

Article

Change Point Detection in Financial Market Using Topological Data Analysis

Jian Yao ^{1,3} , Jingyan Li ³ , Jie Wu ^{2,3,*} , Mengxi Yang ^{1,4}  and Xiaoxi Wang ^{1,3}

¹ School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China; yaojian23@mails.ucas.ac.cn (J.Y.); yangmengxi@ucas.ac.cn (M.Y.)

² School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050024, China

³ Beijing Key Laboratory of Topological Statistics and Applications for Complex Systems, Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101408, China; jingyanli@bimsa.cn (J.L.); wangxiaozi@bimsa.cn (X.W.)

⁴ MOE Social Sciences Innovative Group on Complex Systems Modeling in Economic Management in the Era of Digital Intelligence, MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, Beijing 100190, China

* Correspondence: wujie@bimsa.cn

Abstract

Change points caused by extreme events in global economic markets have been widely studied in the literature. However, existing techniques to identify change points rely on subjective judgments and lack robust methodologies. The objective of this paper is to generalize a novel approach that leverages topological data analysis (TDA) to extract topological features from time series data using persistent homology. In this approach, we use Taken's embedding and sliding window techniques to transform the initial time series data into a high-dimensional topological space. Then, in this topological space, persistent homology is used to extract topological features which can give important information related to change points. As a case study, we analyzed 26 stocks over the last 12 years by using this method and found that there were two financial market volatility indicators derived from our method, denoted as L_1 and L_2 . They serve as effective indicators of long-term and short-term financial market fluctuations, respectively. Moreover, significant differences are observed across markets in different regions and sectors by using these indicators. By setting a significance threshold of 98 % for the two indicators, we found that the detected change points correspond exactly to four major financial extreme events in the past twelve years: the intensification of the European debt crisis in 2011, Brexit in 2016, the outbreak of the COVID-19 pandemic in 2020, and the energy crisis triggered by the Russia–Ukraine war in 2022. Furthermore, benchmark comparisons with established univariate and multivariate CPD methods confirm that the TDA-based indicators consistently achieve superior F1 scores across different tolerance windows, particularly in capturing widely recognized consensus events.

Keywords: change point detection; topological data analysis; persistent homology; financial extreme events; time series analysis



Academic Editor: Zongjun Wang

Received: 1 September 2025

Revised: 30 September 2025

Accepted: 2 October 2025

Published: 6 October 2025

Citation: Yao, J.; Li, J.; Wu, J.; Yang, M.; Wang, X. Change Point Detection in Financial Market Using Topological Data Analysis. *Systems* **2025**, *13*, 875. <https://doi.org/10.3390/systems13100875>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Time series data are widely present in various areas, including finance [1], biology [2], and engineering [3]. Time series analysis plays a crucial role in financial markets. For example, predicting financial market volatility helps assess the risks of asset price fluctuations, while historical stock price and trading volume analysis aids in forecasting future price

trends [4]. Additionally, portfolio optimization can be achieved by analyzing asset return rates, correlations, and covariances in time series data [5]. Beyond financial applications, time series data are also widely used in macroeconomic analysis. By leveraging historical data on economic indicators such as GDP, unemployment rates, and inflation, researchers can analyze economic cycles and predict economic trends and policy effects [6–9]. The inherent characteristics of time series data—such as autocorrelation, trend, periodicity, seasonality, noise, and lag effects—usually make their analysis complicated [10]. Traditional time series models excel at capturing linear dependencies, but struggle with nonlinear structures in financial data. To address these limitations and incorporate the inherent nonlinear patterns present in real-world financial markets, numerous nonlinear models have been developed [11,12].

The financial market is a highly complex and dynamically evolving system, with its fluctuations influenced by various factors such as economic policies, macroeconomic indicators, market sentiment, and international events. As a critical technique for identifying structural shifts in time series data, change point detection (CPD) holds significant theoretical and practical value in financial research. CPD is an important research subject in financial time series analysis with significant theoretical and practical advancements [13]. Over the past several decades, various statistical methods have been developed for CPD [14]. Multivariate CPD not only detects change points in univariate series but also integrates inter-dependencies between multiple variables, enhancing detection accuracy and robustness [15]. However, when conducting CPD on multivariate financial time series data, many existing algorithms are highly sensitive to noise and outliers, leading to excessive false alarms during periods of high-frequency market fluctuations [16]. Additionally, when multiple dimensions must be analyzed simultaneously, computational complexity increases significantly, resulting in low efficiency. Many CPD methods perform well in detecting abrupt short-term changes, such as sudden market crashes, but struggle to effectively capture gradual transitions, such as long-term market trends [17]. In financial time series change point detection, traditional econometric tools primarily rely on linear and parametric methods to analyze relationships between variables, with the final results heavily dependent on data pre-processing. While effective in some cases, these methods struggle with the increasing complexity of high-dimensional data [18,19]. As a result, economists are actively exploring more robust multivariate CPD methods to overcome these challenges. One direction of such research is to use topological data analysis (TDA) in financial market time series data.

In recent years, TDA has been successfully applied in numerous studies as a novel technique for change point detection in time series [20–22]. Unlike conventional methods that rely on aggregated data and model assumptions, TDA focuses on visualization and topological features to identify hidden patterns [23]. Moreover, TDA does not require predefined assumptions about data distributions, making it well suited for capturing non-linear relationships. TDA generally uses persistent homology as its core tool to extract topological features across different scales, thereby identifying the global structure and patterns of the data. TDA regards time series data as a collection of points and maps these points to a high-dimensional space; the persistent homology technology of TDA can help identify topological features in these point clouds, such as changes in connected branches and periodic patterns [24,25]. Edelsbrunner et al. [26] simplified the framework of topological filtering and established an algorithm for computing an individual persistent homology group over arbitrary principal ideal domains in any dimension. Another landmark study Zomorodian and Carlsson [27] described the qualitative characteristics of complex structural data by computing persistent homology, and proved that this method possesses a certain degree of robustness. A comprehensive review study on topology and

data written by Carlsson et al. [28] greatly promoted the research progress of data analysis in topology. Bubenik et al. [29] also gave a statistical explanation for the calculation of persistent homology. Before that, data analysis using topological perspective focused more on geometry and network reconstruction [30]. After that, persistent homology computing came into the researchers' field of vision and was successfully applied in bioinformatics [31], medicine, neuroscience [32], social networks [33], and even finance [34]. Many studies have gradually introduced concepts such as barcodes, persistence diagrams, and persistence landscapes [35], which we will elaborate upon individually in Section 3.

While being a powerful tool in time series data analysis, TDA is not so widely used to mine structural information from time series data in financial markets [36]. This gap is what motivates the research presented in this paper. The key questions we want to answer are the following: 1. Can topological features effectively identify major change points in financial markets? 2. Do topological indicators exhibit significant changes in response to major financial extreme events? 3. Can topological features provide portfolio recommendations for investors with different preferences? In this paper, we try to answer these questions.

To address these questions, we found that considering the TDA norms of persistent landscapes is rather useful. TDA norms quantify the "duration" of topological features, providing a measure of strength and significance of the features [37]. For example, when there are drastic changes in financial markets, we may see peaks in the topological signal [34]. This suggests that the "duration" property of topological features is closely linked to the structural characteristics of time series volatility [38]. We successfully applied this finding to the monitoring of financial extreme events. To the best of our knowledge, this study presents the first comprehensive application of persistent homology in multi-time series analysis for solving the CPD problem. Our key contributions are as follows: We propose a TDA-based change point indicator for multi-stock price data and establish a threshold for detecting global financial change points. We analyze financial market fluctuation across both regional and sector dimensions, providing a novel perspective on investment portfolios. We apply this approach and successfully detect four extreme financial events in the past 12 years: the European debt crisis in 2011, Brexit in 2016, the outbreak of the COVID-19 pandemic in 2020, and the energy crisis triggered by the Russia–Ukraine war in 2022. These extreme events have not been fully detected in past TDA-based CPD studies, which shows that this approach can not only detect internal system risks, but also provide early warning signals for external risk shocks.

The structure of this paper is arranged as follows: Section 2 reviews the existing research methods for CPD and their limitations, as well as the research progress of TDA in detecting change points in financial time series data. Section 3 provides a comprehensive introduction to the background knowledge and analytical framework of TDA. Section 4 presents the technique of mapping multivariate time series into topological space. Section 5 presents the basic process of data collection, as well as preprocessing, and the pipeline of TDA. Sections 6 and 7 present the analysis of our results. Finally, Section 8 is the conclusion.

2. Related Work

2.1. Change Point Detection

Change point detection (CPD) plays a crucial role in financial time series analysis, enabling researchers to identify significant shifts in market trends and volatility [39]. Mature statistical approaches and CPD technologies have been applied to the financial field. The CUSUM method, proposed by Page [40], has become a classic method for detecting mean change points. Subsequently, the multiple structural change point model proposed by Bai and Perron [41] allows for the identification of multiple change points in a regression framework,

which is particularly suitable for the analysis of trend and volatility changes in financial time series. Donald [42] proposed a CPD method based on Wald, LM and LR statistics which allows the detection of heteroscedasticity. However, these methods rely on strong parametric assumptions and often struggle with non-stationary and high-dimensional data.

To address these limitations, Barry and Hartigan [43] proposed a Bayesian CPD method, which estimates the change point location through Bayesian inference and can naturally handle uncertainty. Chib et al. [44] introduced a Bayesian method using Markov Chain Monte Carlo (MCMC) technology to estimate change points which is suitable for high-volatility time series. In the CPD of high-frequency financial data, Carlin [45] proposed a non-parametric method, based on the Bayesian model, which offers flexibility and is widely used in the detection of structural mutations in stock price fluctuations. Energy-based CPD measures the significant change in distribution between two intervals by calculating the energy difference of financial time series in different sub-intervals (before and after segmentation) [46]. With the increasing availability of large-scale financial data, machine learning-based CPD techniques have emerged. Amini and Wainwright [47] introduced a kernel-based CPD algorithm that leverages support vector machines and kernel methods to effectively detect nonlinear structural changes. The sparse model proposed by Harchaoui and Lévy-Leduc [48] can identify multiple change points in high-dimensional time series data. In recent years, deep learning methods have also begun to emerge in CPD. RNN-based methods excel in capturing sequence dependencies, and are particularly suitable for complex pattern detection in financial time series [49]. The deep learning-based CPD methods proposed by Lavielle [50] and Wei and Luo [51] can identify hidden change points in non-stationary sequences. In addition, CNN and Transformer architectures have also achieved initial success in the CPD of high-frequency time series data [52].

Despite these advancements, existing CPD methods still face challenges in handling high-dimensional financial data and capturing non-linear relationships without strong model assumptions. The study of financial time series from the perspective of dynamic systems provides new ideas for CPD. Khasawneh and Munch [53] explored the possibility of using topological data analysis to reconstruct datasets of dynamic systems, and showed that persistent homology can distinguish different types of equilibrium, thus proving that this approach can be used in automated data analysis of change detection and prevention. They also examined the time series by topological data analysis to determine the stability of the autonomous stochastic delay equations in parameter space. The results of this study show that the described approach can be used for analyzing datasets of delay dynamical systems generated both from numerical simulation and experimental data [54]. Kim et al. [55] proposed a simple and efficient method to observe time series using the topological features of the attractor of the underlying dynamic system. The persistence landscapes and silhouettes of the Rips complex were obtained by performing a denoising step on the principal components [56,57]. The denoising step applied a time delay embedding technique based on noisy discrete time series samples, and the results showed that this method is stable and can extract features from noisy time series data [55]. These advancements in research provide the possibility of introducing TDA into CPD studies. For example, Gu et al. [22] applied this method to extract Betti numbers and persistent homology features for multi-agent CPD in high-dimensional data. Leibon et al. [58] proposed a novel method for the scale-dependent topological characterization of network structures in equities market.

2.2. Detection of Extreme Financial Events

The global economy has undergone recurrent crises or other extreme events throughout history. Thus, it is crucial to identify the underlying patterns in recurrent economic crises rather than treating them as isolated events. The detection of extreme financial

events is transformed into a statistical change point detection problem in financial time series. While TDA has been applied to CPD in many fields, using TDA for CPD in financial markets remains a relatively new research direction [13]. Lee et al. [38] investigated how the topological structure of the global macroeconomic network influences the propagation dynamics of economic crises. Gidea [59] proposed a method for detecting early indicators of critical transitions in financial data. By constructing time-dependent networks from multiple stock price time series and analyzing their topological features, researchers can compute persistent homology to track structural changes in financial data. The results indicate that these topological changes can serve as indicators of approaching critical transitions and have been successfully applied in the detection of the 2008 financial crisis. Next, the aforementioned authors applied this method to analyze daily return fluctuations in the four major US stock market indexes during the 2007–2009 financial crisis, and detected the dotcom crash on 10 March 2000 and Lehman bankruptcy on 15 September 2008 [34]. They also combined TDA with the machine learning method to identify early warning signals when the four cryptocurrencies (Bitcoin, Ethereum, Litecoin, and Ripple) approached critical transitions, such as market crashes, and further analyzed multiple Bitcoin mini-crashes Bitcoin between 2016 and 2018 [60]. Similarly, Aguilar and Ensor [61] studied the daily log returns of the four major U.S. stock market indices and 10 ETF sectors, utilizing topological features in high-dimensional time series to characterize stock market dynamics. At a significance level of $\alpha = 0.05$, they identified structural changes in the U.S. stock market between 2019 and 2020. Goel et al. [62] was the first to introduce the application of TDA to asset allocation in the financial sector, proposing an investment portfolio strategy based on a TDA-derived risk index (EI) and discovering that this index is more effective than standard deviation. In 2024, they extended this research by proposing a strategy for stock selection based on TDA and data clustering tools, specifically designed for sparse portfolio construction. The robustness of the method was validated using the S&P 500 index from 2009 to 2020, including data from the COVID-19 period [63]. This two-stage portfolio construction method, involving time series representation generation and clustering analysis, was also adopted in Sokerin et al. [64]'s research. In this paper, the indicators derived from TDA are named L_1 and L_2 , and a detailed explanation will be provided in Section 3.

Related studies have also applied the same methodology to detect the 2008 global financial crisis and the 2010 European debt crisis [65]. Furthermore, they also analyzed China's stock market data and identified three significant market fluctuations since 2013 [66]. Coincidentally, Ismail et al. [67] found that the variance in TDA norms outperformed residuals in providing early warning signals for major financial crises in Bitcoin during 2017 and 2019. Furthermore, they integrated machine learning methods with persistent homology to analyze the Kuala Lumpur stock market, achieving improved performance in stock trend prediction [68]. Additionally, they combined L_1 with three other key indicators to detect early signals of financial crises in the U.S., Singapore, and Malaysia markets, validating the robustness of this method in identifying the dot-com bubble burst and the collapse of Lehman Brothers [69]. Applying this method to the Credit Default Swap (CDS) market, Katz and Biem [70] thought that L_1 could serve as a leading indicator of impending financial crises driven by endogenous market forces and that the stock market lags behind the CDS market. Yen and Cheong [71] explored the application of TDA and persistent homology in analyzing the Singapore and Taiwan markets and identified structural changes during market crashes by computing topological features. Furthermore, they proposed a systematic improvement by integrating TDA with curvature to gain deeper insights into the topological and geometric structure of financial networks and analyzed a stock market crash in the Taiwan market [72]. Related studies, including Majumdar and Laha [73], have proposed the SOM-TDA and RF-TDA methods for time series classification and cluster-

ing and applied them to the classification of stocks from different sectors, which implies that the topological features of stock price time series vary across sectors. Rai et al. [74] not only focused on industry sectors but also linked TDA features with continental plate characteristics, conducting a sector-by-sector analysis of the Indian stock market during the COVID-19 period. The study by Sebestyén and Iloskics [75] employs a topological network approach to analyze the topological characteristics (transitivity, path lengths, skewness of degree distribution and stability of connections) of the economic shock contagion network based on pairwise Granger causality relationships between national economic outputs.

2.3. Recent Studies of TDA with Financial Applications

Beyond the studies above, TDA is also generating innovations across other areas of economics and finance and shows strong promise for practical application. Zhang and Wu [76] use persistence homology on sliding-window embeddings of returns to characterize topological patterns preceding market crashes, illustrating how PH descriptors evolve around turmoil. Guritanu et al. [77] propose a strictly causal early-warning framework: multivariate returns are mapped to point clouds, Vietoris–Rips diagrams are summarized via persistence landscapes and L_p norm signals are generated without look-ahead bias. Nath Sharma et al. [78] combine TDA distances with Granger causality to study crash time interdependence across stocks and commodities. Collectively, these works share a pipeline of windowed point cloud construction, PH features, crisis detection or cross-market diagnostics. In contrast, our study focuses on formal CPD with explicit threshold selection and diagnostics [79], thereby complementing early warning and causality focused approaches with a clear decision rule for dating regime shifts.

Based on the review of the above studies, we find that using persistent homology in TDA to analyze financial market fluctuations has become a relatively mature technique (Table 1). However, there are still many shortcomings in its specific applications. In the detection of extreme financial events, most studies directly analyze stock indices from different regions or sectors to identify financial crises. The events covered are mainly concentrated on the 2008 financial crisis and the 2020 COVID-19 pandemic outbreak, indicating that the detected events are incomplete. Additionally, existing research primarily distinguishes financial markets based on regions or sectors but does not differentiate between long-term and short-term fluctuations. Furthermore, since most studies focus on index data, they fail to provide direct stock-level investment portfolio recommendations in CPD-based asset allocation. To address these gaps, this study aims to apply TDA techniques to CPD analysis at the individual stock level, offering innovative insights into financial market change points and providing more specific investment portfolio recommendations.

Table 1. Literature on finance using persistent homology (SW: sliding window strategy; TDE: Takens' time delay embedding; SV: statistic variables; PL: persistence landscape; ✓ indicates that this method was used in this study; / indicates that this method was not used in this study.)

Reference	Dataset	Time Span	Data Type	SW	TDE	SV	PL	Applications
Gidea [59]	DJIA 29 stocks	23 February 2007–2008	daily closing prices	✓	/	Pearson correlation	/	Detect early warning signals for critical transitions
Gidea and Katz [34]	S&P 500, DJIA, NASDAQ, and Russell2000	23 December 1987–8 December 2016	7301 daily log-returns	✓	✓	/	L1 norm, L2 norm	Detect early warning signals of imminent market crashes
Gidea et al. [60]	Bitcoin, Ethereum, Litecoin and Ripple	2016–2018	daily log-returns	✓	✓	/	C1 norm	Critical transitions in time series of cryptocurrencies
Aguilar and Ensor [61]	4 major US stock market indices 10 ETF sectors indices	January 2010–June 2020	daily log-returns	✓	/	/	L1 norm, L2 norm	Critical transitions are identified by the statistical properties
Goel et al. [62]	Thompson Reuters EIKON data stream	January 2005–November 2018	daily closing price	✓	✓	/	L1 norm	Investment decision
Guo et al. [65]	DAJA, NASDAQ indices	2 January 2003–31 December 2013	daily log-returns	✓	✓	/	L1 norm, L2 norm	Financial crises
Ismail et al. [67]	Bitcoin	24 August 2016–19 February 2020	daily closing prices	✓	/	/	L1 norm	Detect early warning signals
Majumdar and Laha [73]	IBEX35,FTSE MIB, FTSE ATHEX20, PSI20 and ISEQ; Shanghai Composite index and Shenzhen Component index	1 January 2003–31 December 2013	daily log-returns	✓	✓	/	L1 norm	Identified sectors features
Ismail et al. [68]	3 indices from Kuala	15 May 2000–27 March 2020	daily log-returns	✓	✓	Betti sequences	/	Detect early warning signals
Sebestyén and Iloskics [75]	GDP growth Y	1961 Q2–2006 Q4 1996 Q3–2006 Q4	$\Delta Y/Y$	/	/	Network properties	/	Reveal shock contagion
Katz and Biem [70]	CDS spreads with 5-years maturity on senior unsecured debt of 93 North American firms distributed among 10 economic sectors	January 2004–August 2019	daily log-returns	✓	✓	/	L1 norm	Establish a indicator of an approaching financial crash
Yen and Cheong [71]	Singapore Exchange & Taiwan Stock Exchange	1 January 2017–30 April 2019	daily adj closing price	✓	✓	Betti number & Euler characteristics	/	Identify which homology groups becomes less persistence
Yen et al. [72]	Taiwan stock exchange	1 January 2017–30 April 2019	daily-closing price	✓	/	Ricci Curvature, Ricci Flow	/	Understand changes in financial market correlation

Table 1. *Cont.*

Reference	Dataset	Time Span	Data Type	SW	TDE	SV	PL	Applications
Guo et al. [66]	100 stocks from China's markets	3 January 2013–31 August 2020	daily log returns	✓	✓	/	L1 norm, L2 norm	Risk analysis
Ismail et al. [69]	11 indexes of US, Singapore Malaysia	22 December 1987–29 December 2017, 31 August 1991–27 March 2018, 15 May 2000–27 March 2018	daily closing price	✓	✓	/	L1 norm	Detect early warning signals of financial crises
Sokerin et al. [64]	S&P 500 index	2012–2013; 2015–2016; 2018–2019	daily closing price	✓	✓	Bars statistics	1 & 2-dim	Protfolio selection
Goel et al. [63]	S&P 500 index	December 2009–August 2022	daily log returns	✓	✓	/	LP norm	Sparse portfolios
Rai et al. [74]	Forty indices from 40 countries/regions	2006–2010 COVID-19 pandemic era	daily log returns	✓	/	/	L1 norm, L2 norm	Identifying extreme events
Our research	26 stocks from NASDAQ	24 March 2011–15 December 2023	daily log returns	✓	✓	/	L1 norm, L2 norm	Long- and short-term volatility analysis, financial extreme event detection and portfolio advice

3. TDA Concepts and Methods

3.1. A Brief Review of TDA

Topological data analysis (TDA) is a newly developed field derived from algebraic topology in recent years, aimed at discovering data's hidden new structures, which can provide novel and valuable insights that conventional data analysis techniques may fail to capture [23]. The underlying idea of TDA is that data has a shape which can convey valuable meanings [80]. Common data can be viewed as point clouds embedded in high-dimensional Euclidean space or general metric space [24]. These point clouds are not uniformly distributed and usually contain nonlinear geometric information with nontrivial topology [56]. TDA can mine the topological information in point clouds via persistent homology and express it through persistence diagrams and barcodes [81,82]. Replacing the persistence diagram with another robust tool—persistence landscapes, which can be further embedded into the Banach space—gives a natural metric space structure [35]. A persistence landscape comprises a sequence of persistence and piecewise linear functions defined on a re-scaled birth–death coordinate [37]. TDA can effectively explain how objects relate to one another for their qualitative structural properties, such as shape and structure. Its analysis process does not assume a specific data distribution, making it highly suitable for analyzing complex high-dimensional data [83]. For instance, Pereira et al. [25] quantified periodicity in time series in a shape-agnostic manner and with resistance to damping. Without presupposing a particular pattern, they evaluated the circularity of a high-dimensional representation of the signal. Pereira and de Mello [36] proposed a clustering method for time series and spatial data based on topological features, highlighting persistent homology as an effective tool for capturing multi-scale structures and analyzing financial time series structural changes.

3.2. Persistent Homology

In this part, we will introduce the core components of TDA, persistent homology. The key idea of persistent homology is that it allows us to study the shape of data across multiple scales [26]. It provides a multi-scale description of the topological features (such as connected components, loops, and voids) in a dataset by tracking their birth and death as a filtration parameter changes [27,28]. In this way, it can systematically reflect the changes in some hidden structures which are not seen in normal data analysis. Given a dataset, which could be a point cloud or a simplicial complex made from it, the work flow of calculating persistent homology is as follows:

- Firstly, by using proper filtration parameters and the initial data, construct a sequence of simplicial complexes, a process called filtration:

$$S_1 \subset S_2 \subset \cdots \subset S_n \quad (1)$$

Here, each S_i is a simplicial complex.

- Secondly, calculate the simplicial homology of each simplicial complex S_i ;
- Thirdly, from the result of the second step, obtain more concise information like persistent diagrams, barcodes and persistent landscapes;
- Finally, conduct further analysis, and observe what insights can be obtained from the topological quantities calculated above.

In the remaining part of this section, we will conduct a concise review of simplicial complexes, simplicial homologies, and filtration, as well as barcodes and persistent diagrams, to make it easier to understand the work flow adopted by us and detailed above.

3.2.1. Simplicial Complex and Simplicial Homology

For all calculations in TDA, the initial step is always to transform the given data into a simplicial complex S_i (or, more precisely, to conduct filtration of the simplicial complexes) and then calculate topological invariants based on it [84]. A simplicial complex S is a collection of simplices (points, edges, triangles, etc.) such that the following becomes true:

- Every face of a simplex in S is also in S .
- The intersection of any two simplices in S is either empty or a face of both.

A k simplex is usually represented by designating its vertices as $[v_0, v_1, \dots, v_k]$. As a simple example, consider a triangle with vertices v_0, v_1 and v_2 ; this is a simplicial complex with the following:

- Three 0-simplices (vertices): v_0, v_1, v_2 ;
- Three 1-simplex (edges): $[v_0, v_1], [v_1, v_2], [v_2, v_0]$;
- One 2-simplex (triangle): $[v_0, v_1, v_2]$.

For a given simplicial complex S , at each dimension k , one can define the following:

- C_k : The group of k -chains which is nothing but the formal sums of k -simplices:

$$C_k = \sum_i a_i \sigma_i \quad (2)$$

with i runs over all k simplices, σ_i in S , and a_i in either elements in \mathbb{Z} or \mathbb{Z}_2 depending on specific problems.

- ∂_k : The boundary operator mapping k -chains to $(k - 1)$ -chains,

$$\partial_k : C_k \rightarrow C_{k-1} \quad (3)$$

and more specifically, in terms of vertices,

$$\partial_k [v_0, v_1, \dots, v_k] = \sum_{i=0}^k [v_0, v_1, \dots, \hat{v}_i, \dots, v_k] \quad (4)$$

with \hat{v}_i omitted.

- B_k : The subgroup of C_k composed by the image of the ∂_{k+1} boundary map, i.e., $B_k = \text{im} \partial_{k+1}$;
- Z_k : The subgroup of C_k composed by k cycles σ , which satisfy the condition $\partial_k \sigma = 0$. It is easy to see that $Z_k = \ker \partial_k$.

With these definitions at hand, we are now ready to define the simplicial homology for a given simplicial complex S . Note that the boundary of a boundary is always empty, i.e., $\partial_k \partial_{k+1} = 0$, which means that every k -boundary is also a k -cycle, so $B_k \subseteq Z_k$. What about the converse statement? It turns out that the converse is not always true, and one can define the k -th homology group as $H_k(S, \mathbb{Z}) = Z_k / B_k$ to measure the discrepancy. Note that, for elements of H_k , the k -cycles are in fact subject to the equivalence relation $\sigma \equiv \sigma + \tau$ for τ a k -boundary. Following the definition of the homology group, we can define the Betti numbers b_k , which define the ranks of the homology groups. Geometrically, the 0-th Betti number b_0 is the number of connected components, and higher Betti numbers b_i reflect $(i + 1)$ -dimensional holes (by counting the cycles wrapping them) [84].

3.2.2. Filtration and Persistent Homology

In real-world practice, people are more interested in datasets varying as a function of certain parameters like time or space. For example, intuitively, datasets at certain time points will give us simplicial complexes which can reflect information at that time. To

capture structures over a time scale, we need to obtain a series of simplicial complices which are parameterized by time. Furthermore, we need to compute each simplicial complex's homology and monitor how it changes with time [26]. This is the intuitive idea behind filtration and persistent homology.

Recall that the definition of a filtration is a sequence of simplicial complices, which can be written as Equation (1). Here, the inclusion map $S_i \subset S_{i+1}$ will induce a map of chains,

$$C_k(S_i) \rightarrow C_k(S_{i+1}) \quad (5)$$

which can further induce a homomorphism of homology groups,

$$H_k(S_i) \rightarrow H_k(S_{i+1}) \quad (6)$$

as a result of the fact that the inclusion map is commutative with the boundary map. With Equation (6), one can track how each cycle changes while the filter parameter changes, and the output of persistent homology can tell us when a cycle is generated and when it dies. Basically, it can tell us how persistent a given topological feature is as the filter parameter changes. Therefore, it can give us more information than Betti's number, which can only measure topological quantities at a specific value of the filter parameter [85]. For this reason, persistent homology is always referred to as a multi-scale technique.

Depending on the nature of a given dataset and the purpose of analysis, there are several different ways to construct a filtration, including the Čech complex, Vietoris–Rips complex, and Delaunay complex [27,86]. There are several standard approaches to constructing simplicial complexes from point cloud data, with the Vietoris–Rips complex being one of the most commonly used complexes. Given a Vietoris–Rips diameter of $\sigma > 0$, a $(k + 1)$ -simplex is formed if the pairwise distances among its vertices are all less than σ . The collection of all simplices formed under a given scale σ constitutes a simplicial complex S_σ (Figure 1).

Evolution of Vietoris–Rips Complex with Increasing Filtration Parameter

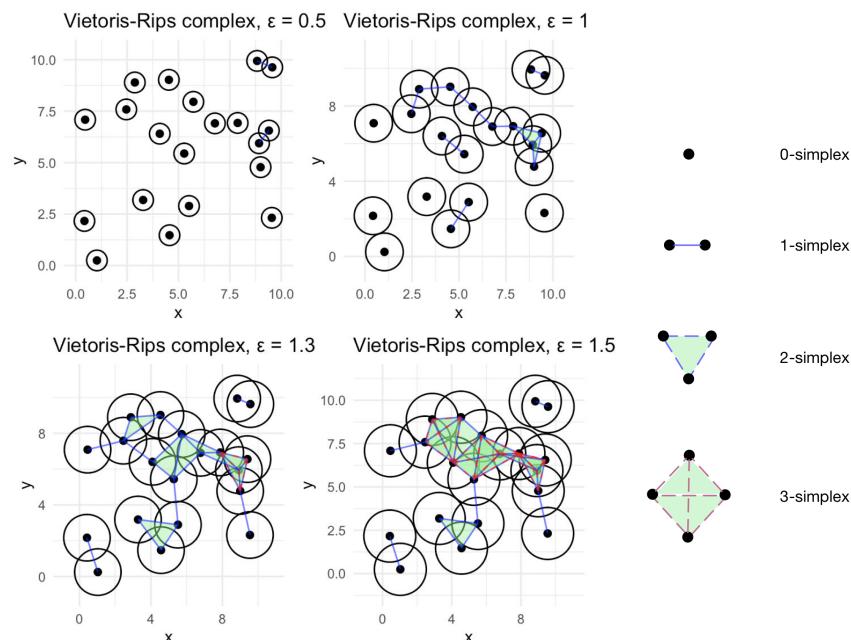


Figure 1. Evolution of Vietoris–Rips complex with increasing filtration parameters ϵ .

3.2.3. Barcode, Persistence Diagram, and Persistence Landscape

A barcode is a graphical representation of persistent homology, visually encoding the birth and death of topological features across different scales [81]. Each horizontal line represents a topological feature, where the left endpoint denotes its birth scale, and the right endpoint marks its death scale [85]. The length of the strip corresponds to the feature's persistence, indicating the scale range over which it remains present (Left panel of Figure 2). This intuitive tool allows for the quick identification of persistent topological structures and short-lived features, which typically correspond to meaningful patterns and noise, respectively. Cohen-Steiner et al. [87] further introduced the persistence diagram as a more succinct visualization tool. As the filtration parameter ϵ increases, we track the birth and death of topological features. A feature that appears in $H_p(S_i)$ and disappears in $H_p(S_j)$ is represented as a point (ϵ_i, ϵ_j) in the persistence diagram. The persistence of the feature is given by persistence = $\epsilon_j - \epsilon_i$ (Right panel of Figure 2). The persistence diagram transforms barcodes into a two-dimensional scatter plot, allowing for a more intuitive quantification of the stability of topological features. In these diagrams, the horizontal axis represents a feature's birth time, while the vertical axis represents its death time. Black dots represent zero-dimensional homology, tracking the birth and merging of connected components, while red triangles denote one-dimensional homology, corresponding to the birth and disappearance of loops. The sizes of the black dots and red triangles reflect their persistence.

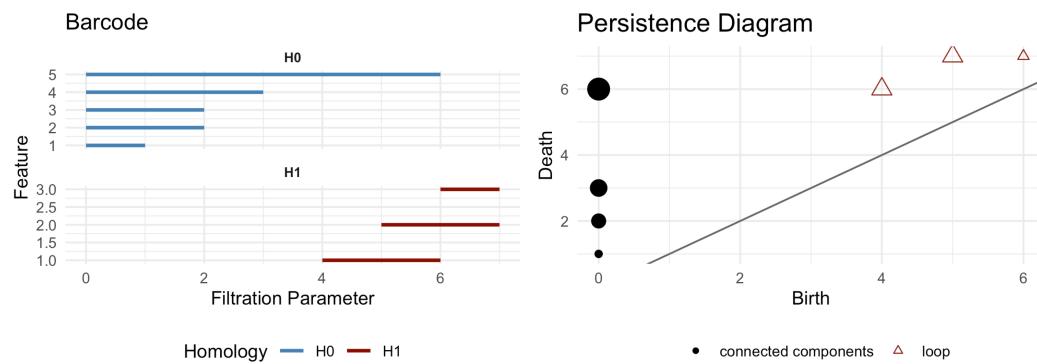


Figure 2. Barcode and persistence diagram.

Persistence diagrams and barcodes provide multi-scale representations of topological features. Although these representations play a crucial role in visualizing topological structures, they do not provide a direct means of quantifying persistence sequences and are not well-suited for statistical analysis. Recent studies have explored methods to transform persistence diagrams into finite-dimensional vector representations, ensuring their stability under small perturbations of the input data. To quantify the persistence diagram, one of the more outstanding studies by Bubenik [35] introduced the concept of a persistence landscape, a topological summary that encodes persistent homology into a sequence of piecewise-linear functions in a vector space. Persistence landscapes transform persistence diagrams into function spaces, typically Banach or Hilbert spaces, enabling their integration with functional analysis techniques. He also proposed an approach that involves mapping the persistence diagram onto the $y = x$ axis to obtain the persistence landscape followed by the computation of the L_p norm. Existing research indicates that the transformation from persistence diagrams to persistence landscapes is both stable and reversible and that these norms are closely related to the oscillatory behavior of time series data [35]. Later, he also introduced a weighted variant of persistence landscapes and defined a single-parameter Poisson-weighted persistence landscape kernel [88]. The following provides the formula

for the conversion of a persistence diagram into a persistence landscape. Given a persistence landscape, λ , it is composed of a series of functions $\lambda_k(t)$, where k represents different levels of the persistence landscape (layers) (Figure 3). The formula for computing the L_p norm is as follows:

$$\|\lambda\|_p = \left(\sum_{k=1}^{\infty} \int_{-\infty}^{\infty} |\lambda_k(t)|^p dt \right)^{\frac{1}{p}} \quad (7)$$

where $\lambda_k(t)$ is the k th layer of the persistence landscape. p is the specified norm parameter (e.g., $p = 1, 2, \infty$). The formula sums up over all layers k and integrates over t .

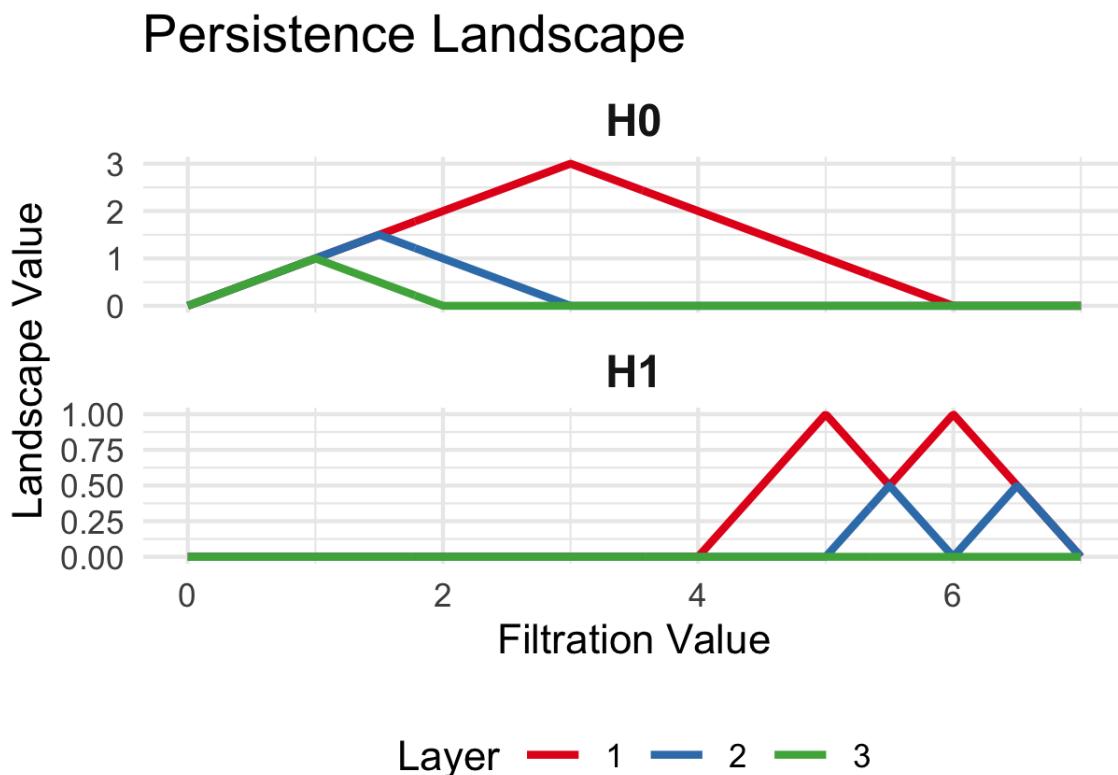


Figure 3. Persistence landscape.

Building on prior research, this study aims to enhance the application of persistent homology in financial time series, with a focus on improving stock volatility detection. To address the research questions outlined in the introduction, this study will focus on two key aspects. First, by analyzing the L_1 and L_2 norms of persistent homology, we examine short-term and long-term market fluctuations across different regional and industrial sectors. Second, using persistent homology results from multivariate stock price time series, we identify major financial extreme events over the past twelve years.

4. From Time Series to Point Cloud

Persistent homology has been widely applied to various types of data, including finite spaces [89], image data [90], and networks [91]. In the analysis of persistent homology, much of the data exists in the form of point clouds [92]. However, time series data itself is one-dimensional, so we need a method to transform it into high-dimensional point clouds to facilitate the application of topological methods. To achieve this, sliding window and time delay embedding techniques are commonly used.

4.1. Sliding Window

The sliding window strategy is commonly employed in time series analysis to generate overlapping subsequences [93]. The basic idea is to select a window length d , and then consecutively extract d continuous data points from the time series to form a high-dimensional vector. The window then slides forward, repeating this process, ultimately generating a point cloud dataset (Figure 4). Perea et al. [25] were the first to combine the sliding window method with persistent homology and demonstrated how to use the point cloud constructed by the sliding window to compute the topological features of time series. They found that this method can quantitatively assess the periodicity of time series and provide robust mathematical theoretical support [94]. The primary purpose of the sliding window technique is to restructure volatile time series data into a matrix representation. The window size plays a crucial role in determining the length of each sample sequence.

Under the framework of delay embedding theory, the window length ω must be sufficiently long to capture the system's characteristic cycles or autocorrelation structures. If the window is too short, the resulting point cloud becomes overly sparse, making it difficult to form meaningful topological cycles; conversely, an excessively long window introduces additional noise and substantially increases computational complexity. In practice, we set the window length to 21 trading days, which corresponds to approximately one calendar month. This is a commonly adopted temporal scale in financial time series analysis, as it allows the capture of a complete market fluctuation pattern within a month while avoiding the structural smoothing effects that typically occur with quarterly or longer windows. Therefore, the choice of 21 days provides a balance that is theoretically grounded and empirically robust.

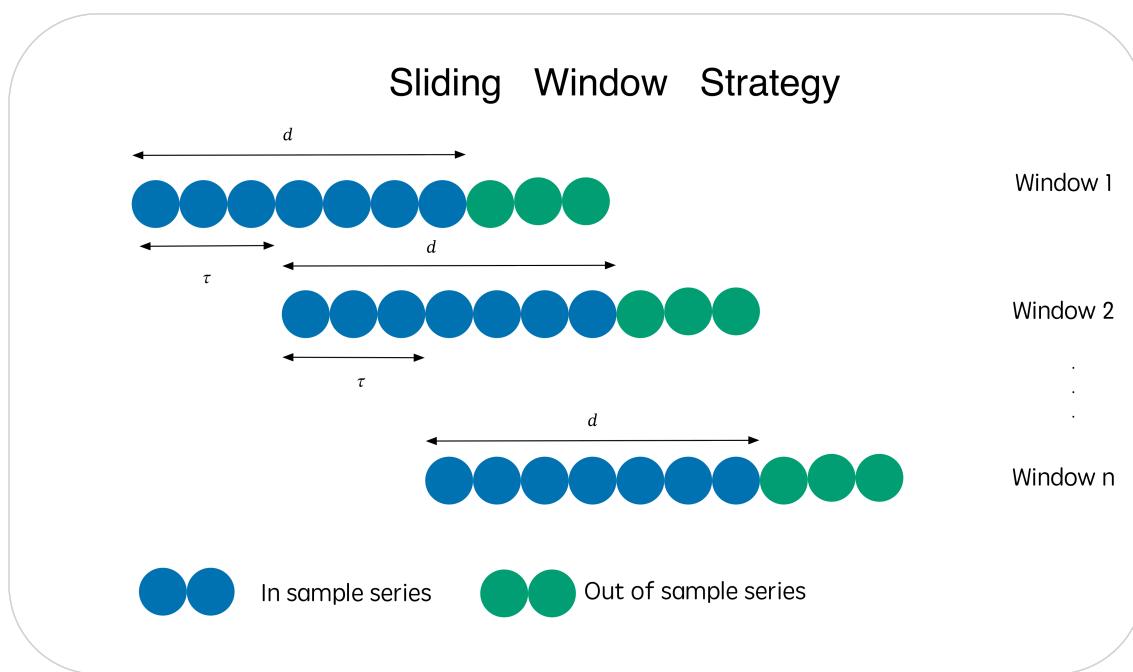


Figure 4. Sliding window strategy.

To verify the robustness of the chosen window size, we compared the TDA norm series under different settings (15, 21, 30, and 50 trading days). As shown in Figure 5, shorter windows (15 days) introduce excessive noise and fragmented fluctuations, while longer windows (30 or 50 days) over-smooth the dynamics and obscure significant events. By contrast, the 21-day window (corresponding to one trading month) achieves a balance between sensitivity and stability, allowing for the detection of market regime shifts with clear topological signatures.

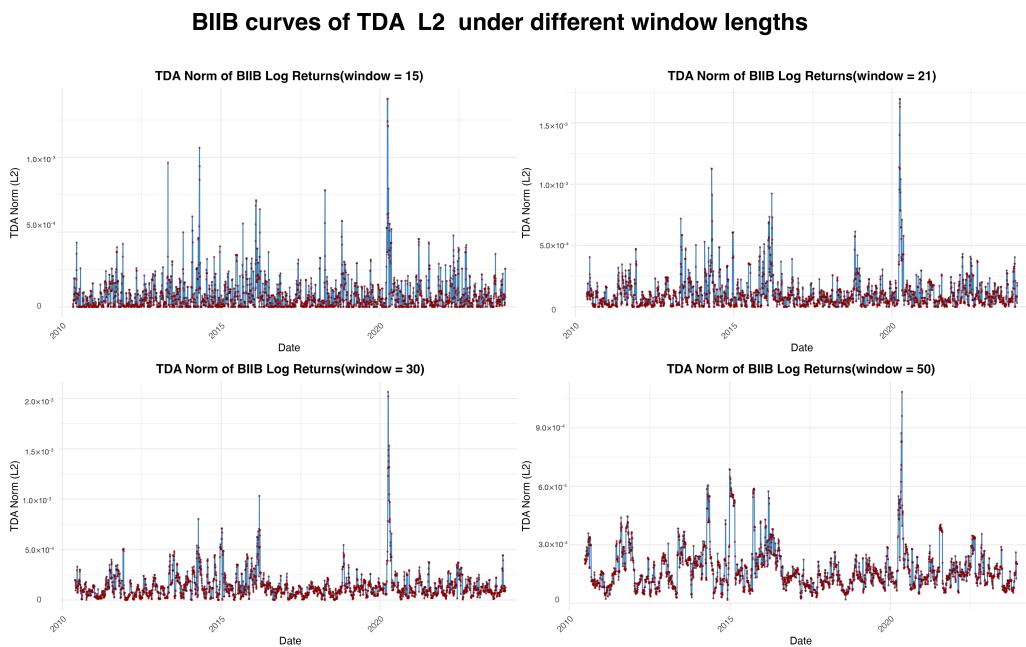


Figure 5. BIIB curves of TDA L_2 under different window lengths.

4.2. Time Delay Embedding

A dynamical system is a mathematical framework used to model phenomena that evolve over time. It defines a set of state transition rules to describe the evolution of the system, revealing its intrinsic relationships and mechanisms, and is often expressed using differential equations [95]. Therefore, constructing equations that describe how states change over time is crucial for understanding its underlying mechanisms. However, constructing these equations is particularly challenging in the absence of prior knowledge about the underlying distribution of the time series data. To address this issue, researchers have drawn inspiration from the study of attractors [96]. In the theory of dynamical systems, an attractor refers to a set of numerical states or trajectories toward which the system evolves over time. It represents a stable state or recurring pattern that the system tends to evolve toward, regardless of initial conditions [97]. Attractors play a key role in time series analysis, as they help in understanding the long-term behavior of the system and can be used for dimensionality reduction, pattern recognition, and predictive analysis. By constructing attractors, valuable information can be extracted from time series data, leading to a better understanding and improved modeling of complex systems [55].

A widely recognized method for reconstructing the state space of a dynamical system is the construction of a quasi-attractor using Takens' time delay embedding theorem [96]. Let \mathcal{M}_0 denote the manifold corresponding to the original dynamical system that generates the observed time series data. According to Takens' theorem, there exists a smooth map, $\Psi: \mathcal{M}_0 \rightarrow \mathcal{M}'$, where \mathcal{M}' is an m -dimensional embedding space (typically a Euclidean space). Furthermore, \mathcal{M}_0 and \mathcal{M}' are homeomorphic, meaning they are topologically equivalent. This embedding is guaranteed provided that $m > 2d_0 + 1$, where d_0 is the box-counting dimension of the attractor in \mathcal{M}_0 . This condition means that if we choose a sufficiently high embedding dimension m , then time delay embedding can preserve the topological structure of the dynamical system without losing important system information.

We adopted a three-dimensional embedding dimension for the following reasons. First, according to Takens' embedding theorem, given a dynamical system, an appropriate embedding dimension m allows the reconstruction of the system's phase space. Theoretically, this requires $m > 2d_0 + 1$, where d_0 denotes the fractal dimension of the underlying system [96]. Second, when the embedding dimension is too low (e.g., $m = 2$),

the reconstruction fails to capture complex structures, leading to the loss of periodic or topological features. Conversely, when the embedding dimension is too high (e.g., $m > 5$), the point cloud becomes sparse, the computational complexity increases exponentially, and noise effects become more pronounced. Third, financial return series often exhibit low-dimensional nonlinear dependencies, for which a three-dimensional embedding is sufficient to reveal transient yet stable topological structures. For these reasons, $m = 3$ has become one of the most commonly employed embedding dimensions in TDA applications to financial data [34,94].

The method applied in this paper first employs sliding window sampling, followed by time delay embedding, to better preserve both the local and global structures of time series and enhance the effectiveness of TDA. The sliding window is used for extracting local structures, smoothing the data, and reducing the impact of noise. Time delay embedding is applied for global dynamic reconstruction, providing a clearer representation of the overall system's evolution. After that, we perform TDA and obtain two metrics related to the persistence landscape: L_1 and L_2 . The entire data processing workflow is shown in Figure 6.

Data Analysis Framework

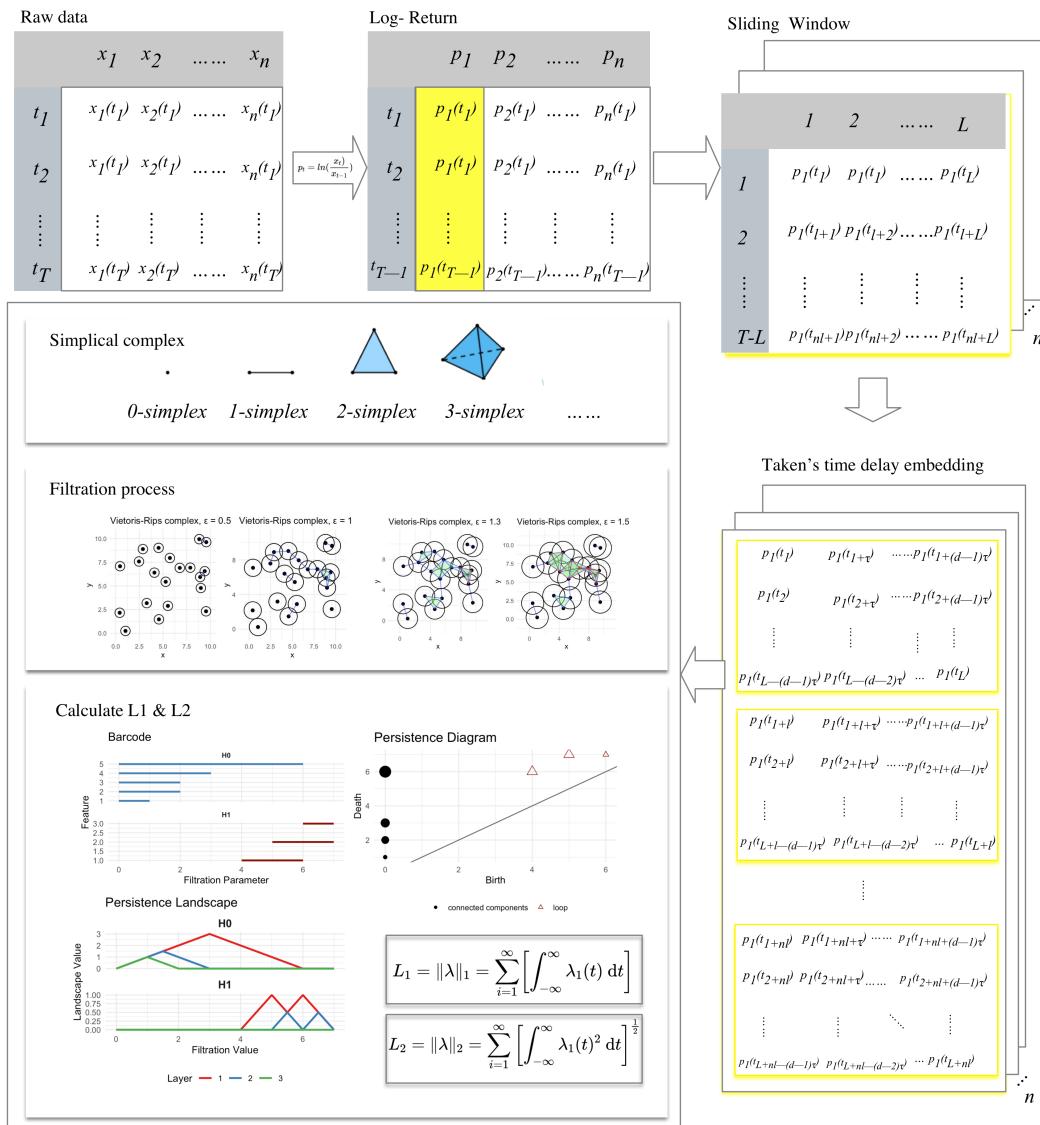


Figure 6. Data analysis framework.

5. Data Preprocess and TDA Pipeline

Our primary objective is to identify features linked to abrupt structural changes in time series data, with a particular emphasis on topological features. First, a TDA-based norm was introduced as a filtering technique, and subsequently, an optimal investment portfolio was constructed from the filtered asset categories [63]. Sokerin et al. [64] demonstrated that this approach outperforms conventional methods and serves as an effective tool for portfolio selection.

5.1. Data Preprocessing

We collected closing price data for 26 stocks across 3423 trading days from Yahoo Finance's historical stock market database, covering the period from March 2011 to December 2023. The dataset spans four regions—United States, China, Japan, and Australia—and covers seven industries: technology, industrial sector, consumer discretionary, finance, energy, healthcare, and telecommunications. The detailed information on the stock data sources is provided in Table A1.

In financial time series analysis, detrending and deseasonalizing are common preprocessing steps. These steps help stabilize the data, ensuring that models can more accurately capture the underlying patterns. During the preprocessing stage of multivariate time series data, this study first applies a logarithmic transformation to normalize price fluctuations on a logarithmic scale. Then, the *auto.arima()* function in the R programming language automatically selects the optimal orders for the autoregressive (AR) term, differencing (I), and the moving average (MA) term, while also removing overall and seasonal trends (Figure 7). Finally, the data undergoes an Augmented Dickey–Fuller (ADF) test to verify stationarity. After data preprocessing, we apply the previously described method for TDA feature extraction.

Decompositions of time series data(XLF)

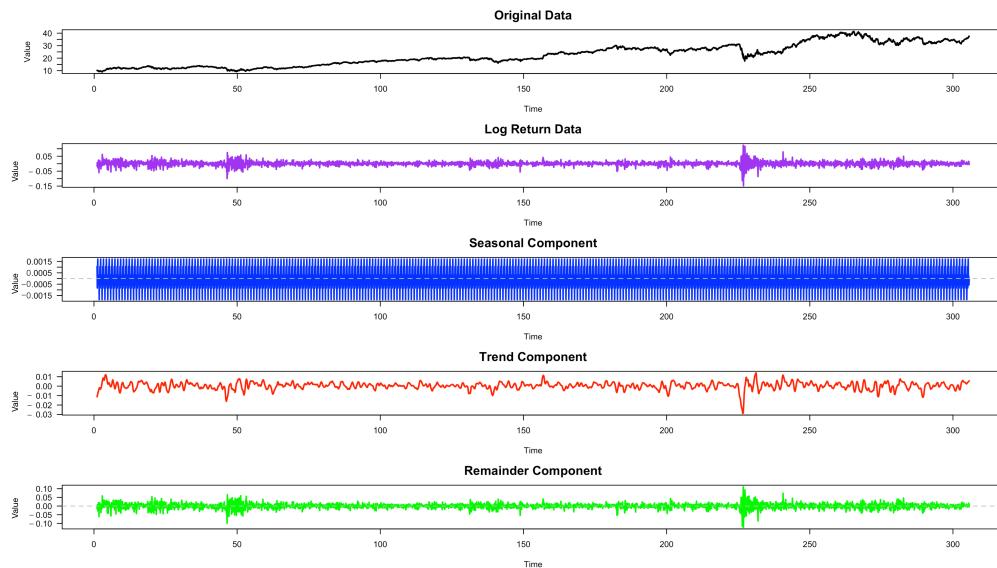


Figure 7. Decomposition of time series data (XLF).

5.2. TDA Pipeline

In financial time series analysis, two low-dimensional homology groups, H_0 and H_1 , are typically used to capture key structural changes in the market [34]. H_0 assesses market connectivity and helps identify isolated asset groups during periods of extreme volatility in financial markets. H_1 identifies risk cycles and anomalous volatility patterns in financial networks [98]. To effectively detect financial extreme events, this study focuses on extracting

H_1 features as the primary research target. In computing the L_p norm, we consider homology up to dimension 1 (maxdimension = 1) to capture both 0-dimensional and 1-dimensional topological features. However, as the first-layer norm in 0-dimensional homology remains constant due to persistent connected components, we focus exclusively on variations in L_1 and L_2 norms within the 1-dimensional homology class. This selection provides more meaningful insights into the topological changes in financial market structures.

In conclusion, this study introduces a novel CPD approach that integrates topological feature extraction from time series data. In this approach, the time series data is transformed into point cloud data, and the topological features, L_1 and L_2 , are extracted using persistent homology. Then, the L_1 and L_2 values from the multivariate time series are aggregated, and a fixed threshold is applied to identify change points. In summary, the process of this approach is composed of the following steps:

- Step 1: We apply the sliding window strategy to resample the time series data. Refer to Section 4.1 for the rationale. Subsequently, a time delay embedding algorithm is applied to the resampled data, using a time delay of one day. To simplify calculations and enhance visualization, we represent the transformed data as a 3D point cloud dataset. Following these transformations, we obtain a sequence of point cloud representations of the original time series.
- Step 2: We use the point cloud dataset to construct the Vietoris–Rips complex and then compute the persistence diagram and barcode representation of the extracted topological features of H_1 .
- Step 3: The persistence diagrams are then transformed into persistence landscapes, from which we compute the L_1 and L_2 norms.

After performing all of the calculations above, we need to find the implications of the results in the financial domain. The analysis of time series data typically focuses on short-term volatility and long-term fluctuations, each corresponding to different market dynamics and influencing factors. If the L_1 and L_2 values of the entire dataset are analyzed together, periods of extreme fluctuations will aggregate, making the analysis more complex. Moreover, financial volatility detection lacks a ground truth dataset. If treated as a binary classification problem, the magnitude of market crashes cannot be properly assessed. As a result, previous evaluations of CPD effectiveness have often been a mix of qualitative judgment and quantitative assessment. In this study, L_1 and L_2 are used as quantitative signals for detecting financial volatility, with specific thresholds set to directly determine the exact timing of financial CPD events. This significantly enhances the practical applicability of financial volatility detection.

The L_1 -norm is particularly sensitive to sparse signals and can be interpreted as the cumulative short-term fluctuation amplitude. It works better when evaluating the total impact of multiple small fluctuations [99]. The L_2 -norm captures the overall intensity or energy of fluctuations and is well suited for characterizing the impact of long-term drastic changes [100]. Therefore, this study uses L_1 and L_2 from persistence landscapes to represent short-term fluctuations and long-term fluctuations, respectively.

6. Fluctuations and Investment Analysis

We applied the TDA method to compute the L_1 and L_2 norms for 26 stocks over a 12-year period. By averaging the multidimensional L_1 and L_2 sequence data, we generated a fluctuation chart illustrating the stock price dynamics alongside the L_1 and L_2 values over the 12-year time frame (Figure 8). It can be observed that stock prices exhibit a long-term upward trend but experience sharp fluctuations during specific periods. The log volatility shows relatively small fluctuations overall but exhibits significant spikes at certain moments. The volatility indicators represented by L_1 and L_2 show a notable increase during

periods of sharp stock price fluctuations, with L_2 amplifying to a significantly greater extent than L_1 .

6.1. Long-Term and Short-Term Fluctuation Analysis

We examined the relationship between L_1 and L_2 norms and financial market change points across different regions, with a focus on regional differences in long-term and short-term financial fluctuations. Additionally, we investigated the relationship between L_1 and L_2 norms and financial market change points across different sectors, emphasizing variations in long-term and short-term financial fluctuations among industry groups (Tables 2 and 3).

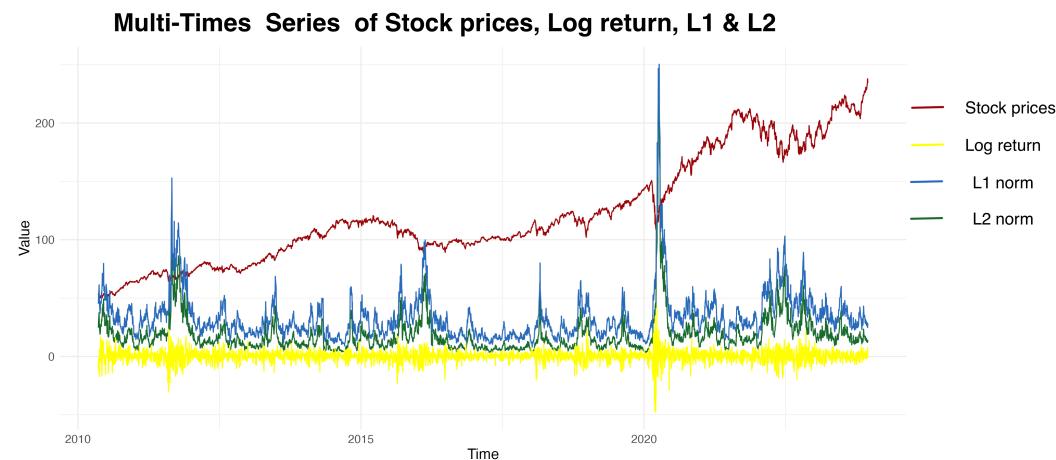


Figure 8. Multi-times series of stock prices (red line), log return (yellow line), L_1 (green line) and L_2 (blue line).

Table 2. Results by regions of L_1 & L_2 .

Region	Stock Tickers	L1 Mean	L2 Mean
USA	CRM, ORCL, QCOM, AAL, CMG, COST, EBAY, KO, NKE, PEP, BRK-B, V, XLF, GLD, COP, USO, AMGN, BIIB, GILD, GSK, MRK, CMCSA	2.58×10^{-6}	3.80×10^{-5}
China	BIDU, TSM	8.30×10^{-6}	1.01×10^{-4}
Australia	BHP	1.48×10^{-5}	1.52×10^{-4}
Japan	TM	8.60×10^{-6}	1.01×10^{-4}

Table 3. Results by sectors of L_1 & L_2 .

Sectors	Stock Tickers	L1 Mean	L2 Mean
Technology	BIDU, CRM, ORCL, QCOM, TSM	8.30×10^{-6}	1.01×10^{-4}
Industry	AAL, TM	5.88×10^{-6}	7.27×10^{-5}
Consumer good	CMG, COST, EBAY, KO, NKE, PEP	6.22×10^{-6}	7.86×10^{-5}
Finance	BRK-B, V, XLF, GLD	9.42×10^{-6}	1.06×10^{-4}
Energy	COP, USO, BHP	6.89×10^{-6}	8.52×10^{-5}
Health	AMGN, BIIB, GILD, GSK, MRK	4.78×10^{-6}	6.64×10^{-5}
Telecommunication	CMCSA	2.56×10^{-6}	4.09×10^{-5}

Our analysis reveals that the US market, characterized by capital concentration and diversification, involves the largest number of stocks, yet exhibits relatively low mean values for L_1 and L_2 . Compared to other regions, the US market exhibits greater stability,

as indicated by its lower L_1 and L_2 mean values. The Chinese market exhibits higher short-term price volatility, as reflected in its elevated L_1 mean. Additionally, its long-term uncertainty, influenced by macroeconomic policies, is highlighted by a relatively high L_2 mean. Australia exhibits the highest L_2 mean among all regions, indicating significant long-term volatility. This aligns with its resource-oriented economic structure, which contributes to a more concentrated and fluctuating market. The Japanese market demonstrates moderate long-term volatility, with an L_2 mean higher than that of the US but lower than Australia's. This pattern is likely influenced by Japan's reliance on core markets and economic concentration.

An analysis of L_1 and L_2 values across different industry sectors reveals that the financial sector has the highest mean L_1 , indicating that stocks within this sector exhibit the most pronounced short-term volatility. This aligns with the financial industry's inherent sensitivity to interest rates, monetary policy, and market sentiment. Additionally, the relatively high L_2 values suggest a significant degree of long-term uncertainty, likely driven by regulatory changes, economic cycles, and financial crises. Similarly, the technology sector also exhibits high L_1 and L_2 values. Due to the influence of innovation cycles and macroeconomic factors, stock prices in this sector experience notable short-term fluctuations, while long-term uncertainty remains elevated. In contrast, the telecommunications sector has the lowest L_1 and L_2 values, indicating that its short-term and long-term volatility are the lowest among all sectors. The sector's stability can be attributed to sustained technological advancements and consistent demand, leading to a relatively fixed profitability model with limited short-term fluctuations. A similar pattern is observed in the pharmaceutical sector, which typically benefits from stable cash flows and strong resilience to economic cycles. As a result, this sector is less likely to experience significant short-term fluctuations, contributing to its relatively low volatility in both the short and long term. The energy sector, on the other hand, is significantly influenced by oil price fluctuations, geopolitical risks, and changes in supply and demand. Consequently, its short-term volatility and long-term uncertainty remain at moderate levels. Meanwhile, the industrial and consumer goods sectors exhibit similar L_1 and L_2 values, suggesting a balanced risk profile. This stability can be attributed to relatively steady market demand and diversified business models.

6.2. Investment Advice

From an investment perspective, investors seeking stability may prioritize the health-care and telecommunications sectors. In contrast, those willing to take on higher risks for potential returns may find the technology and financial sectors more attractive, as their elevated volatility could present greater profit opportunities. This classification framework offers a structured way for investors to align their risk tolerance with appropriate investment options. Conservative investors may prefer stable or long-term sensitive portfolios, while those seeking higher returns and market opportunities may consider short-term or high-volatility portfolios (Table 4).

Table 4. Investment advice for different risk preferences.

Fluctuation Dimension	Short-Term Low Volatility (L1-L)	Short-Term High Volatility (L1-H)
Long-term low volatility (L2-L)	<i>Stable portfolio:</i> CMCSA, AMGN, BIIB, GILD, GSK, MRK	<i>Short-term sensitive portfolio:</i> CRM, ORCL, QCOM, BHP, BIDU, TSM, TM
Long-term high volatility (L2-H)	<i>Long-term sensitive portfolio:</i> COP, USO, GLD, BRK-B, V, XLF	<i>High-volatility portfolio:</i> AAL, CMG, COST, EBAY, KO, NKE, PEP

It provides a risk-based classification of stocks derived from TDA indicators. By combining short-term (L_1) and long-term (L_2) fluctuation dimensions, we construct a two-dimensional risk matrix that naturally divides assets into four categories: stable, short-term sensitive, long-term sensitive, and highly volatile. This categorization offers intuitive investment guidance for different risk preferences. Even without portfolio backtesting, the table demonstrates the practical interpretability of TDA indicators, linking topological features of time series to real-world investment decisions. Future work could extend this framework with systematic backtesting and asset allocation strategies to further validate its performance.

6.3. Back-Testing

In this section, we treat the detected change points as signals of market state transitions, construct a corresponding position-adjustment strategy, and compare its performance against the standard buy-and-hold benchmark. When the TDA index (L_1 or L_2) exceeds the 95th percentile threshold, it is interpreted as elevated market risk and the strategy shifts to a crash; otherwise, the portfolio holds the S&P 500, thereby maintaining returns while reducing large drawdowns.

As illustrated in Figure 9, both TDA-based strategies (L_1 and L_2) provide effective signals for portfolio allocation. While the buy-and-hold benchmark achieves the highest raw cumulative return, it suffers from large drawdowns during crisis periods (e.g., the COVID-19 shock in 2020). In contrast, the TDA-based strategies exhibit substantially reduced drawdowns, indicating their ability to capture structural changes in market topology and enhance downside protection. Moreover, the L_1 and L_2 -based strategies display complementary features: L_1 is more sensitive to early fluctuations, while L_2 is more robust in capturing persistent systemic disruptions. These results demonstrate that TDA indices can serve as valuable risk control tools in financial markets, offering empirical support for their practical applicability in investment strategies.

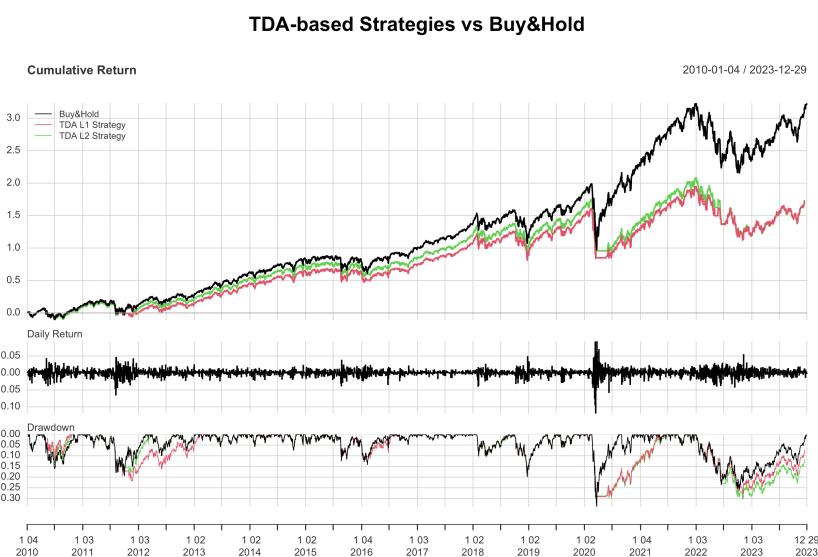


Figure 9. TDA-based strategies vs. buy and hold.

7. Change Point Detection for Financial Extreme Events

We frame CPD as a classification problem, where the objective is to distinguish between 'change points' and 'non-change points'. In this setting, since change points typically occur far less frequently than non-change points, the resulting dataset is inherently imbalanced. If stock price fluctuations are modeled as a dynamic system, then instances where the amplitude surpasses a predefined threshold can be classified as extreme events within

this system. Within the TDA framework, we posit that higher L_p values correspond to greater dispersion of data points in the space, indicating increased volatility in the dataset. Therefore, we establish a threshold on L_p to identify change points. The 98% threshold is a widely used statistical method for identifying extreme events in time series analysis [101]. Appendix D (Threshold Selection) provides a clear description of how this cutoff was determined and why it is appropriate. The change points output by L_1 and L_2 serve as indicators for financial extreme events. Our findings indicate that the change points identified by L_1 and L_2 are consistent, primarily occurring in 2011, 2016, 2020, and 2022. This demonstrates how TDA-based metrics (L_1 and L_2) can effectively capture financial market volatility and detect change points. The peaks align with known financial crises, validating the method's practical use in risk assessment and economic forecasting.

7.1. Financial Extreme Events

From an economic perspective, spikes in the TDA indicators signal abrupt shifts in the geometry of the return state space, implying stronger co-movement, correlation clustering, and cross-asset regime transitions. During systemic episodes, compressed risk premia and temporarily impaired diversification produce market-wide synchronization—conditions under which such topological breaks are expected. Accordingly, L_2 (short-term) operates as a timely shock detector, useful for rapid de-risking and liquidity management at the onset of events, whereas L_1 (long-term) serves as a regime persistence gauge, indicating how long risk remains elevated and when to re-risk. For regulators, these signals can complement macro-prudential dashboards as early-warning indicators of system-wide stress; for asset managers, they inform dynamic position sizing, drawdown control, and stress hedging. The close alignment of TDA signals with the European sovereign debt crisis (2011), Brexit (2016), the COVID-19 outbreak (2020), and the Russia–Ukraine energy shock (2022) suggests that TDA captures systemic, market-wide disturbances rather than idiosyncratic noise, thereby improving the timeliness and reliability of regime identification beyond volatility-only measures (as shown in Figure 10).

Detection of Representative Consensus Events Using TDA-based Indicators

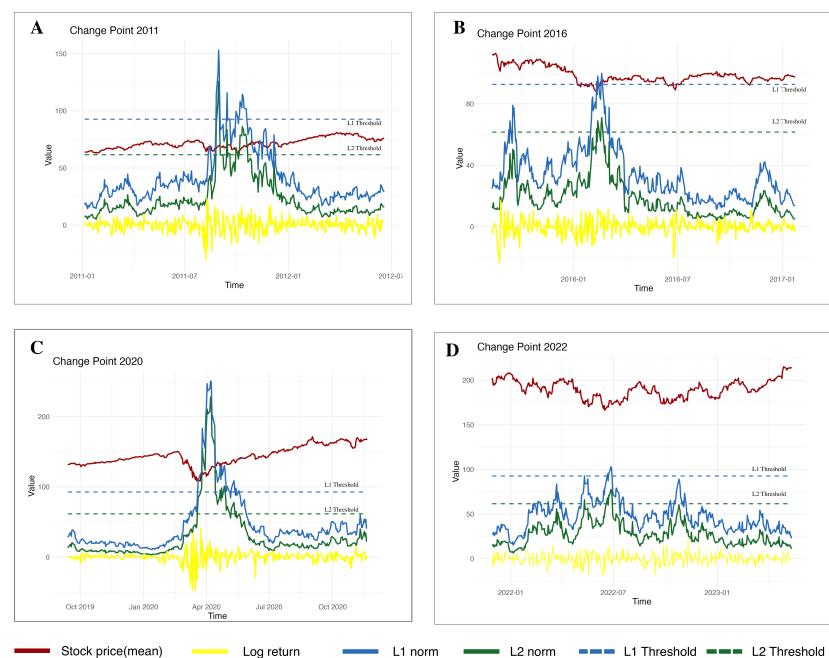


Figure 10. Detection of representative consensus events using TDA-based indicators.

7.1.1. European Debt Crisis in 2011

Among the extreme financial events identified between 2011 and 2023, the first event captured by our TDA-based change point detection method was the market rebound following the late-stage resolution of the European debt crisis in 2011 (Figure 10A). In this context, the short-term fluctuation measure L_1 captures the market's immediate reaction to the European debt crisis, whereas the long-term fluctuation measure L_2 highlights the prolonged influence of expansionary monetary policies on asset prices. The European debt crisis exerted significant pressure on global financial markets, primarily through risk spillovers from the European banking sector, causing heightened short-term volatility across global markets. The United States reintroduced expansionary monetary policies during this period, triggering sustained adjustments in asset valuations. As the European debt crisis gradually subsided, the U.S. stock market experienced a swift recovery. The S&P 500 index embarked on a prolonged bullish trend in 2011, with technology stocks and consumer goods serving as the primary growth drivers. The European debt crisis induced short-term shocks in the global economy and the U.S. stock market, but it also reinforced the U.S. market's role as a safe-haven asset and underscored global capital preferences. A crucial takeaway from this crisis is that the globalization of financial markets facilitates the rapid transmission of regional crises, yet effective policy interventions can play a pivotal role in mitigating financial shocks and fostering market recovery.

7.1.2. Brexit in 2016

The second major financial event identified by our method was the 2016 Brexit referendum and its subsequent global economic adjustments (Figure 10B). The short-term volatility captured by L_1 indicated immediate market panic following the referendum outcome, leading to sharp declines in U.S. stock markets, a surge in safe-haven assets, and a shift in global capital towards low-risk instruments. The long-term fluctuations reflected by L_2 suggest broader structural adjustments, particularly in global trade and currency dynamics.

The Brexit-induced depreciation of the British pound contributed to a stronger U.S. dollar, reducing the competitiveness of U.S. export-oriented firms. Additionally, the uncertainty surrounding Brexit was expected to weaken European demand for U.S. goods and services. However, our findings indicate that while short-term volatility was significant, long-term market adjustments were more nuanced and influenced by multiple factors, including global monetary policies and China's economic restructuring. Compared to the 2011 European debt crisis, Brexit's market impact was more concentrated in currency markets and global trade flows, rather than financial sector stability. Moreover, the market's recovery after Brexit was relatively swift, aided by central bank interventions and policy responses that stabilized investor sentiment. This suggests that while geopolitical risks such as Brexit can trigger short-term turbulence, their long-term financial impact is often moderated by structural economic adjustments and policy interventions.

7.1.3. COVID-19 Pandemic in 2020

The third major financial event identified by our method was the economic shock and subsequent recovery following the outbreak of the COVID-19 pandemic in 2020, alongside the surge in inflation that followed (Figure 10C). In the short term, the market exhibited sharp fluctuations (L_1) driven by inflation concerns and shifting policy expectations. In the long term, persistent uncertainty regarding asset valuations and economic recovery trajectories was captured by L_2 . The global economy was bolstered by widespread vaccination efforts, fiscal stimulus, and loose monetary policy, leading to a global GDP expansion of 5.9%. This strong recovery fueled a broad-based stock market rally, with the S&P 500 index

rising to 26.9% for the year. However, the recovery trajectory varied across industries: technology and healthcare stocks benefited from pandemic-driven demand, whereas cyclical sectors like energy and industrials gained traction during the reopening phase, contributing to heightened market volatility. Meanwhile, supply chain disruptions, labor shortages, and expansionary fiscal policies drove inflation to persistently high levels, with the U.S. CPI exceeding 6% year-over-year (the highest in four decades). As recovery progressed, inflation concerns and expectations of monetary policy tightening, alongside intermittent pandemic resurgences, intensified market volatility. The spillover effects of rising inflation and U.S. monetary policy shifts led to increased synchronization across global stock markets, driven by capital flows, inflation expectations, and risk sentiment.

7.1.4. Energy Crisis in 2022

The fourth major financial event identified by our method was the global energy crisis triggered by the Russia–Ukraine war, compounded by simultaneous monetary tightening. The geopolitical tensions surrounding the conflict severely disrupted global supply chains, triggering extreme volatility in energy markets. The war-induced supply shock caused European natural gas prices to surge, intensifying inflationary pressures and fueling expectations of further monetary tightening. The short-term fluctuation measure (L_1) captured the market's immediate reaction to geopolitical risks, including sharp price swings in commodities and safe-haven assets. Meanwhile, the long-term fluctuation measure (L_2) reflected the prolonged impact of monetary tightening on global asset valuations, as the combination of higher interest rates and geopolitical uncertainty reshaped investor sentiment and capital flows (Figure 10D).

The near-simultaneous occurrence of the Fed's rate hikes and the Russia–Ukraine war amplified market turbulence. The Fed's aggressive tightening cycle drained market liquidity, prompting a flight from riskier assets, while war-induced energy price spikes reinforced inflationary expectations, leading to further policy uncertainty and volatility. Additionally, the energy crisis underscored the increasing inter-connectivity of global financial markets. Rising oil and natural gas prices not only impacted energy-dependent economies but also fueled inflation expectations worldwide, influencing central bank policies and investor sentiment across multiple asset classes.

7.2. Benchmark Test

In this part, we compare our method with five CPD baselines. The baselines comprise three univariate methods, rolling standard deviation [102], PELT [18,103] and a hidden Markov model (Viterbi) [104], and two multivariate methods: E.divisive and E.agglo [46].

As illustrated in Figure 11, the panels present the CPD results for each of the 26 selected assets, with the horizontal axis denoting time (year) and the vertical axis representing asset codes. Each marker corresponds to a detected change point for a given asset. For the Rolling Std method, bubble and color reflect the relative breadth of detection events across assets, while the PELT (MBIC) and HMM (Viterbi) methods identify multiple asset-level breakpoints with higher temporal density. This figure provides a comparative overview of how different univariate detection methods capture market regime shifts at the individual asset level over the sample period. The rolling std captures short-term fluctuations surrounding extreme events but is prone to noise; the PELT (MBIC) approach is highly sensitive to structural breaks but tends to over-segment the data; and the HMM-based method provides a more stable identification of prolonged high-volatility regimes, successfully detecting all four major crises examined.

CPD for 26 Stocks using univariate methods

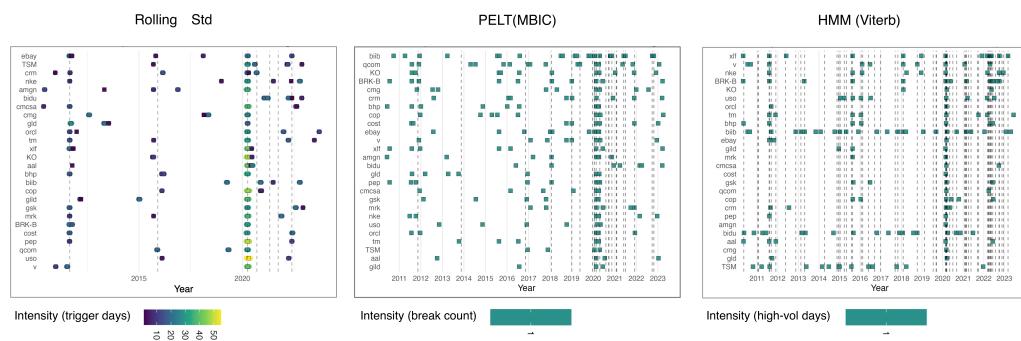


Figure 11. CPD for individual assets using univariate methods.

Next, adjacent detections are merged into contiguous intervals, and event breadth is measured by the number of assets that register a detection; hollow white circles mark the midpoint of each interval (Figure 12). On top of these “global event” bands, we superimpose the outputs of two multivariate methods as vertical markers—gold solid lines for E.aggro and blue dashed lines for E.divisive—to indicate their estimated global change points. Comparing Figures 12 in the manuscript, it is evident that the TDA-based L_1 and L_2 indicators exhibit clear advantages in detecting four globally influential consensus events.

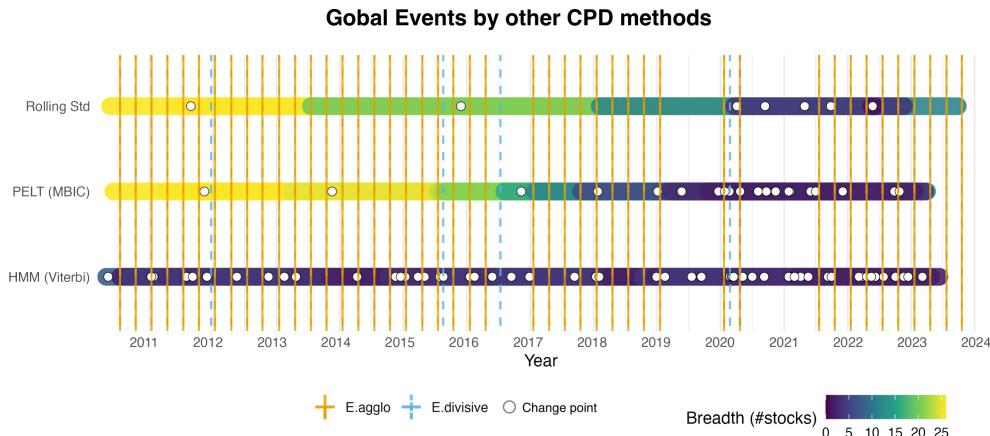


Figure 12. Global events detected by other CPD methods.

To benchmark the performance of different CPD methods, we systematically collected financial news from multiple global sources covering the period March 2011 to December 2023. We constructed news intensity scores by quantifying the frequency and salience of “extreme words” appearing in headlines and leads, which are widely regarded in the literature as reliable proxies for market sentiment shocks. Based on these scores, we established a Global Major Volatility Event Set (Appendix E), which provides an externally validated reference for identifying periods of heightened systemic stress and serves as a benchmark for evaluating the effectiveness of our TDA-based change point detection. To improve our CPD evaluation, we report precision, recall, and F1 because they capture complementary aspects of performance. The tolerance serves to delimit the period over which market states or events are evaluated, thereby ensuring that different methods are compared under consistent temporal conditions. $Precision = TP / (TP + FP)$ measures alert purity and penalizes false positives. It is useful when false alarms are costly. $Recall = TP / (TP + FN)$ measures sensitivity and penalizes missed events. It is critical when misses are costly. $F1 = 2PR / (P + R)$ is the harmonic mean of precision and recall and provides a single score when both error types of matter about equally. Here, TP counts detected change points within the tolerance window. FPs are

spurious detections. And FNs are undetected true changes. All three are threshold-dependent and ignore true negatives. Precision can be inflated at the expense of low recall. Recall can become high due to over-triggering, and F1 assumes equal costs and does not reflect negative-class stability.

This study hypothesizes that TDA methods based on persistent homology can more effectively capture the global structural changes associated with extreme events in financial markets. The empirical results demonstrate that the two TDA-derived indicators (L_1 and L_2) consistently achieve stable and relatively high F1 scores across different tolerance windows (Figure 13). Importantly, the performance is also reflected in balanced Precision (Figure 14) and Recall values (Figure 15). It is indicating that the TDA approach not only maintains accuracy in detecting true change points but also avoids excessive false positives. By contrast, traditional univariate methods (Rolling Std, PELT, HMM) and multivariate methods (E.divisive, E.aggro) tend to exhibit either lagged or fragmented detection, leading to imbalanced Precision-Recall trade-offs and overall low F1 performance. At the cross-asset level, the TDA approach exhibits superior robustness and consistency.

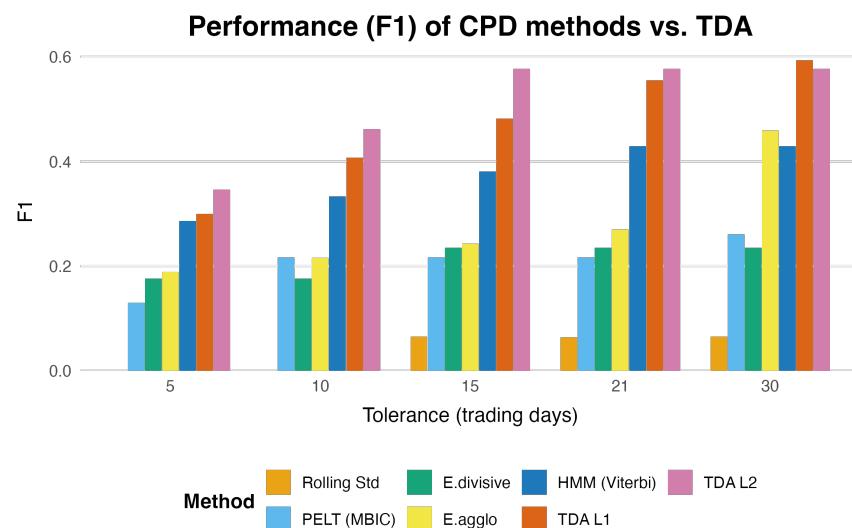


Figure 13. Comparison of other CPD methods with the TDA methods (f1).

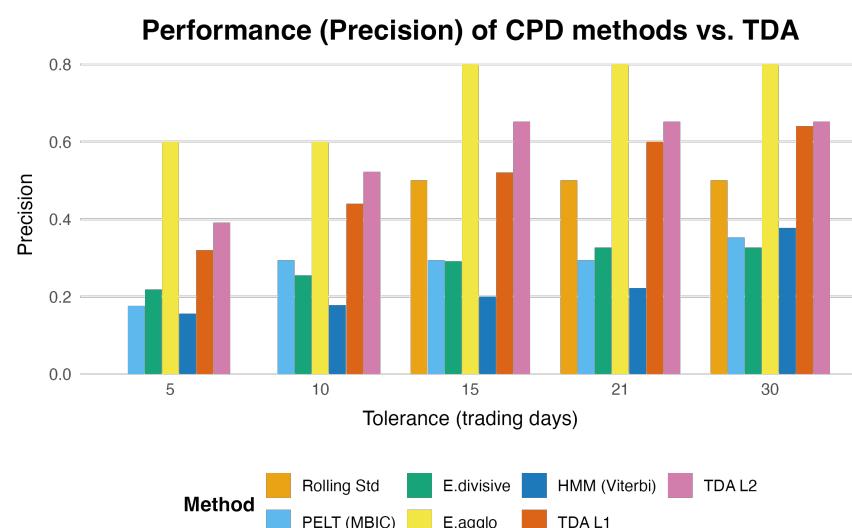


Figure 14. Comparison of other CPD methods with the TDA methods (precision).

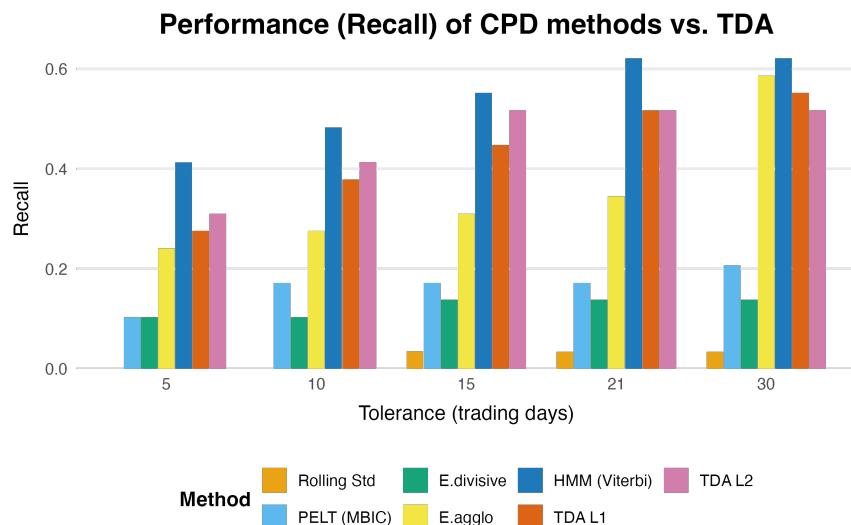


Figure 15. Comparison of other CPD methods with the TDA methods (Recall).

8. Conclusions

8.1. Summary of Findings

This study employs TDA to examine multivariate stock price time series, extracting the L_1 and L_2 indicators to quantify short-term and long-term market fluctuations across different regions and industries. Unlike traditional econometric models, TDA offers a non-parametric approach to capturing structural changes in market dynamics. Our analysis of regional volatility differences, based on TDA-derived L_1 and L_2 measures, reveals distinct market characteristics across different regions. The U.S. stock market exhibits lower historical volatility, with both L_1 and L_2 values at relatively low levels, making it a suitable option for conservative investors seeking stable asset allocation. This reflects the U.S. market's strong liquidity, diversified sector composition, and historically lower systemic risk. In contrast, Japan and Australia demonstrate significantly higher L_2 values, indicating greater long-term market fluctuations. These markets may provide higher risk-adjusted returns for investors with a greater risk appetite. Australia's elevated L_2 suggests that its stock market is highly sensitive to global commodity demand and macroeconomic trends, making it particularly relevant for investors focused on resource price cycles. For investors in Japan, our findings indicate that the country's higher long-term volatility is influenced by its dependence on core industries, such as the automotive sector. Thus, market participants should closely monitor global competitiveness trends and external demand shifts affecting Japan's major export industries. China's market structure, as reflected by its moderate L_1 and L_2 , suggests a balance between short-term stability and long-term policy-driven fluctuations. This market is well-suited for medium- to long-term investors who can navigate regulatory policies and macroeconomic trends.

Our analysis of inter-sector volatility differences, based on TDA-derived L_1 and L_2 measures, reveals distinct risk-return profiles across sectors. The technology sector exhibits the highest short-term volatility (L_1), driven by rapid innovation cycles and shifting market sentiment, leading to complex and unpredictable price movements. The financial industry shows substantial long-term volatility (L_2), as it is highly sensitive to macroeconomic conditions, interest rate policies, and global financial stability. Investors should be aware of cyclical risks and systemic shocks that can influence financial stocks. The telecommunications sector and the healthcare industry demonstrate lower L_1 and L_2 values, indicating relative stability. Telecommunications companies benefit from consistent demand and regulatory protection, while healthcare stocks are less correlated with economic cycles, offering

defensive investment opportunities. In contrast, the energy and consumer goods sectors display higher L_2 values, indicating sensitivity to external macroeconomic conditions, such as commodity price fluctuations and global supply chain dynamics. Energy stocks, in particular, exhibit sharp fluctuations in response to geopolitical events and resource demand shifts. Investors should tailor their strategies based on industry volatility characteristics. Risk-averse investors may prefer defensive sectors like telecommunications and healthcare, while those seeking higher returns could explore cyclical industries such as technology and energy. The application of L_1 and L_2 indicators provides a novel framework for optimizing industry allocation and risk management.

Furthermore, the volatility patterns reflected by L_1 and L_2 successfully identified four major financial events: the European debt crisis in 2011, Brexit in 2016, the COVID-19 market shock in 2020, and the energy crisis triggered by the Russia–Ukraine war in 2022. The ability of TDA to detect these transitions demonstrates its potential as an effective tool for analyzing financial market stability and identifying significant economic turning points. Compared to previous studies, this research not only expands the scope of financial extreme event identification but also demonstrates that TDA can detect not only systemic financial crises but also market turbulence driven by geopolitical and policy factors (Table 5). Furthermore, it validates the effectiveness of topological data analysis in identifying financial market fluctuations and highlights how financial market globalization accelerates the transmission of regional crises. These findings provide important insights for investors, policymakers, and risk managers.

Table 5. Comparison of detected financial crises across studies (✓ indicates a detected extreme event, and - denotes a time interval not covered).

	Dotcom Crash 2000	Financial Crisis 2008	European Debt 2011	Brexit 2016	COVID-19 2020	Russia-Ukraine War 2022
Gidea [59]	-	✓	-	-	-	-
Gidea and Katz [34]	✓	✓	-	-	-	-
Aguilar and Ensor [61]	-	-	-	-	✓	-
Katz and Biem [70]	-	✓	-	-	-	-
Yen and Cheong [71]	-	-	-	-	✓	-
Guo et al. [66]	-	-	-	-	✓	-
Ismail et al. [69]	✓	-	-	-	-	-
Rai et al. [74]	-	✓	-	-	✓	-
Our study	-	-	✓	✓	✓	✓

8.2. Economic and Social Implications

Our study demonstrates that the TDA-based CPD approach provides a unified and robust temporal signal for detecting structural regime shifts in financial markets. The close alignment of the detected change points with globally recognized “consensus events”, including the European debt crisis, Brexit, the COVID-19 outbreak, and the Russia–Ukraine energy crisis, as well as the consistently superior F1 performance across different tolerance windows, validates the effectiveness of this method in identifying systemic disruptions. From an economic perspective, the proposed approach can provide regulatory authorities with early warnings of systemic risk and a basis for countercyclical policy interventions, thereby reducing coordination costs during crises. For asset managers, these signals enable dynamic portfolio rebalancing and risk exposure management, improving both portfolio resilience and capital allocation efficiency. For long-term institutional investors such as pension funds and insurance companies, the approach also has positive implications for tail risk control and the safeguarding of long-term returns. At the same time, we acknowledge that practical implementation may face challenges such as procyclicality,

parameter sensitivity, and computational demands. Future research will therefore explore multi-indicator integration, operational buffer mechanisms, and robustness checks to ensure that the method can maximize both its economic and social benefits in practice.

8.3. Limitation and Future Directions

This study innovatively applies TDA to systematically investigate both short-term and long-term stock market fluctuations across regions and industries, offering a novel perspective beyond conventional statistical or econometric approaches. Moreover, our method successfully identified four major stock market events over the past twelve years, aligning with key turning points documented in existing financial research. These findings strongly validate the effectiveness of TDA in detecting structural changes and change points in financial markets.

While TDA offers a powerful lens for identifying market fluctuations, several limitations remain. First, the intricate cross-sectional and temporal interdependencies in financial data are underexplored here. In particular, we do not fully model linkages between individual stocks and their network dynamics. Second, computational complexity is non-trivial. Computing persistent homology on sliding window point clouds scales rapidly with sample size and ambient dimension. Memory constraints restrict window length and multivariate extensions. Therefore, approximate schemes are left for future work. Third, despite stability theorems, TDA features can exhibit noise sensitivity in practice—short bars may reflect microstructure noise or parameter choices. More systematic denoising, bootstrapping, and multi-parameter sensitivity analyses are needed. Finally, ethical and governance considerations arise for market prediction. Risks include data-snooping/overfitting and unequal access to predictive signals, as well as pro-cyclical feedback or manipulative misuse. Future research should integrate topological concepts with complex-network modeling and machine learning methods to uncover mechanisms behind change points, alongside transparent validation protocols, model risk controls, and compliance safeguards to enhance interpretability and responsible deployment.

Author Contributions: J.Y. conceived the study, collected the data, and finalized the manuscript; J.L. and J.W. contributed to the interpretation of the data analysis; M.Y. and X.W. reviewed the manuscript and provided constructive suggestions. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the High-Level Scientific Research Foundation of Hebei Province, “New Topology Based on GLMY Theory and Its Applications” project Sponsored by Shanghai Institute for Mathematics and Interdisciplinary Sciences and National Natural Science Foundation of China under Grant Nos. 72102220 and 72192843.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available in Yahoo Finance. The proposed method is implemented in R. Code and is available at <https://github.com/Janeyao/Detecting-Change-Points-in-Multivariate-Stock-prices-Using-Topology-data-analysis> (accessed on 2 October 2025).

Acknowledgments: We gratefully acknowledge the Beijing Key Laboratory of Topological Statistics and Applications for Complex Systems for the computational resources provided by them.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Description of Data

Table A1. This table shows the basic information of 26 companies.

	Company	Ticker	Region	Sector
1	American Airlines Group Inc.	AAL	USA	Industry
2	Amgen Inc.	AMGN	USA	Health
3	BHP Group Limited	BHP	Australia	Energy
4	Baidu, Inc.	BIDU	China	Technology
5	Biogen Inc.	BIIB	USA	Health
6	Berkshire Hathaway Inc Class B	BRK-B	USA	Finance
7	Chipotle Mexican Grill, Inc.	CMG	USA	Cousumer good
8	Comcast Corporation	CMCSA	USA	Telecommunication
9	ConocoPhillips	COP	USA	Energy
10	Costco Wholesale Corporation	COST	USA	Consumer good
11	Salesforce, Inc.	CRM	USA	Technology
12	eBay Inc.	EBAY	USA	Consumer good
13	Gilead Sciences, Inc.	GILD	USA	Health
14	SPDR Gold Shares	GLD	USA	Finance
15	GSK plc	GSK	USA	Health
16	The Coca-Cola Company	KO	USA	Consumer good
17	Merck & Co., Inc.	MRK	USA	Health
18	NIKE, Inc.	NKE	USA	Consumer good
19	Oracle Corporation (ORCL)	ORCL	USA	Technology
20	Pepsi Co., Inc.	PEP	USA	Consumer good
21	QUALCOMM Incorporated	QCOM	USA	Technology
22	Toyota Motor Corporation	TM	Japan	Industry
23	Taiwan Semiconductor Manufacturing Company Limited	TSM	China	Technology
24	United States Oil Fund, LP	USO	USA	Energy
25	Visa Inc.	V	USA	Finance
26	The Financial Select Sector SPDR Fund	XLF	USA	Finance

Note : The selection was guided by both methodological and practical considerations. First, our study covers a relatively long time span (2011–2023), during which many stocks were listed or delisted. To ensure temporal consistency and avoid distortions caused by incomplete time series, we restricted the sample to firms with continuous data availability across the entire period. Second, we aimed to achieve broad market coverage by including stocks from multiple industries and geographical regions, to capture heterogeneous dynamics and avoid bias toward any single sector. Third, given the computational complexity of our topological data analysis (TDA) framework, we deliberately limited the number of stocks to a moderate size, striking a balance between representativeness and tractability.

Appendix B. L_p Norm Computation

This section primarily discusses the specific steps of L_p norm computation, the involved parameters, and its application in the topological analysis of financial markets at the algorithmic level. In the computation of the L_p norm, the *maxdimension* parameter in the *ripsDiag* function (R) specifies the maximum-homology dimension to be considered. When *maxdimension* = 1, both 0-dimensional and 1-dimensional homology classes of the point cloud data are computed. In the persistence landscape function, the *dimension* parameter defines which homology dimension is mapped into the persistence landscape representation. When *dimension* = 1, the 1-dimensional homology is converted into a persistence landscape. The *KK* parameter determines the topological feature layer being analyzed. When *KK* = 1, only the first-layer features are considered. The *p* parameter in the TDA norm function specifies the norm to be computed, where L_p norms quantify the persistence of topological features.

In the algorithm for calculating the L_p norm, *maxdimension* of *ripsDiag* function in R indicates the maximum homology dimension calculated. When *maxdimension* = 1, it indicates that the 0-dimensional and 1-dimensional homology of point cloud data are

calculated. The dimension in landscape function indicates the homology of the dimension to be converted into landscape. When $dimension = 1$, it indicates that the 1-dimensional homology is converted into landscape.

After computing the persistence landscape L_p norms, we analyze the results based on the following observations: The norm value of the first layer in the 0-dimensional homology remains constant and does not provide significant information. This phenomenon occurs because, in the persistence diagram, there is always a persistent 0-dimensional homology class that never vanishes. As a result, the norm reflected in the persistence landscape remains unchanged. Given these considerations, this study focuses exclusively on the variations in L_1 and L_2 norms within the 1-dimensional homology class, as they provide more informative insights into topological changes in financial market structures.

Appendix C. Explanation of Taken's Time Delay Embedding

Based on Taken's Time Delay Embedding Theorem , once the map Ψ is obtained, the underlying dynamical structure of the time series can be analyzed through \mathcal{M}' . Previous studies have demonstrated that this method effectively extracts information from time-series data embedded in Euclidean space. Moreover, this embedding theorem has proven effective for analyzing chaotic and noisy time series. In particular, this theorem is widely used for state space reconstruction, enabling the transformation of raw time series data into high-dimensional point cloud representations.

Using Takens'embedding theorem, univariate time series data can be reconstructed in a higher-dimensional space. Since a time series represents a projection of the underlying states of a dynamical system, reconstructing its state space is essential for understanding its evolution. The goal of time series analysis is to predict the future evolution of underlying time-varying phenomena. However, reconstructing the state space and identifying its governing dynamical rules from time series data is often challenging, as the distribution of the data-generating process is unknown a priori. As a result, many techniques rely on attractors to achieve state-space reconstruction. An attractor is a set of values toward which a dynamical system converges, regardless of initial conditions. Since constructing a complete attractor requires an infinite number of points, time delay embedding is commonly employed as a quasi-attractor. Since the publication of Takens' seminal work *On the Nature of Turbulence* (1981), his embedding method has been widely used to attribute the onset of turbulence in experimental data to the presence of strange attractors [96]. This foundational work established a method for state space reconstruction, enabling researchers to embed a single time series into a high-dimensional dynamical system representation. Representing time series data as a point cloud enables the extraction of topological features via persistent homology. This approach ensures that the reconstructed dynamics remain consistent with those of the original phase space, thereby preserving the topological structure of the data.

A univariate time series $x = (x_1, x_2, \dots, x_N)$ can be reconstructed in an embedding space. The reconstructed state space representation is defined by two key parameters: τ , the time delay, and d , the embedding dimension. The total number of points in the phase space is given by $N - (d - 1)\tau$. Each state point corresponds to a row in the embedding matrix. In the process of reconstructing the phase space, the resulting embedding matrix provides a way to transform a one-dimensional time series into a higher-dimensional spatial representation. As the embedding dimension increases, the data points become more dispersed in the high-dimensional space, leading to an increase in the L_p norm.

Appendix D. Threshold Selection

In our study, we tested multiple thresholds (92%, 95%, 98%, 99%) to ensure robustness. The results in Figure A1 show that at thresholds of 92%, 95%, and 98%, the TDA-based

indicators (L_1 and L_2) consistently outperform traditional CPD methods in terms of F1 scores. However, when the threshold is set at 99%, the F1 score of the TDA methods drops sharply. This decline arises because an excessively high threshold leads to oversparsification of signals, where only a very limited number of extreme points are identified, thereby severely reducing recall. Although, in theory, a higher threshold can help filter out noise and increase precision, extreme events in financial markets are heterogeneous in both timing and magnitude. Consequently, at the 99% level, the model misses a substantial number of consensus events, undermining the balance between precision and recall. By contrast, with a 98% threshold, the detected change points align precisely with four widely recognized consensus events demonstrating that this threshold achieves a robust balance between sensitivity and specificity. Compared with this, traditional univariate methods (Rolling Std, PELT, HMM) and multivariate methods (E.divisive and E.aggro) exhibit fragmented or delayed detection, with generally lower F1 scores. Taken together, these findings empirically validate that the choice of a 98% threshold is not only effective in capturing systemic market disruptions but also theoretically well grounded.

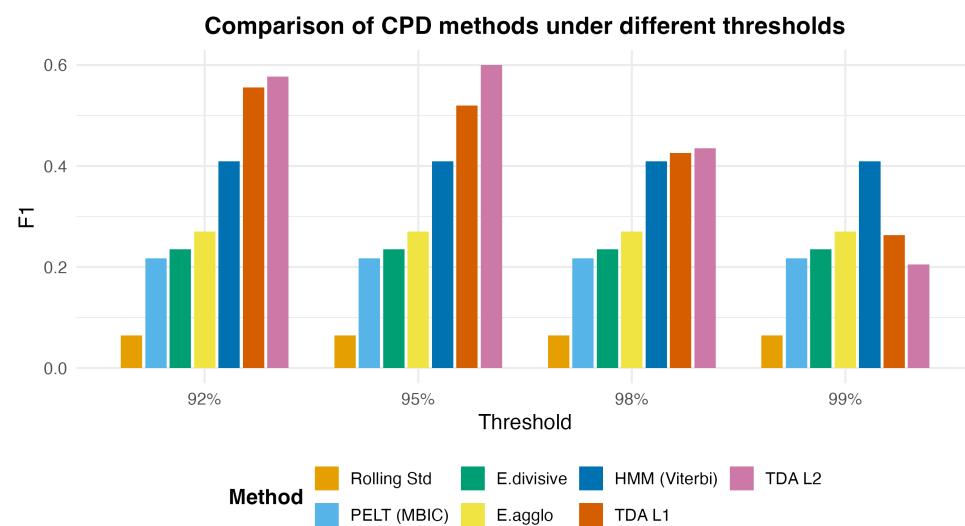


Figure A1. Comparison of CPD methods under different thresholds.

Appendix E. Global Major Volatility Event Set (2011–2023)

Table A2. Major financial and economic events (2011–2023).

Date	Event	Brief
18 April 2011	The aftermath of the nuclear accident	Global energy policy and insurance industries take a hit
6 May 2011	The Dow Jones Index plummeted instantly	Market fragility in the era of high-frequency trading
6 June 2011	The European debt crisis	The worsening European debt crisis triggered a global asset sell-off
5 August 2011	S&P downgrades US sovereign credit rating	The first downgrade of the US credit rating in history, followed by a repricing of global risk assets
8 August 2011	US Debt Downgrade/Peak of European Debt Crisis	Sharp decline in risk appetite
6 July 2012	“Whatever it takes” policy	Temporarily curbed the panic selling in the market

Table A2. *Cont.*

Date	Event	Brief
22 May 2013	Taper Talk (Tapering Expectations)	Rise in term premium and volatility ("Taper Tantrum")
24 August 2015	Global Stock Market Crash Post China's "811" FX Reform	Synchronized decline in global risk assets
24 June 2016	Brexit Referendum Result	Cross-asset repricing led by European equities and GBP
2 November 2016	Before the election	Markets priced in the risk of a very different future for the United States
6 December 2016	Italian referendum	Fluctuations involving European banks
5 February 2018	"Volmageddon" (XIV ETN Implosion)	Surge in both implied and realized volatility
10 October 2018	Onset of Q4 2018 US Stock Market Correction	Medium-term deterioration in risk sentiment
24 February 2020	Initial Global Sell-off due to COVID-19 Pandemic	Extreme risk shock (risk aversion)
23 March 2020	COVID-19 Market Bottom (US Stocks)/Global Policy Bottom	Policy support, inflection point in risk appetite
20 April 2020	Commodity Energy	WTI crude oil settlement price turned negative for the first time, hitting risk appetite again
25 June 2020	Fluctuations in post-pandemic liquidity and "reflation" phase	Huge liquidity from monetary and fiscal policies
23 September 2021	Credit pressure on Evergrande/Chinese real estate companies	Market worried China's real estate risks may spread to the financial system
26 November 2021	Omicron Variant News	Short-term risk repricing
15 December 2021	Signals of Accelerated Fed Tightening	Shift in inflation-policy expectations, rise in volatility
24 February 2022	Outbreak of Russia-Ukraine Conflict	Geopolitical and commodity-driven equity shock
13 June 2022	Above-Expectation Inflation/Accelerated Rate Hike Pricing	Peak of interest rate-valuation recalibration
28 September 2022	Bank of England temporary purchase of long-term government bonds	Rare financial stability operation disrupted global asset rhythm
13 October 2022	Near Market Bottom Post US CPI Report	Inflection point expecting "peak inflation → peak rates"
10 March 2023	SVB Incident/Regional Banking Stress	Transmission of liquidity and credit volatility
16 June 2023	Long-term U.S. Treasury bonds rose	Market expected Fed to maintain high interest rates for longer
24 July 2023	U.S. long and short yields fall	Economic slowdown signals outweighed inflation concerns
27 October 2023	Near 2023 Market Bottom (US Stocks)	Strong rally followed/"Soft Landing" trade
14 November 2023	US CPI Lower Than Expected	Optimistic repricing of rate cut path

References

1. Tsay, R.S. *Analysis of Financial Time Series*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
2. Bar-Joseph, Z. Analyzing time series gene expression data. *Bioinformatics* **2004**, *20*, 2493–2503. [[CrossRef](#)] [[PubMed](#)]
3. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
4. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: A survey. *IEEE Comput. Intell. Mag.* **2009**, *4*, 24–38. [[CrossRef](#)]
5. Dose, C.; Cincotti, S. Clustering of financial time series with application to index and enhanced index tracking portfolio. *Phys. A Stat. Mech. Its Appl.* **2005**, *355*, 145–151. [[CrossRef](#)]
6. James, S.L.; Gubbins, P.; Murray, C.J.; Gakidou, E. Developing a comprehensive time series of GDP per capita for 210 countries from 1950 to 2015. *Popul. Health Metrics* **2012**, *10*, 12. [[CrossRef](#)]
7. Katris, C. Prediction of Unemployment Rates with Time Series and Machine Learning Techniques. *Comput. Econ.* **2020**, *55*, 673–706. [[CrossRef](#)]
8. Inoue, A.; Kilian, L. How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation. *J. Am. Stat. Assoc.* **2008**, *103*, 511–522. [[CrossRef](#)]
9. Hamilton, J.D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econom. J. Econom. Soc.* **1989**, *57*, 357–384. [[CrossRef](#)]
10. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 2020.
11. Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Rojo-Alvarez, J.L.; Martinez-Ramon, M. Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1822–1835. [[CrossRef](#)]
12. Székely, G.J.; Rizzo, M.L. A new test for multivariate normality. *J. Multivar. Anal.* **2005**, *93*, 58–80. [[CrossRef](#)]
13. Amini-khanghani, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367. [[CrossRef](#)]
14. Lavielle, M.; Teyssière, G. Adaptive Detection of Multiple Change-Points in Asset Price Volatility. In *Long Memory in Economics*; Teyssière, G.; Kirman, A.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 129–156. [[CrossRef](#)]
15. Niu, Y.S.; Hao, N.; Zhang, H. Multiple Change-Point Detection: A Selective Overview. *Stat. Sci.* **2016**, *31*, 611–623. [[CrossRef](#)]
16. Burg, G.J.J.v.d.; Williams, C.K.I. An Evaluation of Change Point Detection Algorithms. *arXiv* **2020**, arXiv:2003.06222.
17. Hassler, U.; Scheithauer, J. Detecting changes from short to long memory. *Stat. Pap.* **2011**, *52*, 847–870. [[CrossRef](#)]
18. Killick, R.; Fearnhead, P.; Eckley, I.A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **2012**, *107*, 1590–1598. [[CrossRef](#)]
19. Kim, J.Y. Detection of change in persistence of a linear time series. *J. Econom.* **2000**, *95*, 97–116. [[CrossRef](#)]
20. Tempelman, J.R.; Khasawneh, F.A. A look into chaos detection through topological data analysis. *Phys. D Nonlinear Phenom.* **2020**, *406*, 132446. [[CrossRef](#)]
21. Islambekov, U.; Yuvaraj, M.; Gel, Y.R. Harnessing the power of topological data analysis to detect change points. *Environmetrics* **2020**, *31*, e2612. [[CrossRef](#)]
22. Gu, K.; Yan, L.; Li, X.; Duan, X.; Liang, J. Change point detection in multi-agent systems based on higher-order features. *Chaos Interdiscip. J. Nonlinear Sci.* **2022**, *32*, 111102. [[CrossRef](#)]
23. Carlsson, G.E. Topology and data. *Bull. Am. Math. Soc.* **2009**, *46*, 255–308. [[CrossRef](#)]
24. Carlsson, G. Topological pattern recognition for point cloud data. *Acta Numer.* **2014**, *23*, 289–368. [[CrossRef](#)]
25. Perea, J.A.; Deckard, A.; Haase, S.B.; Harer, J. SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC Bioinform.* **2015**, *16*, 257. [[CrossRef](#)]
26. Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, 12–14 November 2000; pp. 454–463. [[CrossRef](#)]
27. Zomorodian, A.; Carlsson, G. Computing Persistent Homology. *Discret. Comput. Geom.* **2005**, *33*, 249–274. [[CrossRef](#)]
28. Carlsson, G.; Ishkhanov, T.; de Silva, V.; Zomorodian, A. On the Local Behavior of Spaces of Natural Images. *Int. J. Comput. Vis.* **2008**, *76*, 1–12. [[CrossRef](#)]
29. Bubenik, P.; Carlsson, G.; Kim, P.; Luo, Z. Statistical Topology Via Morse Theory Persistence and Nonparametric Estimation. *Contemp. Math.* **2009**, *516*, 75–92. [[CrossRef](#)]
30. Bullmore, E.; Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **2009**, *10*, 186–198. [[CrossRef](#)]
31. Kovacev-Nikolic, V.; Bubenik, P.; Nikolic, D.; Heo, G. Using persistent homology and dynamical distances to analyze protein binding. *Stat. Appl. Genet. Mol. Biol.* **2016**, *15*, 19–38. [[CrossRef](#)]
32. Bendich, P.; Marron, J.S.; Miller, E.; Pieloch, A.; Skwerer, S. Persistent Homology Analysis of Brain Artery Trees. *Ann. Appl. Stat.* **2016**, *10*, 198–218. [[CrossRef](#)]

33. Carstens, C.J.; Horadam, K.J. Persistent Homology of Collaboration Networks. *Math. Probl. Eng.* **2013**, *2013*, 815035. [[CrossRef](#)]
34. Gidea, M.; Katz, Y. Topological data analysis of financial time series: Landscapes of crashes. *Phys. A Stat. Mech. Its Appl.* **2018**, *491*, 820–834. [[CrossRef](#)]
35. Bubenik, P. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **2015**, *16*, 77–102. [[CrossRef](#)]
36. Pereira, C.M.M.; de Mello, R.F. Persistent homology for time series and spatial data clustering. *Expert Syst. Appl.* **2015**, *42*, 6026–6038. [[CrossRef](#)]
37. Bubenik, P.; Dłotko, P. A persistence landscapes toolbox for topological statistics. *J. Symb. Comput.* **2017**, *78*, 91–114. [[CrossRef](#)]
38. Lee, K.M.; Yang, J.S.; Kim, G.; Lee, J.; Goh, K.I.; Kim, I.m. Impact of the Topology of Global Macroeconomic Network on the Spreading of Economic Crises. *PLoS ONE* **2011**, *6*, e18443. [[CrossRef](#)] [[PubMed](#)]
39. Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econom.* **2003**, *18*, 1–22. [[CrossRef](#)]
40. Page, E.S. Controlling the Standard Deviation by Cusums and Warning Lines. *Technometrics* **1963**, *5*, 307–315. [[CrossRef](#)]
41. Bai, J.; Perron, P. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* **1998**, *66*, 47–78. [[CrossRef](#)]
42. Donald, W.K.A. Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica* **1993**, *61*, 821–856. [[CrossRef](#)]
43. Barry, D.; Hartigan, J.A. A Bayesian Analysis for Change Point Problems. *J. Am. Stat. Assoc.* **1993**, *88*, 309–319. [[CrossRef](#)]
44. Chib, S.; Greenberg, E.; Winkelmann, R. Posterior simulation and Bayes factors in panel count data models. *J. Econom.* **1998**, *86*, 33–54. [[CrossRef](#)]
45. Carlin, J.B. Meta-analysis for 2×2 tables: A bayesian approach. *Stat. Med.* **1992**, *11*, 141–158. [[CrossRef](#)]
46. Matteson, D.S.; James, N.A. A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *J. Am. Stat. Assoc.* **2014**, *109*, 334–345. [[CrossRef](#)]
47. Amini, A.A.; Wainwright, M.J. High-dimensional analysis of semidefinite relaxations for sparse principal components. In Proceedings of the 2008 IEEE International Symposium on Information Theory, Toronto, ON, Canada, 6–11 July 2008; pp. 2454–2458. [[CrossRef](#)]
48. Harchaoui, Z.; Lévy-Leduc, C. Multiple Change-Point Estimation With a Total Variation Penalty. *J. Am. Stat. Assoc.* **2010**, *105*, 1480–1493. [[CrossRef](#)]
49. Shi, Z.; Chehade, A. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107257. [[CrossRef](#)]
50. Lavielle, M. Using penalized contrasts for the change-point problem. *Signal Process.* **2005**, *85*, 1501–1510. [[CrossRef](#)]
51. Wei, C.Y.; Luo, H. Non-stationary Reinforcement Learning without Prior Knowledge: An Optimal Black-box Approach. *arXiv* **2021**, arXiv:2102.05406.
52. Jiang, J.; Chen, R.; Zhang, C.; Chen, M.; Li, X.; Ma, G. Dynamic Fault Prediction of Power Transformers Based on Lasso Regression and Change Point Detection by Dissolved Gas Analysis. *IEEE Trans. Dielectr. Electr. Insul.* **2020**, *27*, 2130–2137. [[CrossRef](#)]
53. Khasawneh, F.A.; Munch, E. Exploring Equilibria in Stochastic Delay Differential Equations Using Persistent Homology. In Proceedings of the ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Buffalo, NY, USA, 17–20 August 2014. [[CrossRef](#)]
54. Khasawneh, F.A.; Munch, E. Stability Determination in Turning Using Persistent Homology and Time Series Analysis. In Proceedings of the ASME 2014 International Mechanical Engineering Congress and Exposition, Montreal, QC, Canada, 14–20 November 2014. [[CrossRef](#)]
55. Kim, K.; Kim, J.; Rinaldo, A. Time Series Featurization via Topological Data Analysis. *arXiv* **2018**, arXiv:1812.02987. [[CrossRef](#)]
56. Bubenik, P.; Kim, P.T. A statistical approach to persistent homology. *Homol. Homotopy Appl.* **2006**, *9*, 337–362. [[CrossRef](#)]
57. Chazal, F.; Fasy, B.T.; Lecci, F.; Rinaldo, A.; Wasserman, L. Stochastic Convergence of Persistence Landscapes and Silhouettes. *arXiv* **2013**, arXiv:1312.0308. [[CrossRef](#)]
58. Leibon, G.; Pauls, S.; Rockmore, D.; Savell, R. Topological structures in the equities market network. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20589–20594. [[CrossRef](#)]
59. Gidea, M. Topological Data Analysis of Critical Transitions in Financial Networks. In *3rd International Winter School and Conference on Network Science*; Shmueli, E., Barzel, B., Puzis, R., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 47–59.
60. Gidea, M.; Goldsmith, D.; Katz, Y.; Roldan, P.; Shmalo, Y. Topological recognition of critical transitions in time series of cryptocurrencies. *Phys. A Stat. Mech. Its Appl.* **2020**, *548*, 123843. [[CrossRef](#)]
61. Aguilar, A.; Ensor, K.E. Topology Data Analysis Using Mean Persistence Landscapes in Financial Crashes. *J. Math. Financ.* **2020**, *10*, 648–678. [[CrossRef](#)]
62. Goel, A.; Pasricha, P.; Mehra, A. Topological data analysis in investment decisions. *Expert Syst. Appl.* **2020**, *147*, 113222. [[CrossRef](#)]
63. Goel, A.; Filipović, D.; Pasricha, P. Sparse Portfolio Selection via Topological Data Analysis Based Clustering. *arXiv* **2024**, arXiv:2401.16920. [[CrossRef](#)]
64. Sokerin, P.O.; Kuznetsov, K.; Makhneva, E.; Zaytsev, A. Portfolio selection via topological data analysis. In Proceedings of the International Conference on Machine Vision, Yerevan, Armenia, 15–18 November 2023.

65. Guo, H.; Xia, S.; An, Q.; Zhang, X.; Sun, W.; Zhao, X. Empirical study of financial crises based on topological data analysis. *Phys. A Stat. Mech. Its Appl.* **2020**, *558*, 124956. [[CrossRef](#)]
66. Guo, H.; Yu, H.; An, Q.; Zhang, X. Risk analysis of China's stock markets based on topological data structures. *Procedia Comput. Sci.* **2022**, *202*, 203–216. [[CrossRef](#)]
67. Ismail, M.S.; Hussain, S.I.; Noorani, M.S.M. Detecting Early Warning Signals of Major Financial Crashes in Bitcoin Using Persistent Homology. *IEEE Access* **2020**, *8*, 202042–202057. [[CrossRef](#)]
68. Ismail, M.S.; Md Noorani, M.S.; Ismail, M.; Abdul Razak, F.; Alias, M.A. Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange. *Appl. Soft Comput.* **2020**, *93*, 106422. [[CrossRef](#)]
69. Ismail, M.S.; Noorani, M.S.M.; Ismail, M.; Razak, F.A.; Alias, M.A. Early warning signals of financial crises using persistent homology. *Phys. A Stat. Mech. Its Appl.* **2022**, *586*, 126459. [[CrossRef](#)]
70. Katz, Y.A.; Biem, A. Time-resolved topological data analysis of market instabilities. *Phys. A Stat. Mech. Its Appl.* **2021**, *571*, 125816. [[CrossRef](#)]
71. Yen, P.T.W.; Cheong, S.A. Using Topological Data Analysis (TDA) and Persistent Homology to Analyze the Stock Markets in Singapore and Taiwan. *Front. Phys.* **2021**, *9*, 572216. [[CrossRef](#)]
72. Yen, P.T.W.; Xia, K.; Cheong, S.A. Understanding Changes in the Topology and Geometry of Financial Market Correlations during a Market Crash. *Entropy* **2021**, *23*, 1211. [[CrossRef](#)] [[PubMed](#)]
73. Majumdar, S.; Laha, A.K. Clustering and classification of time series using topological data analysis with applications to finance. *Expert Syst. Appl.* **2020**, *162*, 113868. [[CrossRef](#)]
74. Rai, A.; Nath Sharma, B.; Rabindrajit Luwang, S.; Nurujjaman, M.; Majhi, S. Identifying extreme events in the stock market: A topological data analysis. *Chaos Interdiscip. J. Nonlinear Sci.* **2024**, *34*, 103106. [[CrossRef](#)]
75. Sebestyén, T.; Illoskics, Z. Do economic shocks spread randomly?: A topological study of the global contagion network. *PLoS ONE* **2020**, *15*, e0238626. [[CrossRef](#)]
76. Zhang, F.; Wu, Y. Topological Time Series Analysis of Market Crashes: A Persistence Homology Approach. In Proceedings of the 2025 5th International Conference on Applied Mathematics, Modelling and Intelligent Computing, Shanghai, China, 21–23 March 2025. [[CrossRef](#)]
77. Guritanu, E.; Barbierato, E.; Gatti, A. Topological Machine Learning for Financial Crisis Detection: Early Warning Signals from Persistent Homology. *Computers* **2025**, *14*, 408. [[CrossRef](#)]
78. Nath Sharma, B.; Rai, A.; Luwang, S.; Nurujjaman, M.; Majhi, S. Causality Analysis of COVID-19 Induced Crashes in Stock and Commodity Markets: A Topological Perspective. *arXiv* **2025**, arXiv:2502.14431. [[CrossRef](#)]
79. de Jesus, L.C.; Fernández-Navarro, F.; Carbonero-Ruz, M. Enhancing financial time series forecasting through topological data analysis. *Neural Comput. Appl.* **2025**, *37*, 6527–6545. [[CrossRef](#)]
80. Biasotti, S.; Floriani, L.D.; Falcidieno, B.; Frosini, P.; Giorgi, D.; Landi, C.; Papaleo, L.; Spagnuolo, M. Describing shapes by geometrical-topological properties of real functions. *ACM Comput. Surv.* **2008**, *40*, 12. [[CrossRef](#)]
81. Carlsson, G.; Zomorodian, A.; Collins, A.; Guibas, L. Persistence barcodes for shapes. In Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, Nice, France, 8–10 July 2004. [[CrossRef](#)]
82. Ghrist, R. Barcodes: The persistent topology of data. *Bull. Am. Math. Soc.* **2007**, *45*, 61–75. [[CrossRef](#)]
83. Chazal, F.; Michel, B. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Front. Artif. Intell.* **2021**, *4*, 667963. [[CrossRef](#)] [[PubMed](#)]
84. Munkres, J.R.; Munkres, J.W. *Elements of Algebraic Topology*; CRC Press: Boca Raton, FL, USA, 2018.
85. Carlsson, G.; Zomorodian, A. The Theory of Multidimensional Persistence. *Discret. Comput. Geom.* **2009**, *42*, 71–93. [[CrossRef](#)]
86. Edelsbrunner, H.; Harer, J. Computational Topology: An Introduction. In *Effective Computational Geometry for Curves and Surfaces*; Springer: Berlin/Heidelberg, Germany, 2010. [[CrossRef](#)]
87. Cohen-Steiner, D.; Edelsbrunner, H.; Harer, J. Stability of Persistence Diagrams. *Discret. Comput. Geom.* **2007**, *37*, 103–120. [[CrossRef](#)]
88. Bubenik, P. The Persistence Landscape and Some of Its Properties. In *Topological Data Analysis*; Baas, N.A., Carlsson, G.E., Quick, G., Szymik, M., Thaule, M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 97–117.
89. Feng, M.; Porter, M.A. Persistent Homology of Geospatial Data: A Case Study with Voting. *SIAM Rev.* **2021**, *63*, 67–99. [[CrossRef](#)]
90. Bleile, B.; Garin, A.; Heiss, T.; Maggs, K.; Robins, V. The Persistent Homology of Dual Digital Image Constructions. *arXiv* **2021**, arXiv:2102.11397. [[CrossRef](#)]
91. Myers, A.; Munch, E.; Khasawneh, F.A. Persistent homology of complex networks for dynamic state detection. *Phys. Rev. E* **2019**, *100*, 022314. [[CrossRef](#)]
92. Wang, R.; Nguyen, D.D.; Wei, G.W. Persistent spectral graph. *Int. J. Numer. Method Biomed. Eng.* **2020**, *36*, e3376. [[CrossRef](#)]
93. Ziv, J.; Lempel, A. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory* **1977**, *23*, 337–343. [[CrossRef](#)]
94. Perea, J.A.; Harer, J. Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. *Found. Comput. Math.* **2015**, *15*, 799–838. [[CrossRef](#)]

95. Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Geometry from a Time Series. *Phys. Rev. Lett.* **1980**, *45*, 712–716. [[CrossRef](#)]
96. Takens, F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*; Springer: Berlin/Heidelberg, Germany, 1981.
97. Sauer, T.; Yorke, J.A.; Casdagli, M. Embedology. *J. Stat. Phys.* **1991**, *65*, 579–616. [[CrossRef](#)]
98. Lum, P.Y.; Singh, G.; Lehman, A.; Ishkanov, T.; Vejdemo-Johansson, M.; Alagappan, M.; Carlsson, J.; Carlsson, G. Extracting insights from the shape of complex data using topology. *Sci. Rep.* **2013**, *3*, 1236. [[CrossRef](#)] [[PubMed](#)]
99. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **2018**, *58*, 267–288. [[CrossRef](#)]
100. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **2000**, *42*, 80–86. [[CrossRef](#)]
101. McPhillips, L.E.; Chang, H.; Chester, M.V.; Depietri, Y.; Friedman, E.; Grimm, N.B.; Kominoski, J.S.; McPhearson, T.; Méndez-Lázaro, P.; Rosi, E.J.; et al. Defining Extreme Events: A Cross-Disciplinary Review. *Earth's Future* **2018**, *6*, 441–455. [[CrossRef](#)]
102. Montgomery, D. *Introduction to Statistical Quality Control*; Wiley: Hoboken, NJ, USA, 2020.
103. Zhang, N.R.; Siegmund, D.O. A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data. *Biometrics* **2007**, *63*, 22–32. [[CrossRef](#)]
104. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.