Physica A 491 (2018) 820–834

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

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

Article history:
Received 7 April 2017
Available online 9 October 2017

Keywords:
Topological data analysis
Financial time-series
Early warning signals

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.

* 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).

https://doi.org/10.1016/j.physa.2017.09.028
0378-4371/

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.