Analyzing Financial Time Series With Persistent Homology

Nishant DAVE
Ulysse WOLFF

Abstract. Our aim is to analyze the evolution of daily returns of four key US stock markets indices – DowJones, Nasdaq, Russell 2000, S&P500 – over the period from 1989 to 2016 using persistent homology. The objective of this project is to reproduce the experiments of Marian Gidea and Yuri Katz and discuss variants in the approach.

1 Article Overview

1.1 Context

Financial markets, characterized by their intricate dynamics and inherent unpredictability, have profound ramifications on global economies and societies. Periodic disruptions, such as the Dotcom Bubble (2000-2001), the 2008 Financial Crisis, and the 2012 Euro Debt Crisis, underscore the pressing need for proactive and predictive strategies to manage market volatility. Despite strides in econometrics and risk modeling, accurately forecasting market behavior, especially during volatile periods, remains a formidable challenge.

In this context, Marian Gidea and Yuri Katz have experimented using persistence homology to evidence changes in financial markets' behavior.

1.2 Methodology

Marian Gidea and Yuri Katz propose to analyze time series using daily log-returns of 4 major US stock market indexes: S&P500, Dow Jones, NASDAQ, Russell 2000.

The methodology consists of:

1) *Pre-processing data to obtain log-returns:* The adjusted closing prices from the selected US stock market indexes (a $n \times 4$ time series, with n the number of days covered) is processed into a a $(n - 1) \times 4$ matrix of log-returns.

2) *Selection of TDA parameters:* 2 parameters are selected: the sliding/scaling window w and the parameter p of the end L^p norm time series. In the article, $w \in \{50, 100\}$, $p \in \{1, 2\}$

3) *Extraction of time-dependent point cloud data sets:* As defined in the article, a point cloud data set Xi is a $w \times d$ matrix. Each of the d columns of a point cloud corresponds to a w-window-sized slice of one of the 4 stock market log-return time series.

4) *Measuring topological persistence:* For each point cloud, a Vietoris-Rips complex $R(X, \epsilon)$ is constructed at each ϵ-resolution (with $\epsilon > 0$)
Each complex $R(X, \epsilon)$ form a filtration and we can compute k-dimensional homology classes, part of the homology group $Hk(R(X, \epsilon))$.
In the article, a Vietoris-Rips complex filtration is computed for each of the $n - w$ point clouds.

5) *Encoding on a persistence diagram:* For each Vietoris-Rips complex filtration of the $n - w$ point clouds, a persistence diagram is computed by projecting the points $z\alpha$ of the homology group $Hk(R(X, \epsilon))$ on a 2D plot. The points $\{(x, y)|x = y\}$ are also projected on the same 2D plot. They represent the classes instantly born and dying at each level ϵ.

6) *Encoding on a persistence landscape* : The next step is the embedding of the n - w persistent diagrams a space of persistence landscapes.

7) *Computing L^p norms*: A L^p norm is finally computed for each persistence landscape, producing a time series of L^p norms. As stated previously, the article focuses only on L^1 and L^2.

2 Proposed approach

2.1 Data pre-processing

The first step of our work consisted of a data preprocessing. We used data of the four main US indexes and applied two preprocessing steps. The first was reversing the order of datasets to have data from the oldest entry to the most recent. The second was applying logarithmic returns to the adjusted close prices to follow the methodology defined in the article.

2.2 Computation of persistence diagrams

To analyze the topological features of the time series data, we employed time-delay embedding. This technique involves transforming the time series into a higher-dimensional space, where each point represents a snapshot of the system at a specific time.

Persistent homology captures the evolution of topological features across different scales. We utilized the Vietoris-Rips complex, a method for constructing simplicial complexes from point cloud data, to compute persistence diagrams (See Fig. 1). These diagrams represent the birth and death times of topological features (connected components, loops, etc.) in the data.

2.3 Replication of Section 4. experiments

Replicating experiments of Gidea and Katz using window sizes w of 40, 80, and 120 (instead of 50 or 100), we obtain very similar results to that of the article (See Fig. 2 in Annex for the resulting L1 and L2 norm time series).

While scales are not identical, we can identify similar trends on the 1000-trading-day snapshots the article highlighted (See Fig. 3 in Annex for the resulting L1 and L2 norm time series with a focus on the Dotcom Bubble Crash period), offering evidence of the article's reproducibility and validating the robustness of the methodology.

We can highlight the identified trends at or around market volatility events such as the Dotcom Bubble or the 2008 Financial Crisis by computing several statistics: 500-day rolling variance, low-pass filtering, first lag of autocorrelation, and the state identified by a Mann-Kendall test (See Fig. 4, 5, 6 in Annex for the resulting statistics for each window).

2.4 Other explorations

2.4.1 Bottleneck Distance Between Diagrams

Another interesting exploration is the computation of the bottleneck distance between consecutive diagrams, yielding a 1-dimensional time series that evidences large movements in terms of diagrams day-by-day.

Based on our results (Fig. 7) we find here that the bottleneck distance, though catching switches in behaviors at the time of a crisis does not seem to display much, if any, predictive capability as most spikes seem to happen after a financial trigger event.

2.4.2 Norm of the Difference Between Landscapes

We can also compute the norm of the difference between consecutive landscapes, yielding a 1-dimensional time series that evidences large movements in terms of landscapes day-by-day (See Fig. 8)

Based on the hypothesis that such a norm displays volatility behavior on financial markets, we obtain results similar to that of the previously introduced L^1 and L^2 norms of the persistence landscapes. We also find that the norm of the difference seems to highlight increases in volatility. However, the upward trends prior to an eventual financial crash, if these trends ever existed, are milder or absent compared to the article's method.

3 Conclusion

In summary, the article introduces a fresh perspective to the field of econometrics by devising metrics capable of identifying emerging adverse events, commonly referred to as early warning signs. Through a multi-step procedure leveraging Topological Data Analysis, it becomes possible to extract valuable insights, particularly volatility, from inherently noisy, multi-dimensional time series data. Our research demonstrates that persistence landscapes can uncover underlying patterns in financial data, which hold potential for offering insights into market volatility.

However, while the proposed methodology adeptly captures shifts in market conditions, its ability to predict future outcomes remains somewhat uncertain—a significant aspiration in econometrics. Although our method effectively detects volatility, its predictive capacity concerning the timing of crises, such as the Dotcom Bubble and the 2008 Crisis, is limited. Therefore, interpreting increases in volatility metrics as early indicators of impending crises should be approached cautiously.

Furthermore, as an avenue for further investigation, access to intraday data (such as stock indexes or individual company data) could enable the application of the proposed persistence landscape method to examine and highlight changes in log-return tail behavior. This aspect holds particular relevance in econometrics concerning regulatory metrics like Value-at-Risk or Expected Shortfall, which are often subject to mandatory reporting in North America and the European Union

Annex

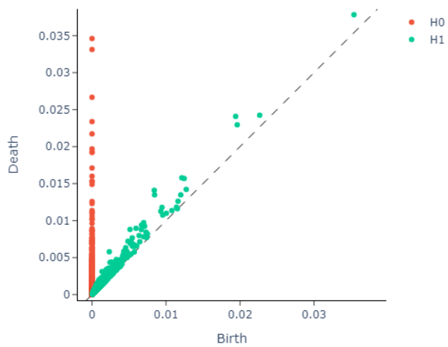

Fig. 1. Example of persistence diagram computed

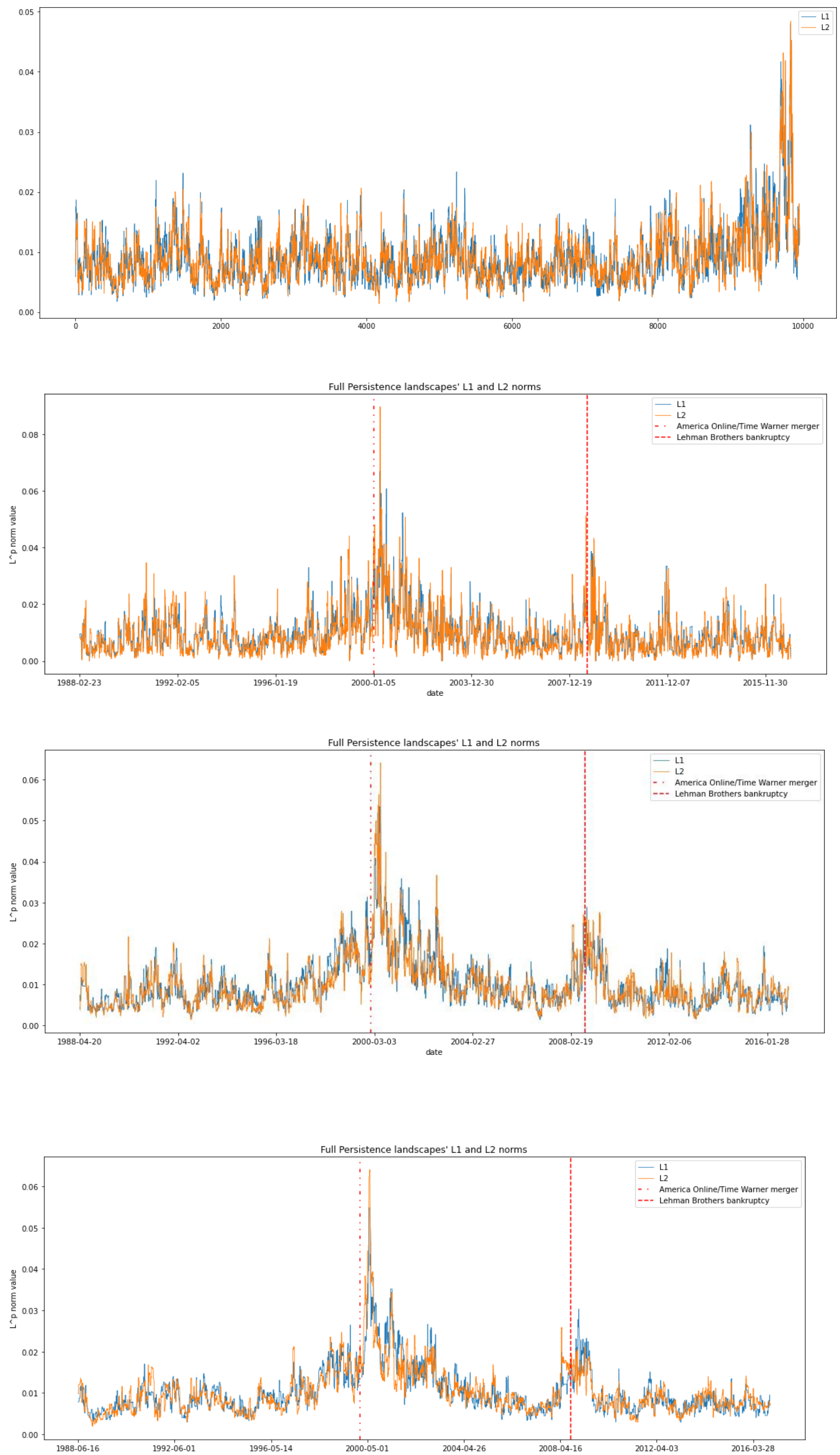

Fig. 2. Resulting $L^1$and L^2norms from performing the paper's persistence landscape workflow with window sizes, in order, of 40, 80, and 120

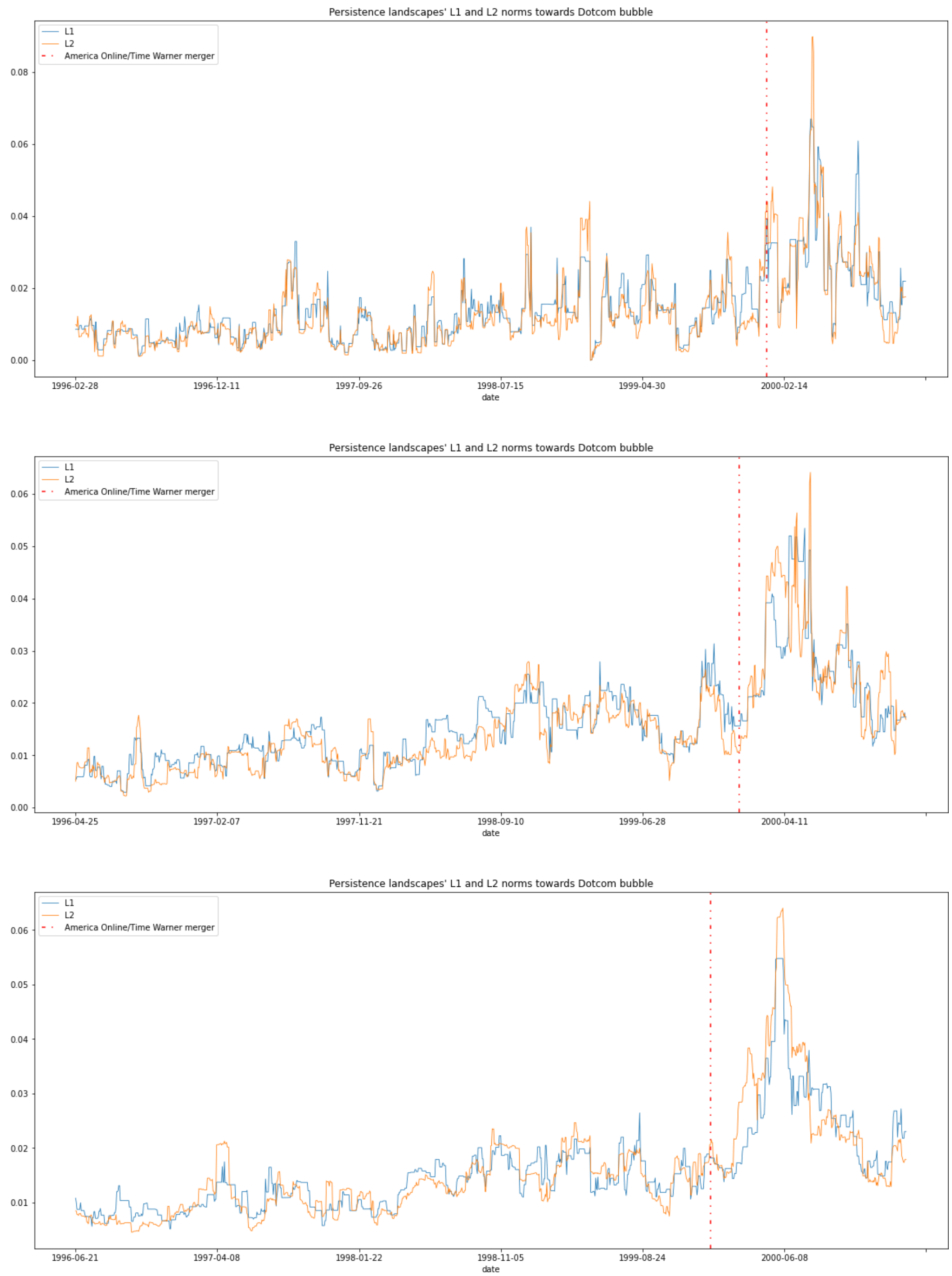

Fig. 3. Zoom on the period prior and during the Dotcom Bubble Crisis on the resulting L^1 and L^2 norms from performing the paper's persistence landscape workflow with window sizes, in order, of 40, 80, and 120

Fig. 4. 500-day rolling variance, low-pass filtering, first autocorrelation lag, Mann-Kendall-identified state of the L^1 and L^2 norms resulting from the paper's workflow with a window size of 40

Fig. 5. 500-day rolling variance, low-pass filtering, first autocorrelation lag, Mann-Kendall-identified state of the L^1 and L^2 norms resulting from the paper's workflow[1] with a window size of 80.

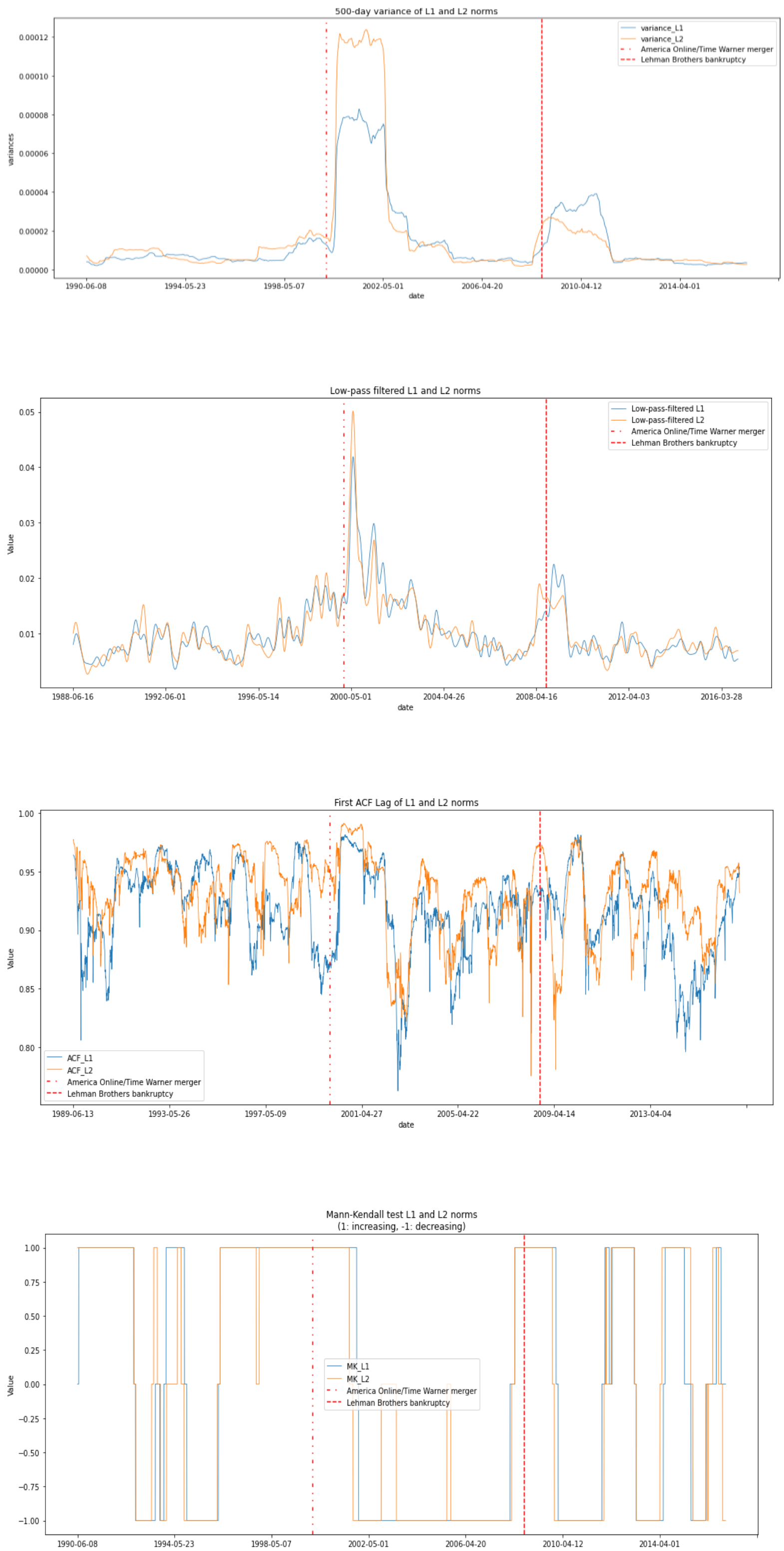

Fig. 6. 500-day rolling variance, low-pass filtering, first autocorrelation lag, Mann-Kendall-identified state of the L^1 and L^2 norms resulting from the paper's workflow with a window size of 120.

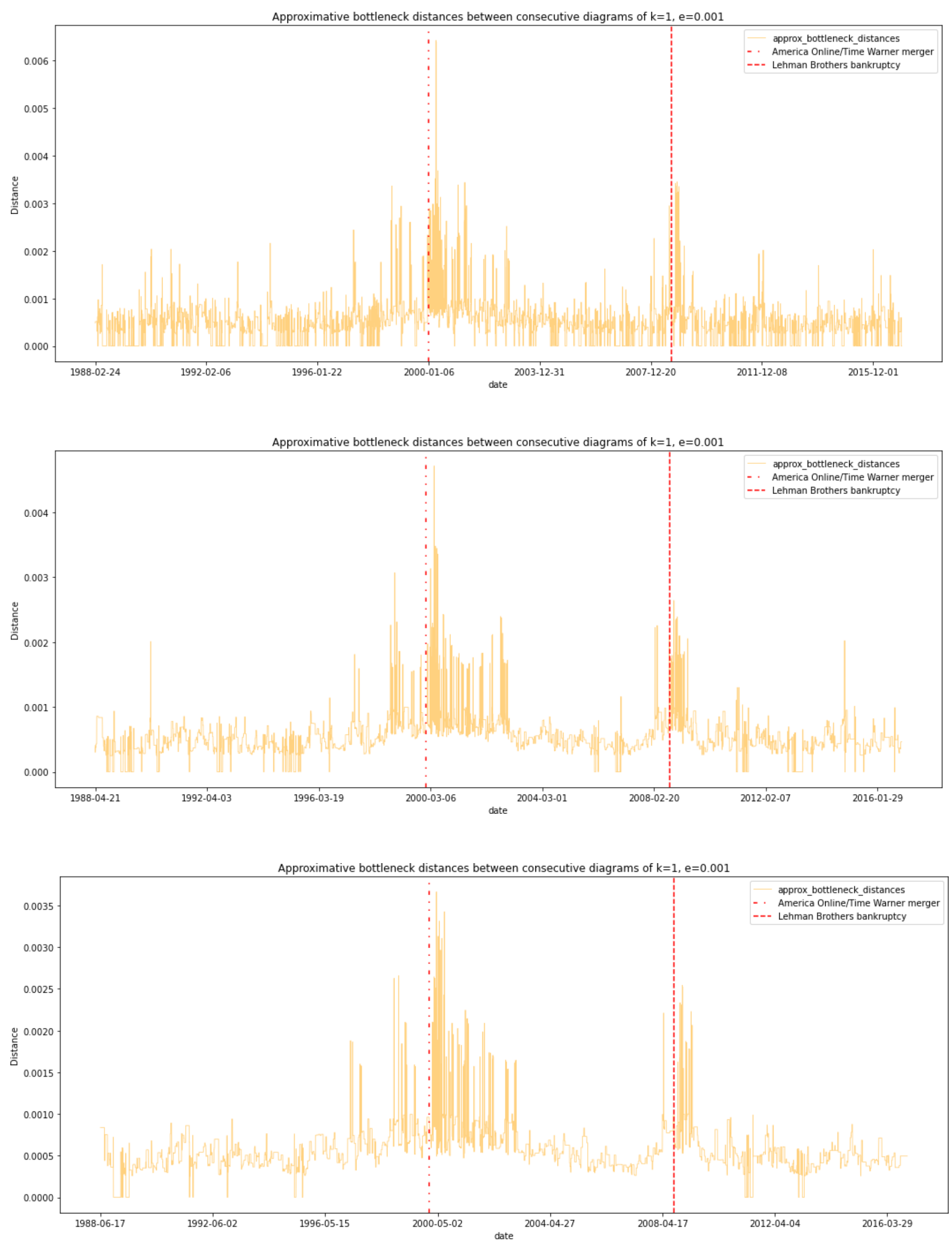

Fig. 7. Resulting bottleneck distance between persistence diagrams computed with window sizes, in order, of 40, 80, and 120

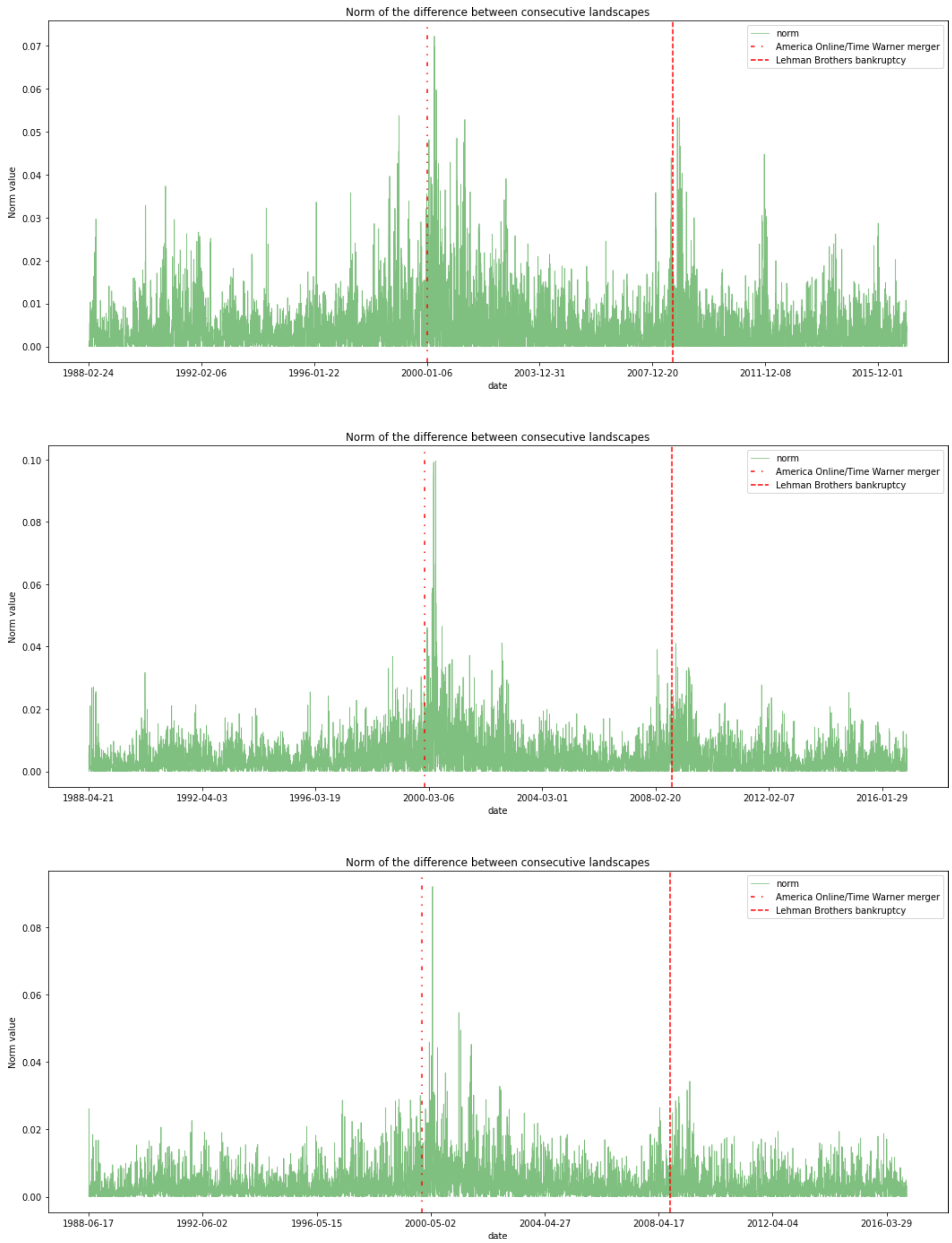

Fig. 8. Resulting norm of the difference between persistence landscapes computed with window sizes, in order, of 40, 80, and 12