International Journal of Informatics and Information System ISSN 2579-7069
Vol. 4, No. 1, March 2021, pp. 63-70

With topological data analysis, predicting stock market crashes

Nugroho Agung Prabowo [1,*], R Arri Widyanto [2], Mukhtar Hanafi [3],
Andi Widiyanto [4], Bambang Pujiarto [5], Meidar Hadi Avizenna [6]

[1-6] Universitas Muhammadiyah Magelang, Indonesia
[1,*] naprabowo@ummgl.ac.id; [2] arri_w@ummgl.ac.id; [3] hanafi@ummgl.ac.id, [4] andi.widiyanto@ummgl.ac.id, [5]bpujiarto@ummgl.ac.id,
[6] meidar.hadi@ummgl.ac.id
* corresponding author

(Received February 11, 2021 Revised February 22, 2021 Accepted February 27, 2021, Available online March 1, 2021)

Abstract

We are investigating the evolution of four big US stock market indexes' regular returns after the 2000 technology crash and the 2007-2009 financial crisis. Our approach is based on topological data processing (TDA). To identify and measure topological phenomena occurring in multidimensional time series, we use persistence homology. We obtain time-dependent point cloud data sets using a sliding window, which we connect a topological space for. Our research indicates that a new method of econometric analysis is offered by TDA, which complements the traditional statistical tests. The tool may be used to predict early warning signs of market declines that are inevitable.

Keywords: TDA, Market Crashes, Stock Detection, Topology

1. Introduction

As long as the financial markets are in operation, there will be financial crises. When the market falls most of it loses, anyone who can forecast it can protect their investments or take aggressive short positions to make a profit (a nevertheless stressful situation to be in, as depicted in the Big-short). An asset in the market is associated with a competitive mechanism, the price of which varies according to the details available. A wide variety of information is used to assess the price of an asset on a stock exchange and, under the effective market theory, a simple adjustment in the information would be automatically priced in.

“The dynamics of financial systems are comparable to those of physical systems” Often as phase transformations between solids, liquids, and gases happen, we can differentiate a normal market regime from a chaotic one. Observations indicate that a cycle of intensified asset price oscillation precedes market crashes[1]. This anomaly transforms into an abnormal change in the time series' geometric structure. In this research, in order to create an accurate detector for stock market crashes, we use topological data analysis (TDA) to capture these geometric changes in the time series.

2. The Past Literature

Topological Data Analysis (TDA) is a new field that arose during the first decade of the century from separate works in applied(algebraic) topology and computational geometry. Although geometric methods for data analysis can be traced back very far in the past, it began as a discipline of persistent homology with the groundbreaking works of Edelsbrunner et al [1]. and Zomorodianand Carlsson [2] was popularized in a seminal paper in 2009 [3]. Tda is primarily inspired by the concept that topology and geometry offer an effective approach to infer strong qualitative and often quantitative data structure knowledge from Chazalal [4].

Tda seeks to include well-founded mathematical, statistical and algorithmic techniques to infer, evaluate and manipulate the raw data of complex topological and geometric systems that are frequently described in Euclidean or more common metric spaces as point cloud data. A major effort has been made over the last few years to provide tda

with stable and effective data structures and algorithms that are now integrated and usable and easy to use via standard libraries such as the Gudhi library (C++ and Python) Maria et al [5] and its R program interface Fasy et al [6].

2.1. TDA Pipeline

Tda has recently acknowledged advances in different directions and fields of implementation. A wide range of methods driven by topological and geometric approaches are now available. It is beyond the reach of this introductory survey to provide a full review of all these current methods. Most of them, however, depend on the following fundamental and normal pipeline that will act as the foundation of this paper:

1. A finite set of points with a notion of distance or resemblance between them is considered to be the input. The metric in the ambient space (e.g. the Euclidean metric when the data is embedded in R^d) will induce this distance or appear as an inherent metric described by a distance matrix pairwise. As an input or directed by the program, the description of the metric on the data is generally given. However, it is important to note that it could be critical to use a metric to uncover fascinating topological and geometric data characteristics.

2. In order to illustrate the underlying topology or geometry, a "continuous" form is constructed on top of the results. This is also a simplicial complex or a nested family of simplicial complexes, referred to as a filtration, which represents the data structure at various scales. Simplicial complexes can be used in many standard data processing or learning algorithms as higher dimensional generalizations of adjacent graphs that are classically constructed on top of data. The task here is to identify constructs that have been shown to represent important knowledge about the data structure and that can be built and easily manipulated successfully in actual practice.

3. From the structure constructed on top of the results, topological or geometric knowledge is extracted. This can either result in a complete reconstruction, usually a triangulation, of the shape of the data from which topological/geometric features could become effectively extracted, or in crude summaries or estimations from which complex methods, such as persistent homology, are needed for the extraction of relevant details. In addition to identifying and visualizing and interpreting interesting topological/geometric details, the difficulty at this stage is to demonstrate its importance, particularly its stability with regard to disturbances or the presence of noise in the input data.

4. The topological and geometric knowledge collected offers new groups of data features and descriptors. They can be used, in particular by visualization, to further interpret the data or can be paired with other forms of features for more study and machine learning activities. A significant problem at this point is to demonstrate the added-value and the complementarity (with regard to other features) of the details given by the tda tools.

2.2. TDA on Data Science

On the application side, the role of topological and geometric methods in a growing number of fields has been seen by several recent promising and successful findings, such as material sciences [7-8], 3D shape analysis [9-10], multivariate time series analysis [11], biology [12], chemistry [13], network sensors [14]. An comprehensive list of applications for tda is outside the scope. On the other hand, much of TDA's contributions are attributed to its conjunction with other types of study or learning.

3. Method

In this Study, we evaluate daily S&P 500 index prices from 1980 to the present. The S&P is an index that is widely used to assess the state of the financial market as it calculates the 500 large-cap US companies' stock performance. We find that topological signals appear to be resilient to noise and therefore less likely to generate false positives compared to a simple baseline. This study is one among several main reasons behind TDA, namely that topology and geometry in complex data can provide a powerful way to abstract subtle structure.

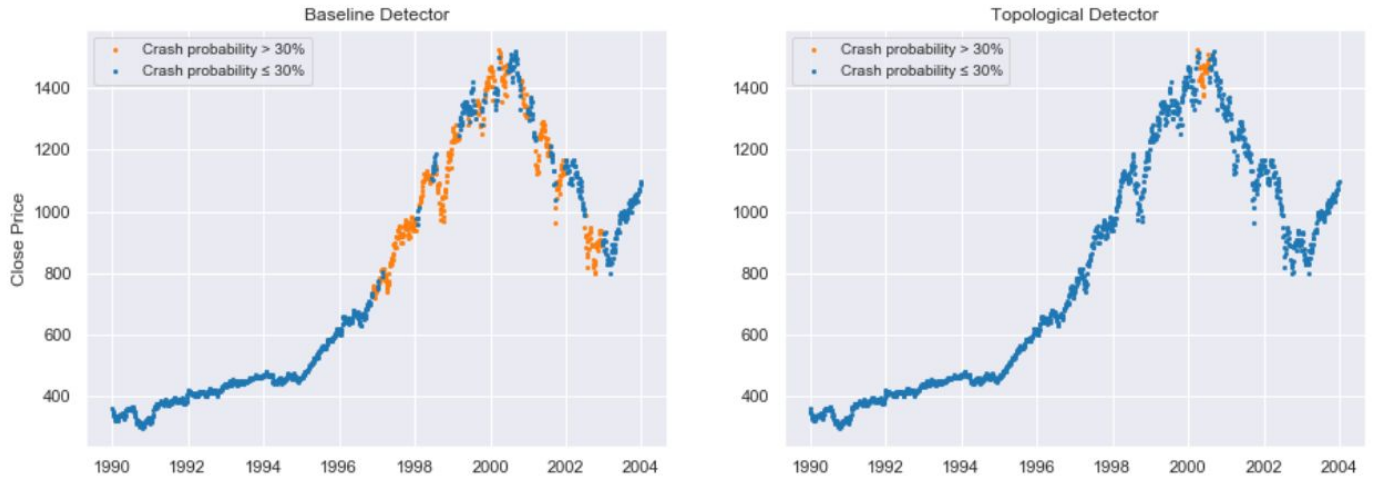

Fig. 1. Baseline and topological model detection of stock market crashes.

3.1. Baseline Model

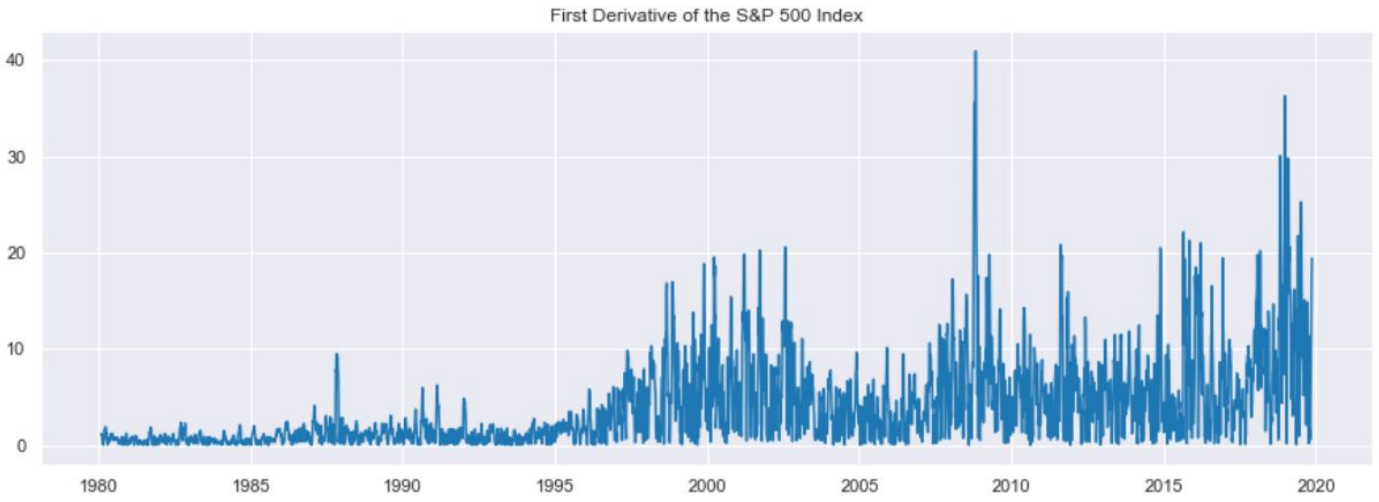

Fig. 2. Magnitude between the first median near price derivative among successive windows.

Considering that market crashes involve a sudden fall on stock prices, monitoring the first derivative of average price values over a rolling window is one easy approach to detecting these changes. Indeed, we can see in the figure above that the Black Monday crash (1987), the burst of the dot-com bubble (2000-2004), and the financial crisis (2007-2008) are already captured in this naïve approach. We may add a threshold to mark points on our original time series where a crash occurred by normalising this time series to take values in the [0,1] interval. Following this guideline could lead to over-panicking and selling our assets too fast, with several points labeled as a crash.

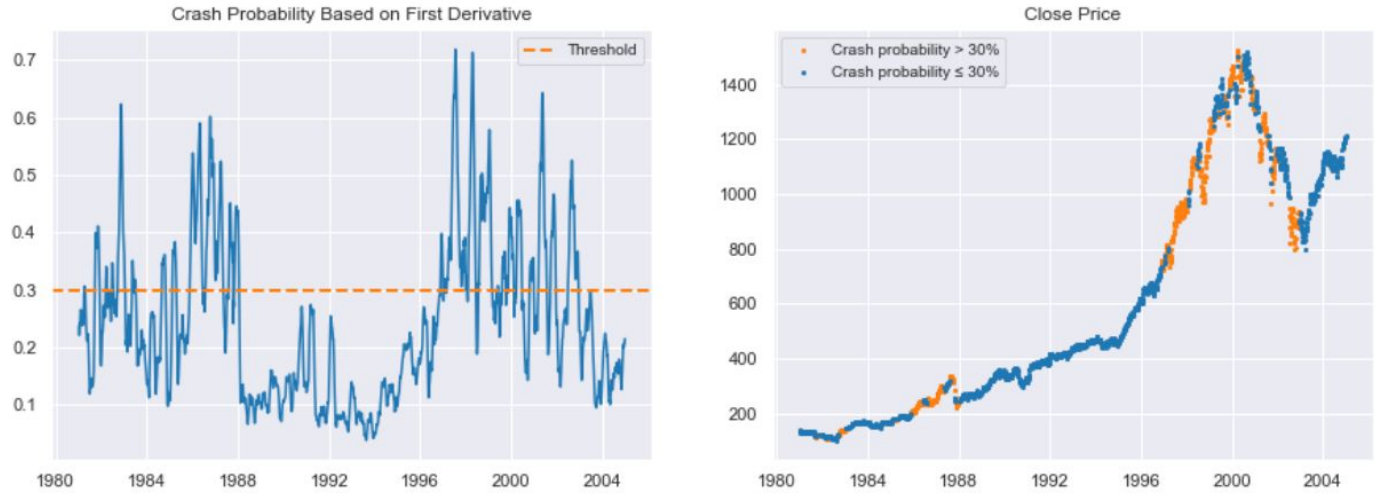

Fig. 3. Baseline model crash probability (left), with points above the threshold shown in the original time series (right).

3.2. Topological Model

The underlying TDA mathematics is detailed and will not be discussed in this article. It is enough for our purposes to think of TDA as a way of extracting descriptive characteristics that can be used for downstream modeling. The pipeline we have built is constructed from:

- Embedding of the time series into a point cloud and construction of point cloud sliding windows
- To create a filtration on each window to provide a developing structure encoding each window's geometrical shape
- Using persistence homology, extracting the related features of those windows
- By measuring the difference between these features from one window to the next, comparing each window
- Constructing an indicator of crash based on this difference

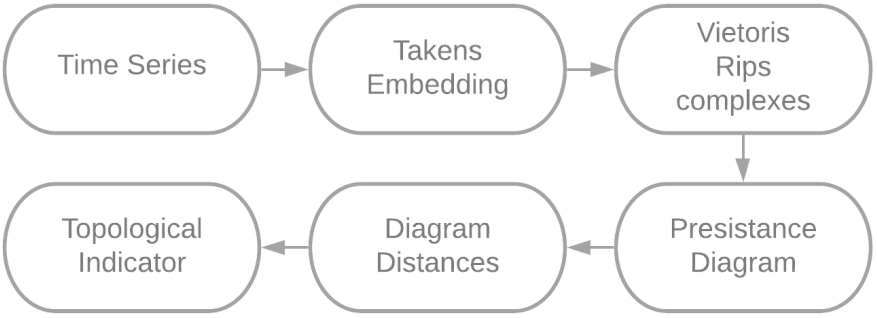

Fig. 4. TDA Pipeline

3.3. Time series as point clouds

In a TDA pipeline, a typical starting point is to generate a simplified complex from a point cloud. In time series applications, therefore, the crucial question is how to generate such point clouds? Typically, discrete time series, like those we are considering, are visualized in two dimensions as scatter plots. By scanning the plot from left to right, this representation makes the local behavior of the time series easy to track. But it is often ineffective in transmitting significant effects that may occur over larger timeframes.

Fourier analysis provides one well-known set of methods for capturing periodic behaviour. For example, the discrete Fourier transformation over the time series of a temporal window provides information on whether the signal in that

window occurs as the total amount of several basic periodic signals. We see a different way of encoding a time-evolving process for our purposes. It is based on the idea that some of the dynamics' key properties can be effectively unveiled in higher dimensions. We begin by illustrating a way of representing as a point cloud a univariate time series, i.e. a set of vectors in an arbitrary dimensional Euclidean space.

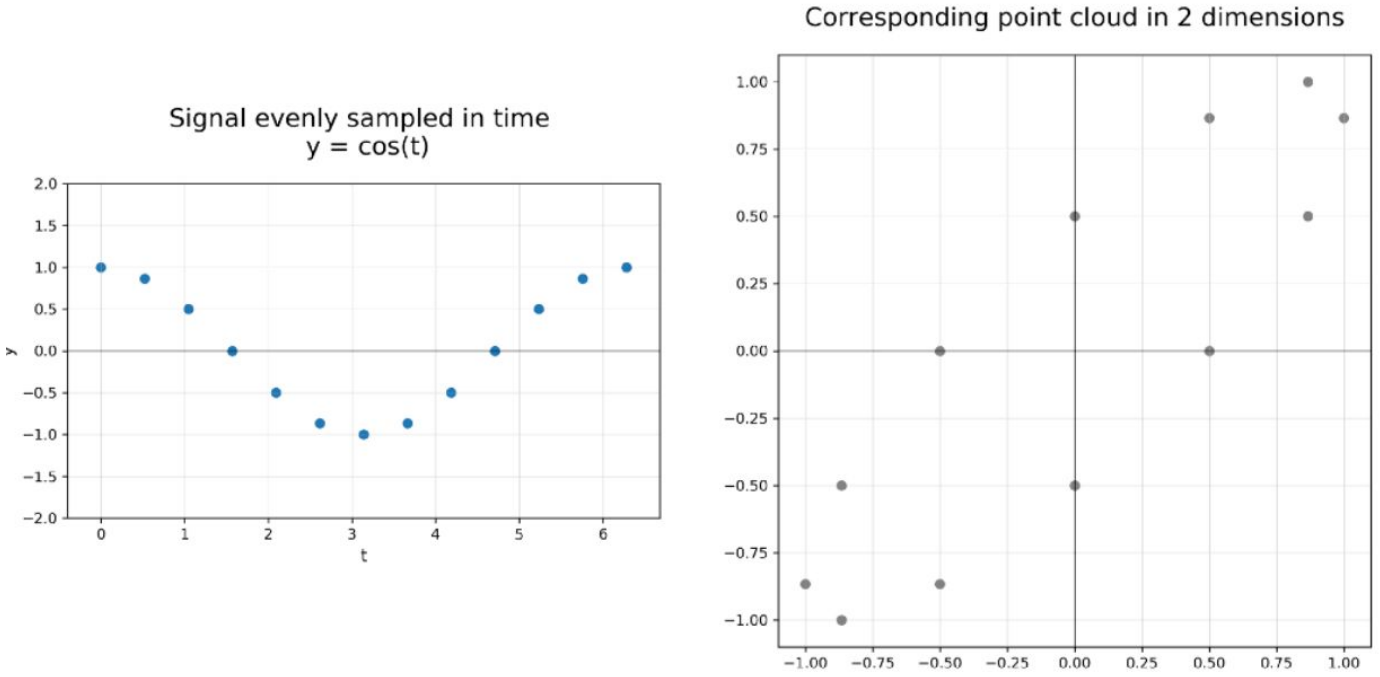

Fig. 5. Embedding with embedding dimension d=2 and time delay τ=1

The procedure works as follows: we pick two integers d and τ. For each time $t_i \in (t_0, t_1, \ldots)$, At different times, we collect the values of the variable y at d, evenly spaced by τ and starting at t_i, and describe them as a vector with d entries, notably:

$$Y_{t_i} = (y_{t_i}, y_{t_i+\tau}, \ldots, y_{t_i+(d-1)\tau}).$$

A set of vectors in d-dimensional space is the result, The time delay parameter is called τ, and the embedding dimension is called d. After Floris Takens, which demonstrated its significance with a celebrated theorem in the context of nonlinear dynamical systems, this time-delay embedding technique is also called Takens' embedding. Finally, it leads to a time series of point clouds (one per sliding window) with potentially interesting topologies to apply this procedure separately on sliding windows over the full time series. How such a point cloud is generated in 2 dimensions is shown in the figure 5 above.

3.4. point clouds to persistence diagrams

What can we do with this data now that we know how to generate a time series of point clouds, Enter persistent homology, which looks for topological features that persist over a certain range of parameter values in a simplified complex. Typically, a feature, such as a hole, will not be observed at first, then it will appear, and the parameter will disappear again after a range of values.

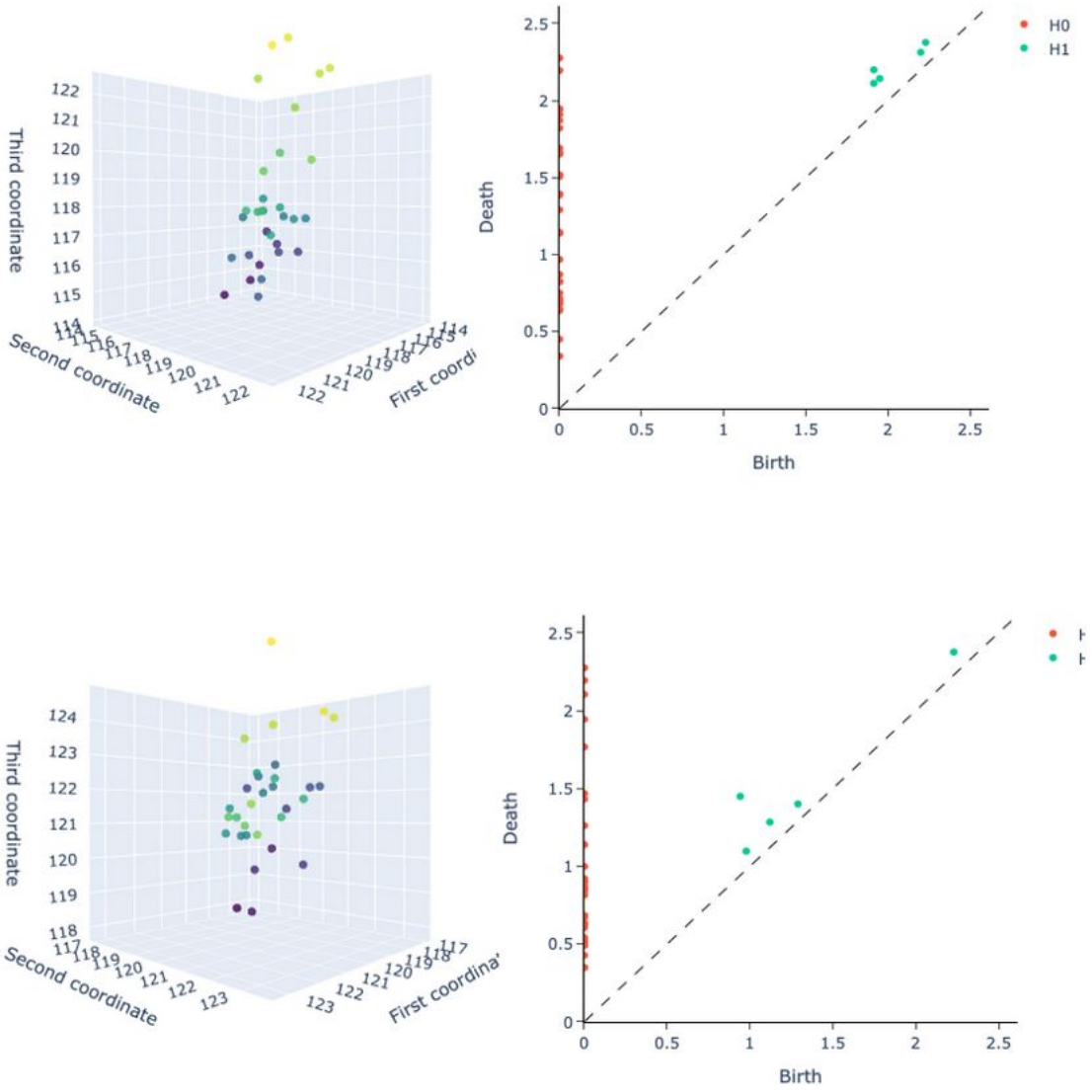

Fig. 6. Point clouds and their associated persistence diagram from two successive windows

3.5. Distances between persistent diagrams

We can measure a set of distance metrics, given two windows and their corresponding persistence diagrams. We compare two distances here, one based on the notion of a landscape of persistence, the other on Betti curves. We can infer from these figures that the metric is less noisy than the Betti curves, based on landscape distance.

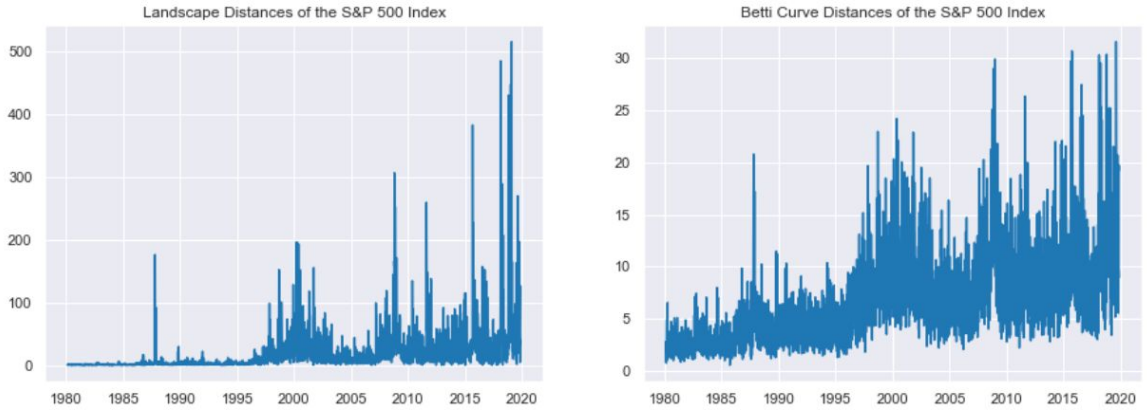

Fig. 7. Magnitude between successive windows of the landscape and Betti curve distances.

4. Results

It is indeed a good method to normalize it, as we did for the baseline model, using the landscape distance between windows as our topological feature. The subsequent detection of stock market declines due to the dot-com bubble and global financial crisis as seen below. We can see that using topological features tends to reduce the noise in the signal of interest relative to our basic baseline.

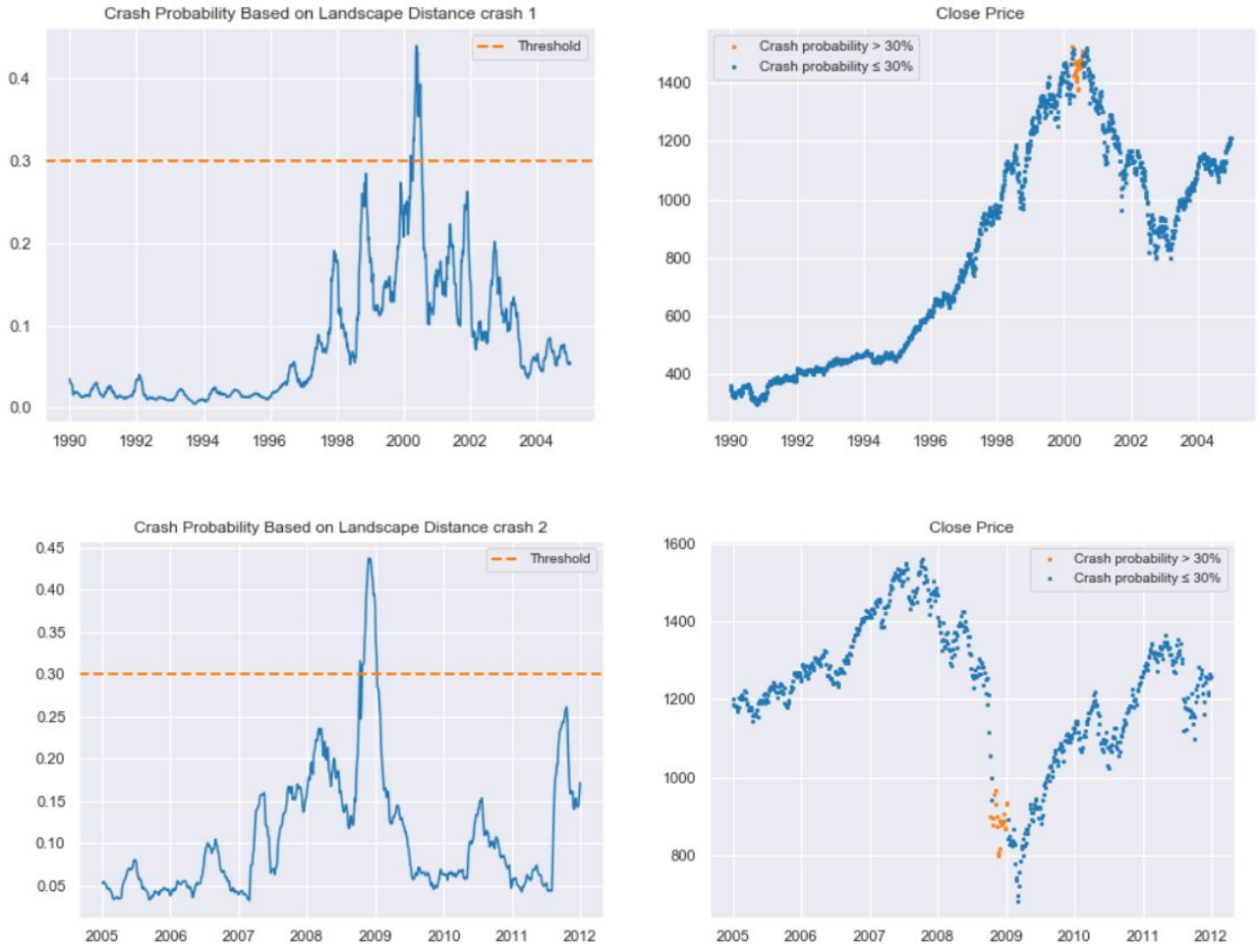

Fig. 8. Using topological features, crash probabilities and detections. The time periods correspond to the 2000 (upper) dot-com bubble and the 2008 global financial crisis (lower).

5. Conclusion

Our results suggest that geometric signatures that can be more robustly identified using topological data analysis are produced by the periods of high volatility preceding a crash. These results, however, affect only a single industry and for a brief period of time, so the robustness of the process in various markets and differing thresholds should be further studied. However, the findings are promising and open up some exciting ideas for future growth.

References

[1] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," *Discret. Comput. Geom.*, vol. 28, no. 4, pp. 511–533, 2002, doi: 10.1007/s00454-002-2885-2.

[2] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Proc. Annu. Symp. Comput. Geom.*, vol. 274, pp. 347–356, 2004, doi: 10.1145/997817.997870.

[3] G. Carlsson, *Topology and data*, vol. 46, no. 2. 2009.

[4] F. Chazal, H. T. Data, and A. Handbook, "High-Dimensional Topological Data Analysis Frédéric Chazal To cite this version : HAL Id : hal-01316989 HIGH-DIMENSIONAL TOPOLOGICAL DATA ANALYSIS," 2016.

[5] C. Maria, J. D. Boissonnat, M. Glisse, and M. Yvinec, "The Gudhi library: Simplicial complexes and persistent homology," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8592 LNCS, pp. 167–174, 2014, doi: 10.1007/978-3-662-44199-2_28.

[6] B. T. Fasy, J. Kim, F. Lecci, and C. Maria, "Introduction to the R package TDA," no. January 2015, pp. 1–16, 2014, [Online]. Available: http://arxiv.org/abs/1411.1830.

[7] M. Kramar, A. Goullet, L. Kondic, and K. Mischaikow, "Persistence of force networks in compressed granular media," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 87, no. 4, pp. 1–8, 2013, doi: 10.1103/PhysRevE.87.042207.

[8] T. Nakamura, Y. Hiraoka, A. Hirata, E. G. Escolar, and Y. Nishiura, "Persistent homology and many-body atomic structure for medium-range order in the glass," *Nanotechnology*, vol. 26, no. 30, pp. 1–22, 2015, doi: 10.1088/0957-4484/26/30/304001.

[9] P. Skraba, S. Univerity, S. Ca, and L. Guibas, "Persistence-based Segmentation of Deformable Shapes Fr ´," *Work*, no. 5, 2006.

[10] K. Turner, S. Mukherjee, and D. M. Boyer, "Persistent homology transform for modeling shapes and surfaces," *Inf. Inference*, vol. 3, no. 4, pp. 310–344, 2014, doi: 10.1093/imaiai/iau011.

[11] L. M. Seversky, S. Davis, and M. Berger, "On Time-Series Topological Data Analysis: New Data and Opportunities," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 1014–1022, 2016, doi: 10.1109/CVPRW.2016.131.

[12] Y. Yao *et al.*, "Topological methods for exploring low-density states in biomolecular folding pathways," *J. Chem. Phys.*, vol. 130, no. 14, pp. 1–10, 2009, doi: 10.1063/1.3103496.

[13] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit, "Quantifying similarity of pore-geometry in nanoporous materials," *Nat. Commun.*, vol. 8, no. May, 2017, doi: 10.1038/ncomms15396.

[14] V. De Silva and R. Ghrist, "Homological Sensor Networks Sensors and Sense-ability."