A Topological Data Analysis of Stock Market Time Series Using Persistent Homology

Yanbing Hou[1], Shijin Xu[2], Zheng Yang[3(✉)], Xiangrong Jia[3], Yiming Zhang[4], and Kuofu Liu[5]

[1] Business School, Sichuan University, Chengdu 610065, People's Republic of China
[2] Management School, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China
[3] Pittsburgh Institute, Sichuan University, Chengdu 610225, People's Republic of China
zhengyang2018@scu.edu.cn
[4] Mechanical and Electrical Engineering School, Harbin Institute of Technology, Harbin 150001, People's Republic of China
[5] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, USA

Abstract. Topological Data Analysis (TDA) has gained considerable attention as a robust tool for unraveling intricate data across diverse application domains. Central to TDA is the utilization of persistent homology, a method aimed at extracting and analyzing low-dimensional topological features from high-dimensional datasets. Recently, TDA has found novel application in financial research, where the viability of its application to stock market data analysis is underscored through quantitative analysis. The analysis of stock market crashes constitutes a pivotal topic within contemporary stock market research. However, due to the volatility of stock data and the interference of noise, the detection and prediction of such crashes often pose challenges. This paper endeavors to employ the persistent homology methodology to distill topological features from the data of China's Shanghai Stock Exchange 50 (SSE 50) stocks, including persistent diagram, persistent landscape, L^2-norm and persistent entropy. By merging statistical and machine learning methodologies, rigorous quantitative analysis is conducted. Empirical findings consistently demonstrate a high classification accuracy in distinguishing stock market crashes from normal periods. This substantiates its potential to offer crucial insights for financial market risk management and investor guidance, consequently contributing to the mitigation of investment risks.

Keywords: Persistent homology · Topological data analysis · Stock cash analysis · Topological features

J. Xu et al. (Eds.): ICMSEM 2025, LNDECT 264, pp. 247–258, 2025.
https://doi.org/10.1007/978-981-95-1595-0_17

1 Introduction

In the field of data analysis, TDA harnesses the intrinsic power of topology to decipher intricate patterns within complex datasets. Its versatile applications have garnered widespread attention across various domains. At its core lies the concept of persistent homology, a method aimed at extracting and analyzing low-dimensional topological features from high-dimensional datasets. This approach offers novel insights and methods for exploring the topology and shapes of data [9].

In the realm of financial research, persistent homology has forged a new path, particularly within the intricate landscape of stock market data analysis. The dynamic and multifaceted nature of financial markets provides a captivating arena for its application. The analysis of stock market crashes constitutes a pivotal topic within contemporary stock market research. However, as elucidated by Deboeck and Yaser [2,13], stock market data inherently possess non-stationarity, noise, and chaos. This temporal variation challenges the detection and anticipation of stock market crashes [14]. The allure of persistent homology lies in its general use for extracting low-dimensional topological information from high-dimensional data without information loss, maintaining stability even in the presence of noise and missing samples [3]. However, a distinct gap exists - existing studies in this domain often lack the necessary quantitative analysis to substantiate the efficacy of TDA analyses. This gap, more indispensable in today's business practices than ever before, underscores the need for rigorous quantitative validation.

Gidea [5,6] conducted exploratory observations on the differences between two stock market crash periods in the U.S. by analyzing the topological features. Yet, his approach relied on subjective judgments from the obtained persistence landscape (PL), which lacked distinct and objective differentiation between pre-crash and normal periods. Goel et al. [7] (2020) compared the TDA paradigm obtained in a continuous landscape with the standard deviation, analyzed the difference between the two, and used the TDA paradigm to build optimization equations for investment decisions.

However, differences exist in market indices and volatility between Chinese and American stock data[8] This paper explores the relationship between topological features and Chinese stock data. Based on the work of Fang Xuansu [4], who used persistent diagrams (PD) and statistical analyses to differentiate crash and normal periods, we extend his analysis by extracting persistence landscapes and persistent entropy from stock data, and employing statistical exploration and Support Vector Machine (SVM) classification.

Against this backdrop, this study embarks on a quantitative exploration of the relationship between relevant topological features and the analysis of Chinese stock data. Our focus lies in the quantitative analysis of stock market crashes on China's Shanghai Stock Exchange 50 (SSE 50) data using the extracted persistence landscapes and persistent entropy, coupled with statistical and machine learning methodologies. Empirical results demonstrate our model is effective in the sense that it can distinguish stock market crashes from normal periods, consistently exhibiting high classification accuracy, which has significant implications for future stock market regulation and investor guidance.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical background and methodologies employed in this study. Section 3 describes the data collection, preprocessing, and feature extraction processes in detail. In Sect. 4, we present the experimental results and perform a comprehensive analysis of the classification performance. Finally, Sect. 5 concludes the paper and outlines directions for future research.

2 Methodologies

2.1 Persistent Homology

Persistent homology [3] is a powerful tool for extracting and understanding geometric and topological features from complex data sets. Firstly, the data is transformed into a structure called simple complex. The data points are regarded as vertices, and the distance measurement between data points is regarded as edges. Based on the establishment of simple complex, the Filtering is further constructed. Filtering is a series of gradually increasing simplified complexes, starting from empty complexes to complete simple complexes. It reflects the process of gradually increasing the relationship of data points according to a certain measure. At each filtering level, we can calculate the corresponding homology group, which can capture the topological characteristics of the data set at that level. With the improvement of the filtering level, some topological features will disappear and some new features will appear. This process of emergence and disappearance constitutes a persistent homology.

In order to visualize the evolution of the persistent homology, we construct Persistent Diagram (PD). In the PD, each point represents a feature, whose horizontal coordinate is the filtering level at which the feature appears, and the vertical coordinate is the filtering level at which the feature disappears. This representation provides an intuitive way to understand the topology of data.

2.2 Persistent Landscape

Because the PD could not directly quantify changes in data, Bubenik [1] introduced persistent landscape (PL) and it L^p norm and applied them to complex shapes to quantify the changing shape of point cloud data. The persistent landscape is constructed by a series of persistent real-valued piecewise linear functions defined in the coordinates of the PD. The PL can still maintain the robustness of the graph under data perturbation.

Each point (a_i, b_i) in the PD corresponds to a piecewise function as follows [12],

$$f_{(a_i,b_i)}(x) = \begin{cases} x - a_i & a_i < x \leq \dfrac{a_i + b_i}{2} \\ -x + b_i & \dfrac{a_i + b_i}{2} < x \leq b_i \\ 0 & 0 \leq x \leq a_i \text{ or } x > b_i \end{cases} \tag{1}$$

Then define

$$\lambda_k(x) = k\max\left\{f_{(a_i,b_i)}(x) : i \in I\right\}. \tag{2}$$

where $k\max$ denotes the k most prominent element of $k \in \mathbb{N}^+$. λ_k called the kth persistent landscape function.

A persistent landscape is a sequence of functions $\lambda_1, \lambda_2, \ldots : \mathbb{R} \longrightarrow \mathbb{R}$. The critical point of λ_k is the t value when the slope changes. For two persistent landscapes $\lambda = (\lambda_k(x))_{k\in\mathbb{N}^+}$ and $\eta = (\eta_k(x))_{k\in\mathbb{N}^+}$, for any $x \in \mathbb{R}$ and $k \in \mathbb{N}^+$, define

$$(\lambda + \eta)_k(x) = \lambda_k(x) + \eta_k(x) \quad \text{and} \quad (c \cdot \lambda)_k(x) = c \cdot \lambda_k(x). \tag{3}$$

The L^p norm of the persistent landscape $\lambda = (\lambda_k)_{k\in\mathbb{N}^+}$ is defined as follows,

$$\|\lambda\|_p = \left(\sum_{k=1}^{\infty} \|\lambda_k\|_p^p\right)^{1/p}, \quad 1 \le p < \infty \tag{4}$$

where $\|\lambda_k\|_p$ denotes the L^p norm of λ_k. The space consisting of persistent landscapes with L^p norms constitutes the Banach space $L^p(\mathbb{N}, \mathbb{R})$.

2.3 Persistent Entropy

Persistent Entropy [11] characterizes the degree of randomness at various points in a PD. The higher the entropy, the more uncertain the values of the variables are, and vice versa.

Denote the PD from now on we set a given PD as:

$$D = \{(b_i, d_i)\}, \tag{5}$$

where b_i denotes the birth time of the ith feature and d_i the death time of the ith feature. Then the persistent entropy can be expressed as:

$$E(D) = -\sum_{i\in I} p_i \log(p_i), \tag{6}$$

where,

$$p_i = \frac{(d_i - b_i)}{L_D}, \tag{7}$$

$$L_D = \sum_{i\in I}(d_i - b_i). \tag{8}$$

3 Feature Extraction by Persistent Homology

3.1 Data Preprocessing

The stock market crash refers to the sudden collapse of stock prices due to some accidental factors when the internal contradictions of the stock market accumulate to a certain extent, which causes great social and economic turmoil and huge losses. Four primary stock market crashes in China occurred in 1992, 1993, 2007, and 2015 [14]. Since the first two crashes happened too long ago and the Chinese stock market has not been perfected yet, this paper selects the stock data during the two crashes in 2007 and 2015 for analysis, and the specific data are the daily closing prices of all the stocks in the SSE 50 in the corresponding periods.

After collecting the original data from the **CSMAR**, it is imperative to clean the data meticulously. First, duplicate stock code symbols are removed, keeping only one instance. Next, stock codes that have been flagged by the database as erroneous or disruptive are excluded from the dataset. Additionally, any missing values (NaN) in the time series are replaced with 0. It is important to note that the presence of numerous NaN values is not due to data errors but rather because some stocks within the SSE 50 were not yet listed during the corresponding period. In such cases, the 0 value is used as a placeholder to indicate the absence of trading activity rather than an actual zero price. This approach is justified since these 0 values accurately reflect the non-trading status of the stock at that time and ensure that the overall characteristics of market volatility are not distorted. If a time series comprises more than 50% zeros, the corresponding stock symbol is excluded from further analysis. In order to use the daily closing prices of stocks in the normal period as a comparison, this paper collects two sets of data in the normal period in the same way. So, we collect the daily closing price data of SSE 50, as listed in Table 1, from China Stock Market & Accounting Research Database using Python pandas.

Table 1. Summary of data on stock market crashes and normal periods

phase	timing	duration
Stock market crash 1	November 01, 2007–October 30, 2008	364 days
Normal 1	November 01, 2008–October 30, 2009	363 days
Stock market crash 2	June 01, 2015–January 30, 2016	243 days
Normal 2	January 01, 2015–May 30, 2015	149 days

After collecting the original data, it's important to clean it. Firstly, we eliminate duplicate stock code symbols, retaining only one instance. Next, we identify and remove stock code symbols causing errors or disruptions in the dataset as flagged by China Stock Market & Accounting Research Database. Additionally, any "NaN" entries are replaced with "0". If a time series comprises over 50% "0" values, the corresponding stock symbol is excluded.

The preprocessing steps in this paper are mainly divided into the following steps:

First, we standardized the daily prices to mitigate situations where smaller data with relatively lower absolute values could be overshadowed by larger data due to the intrinsic characteristics of the features when arranged together [14]. We use $S_i(t)$ to denote the closing price of a stock i in the t period.

$$g_i(t) = \frac{s_i(t) - \overline{s_i(t)}}{\sqrt{\left(s_i(t) - \overline{s_i(t)}\right)^2}} \tag{9}$$

Second, In this paper, the data is smoothed using a moving time window to select the appropriate sliding event window. Based on Yen, we use a sliding window of 60 days for data smoothing to ensure the quality of subsequent data analysis.

Thirdly, we use sliding windows to divide each period into multiple windows and calculates each window's corresponding correlation matrix and distance matrix. The Pearson correlation coefficient can reflect the connection or difference between two stocks to analyze the general market. As a result, the matrix of correlation coefficients for the logarithmic returns of stocks can be calculated using the algorithm of Pearson's correlation coefficient C, and the elements of the matrix are the stocks i and the correlation coefficients in the time interval T, denoted as.

$$C_{ij} = \rho_{ij} = \frac{\mathrm{Cov}[g_i, g_j]}{\sqrt{\mathrm{Var}[g_i]\mathrm{Var}[g_j]}} \tag{9}$$

Lastly, We converted the correlation matrix to a distance matrix using the formula

$$d_{ij} = \sqrt{2 * (1 - \rho_{ij})} \tag{10}$$

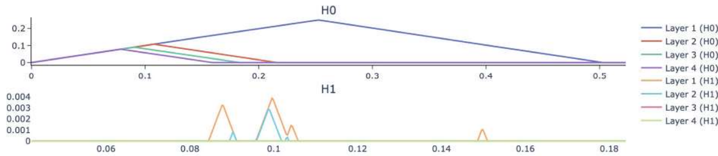
Then, we can use the distance matrices as input data for subsequent TDA calculations.

3.2 The Construction of Persistent Homology

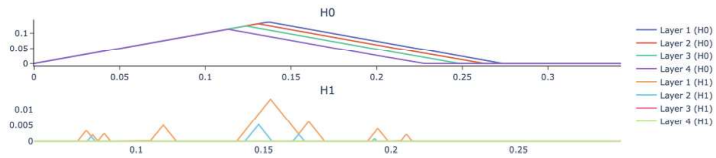
(1) Persistent Landscapes (PL)

We take the distance matrices as input to generate the PL. The horizontal coordinate of the persistence landscape image is the average parameter value for which the feature exists. The vertical coordinate is the half-life of the feature. The results are shown in Fig. 1.

Figure 1(a) shows the persistence landscape of the daily closing price of the SSE 50 stocks in the 0-dimensional, 1-dimensional, and 2-dimensional cases during the crash period, and Fig. 1(b) shows the persistence landscape of the daily closing price of the SSE 50 stocks in the persistent landscapes of daily closing prices of SSE 50 stocks in 0-dimensional, 1-dimensional, and 2-dimensional cases during normal periods. The horizontal axis represents the average parameter value at which a topological feature exists, while the vertical axis indicates the half-life of the corresponding feature. It can be seen that there is a significant difference between the two periods. For 0-dimensional and 1-dimensional space, there are obvious differences between the two. It can be seen from the figure that during the stock market crash, the horizontal and vertical coordinates at the highest point of the zero-dimensional continuous landscape image are



(a) Crash period



(b) Normal period

Fig. 1. The persistence landscape of the daily closing price of the SSE 50 stocks

lower, and the one-dimensional continuous landscape image has more fluctuations, which indicates that with the fluctuation of the stock market, the distribution of related point clouds is more intensive, and the holes in the point cloud become more obvious. For both 0- and 1-dimensional space, there are more obvious differences between the two.

(2) L^2 - Norm

We first use the method of multiple analysis to analyze the difference between the L^2 - Norm and whether the crash occurs. The analysis results are as follows (Tables 2 and 3).

Table 2. Comparison of L^2 - norm for 0-dimensional persistence landscapes in stock market crashes and normal times

	Phase	T-statistic	P-value	Is there a significant difference
1	Catastrophe 1 and Normal 1	−33.0261	0	significant difference
2	Crash 2 and Normal 2	−14.302	0.0049	significant difference
3	Normal 1 and normal 2	−5.231	0.08	No significant difference

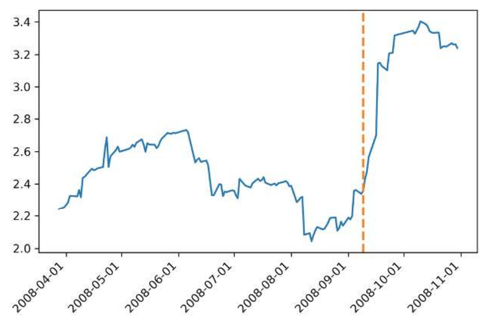
The results show that there is no significant difference in the number of L^2 norms of the 1-dimensional persistence landscape in the crash and normal periods. At the same time, there is a significant difference in the number of L^2 norms of the 0-dimensional persistence landscape in the crash and normal periods. No significant difference exists in the number of L^2 norms of the 0-dimensional persistence landscape in the normal periods. According to the definition, the L^2 - norm of the persistence landscape is equal to the value of the Riemann integral at the second higher level of the persistence landscape. From the results of the persistence landscape, the 1-dimensional persistence landscape L^2 norms are

Table 3. Comparison of L^2 - norm for 1-dimensional persistence landscapes in stock market crashes and normal times

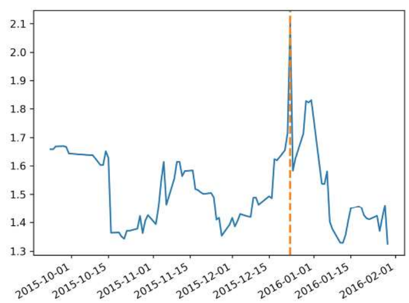
	Phase	T-statistic	P-value	Is there a significant difference
1	Catastrophe 1 and Normal 1	−4.293	0.113	No significant difference
2	Crash 2 and Normal 2	−6.312	0.092	No significant difference
3	Normal 1 and normal 2	−5.231	0.162	No significant difference

not significant in the stock market crash and typical periods, so the statistical analysis is considered reasonable and interpretable.

So to gain insight into the differences between the crash and normal periods, we quantified the numerical characteristics of the persistence landscape using the 0-dimensional persistence landscape L^2 - norm.



(a) Crash period 1

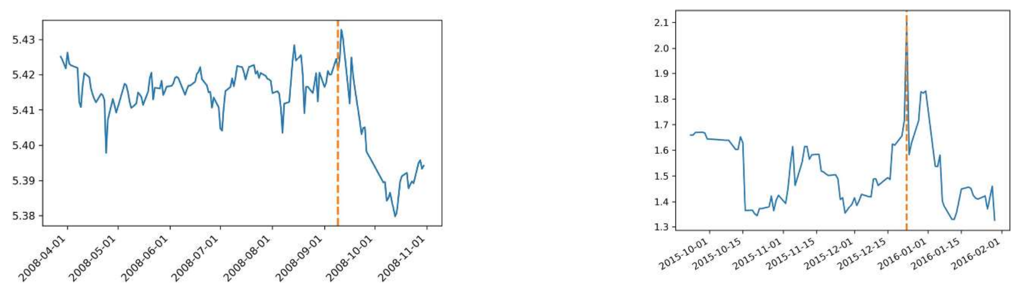


(b) Crash period 2

Fig. 2. Changes in norms L^2 during the stock market crash

Figure 2 shows the L^2-norm change when the Chinese stock market shook in 2008 and 2015, respectively. The Fig. 2(a) shows that before August 2008, the L^2-norm performed exceptionally mildly. The market was stable, while from September 2008 onwards, under the influence of the high market volatility, the L^2-norm increased rapidly and reached its peak. It began to fluctuate at a higher value until November of the same year when it declined significantly. By searching for accurate data, we found that on September 09, 2008, the financial crisis began to spiral out of control and led to the collapse of several reasonably large financial institutions since the subprime housing credit crisis broke out and investors began to lose faith in the value of mortgage securities or were taken over by the government. The results of the experiment are consistent with the actual situation. As can be seen from Fig. 2(b), the L^2 norms remained high between November 2015 and December 2015 until after December 15, 2015, when the market volatility gradually decreased, and the L^2 norms started to decrease significantly and remained at a lower value.

(3) Persistent Entropy
Persistent entropy can be a robust measure of uncertainty in financial markets, especially when a stock market crash is coming. This uncertainty will be elevated for some time when a stock market crash is imminent. Thus the persistent entropy fluctuates within a high level until a financial crisis occurs, and the persistent entropy gradually diminishes. The final result of sustained entropy in dimension 0 is shown in Fig. 3.



(a) Crash period 1 (b) Crash period 2

Fig. 3. Changes in norms L^2 during the stock market crash

From Fig. 3(a) and Fig. 3(b), it is not difficult to see that the 0-dimensional sustained entropy and L^2 norms show opposite trends, but the time nodes are entirely consistent. When the market is in the normal period, the sustained entropy is in the higher value, which indicates that the stock market is more chaotic; while when the market collapses, the sustained entropy will decrease rapidly and fall into the trough, which indicates that the stock market is less chaotic and generally depressed. So based on the results of the L^2 - norm, the persistent entropy can also reflect the changes in the structure of stock data. Moreover, its algorithm is easier and more efficient, so we can innovatively propose to apply it in financial crisis warnings.

4 Classification Results

We use supervised machine learning to label the PD formed by different sliding windows according to whether a stock market crash sample or not in order to make the final classification prediction result. In this paper, we use classical supervised learning SVM (support vector machine) for binary classification when we consider that not much data available and the data are binary-labeled.

As mentioned above, to use the SVM method, we consider using a suitable topological kernel function instead of features to link PD with the SVM. Since PD turn out to be linearly indistinguishable, for the nonlinear case, SVM is handled by choosing a nonlinear mapping (kernel function) that maps the input variables to a high - dimensional feature space, turning them into linearly distinguishable in the high - dimensional space. For persistent homology, computing the topological kernel in this high-dimensional space is necessary. The persistent homology, or using the topological kernel related to persistent homology, we use persistent scale space kernel (PSSK) [10] and the PSSK kernel function is defined as follows:

$$k_{\mathrm{PSSK}}(\mathbf{F}, \mathbf{G}, \sigma) = \frac{1}{8\pi\sigma} \sum_{\mathbf{y}\in\mathbf{F}, \mathbf{z}\in\mathbf{G}} \left(e^{-\frac{\|\mathbf{y}-\mathbf{z}\|^2}{8\sigma}} - e^{-\frac{\|\mathbf{y}-\bar{\mathbf{z}}\|^2}{8\sigma}} \right) \tag{11}$$

where F and G denote the set of point coordinates in two PD: $Z = (b, a)$ and $\bar{Z} = (a, b)$, symmetric about the diagonal, this kernel function is 1-Wasserstein stable [10] and thus theoretically provides more stable performance for topological machine learning.

We processed the data using Python 3.9, the gtda package, and the scikit-learn package. The results were averaged over six runs, with 70%/30% data splits randomly used for training and testing, and the cost factor C of the final SVM was cross-validated in the following grid, as shown in Table 4.

Table 4. Grid cross validation value range

C_values	500	800	900	950	1000	1500	2000	3000	5000
Sigma_values	0.0001	0.0005	0.001	0.005	0.01	0.05	0.1	0.5	1

Finally, after performing the model construction and refinement, we used the following subjects as a test set. We chose a C value of 900 and a sigma value of 0.001 when we obtained a model classification prediction of 96%, which has some effect (Table 5).

Table 5. SVM classification machine learning result index table based on PD and topological kernel

Stock market crash	Precision	Recall	F1-score
Stock market crash	0.97	0.97	0.97
Normal	0.93	0.93	0.93

It can be seen from the above table that the three indicators of the model reach 0.9, which shows that the SVM model based on persistent homology and topological kernel function can well distinguish the PDs between the stock market disaster periods and the normal periods, and has strong practicality.

5 Conclusion

This paper focuses on extracting low-dimensional topological features from stock data using topological data analysis methods, including persistent diagram, persistent landscape, L^2-norm and persistent entropy. These topological summaries are employed to demonstrate changes in the internal structure of the stock market, illustrating their capability to reflect these changes. Experimental results reveal that all aforementioned topological features can differentiate between stock market crashes and regular periods.

L^2 - norm and persistent entropy more accurately capture the abruptness of stock market crashes and align with real-world situations. In the 0-dimensional topological data: during stock market crash periods, paradigm values are notably higher, while sustained entropy values are lower, indicating greater data significance and lower chaos levels. Conversely, during normal periods, paradigm values are lower, and sustained entropy values are higher, suggesting less significant data features and higher chaos levels. Consequently, L^2 - norm and persistent entropy effectively capture market turbulence and reveal the stock market's internal structure. This forms the basis for establishing an early warning model for financial crises, serving as a reference for regulators and investors.

Additionally, this paper conducts preliminary tests for stock market crash recognition. We employ SVM classification machine learning with a topological kernel function to classify PDs before and after stock market crashes, achieving an accuracy rate of 96%, signifying strong model performance. Nonetheless, stock market crash identification involves complex and variable causes, including bubble phenomena, decline rate quantitative characteristics, decline speed quantitative characteristics, and policy interventions, all impacting crash identification. Empirical results demonstrate the model's effectiveness in distinguishing stock market crashes from normal periods, consistently exhibiting high classification accuracy. This holds significant implications for future stock market regulation and investor guidance. Therefore, a comprehensive factor system will be established in the future to enhance more effective crash identification, aiding regulators in implementing relevant rescue efforts and averting more severe developments.

References

1. Bubenik, P., Dłotko, P.: A persistence landscapes toolbox for topological statistics. J. Symb. Comput. **78**, 91–114 (2017)
2. Deboeck, G.J.: Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets. Wiley (1994)
3. Edelsbrunner, Letscher, Zomorodian: Topological persistence and simplification. Discrete Comput. Geom. **28**, 511–533 (2002)
4. Fang, X.: Research on stock market crashes and price fluctuations based on topological data analysis and machine learning. Ph.d. dissertation, Zhejiang University (2021). (in Chinese)
5. Gidea, M.: Topological data analysis of critical transitions in financial networks. In: Shmueli, E., Barzel, B., Puzis, R. (eds.) NetSci-X 2017. SPC, pp. 47–59. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55471-6_5
6. Gidea, M., Katz, Y.: Topological data analysis of financial time series: landscapes of crashes. Physica A Stat. Mech. Appl. **491**, 820–834 (2018)
7. Goel, A., Pasricha, P., Mehra, A.: Topological data analysis in investment decisions. Expert Syst. Appl. **147**(113), 222 (2020)

8. Jin, Z., Guo, K.: The dynamic relationship between stock market and macroeconomy at sectoral level: evidence from Chinese and us stock market. Complexity **1**, 6645570 (2021)
9. Khashei, M., Bijari, M., et al.: Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). Neurocomputing **72**(4–6), 956–967 (2009)
10. Reininghaus, J., Huber, S., et al.: A stable multi-scale kernel for topological machine learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4741–4748 (2015)
11. Rucco, M., Castiglione, F., Merelli, E., Pettini, M.: Characterisation of the idiotypic immune network through persistent entropy. In: Battiston, S., De Pellegrini, F., Caldarelli, G., Merelli, E. (eds.) Proceedings of ECCS 2014. SPC, pp. 117–128. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29228-1_11
12. Shim, J.S., Kim, H.C., et al.: Comparison of various implant provisional resin materials for cytotoxicity and attachment to human gingival fibroblasts. Int. J. Oral Maxillofac. Implant. **34**(2) (2019)
13. Swanson, N.R., White, H.: Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. Int. J. Forecast. **13**(4), 439–461 (1997)
14. Tao, X., Renmu, H., et al.: Input dimension reduction for load forecasting based on support vector machines. In: 2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies. Proceedings, IEEE, vol. 2, pp. 510–514 (2004)
15. Zhang, N., Lin, A., Shang, P.: Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. Physica A Stat. Mech. Appl. **477**, 161–173 (2017)