach can be used beyond the analysis of financial time series presented here.

More specifically, we study persistence of loops (1D persistence homology) in a 4D point cloud formed by a *single* sliding window and *four* 1D time series of daily log-returns of the four stock-market indices. Thus, contrary to the traditional two-parametric time-delay embedding method, our approach has only one parameter: the size of the window, w. Our study

reveals that the *shape* of financial time series strongly depends on the state of the market. The empirical analysis shows that the time series of the L^p-norms exhibit strong growth around the primary peak emerging during a crisis. This behavior reflects an increased persistence of loops appearing in point clouds as the market undergoes transition from the ordinary to the "heated" state. Remarkably, the variability of the time series of the L^p-norms of persistence landscapes, quantified by the variance and the spectral density at low frequencies, demonstrates a strong rising trend for 250 trading days prior to either dotcom crash on 03/10/2000 or to the Lehman Bankruptcy on 09/15/2008. Our study suggests that TDA offers a novel econometric method, and is yielding the new category of early warning signals of an imminent market crash.

Acknowledgments

References

[1] G. Carlsson, Topology and data, Bull. Amer. Math. Soc. 46 (2009) 255.
[2] H. Edelsbrunner, J. Harer, Computational Topology: An Introduction, Amer. Math. Soc., 2010.
[3] H. Edelsbrunner, D. Letscher, A. Zomordian, Topological persistence and simplification, Descrete Comput. Geom. 28 (2002) 511.
[4] P.Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, G. Carlsson, Extracting insights from the shape of complex data using topology, Sci. Rep. 3 (2013) 1236.
[5] P. Bubenik, Statistical topological data analysis using persistence landscapes, J. Mach. Learn. Res. 16 (2015) 77.
[6] P. Bubenik, P. Dlotko, A persistence landscapes toolbox for topological statistics, J. Symbolic Comput. 78 (2016) (2016) 91.
[7] F. Chazal, B. Terese Fasy, F. Lecci, A. Rinaldo, L. Wasserman, Stochastic convergence of persistence landscapes and silhouettes, J. Comput. Geom. 6 (2015) 140–161.
[8] 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, Proc. Natl. Acad. Sci. 108 (2011) 7265.
[9] G. Carlsson, T. Ishkhanov, V. de Silva, A. Zomorodian, On the local behavior of spaces of natural images, Int. J. Comput. Vis. 76 (2008) 1.
[10] J. Berwald, M. Gidea, Critical transitions in a model of a genetic regulatory system, Math. Biol. Eng. 11 (2014) 723.
[11] J. Berwald, M. Gidea, M. Vejdemo-Johansson, Automatic recognition and tagging of topologically different regimes in dynamical systems, Discontinuity Nonlinear. Complex. 3 (2015) 413.
[12] J.A. Perea, J. Harer, Sliding windows and persistence: an application of topological methods to signal analysis, Found. Comput. Math. 15 (2015) 799.
[13] F. Khasawneh, E. Munch, Chatter detection in turning using persistent homology, Mech. Syst. Signal Process. 70–71 (2016) 527.
[14] P. Marco, M. Rucco, E. Merelli, Topological classifier for detecting the emergence of epilepsy seizures, 2016. arXiv:1611.04872 [q-bio.NC].
[15] L.M. Seversky, S. Davis, M. Berger, On time-series topological data analysis: new data and opportunities, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, p. 59.
[16] S. Emrani, T. Gentimis, H. Krim, Persistent homology of delay embedding and its application to wheeze detection, IEEE Signal Process. Lett. 21 (2014) 459.
[17] J.A. Perea, A. Deckard, S.B. Haase, J. Harer, SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data, BMC Bioinformatics 16 (2015) 257.
[18] S. Maletic, Y. Zhao, M. Rajkovic, Persistent topological features of dynamical systems, Chaos 26 (2016) 053105.
[19] M. Kramar, R. Levanger, J. Tithof, B. Suri, M. Xu, M. Paul, M.F. Schatz, M. Paul, M.F. Schatz, K. Mischaikow, Analysis of Kolmogorov flow and rayleigh-benard convection using persistent homology, Physica D 334 (2016) 82.
[20] M. Gidea, Topological data analysis of critical transitions in financial networks, in: E. Shmueli, et al. (Eds.), 3rd International Winter School and Conference on Network Science, in: Springer Proceedings in Complexity, 2017. http://dx.doi.org/10.1007/978-3-319-55471-6_5.
[21] V. Guttal, S. Raghavendra, N. Goel, Q. Hoarau, Lack of critical slowing down suggests that financial meltdowns are not critical transitions, yet rising variability could signal systemic risk, PLoS One 11 (2016) e0144198.
[22] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, Discrete Comput. Geom. 37 (2007) 103.
[23] Dionysus: a C++ library for computing persistent homology, 2012. http://www.mrzv.org/software/dionysus/.;
C. Maria, GUDHI, Simplicial Complexes and Persistent Homology Packages, 2014. https://project.inria.fr/gudhi/software/.;
B.T. Fasy, J. Kim, F. Lecci, C. Maria, Introduction to the R package TDA, 2014. arXiv:1411.1830..
[24] B.T. Fasy, J. Kim, F. Lecci, C. Maria, V. Rouvreau, Statistical tools for topological data analysis, 2016. https://cran.r-project.org/web/packages/TDA/TDA.pdf.
[25] F. Takens, Detecting Strange Attractors in Turbulence, in: Lecture Notes in Mathematics, vol. 898, Springer-Verlag, 1981.
[26] T. Sauer, J.A. Yorke, M. Casdagli, Embedology, J. Stat. Phys. 65 (1991) 579.
[27] H.D.I. Abarbanel, Analysis Of Observed Chaotic Data, Springer, 1996.
[28] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712.
[29] M.R. Muldoon, D.S. Broomhead, J.P. Huke, R. Hegger, Delay embedding in the presence of dynamical noise, Dyn. Stab. Syst. 13 (1998) 175.
[30] K. Mischaikow, M. Mrozek, J. Reiss, A. Szymczak, Construction of symbolic dynamics from experimental time series, Phys. Rev. Lett. 82 (1999) 1144.
[31] M. Hénon, A two-dimensional mapping with a strange attractor, Comm. Math. Phys. 50 (1976) 69.
[32] M. Ausloos, K. Ivanova, Dynamical model and nonextensive statistical mechanics of a market index on large time windows, Phys. Rev. E 68 (2003) 046122.
[33] J. Miśkiewicz, M. Ausloos, A logistic map approach to economic cycles. (I). The best adapted companies, Physica A 336 (2004) 206.
[34] D. Sornette, Critical Phenomena in Natural Sciences — Chaos, Fractals, Self-organization and Disorder: Concepts and Tools, second ed. in: Springer Series in Synergetics, Heidelberg, 2004.
[35] Y. Choi, R. Douady, Financial crisis dynamics: attempt to define a market instability indicator, Quant. Finance 12 (2012) 1351.
[36] B. Mandelbrot, The variation of certain speculative prices, J. Bus. 36 (1963) 394.
[37] E.F. Fama, The behavior of stock-market prices, J. Bus. 38 (1965) 34.
[38] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, 2000.
[39] J.P. Bouchaud, M. Potters, Theory of Financial Risks: From Statistical Physics to Risk Management, Cambridge University Press, Cambridge, 2000.

[40] S. Shephard, Stochastic Volatility: Selected Readings, Oxford University Press, USA, 2005.
[41] L. Borland, A theory of non-Gaussian option pricing, Quant. Finance 2 (2002) 415.
[42] C. Tsallis, C. Anteneodo, L. Borland, R. Osorio, Non-Extensive statistical mechanics and economics, Physica A 324 (2003) 89.
[43] L. Borland, J.-P. Bouchaud, A non-Gaussian option pricing model with skew, Quant. Finance 4 (2004) 499.
[44] T.S. Biro, R. Rosenfeld, Microscopic origin of non-Gaussian distributions of financial returns, Physica A 387 (2007) 1603.
[45] C. Tsallis, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, Springer, 2009.
[46] A. Gerig, J. Vicente, M.A. Fuentes, Model for non-Gaussian intraday stock returns, Phys. Rev. E 80 (2009) 065102(R).
[47] E. Van der Straeten, C. Beck, Superstatistical fluctuations in time series: Applications to share-price dynamics and turbulence, Phys. Rev. E 80 (2009) 036108.
[48] C. Vamos, M. Craciun, Separation of components from a scale mixture of Gaussian white noises, Phys. Rev. E 81 (2010) 051125.
[49] T. Takaishi, Analysis of realized volatility in superstatistics, Evol. Inst. Econ. Rev. 7 (2010) 89.
[50] S. Camargo, S.M. Duarte Queiros, C. Anteneodo, Nonparametric segmentation of nonstationary time series, Eur. Phys. J. B 86 (2013) 159.
[51] Y.A. Katz, L. Tian, q-Gaussian distributions of leverage returns, first stopping times, and default risk valuations, Physica A 392 (2013) 4989.
[52] Y.A. Katz, L. Tian, Superstatistical fluctuations in time series of leverage returns, Physica A 405 (2013) 326.
[53] C. Beck, Dynamical foundations of nonextensive statistical mechanics, Phys. Rev. Lett. 87 (2001) 180601;
     C. Beck, E.G.D. Cohen, Superstatistics, Physica A 322 (2003) 267.
[54] P.D. Praetz, The distribution of share price changes, J. Bus. 45 (1972) 49.
[55] S. Micciche, G. Bonanno, F. Lillo, R.N. Montegna, Physica A 314 (2002) 756.
[56] M.G. Kendall, Rank Correlation Methods, fourth ed. Charles Griffin, London, 1975.
[57] A.I. McLeod, Kendall rank correlation and Mann-Kendall trend test, 2011. https://cran.r-project.org/web/packages/Kendall/Kendall.pdf.