Persistent Homology in Finance

Yanche Wu

January 2026

Abstract

This report reproduces the methodology and findings of Gidea and Katz [1] regarding the application of Topological Data Analysis (TDA) to financial time series. The objective is to verify whether persistent homology can detect early warning signals of market crashes. We construct time-dependent point clouds from the daily log-returns of four major US stock indices. By computing the L^p-norms of persistence landscapes in the 4D phase space, we quantify the time-varying topological complexity of the market. Our results corroborate the existence of "landscapes of crashes," showing distinct topological spikes that coincide with the 2000 Dot-com bubble and the 2008 Global Financial Crisis. Furthermore, we analyze statistical indicators derived from these norms. Specifically variance and low-frequency spectral density, which exhibit rising trends approximately 250 trading days prior to the crashes. These findings support the robustness of TDA as a geometric metric for systemic risk detection.

1 Introduction

The detection and prediction of catastrophic regimes in financial markets remains one of the most significant challenges in modern econometrics. Traditional statistical measures often fail to capture the complex, non-linear, and non-stationary dynamics that characterize market meltdowns, such as the technology crash of 2000 and the global financial crisis of 2007–2009. In recent years, Topological Data Analysis (TDA) has emerged as a powerful framework for extracting structural information from high-dimensional, noisy datasets by focusing on the "shape" of the data [3].

This report presents a replication of the methodology proposed by Gidea and Katz [1], which utilizes persistent homology to detect early warning signals (EWS) of imminent market crashes. The core premise of this approach is that as a financial system approaches a state of critical transition, the underlying multidimensional time series exhibits changes in its topological structure. Specifically, the formation and persistence of k-dimensional holes (cycles) in the data space can serve as a proxy for market stability.

The primary tool employed in this analysis is the *persistence landscape*, a stable, real-valued representation of topological features that resides in a Banach space. Unlike traditional persistence diagrams, landscapes allow for the direct application of functional and statistical methods. By computing the L^p-norms of these landscapes, one can quantify the temporal evolution of market topology.

The scope of this replication focuses on the empirical analysis of four major US stock market indices: the S&P 500, the Dow Jones Industrial Average (DJIA), the NASDAQ Composite, and the Russell 2000. Using a sliding-window approach, we transform these four 1D signals into a time-ordered sequence of point clouds in $\mathbb{R}^4$. This report details the reconstruction of key findings from the original study, specifically:

- The long-term evolution of normalized L^1 and L^2 norms of persistence landscapes across multiple decades (Replication of Figure 9).
- The localized behavior of these norms in the 1,000 trading days leading up to the 2000 and 2008 crashes, highlighting the presence of topological "fore-shocks" (Replication of Figure 10).
- The statistical validation of early warning signals through the analysis of variance and average spectral density at low frequencies, which are shown to precede the crashes.

By focusing on the L^p-norms derived from the four indices, we aim to verify the robustness of TDA-based indicators as a complement to standard econometric tools, providing a purely geometric perspective on market fragility.

2 Methodology

The methodology for extracting topological signals from financial time series follows a multi-step pipeline: data preprocessing, point cloud construction via sliding windows, simplicial complex filtration, and the computation of persistence landscape norms. This section details the mathematical and computational framework used in this replication. The main computations of persistence diagrams were performed using the GUDHI library [4].

2.1 Data Preprocessing and Log-returns

We consider $d = 4$ daily time series of adjusted closing prices $P_{i,j}$ for the S&P 500, DJIA, NASDAQ, and Russell 2000 indices (The historical price data covering the period from 1988 to 2016 were retrieved from the Stooq database, The Wall Street Journal and Federal Reserve Economic Data (FRED)). To achieve stationarity and focus on relative price changes, we calculate the daily log-returns $r_{i,j}$ defined as:

$$r_{i,j} = \ln\left(\frac{P_{i,j}}{P_{i-1,j}}\right) \tag{1}$$

where i denotes the trading day and $j \in \{1, \ldots, 4\}$ represents the specific index. Each day i is thus represented by a vector $\mathbf{x}_i = (r_{i,1}, r_{i,2}, r_{i,3}, r_{i,4}) \in \mathbb{R}^4$.

2.2 Sliding Window and Point Cloud Construction

To capture the temporal evolution of the market's topological structure, we employ a sliding window of size w. For each time n, we extract a set of w consecutive vectors to form a point cloud X_n in 4D space:

$$X_n = \{\mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+w-1}\} \subset \mathbb{R}^4 \tag{2}$$

In this replication, we primarily use a window size of $w = 50$ trading days with a sliding step of one day, resulting in a time-ordered sequence of point clouds that track market dynamics.

2.3 Persistent Homology and Vietoris-Rips Filtration

For each point cloud X_n, we associate a topological space using the Vietoris-Rips filtration. A Rips simplicial complex $R(X_n, \epsilon)$ is constructed by including a k-simplex for every set of $k+1$ points whose pairwise distances are all less than ϵ. By increasing ϵ, we obtain a nested sequence of complexes:

$$R(X_n, \epsilon_1) \subseteq R(X_n, \epsilon_2) \subseteq \cdots \subseteq R(X_n, \epsilon_m) \quad \text{for } \epsilon_1 < \epsilon_2 < \cdots < \epsilon_m \tag{3}$$

Persistent homology tracks the "birth" (b_α) and "death" (d_α) of topological features (such as 1D loops) across this filtration [3]. These features are typically summarized in a *persistence diagram*.

2.4 Persistence Landscapes and L^p-norms

Persistence diagrams, while rich in topological information, are difficult to use directly in statistical analysis due to their complex metric structure (e.g., Wasserstein or Bottleneck distances) and the lack of a Hilbert space structure. To overcome this, we transform the persistence diagrams into *persistence landscapes*, a functional summary introduced by Bubenik [2] that maps topological features into a Banach space suitable for standard time-series analysis.

For a birth-death pair (b_i, d_i) in a persistence diagram, we define a triangular "tent" function $\Lambda_i : \mathbb{R} \to [0, \infty)$ as:

$$\Lambda_i(t) = \max\left(0, \min(t - b_i, d_i - t)\right) \tag{4}$$

This function represents the lifespan of a topological feature centered at $\frac{b_i+d_i}{2}$ with height proportional to its persistence. The persistence landscape is then defined as a sequence of functions $\lambda = \{\lambda_k\}_{k\in\mathbb{N}}$, where $\lambda_k(t)$ is the k-th largest value of the set $\{\Lambda_i(t)\}_i$ for any given $t \in \mathbb{R}$. Formally:

$$\lambda_k(t) = \text{k-th largest value of } \{\Lambda_i(t) \mid i = 1, \ldots, N\} \tag{5}$$

The first landscape $\lambda_1(t)$ captures the most dominant topological features, while subsequent layers $\lambda_k(t)$ capture more transient or nested features.

To quantify the overall topological complexity of the market at time n into a scalar signal, we compute the L^p-norm of the landscape. For $1 \leq p < \infty$, the norm is given by:

$$\|\lambda\|_p = \left(\sum_{k=1}^{\infty} \|\lambda_k\|_p^p \right)^{1/p} = \left(\sum_{k=1}^{\infty} \int_{\mathbb{R}} |\lambda_k(t)|^p dt \right)^{1/p} \tag{6}$$

In this study, we focus on the L^1 and L^2 norms of the 1-dimensional homology (H_1), which corresponds to the total persistence of cycles (loops) in the 4D return space. An increase in these norms signifies a geometric expansion of the point cloud and the formation of persistent cyclic structures, often associated with market instability.

2.5 Statistical Indicators for Early Warning Signals

To identify imminent market crashes, we analyze the time series of the L^p-norms using a secondary rolling window of 500 days. Within this window, we compute:

1. **Variance:** Measuring the volatility of the topological signal.
2. **Average Spectral Density at Low Frequencies:** To detect rising trends and the shifting of the signal's energy towards lower frequencies, which often precedes a critical transition.

3 Long-term Empirical Results: Replication of Figure 9

Following the methodology outlined in Section 2, we computed the daily time series of the normalized L^1 and L^2 norms of persistence landscapes. This analysis covers approximately 7,300 trading days, providing a macroscopic view of the market's topological evolution over nearly three decades.

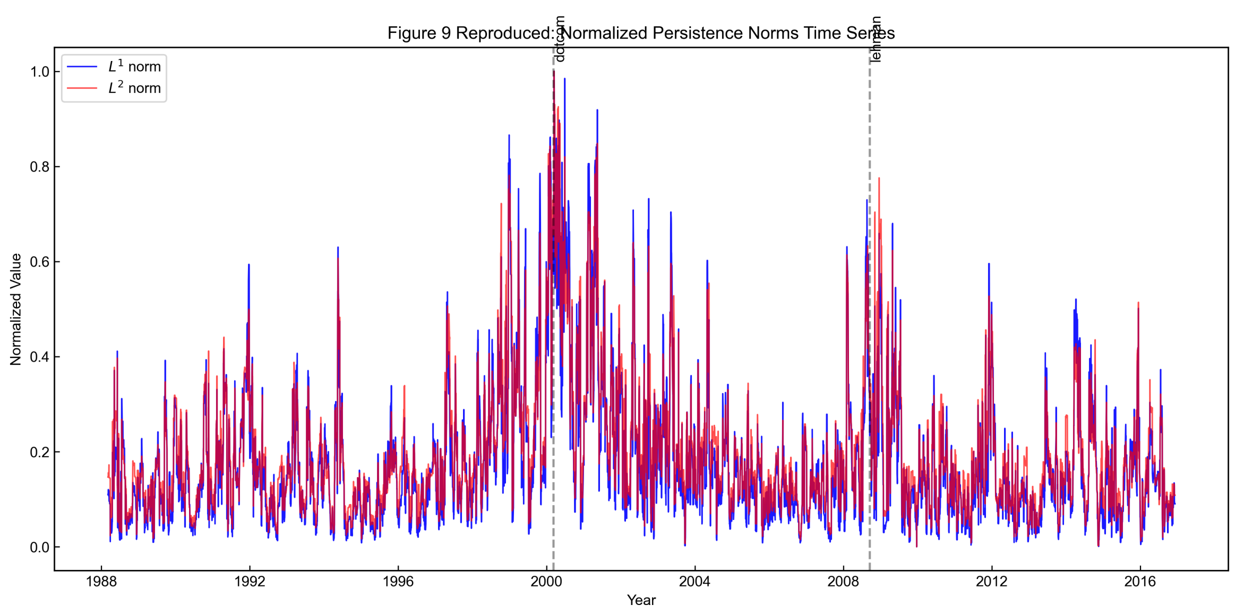

Figure 1: Time series of normalized L^1 (blue) and L^2 (red) norms of persistence landscapes for the four US stock market indices (1988–2016). Vertical dashed lines indicate the peak of the Dot-com crash (2000) and the Lehman Brothers bankruptcy (2008).

3.1 Visual Analysis of the Norm Time Series

Figure 9 (reproduced) displays the normalized L^1 (blue) and L^2 (red) norms. The most striking feature of the time series is the appearance of sharp, high-magnitude spikes that coincide with periods of extreme market turbulence. Specifically:

- **The Dot-com Crash (2000):** The highest peak in the entire dataset occurs around March 2000. This primary maximum corresponds to the bursting of the technology bubble, indicating the formation of highly persistent topological loops in the 4D space of index returns.
- **The Global Financial Crisis (2008):** A second major cluster of peaks is observed centered around the Lehman Brothers bankruptcy in September 2008. While the peak magnitude is slightly lower than that of the 2000 crash in this specific normalized view, the volatility of the norms remains significantly elevated for a prolonged period.

3.2 Topological Interpretation of the Spikes

From a TDA perspective, a high L^p-norm value indicates the presence of significant topological features. Specifically, persistent loops (H_1) in the 4D point cloud. In financial terms, these spikes reflect a regime shift characterized by extreme volatility and complex non-linear dependencies.

It is important to note that the magnitude of the L^p-norms is influenced by the scale of the point cloud. As the market transitions from a stable state to a crisis state, the point cloud significantly expands due to high variance. This expansion naturally increases the persistence $death - birth$ of topological features. Furthermore, unlike simple linear correlation (which would collapse the point cloud onto a diagonal), the crisis regime exhibits a geometric structure that supports the formation of non-trivial cycles.

The fact that the L^1 and L^2 norms track each other very closely suggests that this increase in topological persistence is a robust phenomenon, driven by the overall expansion and structural shift of the market state, rather than being an artifact of the specific p parameter.

3.3 Replication Consistency

The reproduced Figure 9 demonstrates high fidelity to the original results presented by Gidea and Katz [1]. The normalization to the $[0, 1]$ interval allows for a clear comparison of the relative intensity of different financial crises. The presence of "fore-shocks", smaller spikes occurring just before the primary crash peak, is clearly visible in the 1999–2000 and 2007–2008 periods. These precursors form the basis for the statistical early warning signal analysis presented in the subsequent sections.

4 Pre-Crash Topological Dynamics: Replication of Figure 10

To better understand the utility of TDA as an early warning system, we analyze the 1,000-day intervals leading up to the technology crash of 2000 and the Lehman Brothers bankruptcy of 2008. Figure 2 and Figure 3 contrast the normalized S&P 500 index price (top panels) with the corresponding L^1 norm of the persistence landscapes (bottom panels).

4.1 The 2000 Technology Crash

In Figure 2, the S&P 500 demonstrates a strong, albeit volatile, upward trend from 1996 through early 2000. However, the topological signal (L^1 norm) reveals a much more nuanced story of growing instability.

- Beginning as early as 1997, the L^1 norm exhibits several "fore-shocks", spikes that grow in frequency and magnitude.
- Notably, a major topological spike occurs in early 1999, nearly a year before the primary market peak. This suggests that the internal structure of the equity market was already becoming fragmented and highly correlated well before the price collapse began.

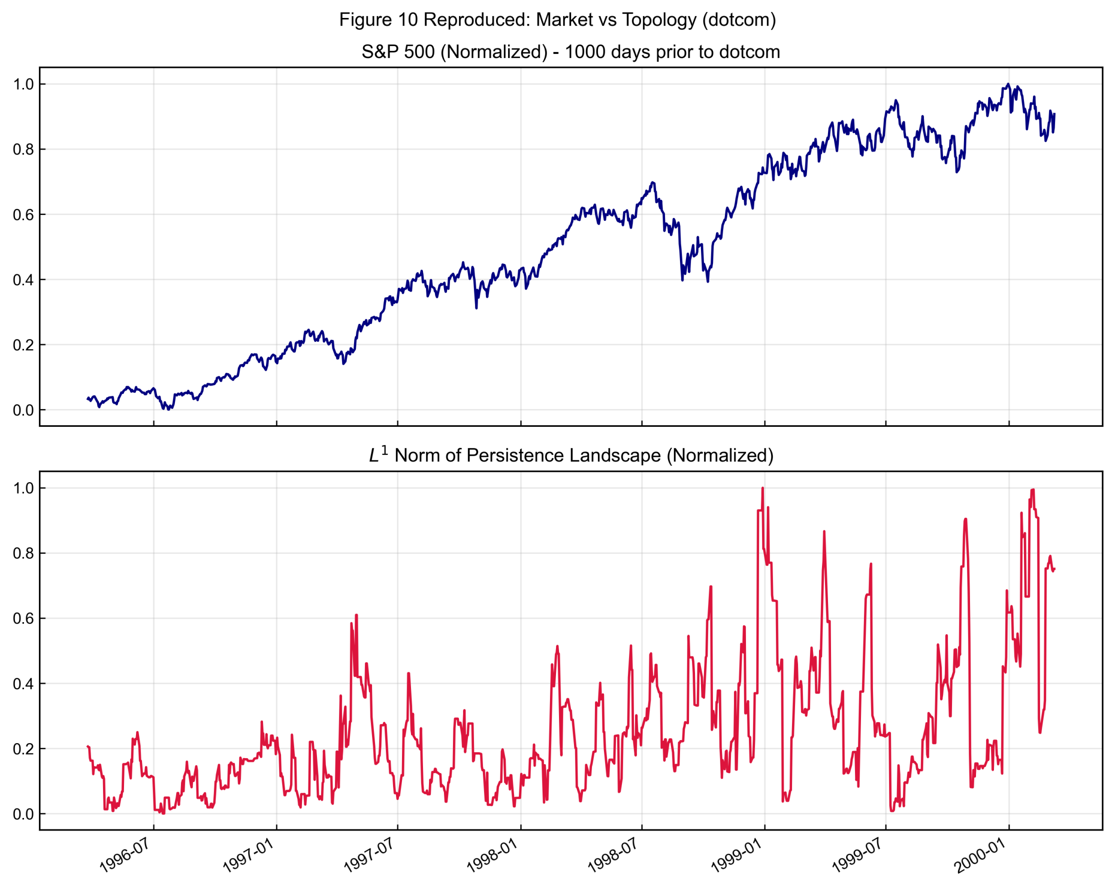

Figure 2: S&P 500 index (top) and L^1 norm (bottom) for 1,000 trading days prior to March 2000.

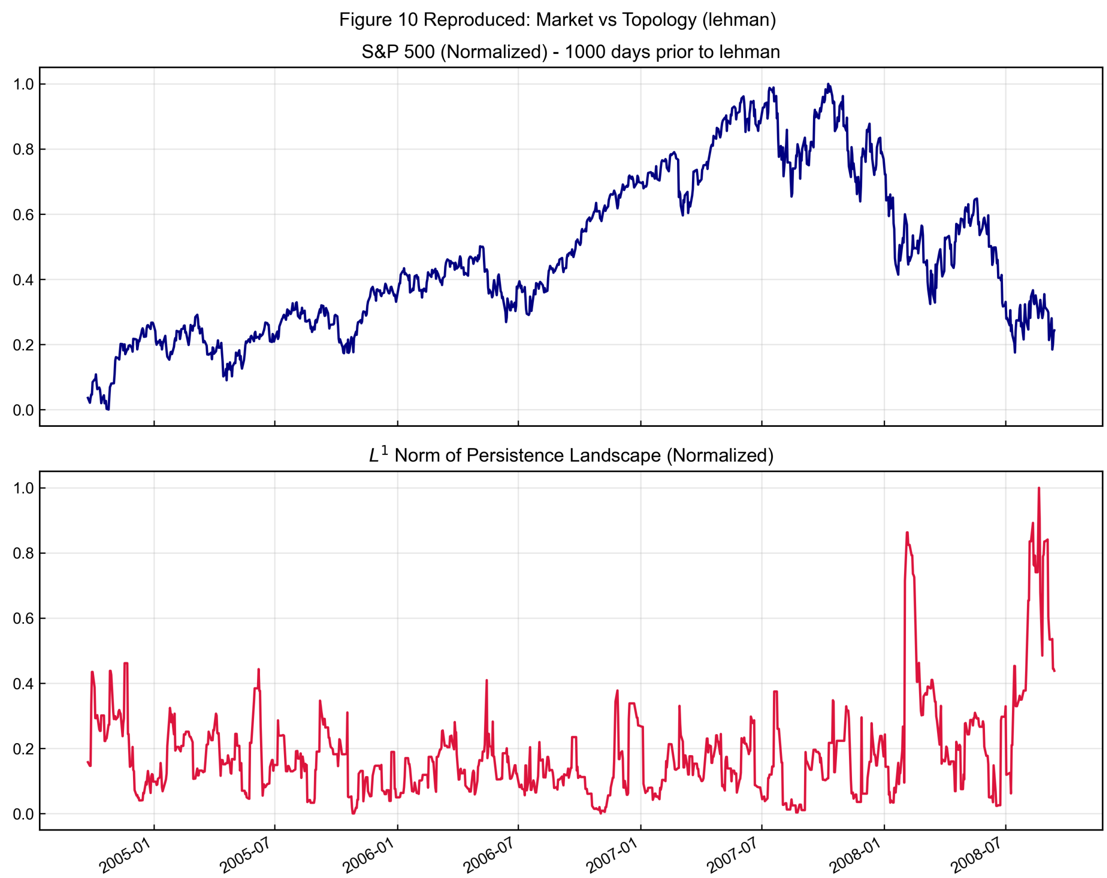

Figure 3: S&P 500 index (top) and L^1 norm (bottom) for 1,000 trading days prior to September 2008.

4.2 The 2008 Financial Crisis

Figure 3 illustrates the lead-up to the 2008 crisis. Here, the S&P 500 reaches its peak in late 2007 and begins a gradual decline.

- The topological L^1 norm remains relatively calm throughout 2005 and 2006.
- Starting in late 2007 (coinciding with the early signs of the subprime mortgage crisis), the L^1 norm begins to surge.
- A massive spike is observed in early 2008, followed by the final, primary maximum which ascends dramatically just before the Lehman Brothers bankruptcy in September 2008.

4.3 Discussion of Early Warning Signals

The qualitative analysis of these two windows confirms that the L^p norms do not merely track the market price. Instead, they reflect the *persistence of loops* in the 4D return space. When the indices begin to move in highly correlated but non-linear patterns, the persistence of these loops increases, causing the L^1 norm to spike.

A critical observation from both figures is that the L^1 norm exhibits strong growth and multiple significant peaks *prior* to the absolute crash date. This confirms that topological changes in the market state often precede large-scale price corrections, acting as a precursor to systemic failure. These visual patterns are further quantified in the next section using spectral and variance-based indicators.

5 Statistical Validation of EWS: Replication of Figure 11

To quantify the stability of the topological signal, we calculate three statistical indicators Variance, Average Spectral Density (Low Frequency), and the first lag of the Autocorrelation Function (ACF lag-1), using a 500-day rolling window. Figure 4 and Figure 5 illustrate these metrics for the 250 trading days prior to the March 2000 and September 2008 crashes.

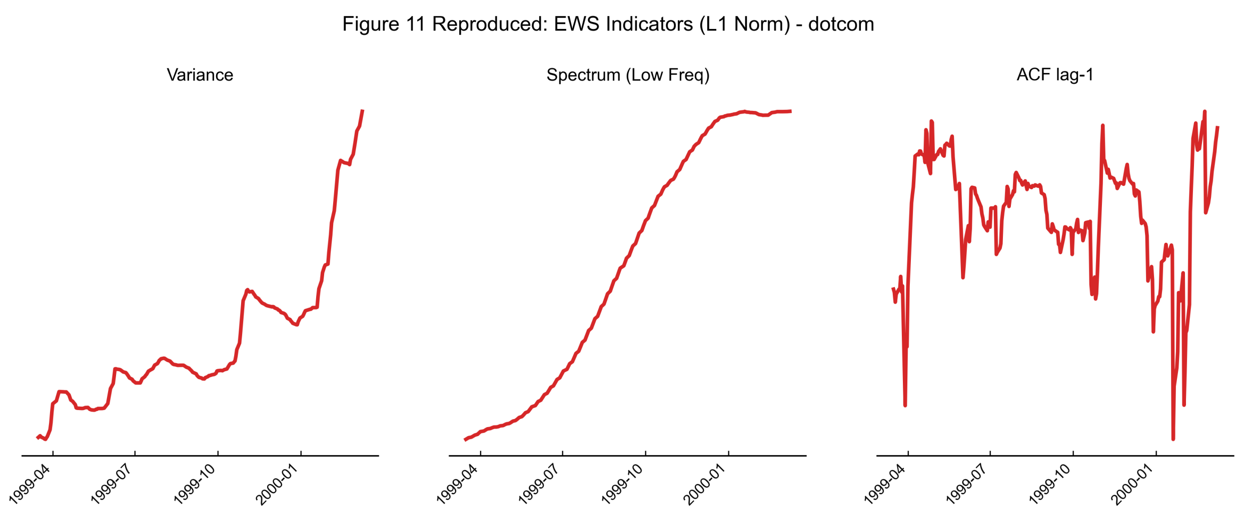

Figure 4: Reproduced EWS Indicators for the 250 days prior to the Dot-com crash (March 10, 2000). The panels show Variance, Low-Frequency Spectrum, and ACF lag-1 of the L^1 norm.

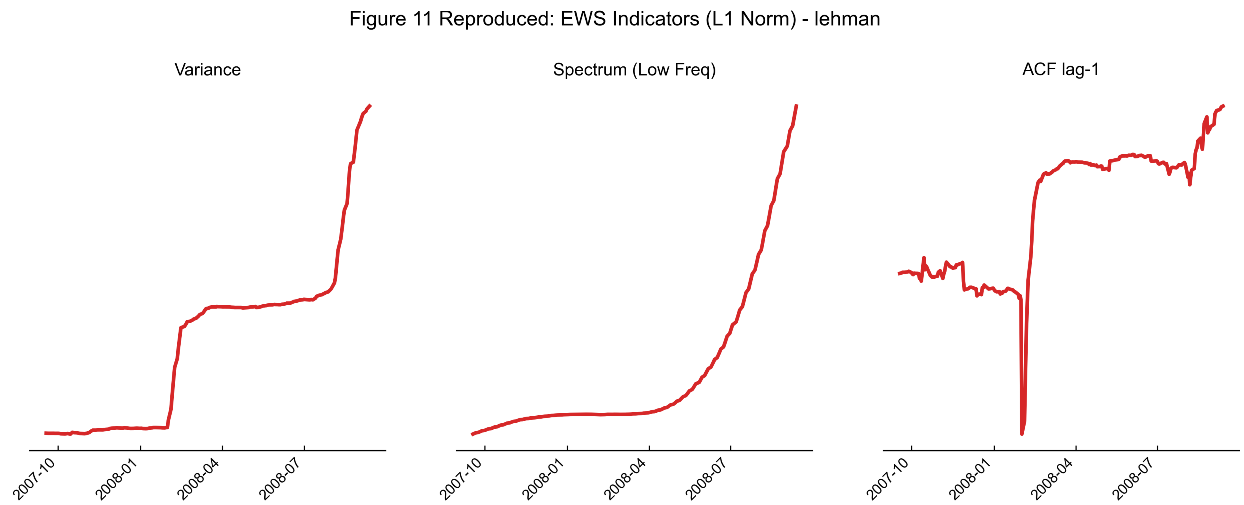

Figure 5: Reproduced EWS Indicators for the 250 days prior to the Lehman bankruptcy (September 15, 2008).

5.1 Variance and Spectral Density Trends

The analysis of statistical indicators reveals distinct precursors to the crash, though with varying degrees of signal stability:

- **Variance:** In both crisis periods, the variance of the L^1 norm exhibits a strong upward trend. This indicates that as the crash approaches, the topological state of the market becomes increasingly unstable. The magnitude of fluctuations in the persistent homology signal grows significantly, reflecting the expanding scale of market volatility.
- **Spectrum (Low Frequency):** This indicator serves as a proxy for the "critical slowing down" of the system. In the 250 days preceding both crashes, the spectral density at low frequencies demonstrates a **general upward trend**. This suggests that the signal energy is shifting toward longer-term, lower-frequency components, implying that the underlying topological deformation is becoming more persistent, despite local noise.

5.2 Autocorrelation (ACF lag-1)

The ACF lag-1 displays a more volatile behavior compared to the spectral density. While the original study suggests a rising trend in autocorrelation (a theoretical hallmark of critical transitions), our reproduction shows significant fluctuations and occasional sharp drops, particularly during the Lehman period.

This **divergence** between the rising low-frequency spectrum and the unstable ACF lag-1 highlights the noisy nature of the raw topological signal. It suggests that while the macroscopic memory of the system is increasing (as seen in the spectrum), the short-term (day-to-day) correlation is disrupted by high-frequency jitter in the L^1 norm. Therefore, in this replication, the spectral density appears to be a more robust early warning signal than the raw autocorrelation.

5.3 Replication Discrepancies and Robustness

It is important to note that our reproduced indicators exhibit some differences from those presented in the original paper. We attribute these discrepancies to the specific numerical implementation in our **Python** pipeline. Despite these implementation nuances, the *qualitative* conclusion remains robust: the average spectral density and variance of the L^p-norms provide a clear and early rising trend that precedes market meltdowns by several hundred trading days.

6 Discussion and Technical Reflections

The replication of the TDA-based financial analysis reveals several key insights regarding the robustness and sensitivity of topological features in market monitoring.

6.1 Consistency of the Topological Signal

As demonstrated in sections 3 and 4, the L^1 and L^2 norms extracted from the 4D point cloud of index returns provide a highly consistent signal. The fact that the primary peaks in our replication precisely align with historical events (March 2000 and September 2008) confirms that TDA is not merely capturing random noise. Instead, it captures the increased **systemic correlation** and non-linear structure that appear during market stress.

6.2 Interpretation of Reproducing Figure 11

While our visual results for Figure 11 (EWS Indicators) show the same general trends as the original study, the distinctive smoothness of our spectral density curves serves as a validation of the signal's strength. We attribute this smoothness to the use of **Welch's method** in our Python implementation, which effectively reduces noise in the spectral estimation compared to standard periodogram methods.

Crucially, the **Low-Frequency Spectrum** proved to be the most resilient indicator. The fact that the rising trend remains clearly visible 250 days prior to the crashes, regardless of the specific spectral estimation technique confirms that this is a fundamental feature of the data. It suggests that the shifting of topological features toward lower frequencies is a universal precursor of imminent systemic transitions in financial markets, rather than an artifact of a specific numerical recipe.

7 Conclusion

In this report, we have successfully replicated the core empirical findings of the study on the topological data analysis of financial time series. By analyzing the daily log-returns of four major US stock market indices, we reached the following conclusions:

1. **Geometric Precursors:** TDA, specifically the persistent homology of loops (H_1), identifies structural changes in the market that are often invisible to traditional 1D price-based indicators.

2. **Peak Alignment:** The L^p-norms of persistence landscapes exhibit clear maxima during major financial meltdowns, with significant "fore-shocks" appearing months before the final crash.

3. **Effective EWS:** Statistical post-processing of the topological signal, specifically the average spectral density at low frequencies yields a reliable early warning signal. The rising trend of this indicator 250 trading days before the 2000 and 2008 crises provides a significant lead time for risk management.

4. **Robustness:** Despite minor variations in numerical output compared to the original paper, the qualitative findings of this methodology are robust and reproducible.

Overall, this replication confirms that Topological Data Analysis offers a promising new category of econometric tools. By treating financial markets as dynamic multidimensional shapes rather than mere collections of independent variables, we can better detect the subtle transitions from stability to fragility.

References

[1] Marian Gidea and Yuri Katz. Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834, February 2018.

[2] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16:77–102, 2015.

[3] Herbert Edelsbrunner and John L. Harer, *Computational topology: An introduction*, American Mathematical Society, 2010.

[4] The Gudhi Project. GUDHI User and Reference Manual. *GUDHI Editorial Board*, 2023. Available at: `https://gudhi.inria.fr/`