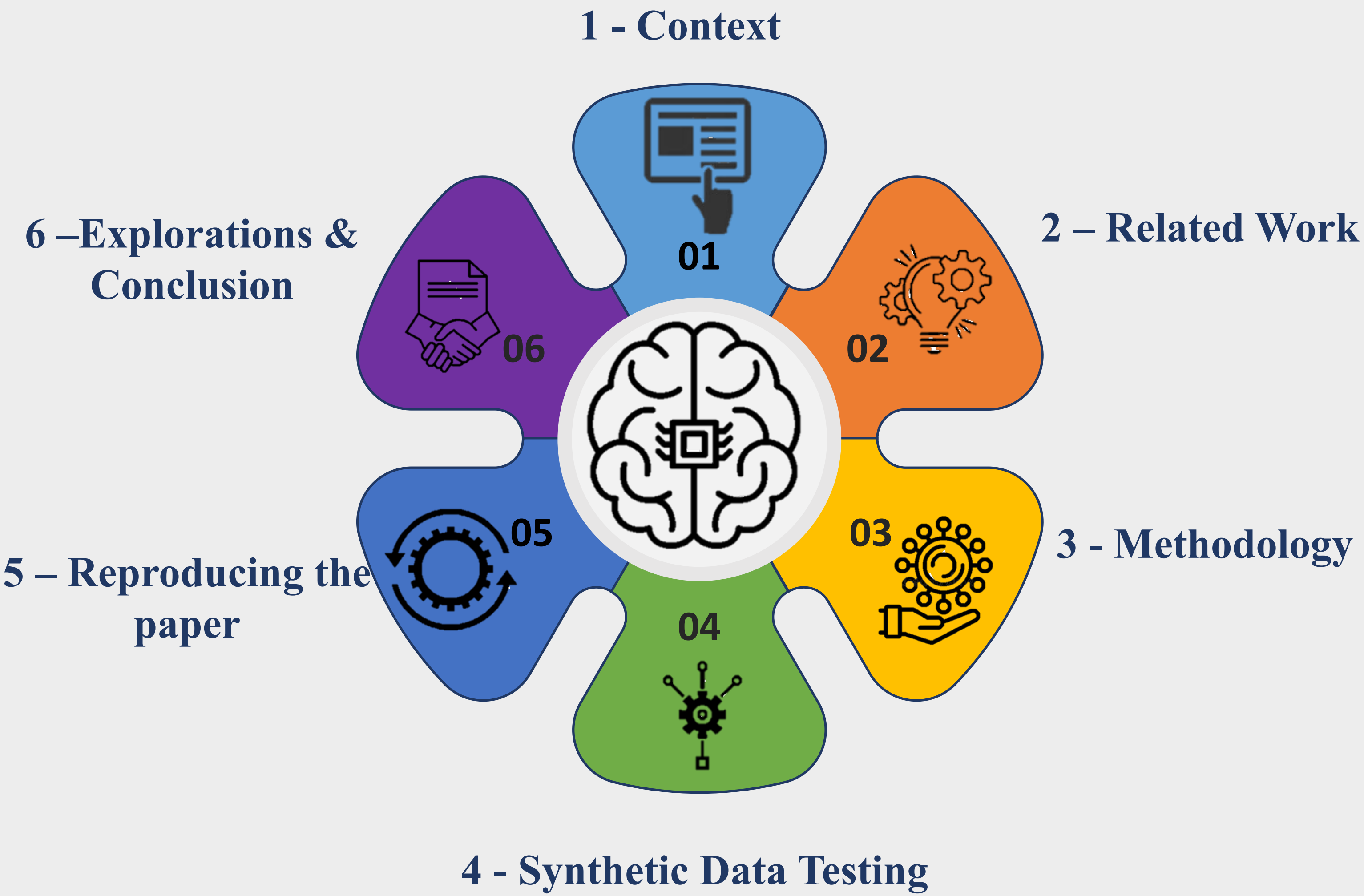
Analyzing Financial Time Series With Persistent Homology

Presented by
Elias Boughosn and Quentin Le Roux

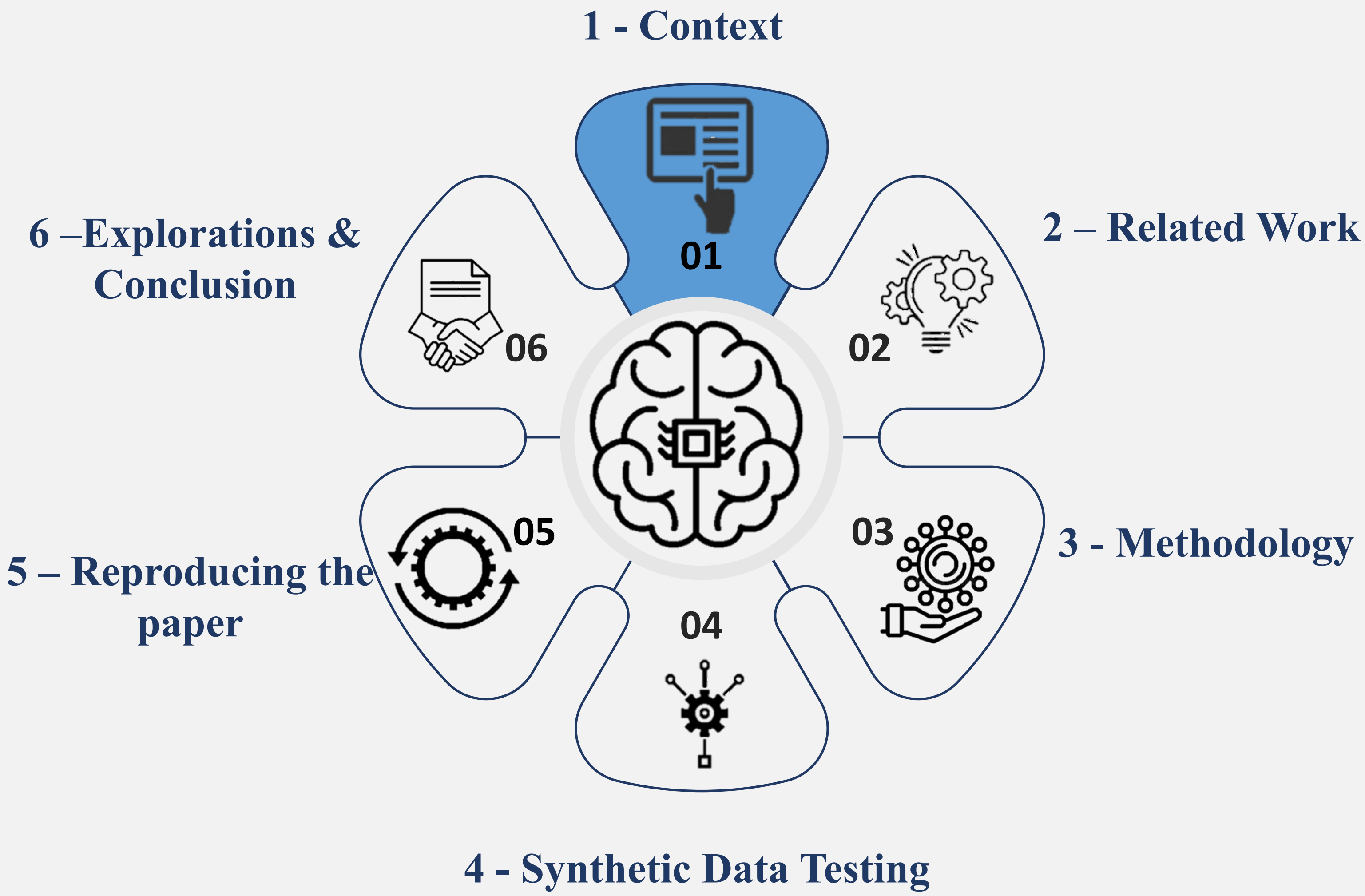
Course: Geometric And Topological Methods In Machine Learning
Instructors: Frederic Cazals, Mathieu Carriere, and Jean-Daniel Boissonnat

Summary



1 - Context
2 – Related Work
3 - Methodology
4 - Synthetic Data Testing
5 – Reproducing the paper
6 –Explorations & Conclusion
01
02
03
04
05
06

Context

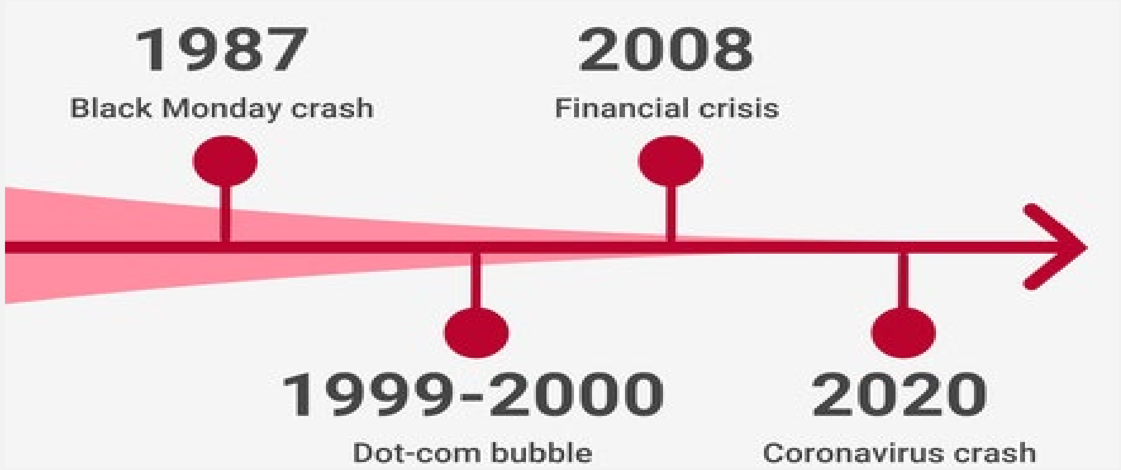


1 - Context
01
2 – Related Work
02
3 - Methodology
03
4 - Synthetic Data Testing
04
5 – Reproducing the paper
05
6 –Explorations & Conclusion
06

1 - Context

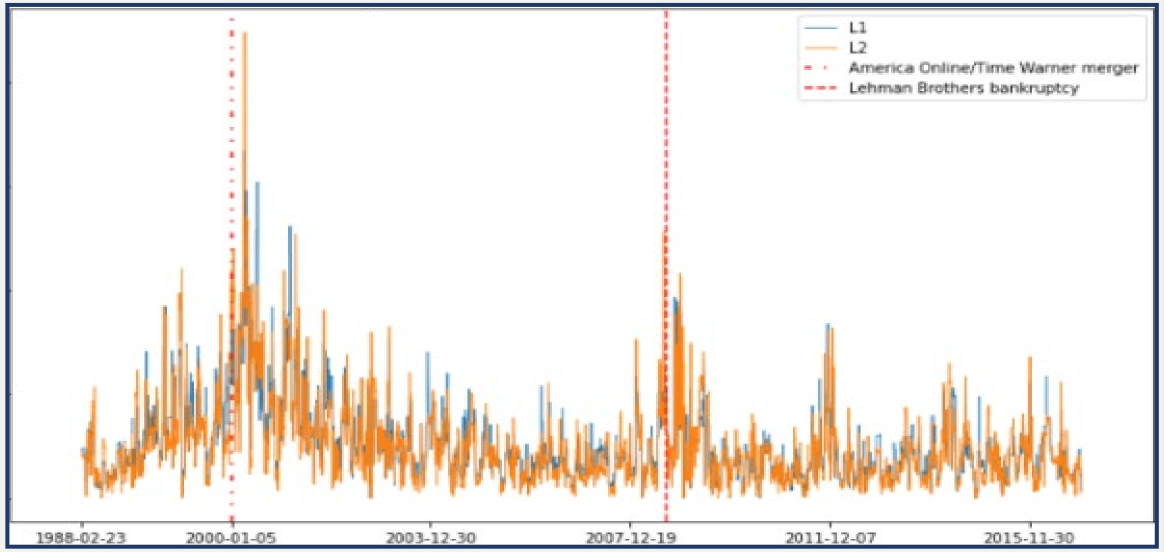
Motivation

- Sharp changes in the behavior of financial markets
- Destabilization of economies, countries, and people’s lives
- Absence of a short-term forecasting of approaching financial disasters



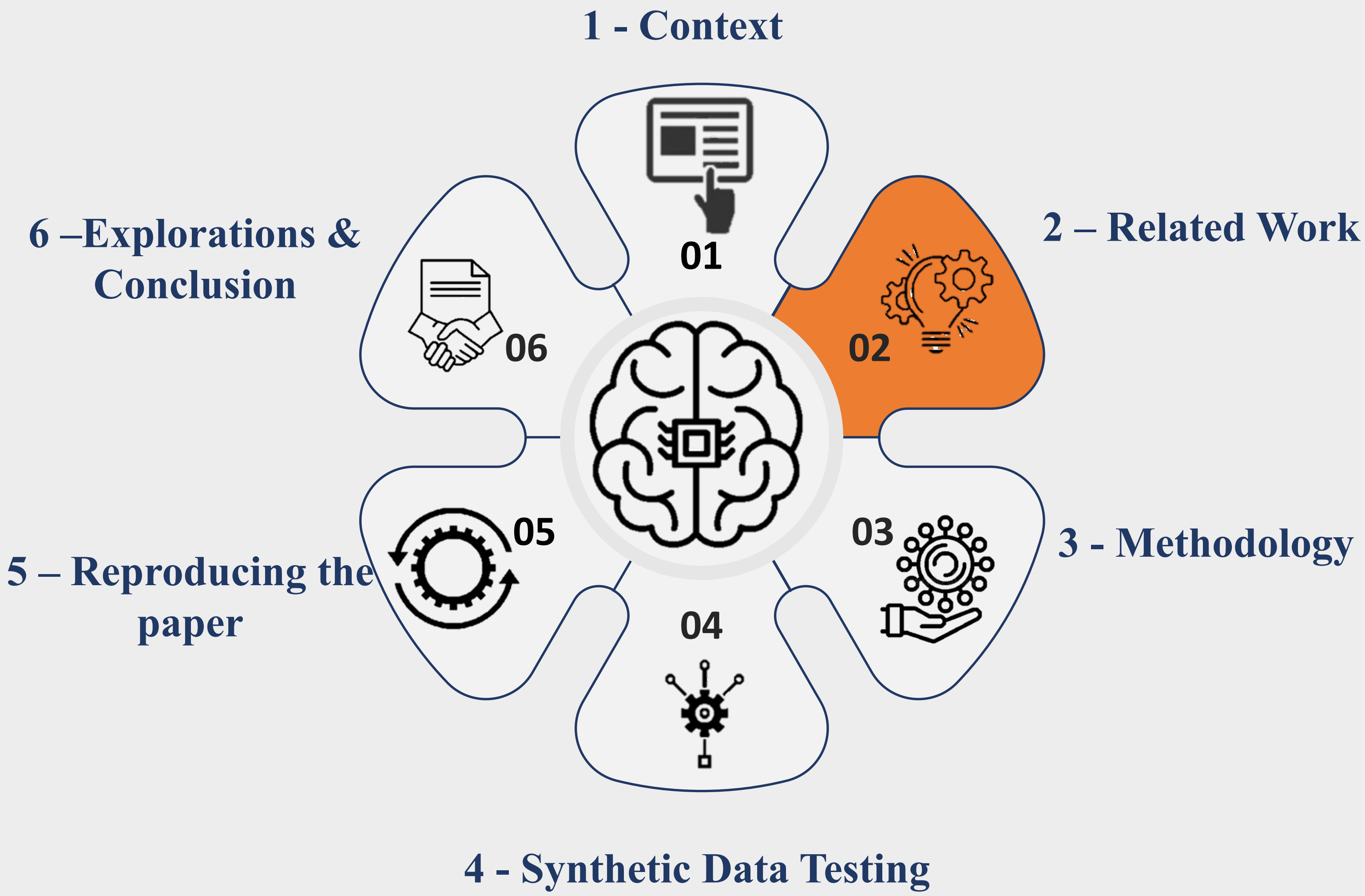
Solution

Topological Data Analysis (TDA) for econometrics



- Predicting rises in volatility
- Anticipating subsequent market crashes

Related Work



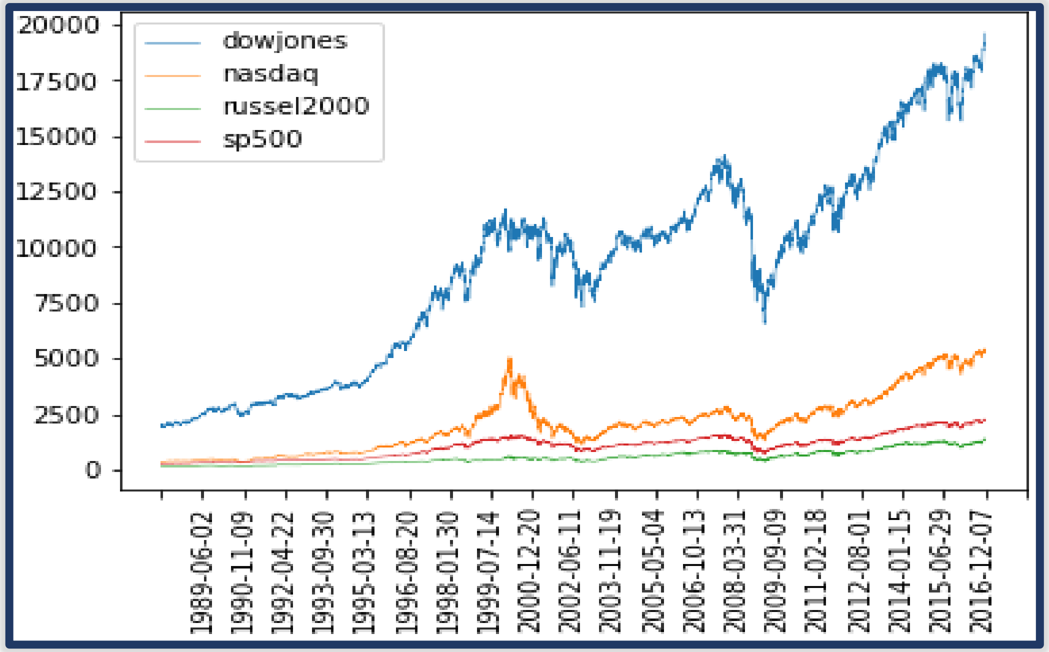
2 – Related Work

Topological Data Analysis of Financial Time Series: Landscapes of Crashes
By Marian Gidea and Yuri Katz

Dataset

Workflow

➢ **Time series** from the daily log-returns of 4 key US stock market indexes
- **Source:** daily adjusted closing prices from Yahoo Finance
- **Name:** S&P500, Dow Jones, NASDAQ, Russell 2000



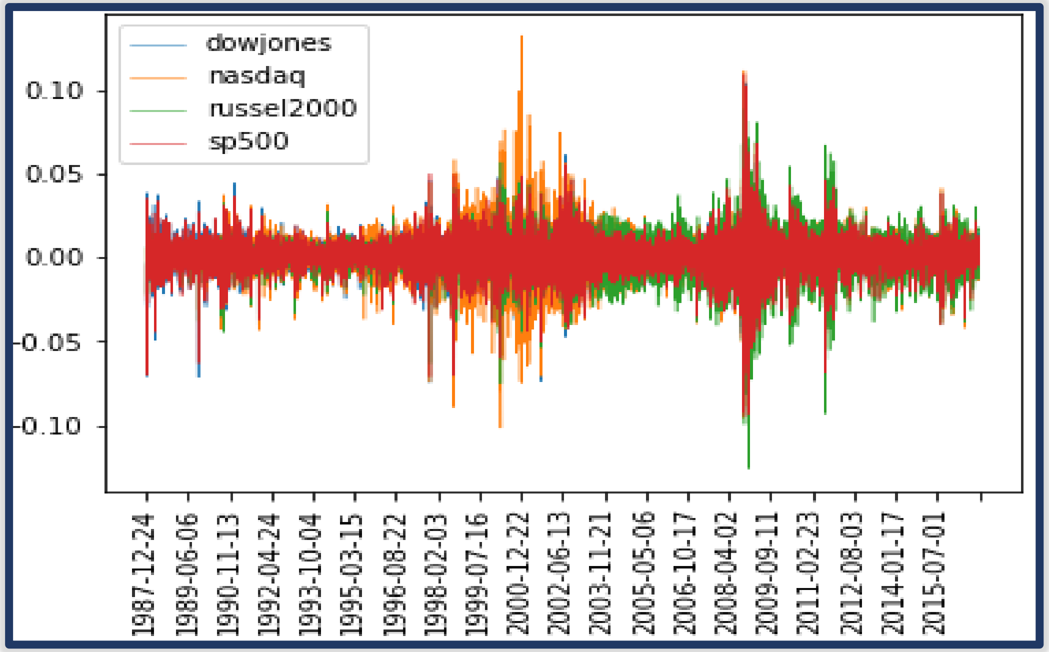


Fig1: Adjusted Closing prices and corresponding log-returns of the 4 key US stock indexes from 1989 to 2016.

2 – Related Work

Topological Data Analysis of Financial Time Series: Landscapes of Crashes
By Marian Gidea and Yuri Katz

Dataset | **Workflow**

➢ **From price to log-return:**

The adjusted closing prices is processed into a $(n-1)\times 4$ matrix of log-returns.

We build the times series R of returns r_i such that:

$$\forall i \in \{1,\ldots,n\},\ r_i = \frac{p_i - p_{i-1}}{p_{i-1}}$$

$$r_i^{log} = log(p_i) - log(p_{i-1})$$

➢ **TDA parameters selection:**

- The sliding/scaling window **w ∈{50,100}**
- L^p norm time series **(p = 1 and p = 2)**.

2 – Related Work

Topological Data Analysis of Financial Time Series: Landscapes of Crashes
By Marian Gidea and Yuri Katz

Dataset

Workflow

- **Extraction of time-dependent point cloud data sets:**

1. A point cloud data set $\mathbf{X_i}$ as a $\mathbf{w \times d}$ matrix
2. The **4D** log-return time-series matrix is split into subsets (point cloud data sets) of size $\mathbf{w \times d}$ with a time step of **1 day**
3. Time-ordered sequence of $\mathbf{n - w}$ point clouds

- **Measuring topological persistence with persistence homology:**

  - Vietoris-Rips complex filtration for each of the $\mathbf{n - w}$ point clouds

Fig 2: Rips filtration of simplicial complexes.

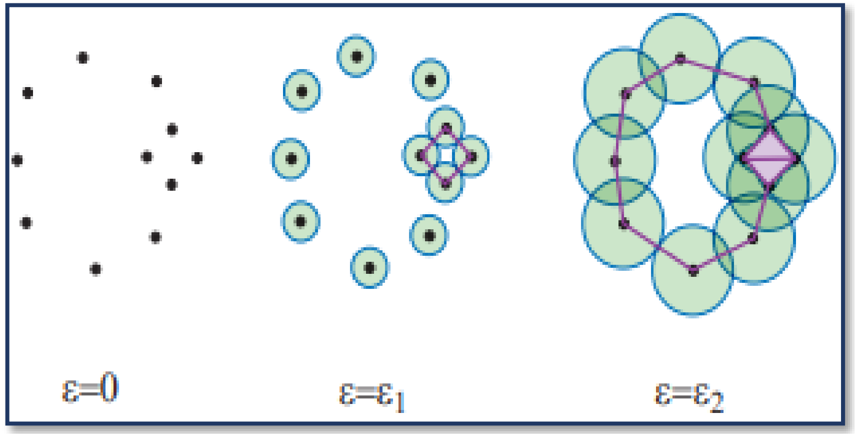
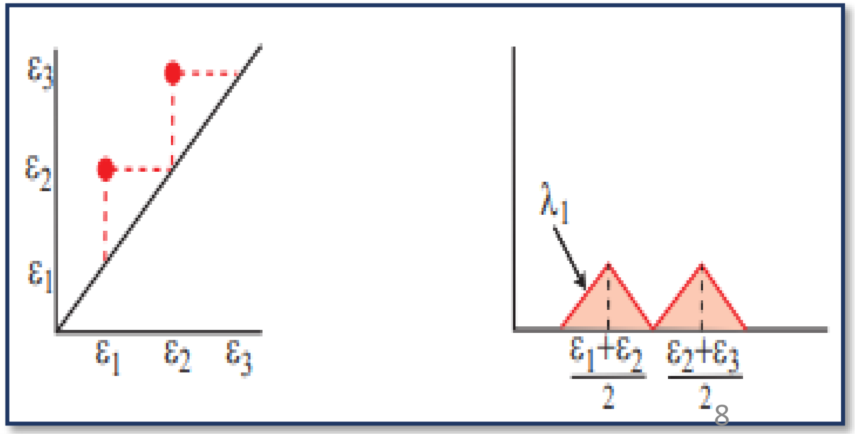


Fig 3: Corresponding persistence diagram and persistence landscape.



2 – Related Work

Topological Data Analysis of Financial Time Series: Landscapes of Crashes
By Marian Gidea and Yuri Katz

Dataset

Workflow

➢ **Encoding in a persistence diagram:**

- From each complex, one can compute k-dimensional homology classes, part of the homology group $H_{\{k\}}(R(X, \epsilon))$ and construct a persistence diagram $\mathbf{P_k}$

➢ **Encoding in a persistence landscape:**

- The persistence landscape of $\mathbf{P_k}$ is the sequence of functions $\boldsymbol{\lambda(x)}$.
- The x and y coordinates of landscapes correspond to the rescaled axes: **x = (d+b)/2** and **y=(d-b)/2**

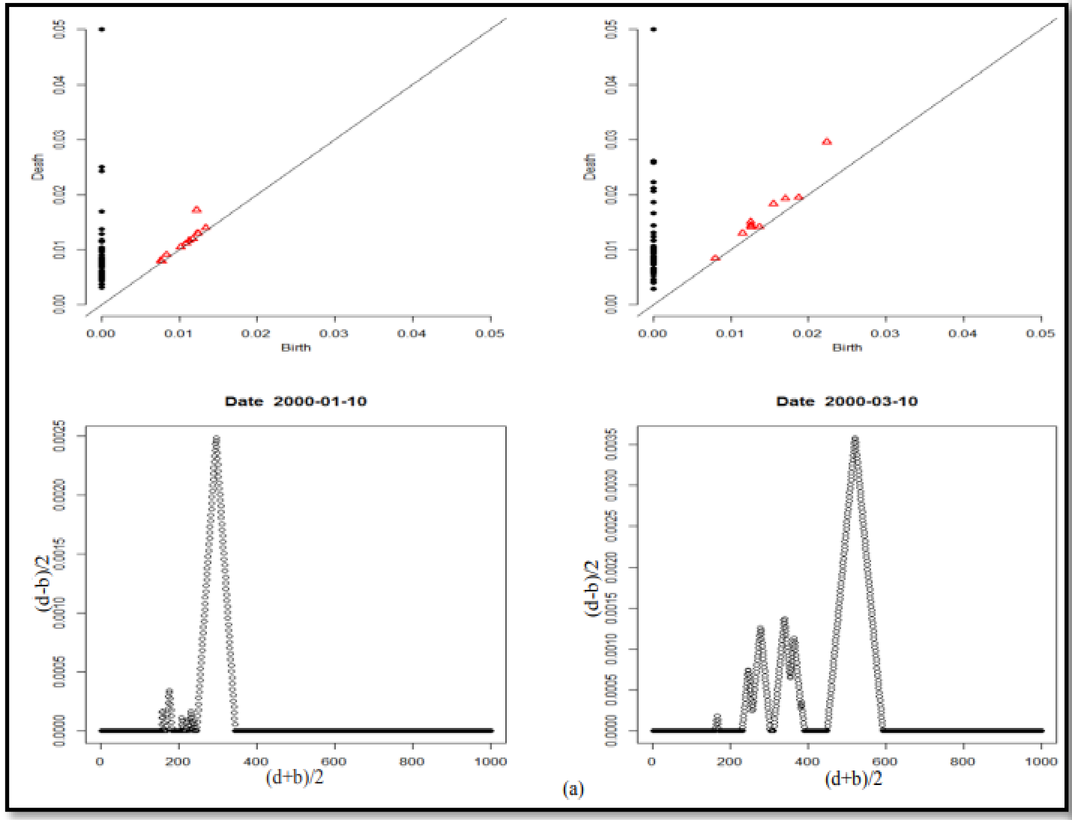


Fig 4: The Rips persistence diagrams and the corresponding persistence landscapes.

2 – Related Work

Topological Data Analysis of Financial Time Series: Landscapes of Crashes
By Marian Gidea and Yuri Katz

Dataset

Workflow

- **Computing of L^P norm times series:**
  - **L^P norm** can be computed for each obtained persistence landscape (one per window)

$$L^p\text{-norm of } \eta = ||\eta||_p = \left(\sum_{i=1}^{\infty} ||\eta_k||_p^p\right)^{\frac{1}{p}}$$

- **Visualization of potential trends with statistics:**
  - The Mann-Kendall test is proposed to assert monotonic up-ward or downward movement trends in the L1 and L2 time series

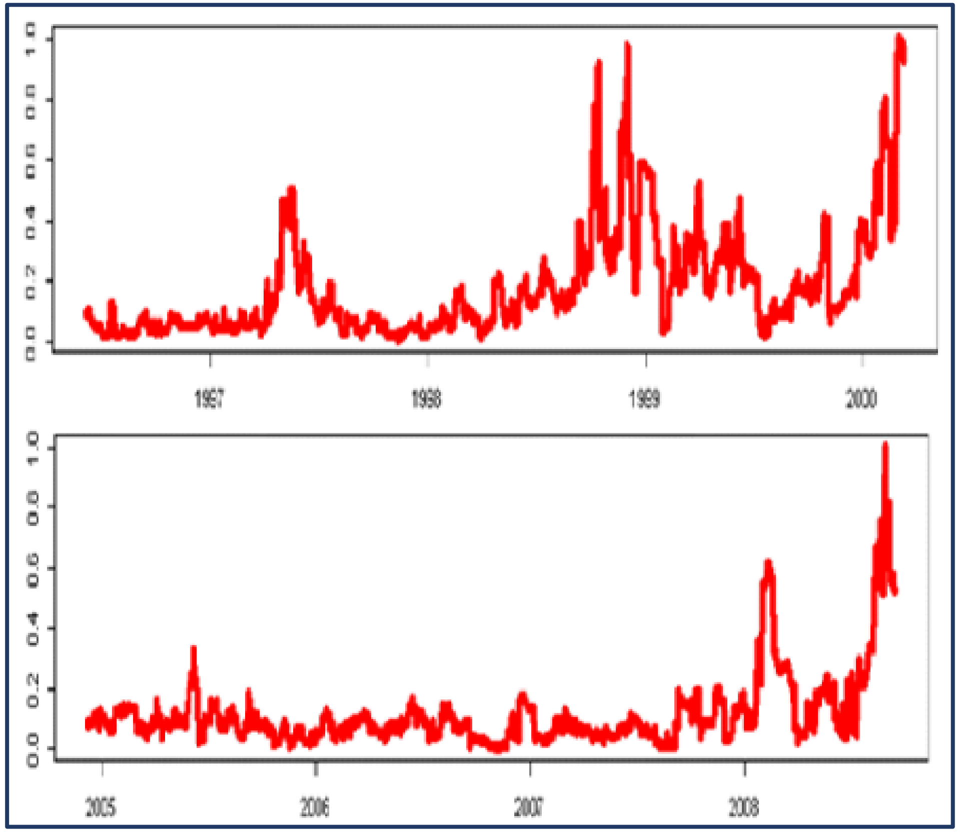
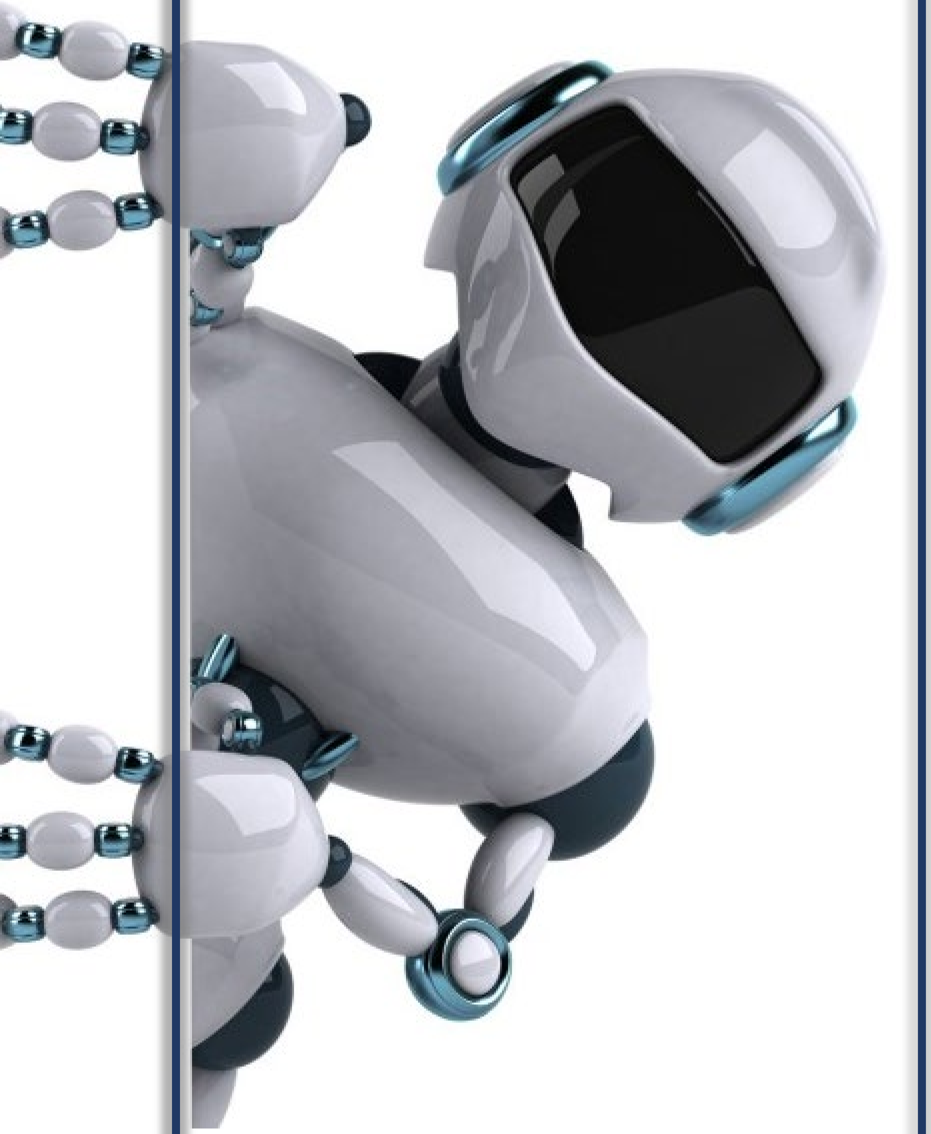


Fig 5: Normalized L 1 -norm of topological landscapes calculated with the sliding window of 50 days for dotcom crash and Lehman bankruptcy.

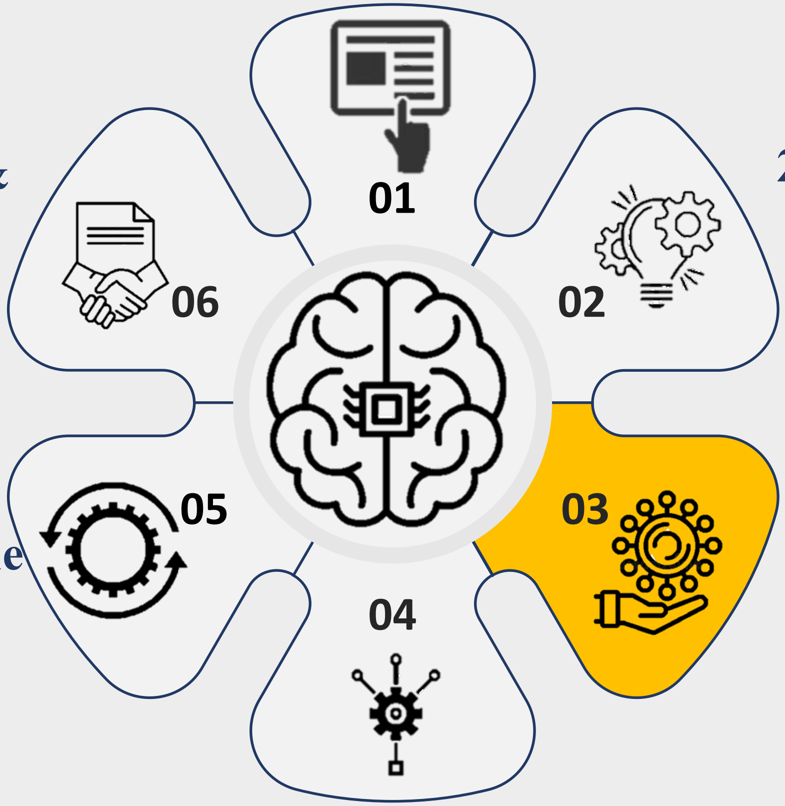


Methodology

1 - Context

6 –Explorations & Conclusion

2 – Related Work



5 – Reproducing the paper

3 - Methodology

4 - Synthetic Data Testing

3 - Methodology

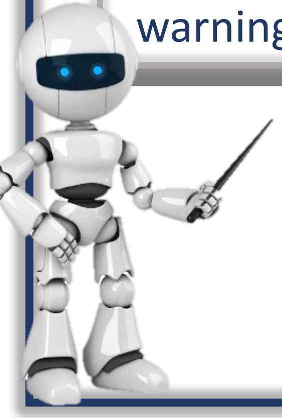
Topological Data Analysis of Financial Time Series: Landscapes of Crashes
By Marian Gidea and Yuri Katz

Objective: Use persistence homology to detect topological patterns in multidimensional time series

Steps:

1. Extract time-dependent point cloud data sets.
2. Detect transient loops and measure their persistence
3. Quantify the temporal changes via $\mathbf{L^P}$ **norm**
4. Test on multidimensional time series generated by non-linear and non-equilibrium models

Conclusion: The method can be used to detect early warning signals of imminent market crashes.



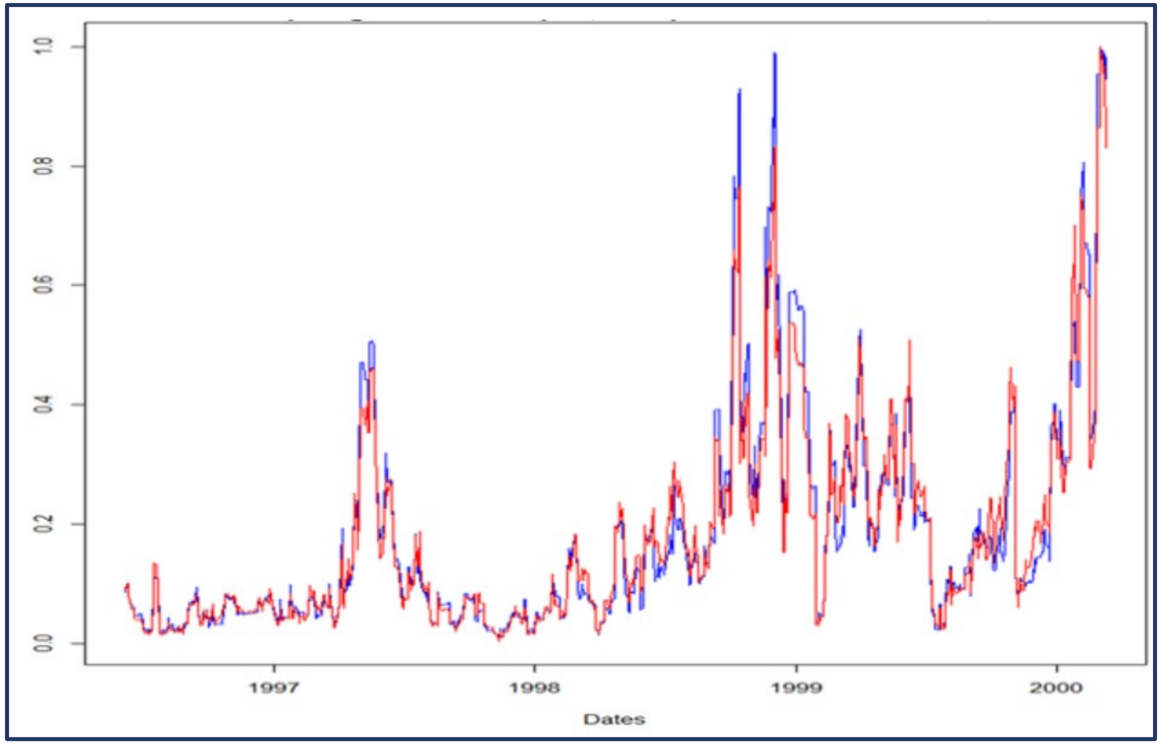


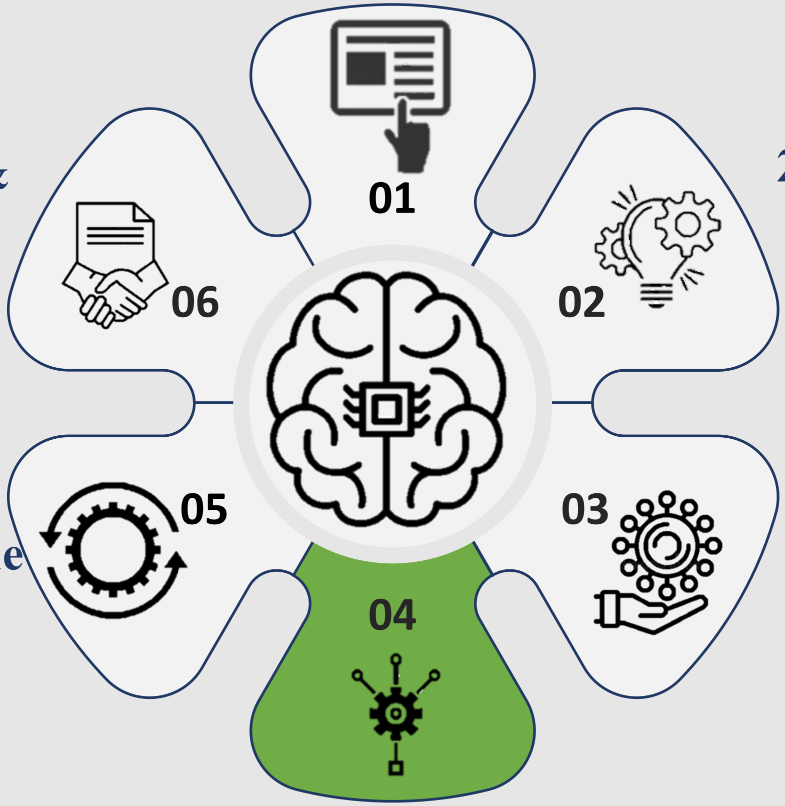
Fig 6: Resulting L1 and L2 norms from performing the paper's persistence landscape workflow with window sizes of 50.

Synthetic Data Testing

1 - Context

6 –Explorations & Conclusion

2 – Related Work



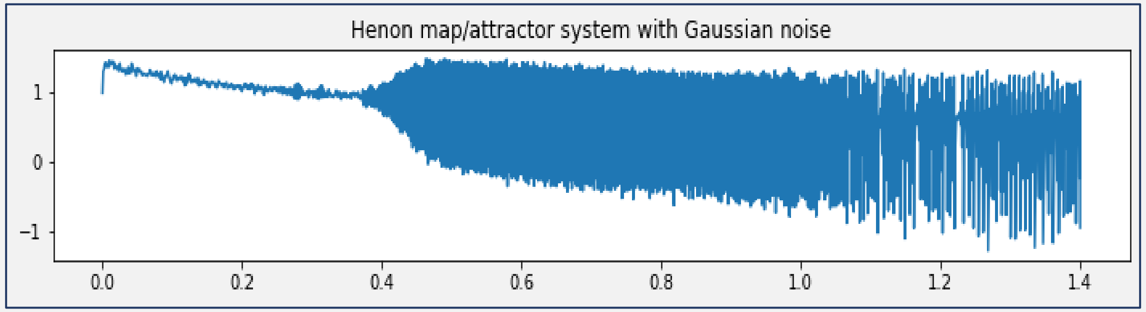
5 – Reproducing the paper

3 - Methodology

4 - Synthetic Data Testing

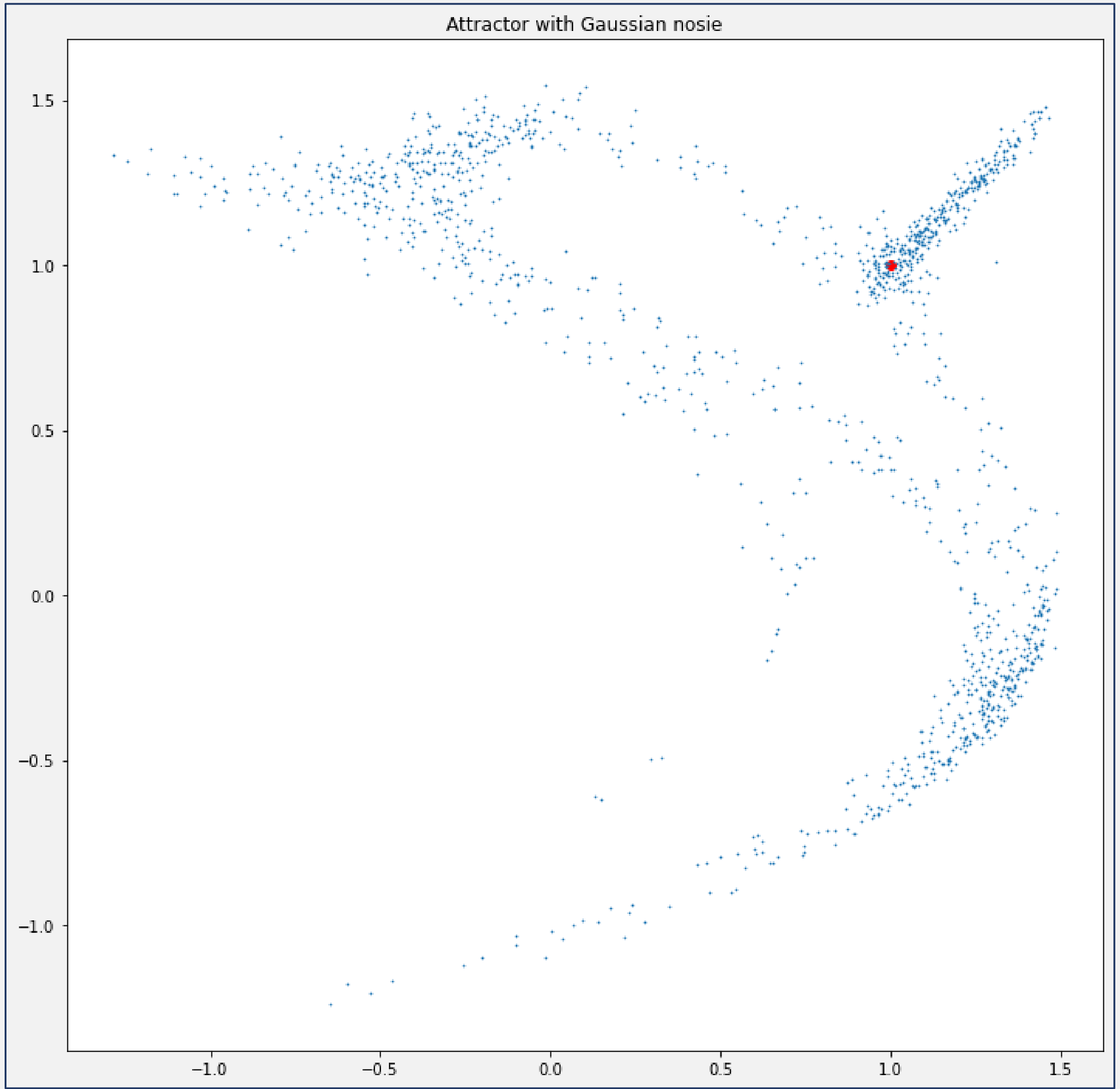
4 - Synthetic Data Testing

Hénon Maps with added noise



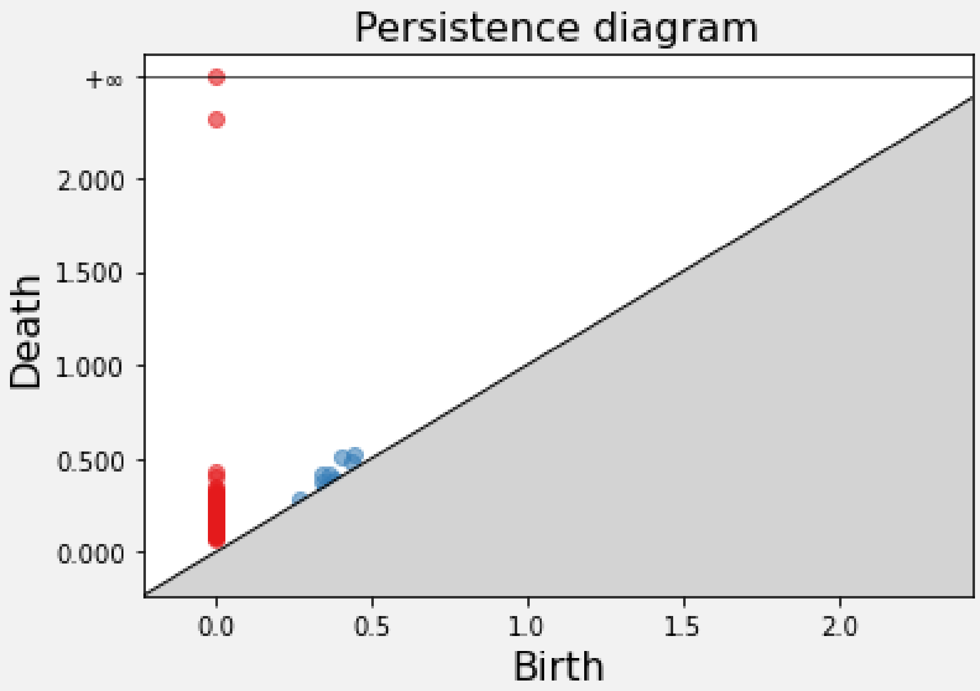
$$x_{n+1} = 1 - ax_n^2 + by_n$$
$$y_{n+1} = x_{n+1}$$

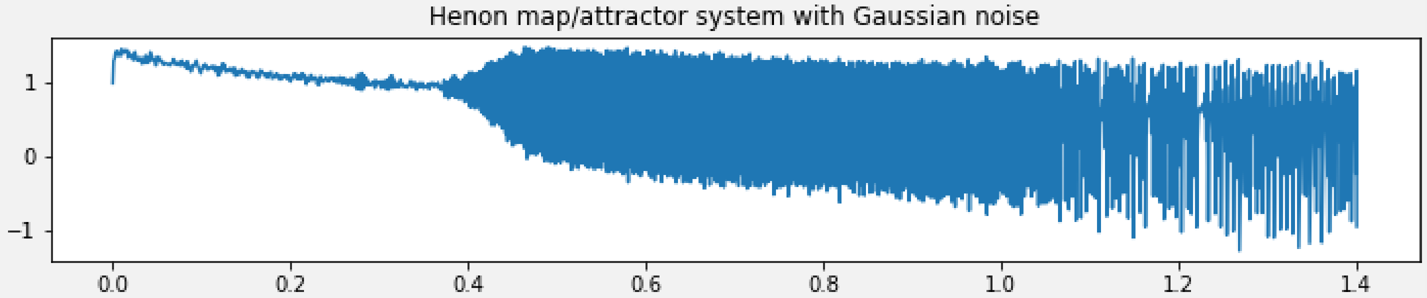
$$a_0 = 0$$
$$\sigma > 0$$
$$W_n \sim \mathcal{N}(0, 1)$$
$$x_{n+1} = 1 - a_n x_n^2 + by_n + \sigma W_n \sqrt{\Delta t}$$
$$y_{n+1} = x_{n+1} + \sigma W_n \sqrt{\Delta t}$$
$$a_{n+1} = a_n + \sqrt{\Delta t}$$

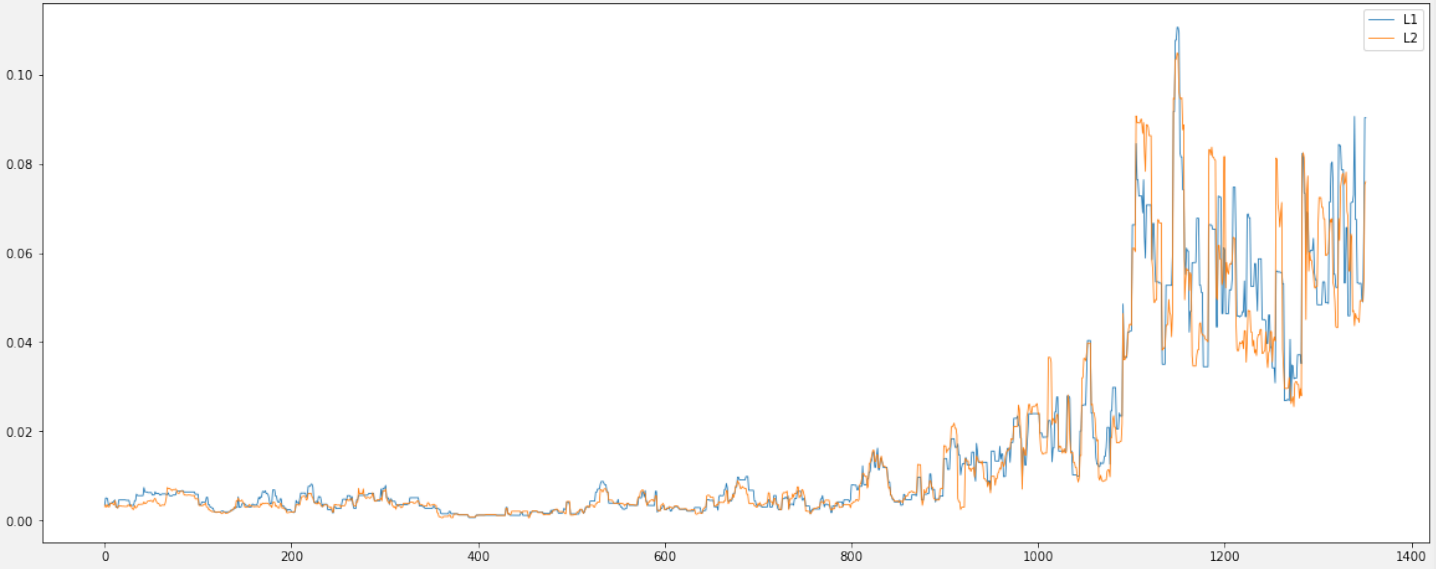


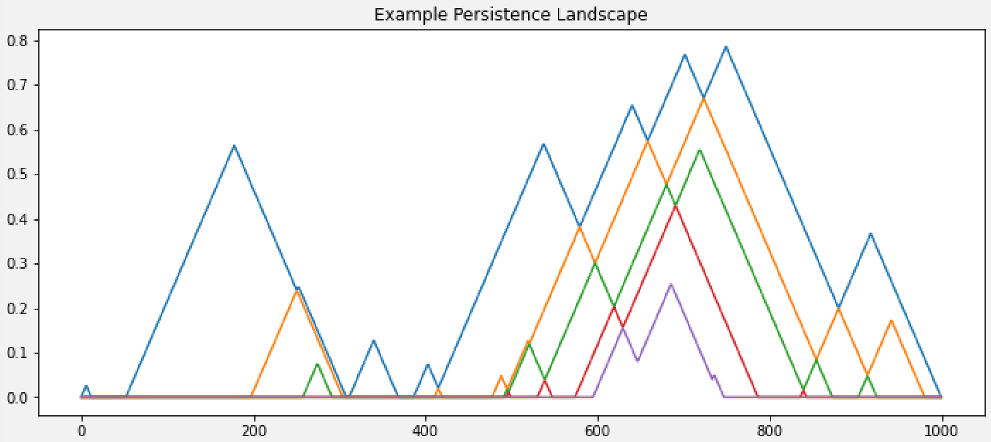
4 - Synthetic Data Testing

Hénon Maps with added noise



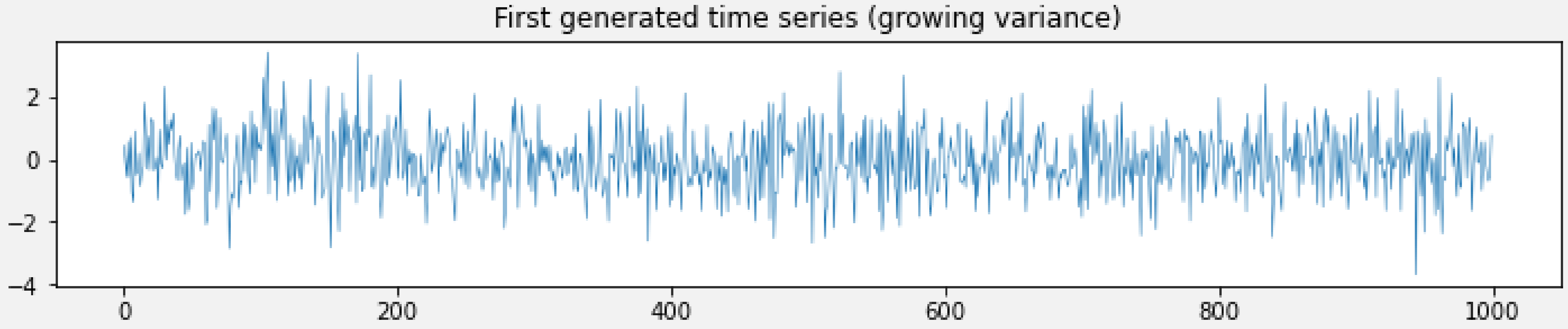






4 - Synthetic Data Testing

White Noise with Growing Variance



Number of simulations: 10 such that $\sigma \in \{1, 2, \dots, 10\}$

The below procedure is repeated for as many times as needed:

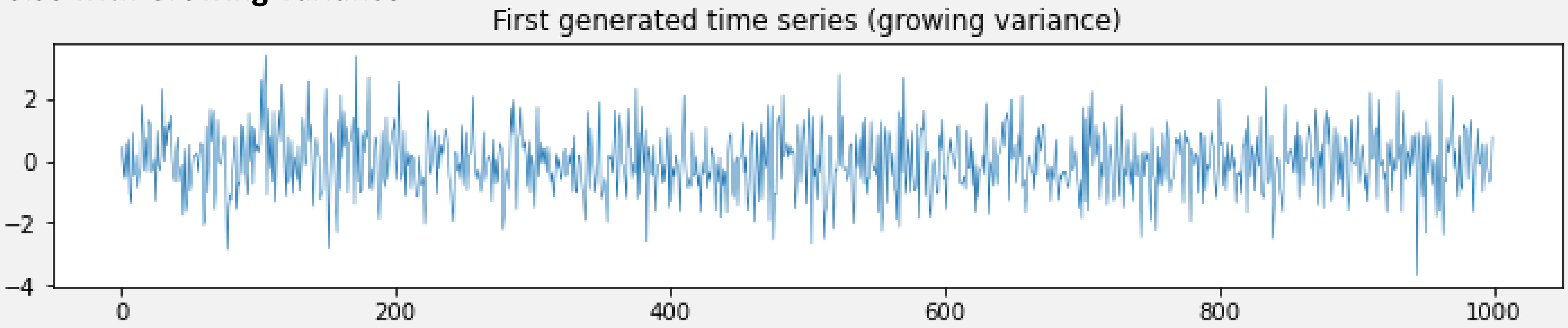
- Generate 4 time series of 100 data points. This results in a 400-point cloud data set is called a "realization" X_i such that:

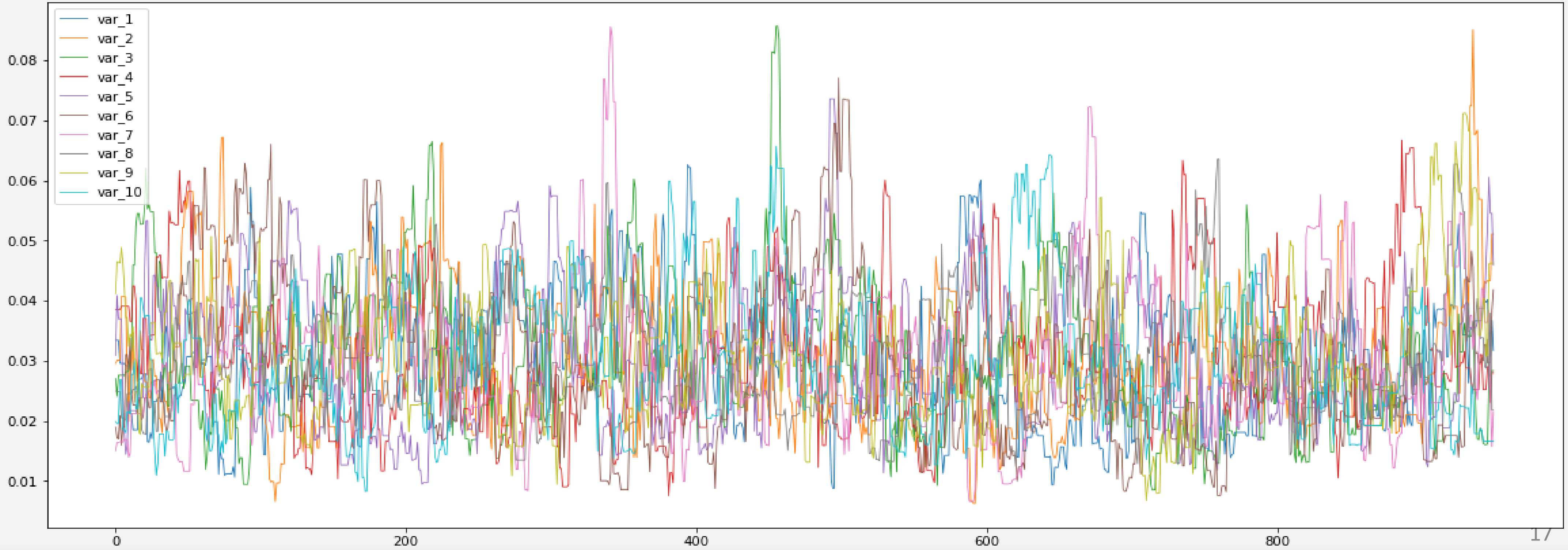
$$\forall i \in \{1, 2, 3, 4\}$$
$$\forall j \in \{1, 2, \dots, 100\}$$
$$\delta_i \sim \text{Unif}[-0.1, 0.1]$$
$$\forall x_{i,j} \in X_i, x_{i,j} \sim \mathcal{N}(0, (\sigma + \delta_i)^2)$$

- Perform the paper's workflow on the realization and yield the corresponding L^1 and L^2 norms
- Collect the L^1 and L^2 norms per realization and compute their mean-value

4 - Synthetic Data Testing

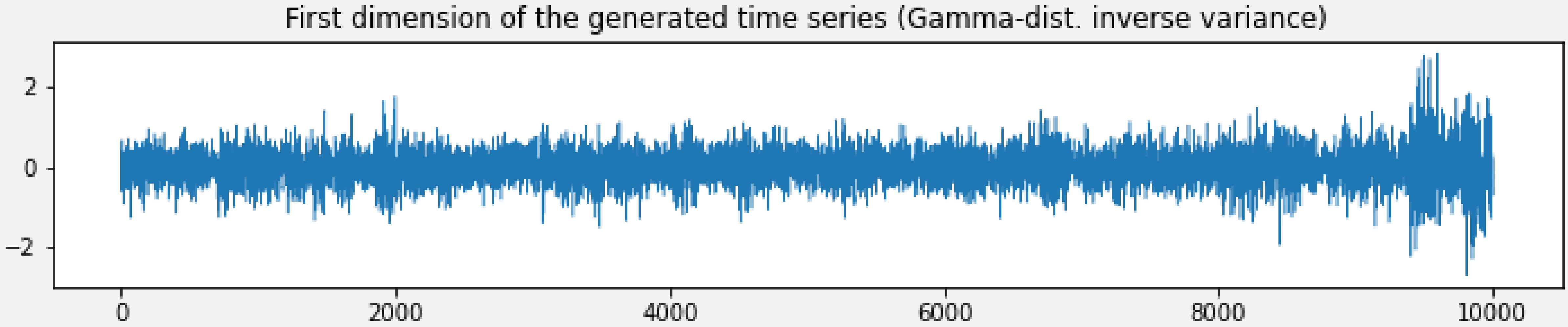
White Noise with Growing Variance





4 - Synthetic Data Testing

White Noise with Gamma-distributed inverse variance



$$\forall x \in X, x \sim \mathcal{N}(0, \frac{1}{\sqrt{\gamma}})$$

$$\gamma \sim \Gamma(\alpha, \beta)$$

$$f_{\Gamma(\alpha,\beta)}(\gamma, \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\gamma^{\alpha-1}exp(-\beta\gamma), \;\; \gamma \geq 0, \alpha, \beta > 0$$

The following procedure is repeated 100 times (producing 100-point cloud data set of 100 data points each):

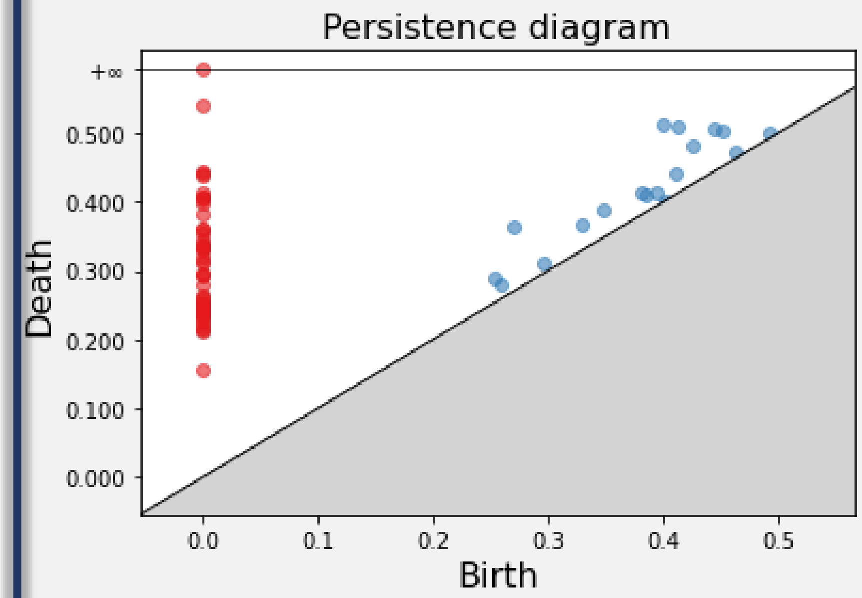
- Generate a 4D time series of 100 data points. This results in a 400-point cloud data set is called a "realization" X_i such that:

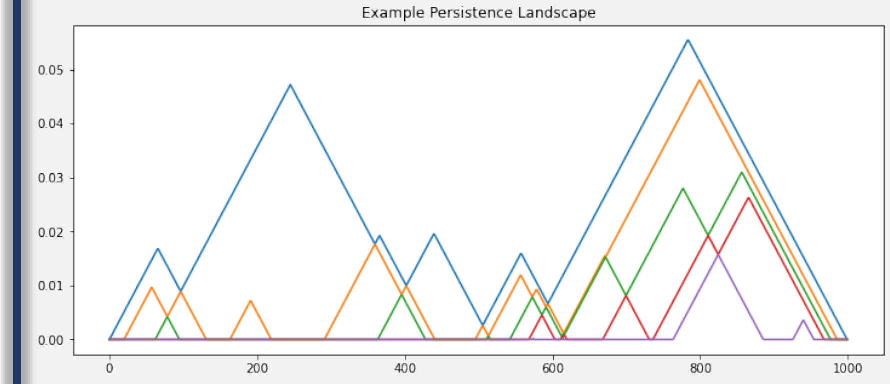
$$\begin{aligned} &\forall i \in \{1, 2, 3, 4\} \\ &\forall j \in \{1, 2, \ldots, 100\} \\ &\gamma_i \sim \Gamma(\alpha, \beta) \\ &\beta = 1 \\ &\alpha = 8 \text{ for the first 75 realizations, then decreases by 0.25-step} \end{aligned}$$

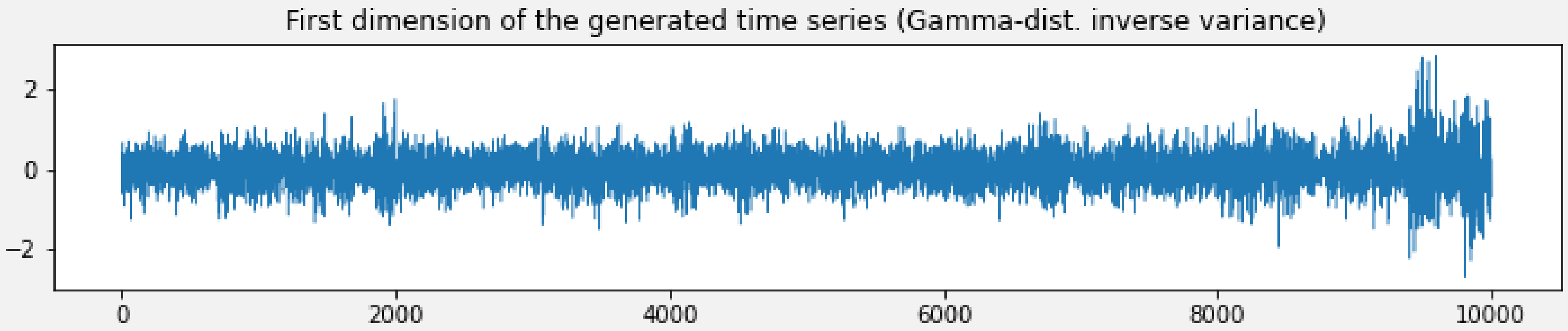
$$\forall x_{i,j} \in X_i, x_{i,j} \sim \mathcal{N}(0, \frac{1}{\sqrt{\gamma}})$$

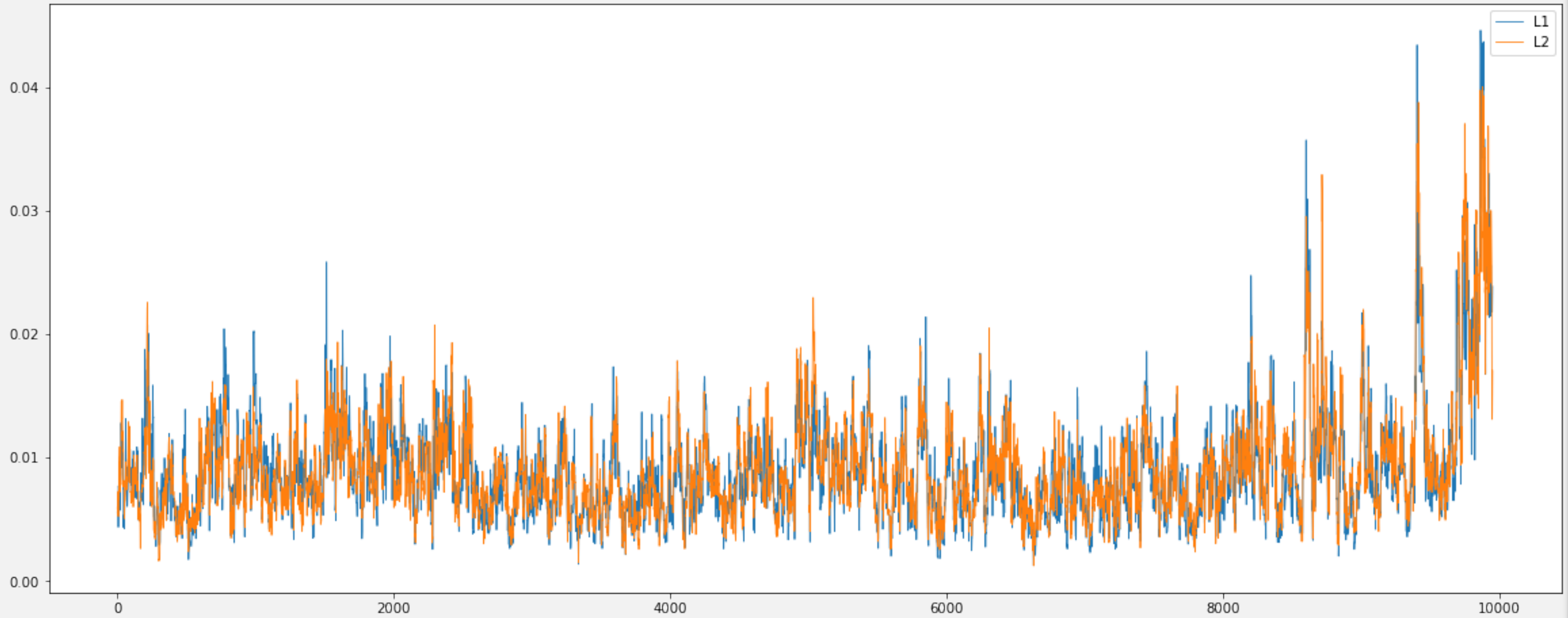
- Perform the paper's workflow on the realization and yield the corresponding L^1 and L^2 norms
- Collect the L^1 and L^2 norms per realization and compute their mean-value

White Noise with Gamma-distributed inverse variance









Reproducing the paper

1 - Context

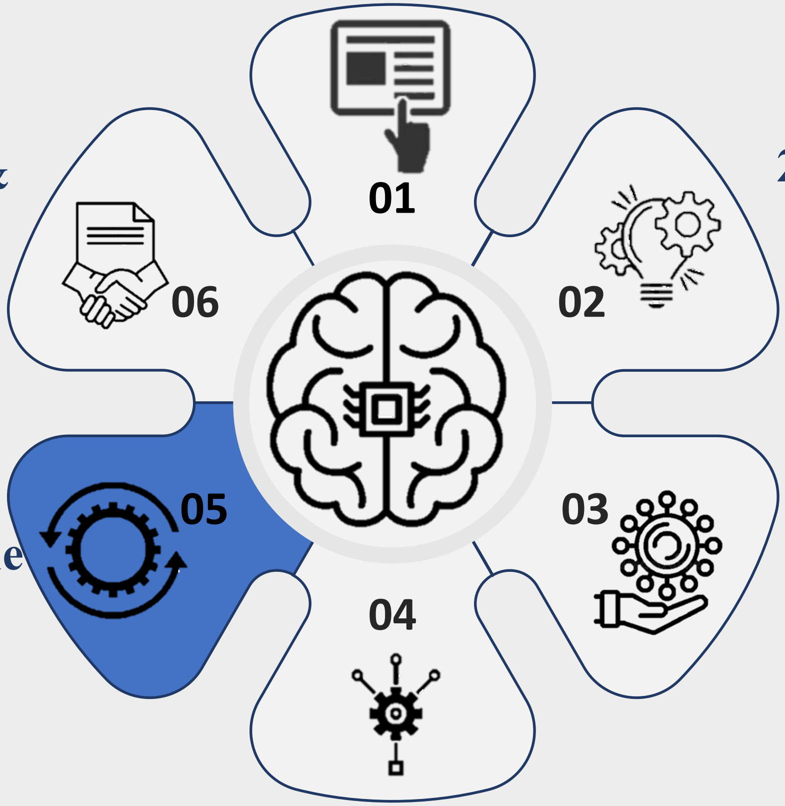
2 – Related Work

3 - Methodology

4 - Synthetic Data Testing

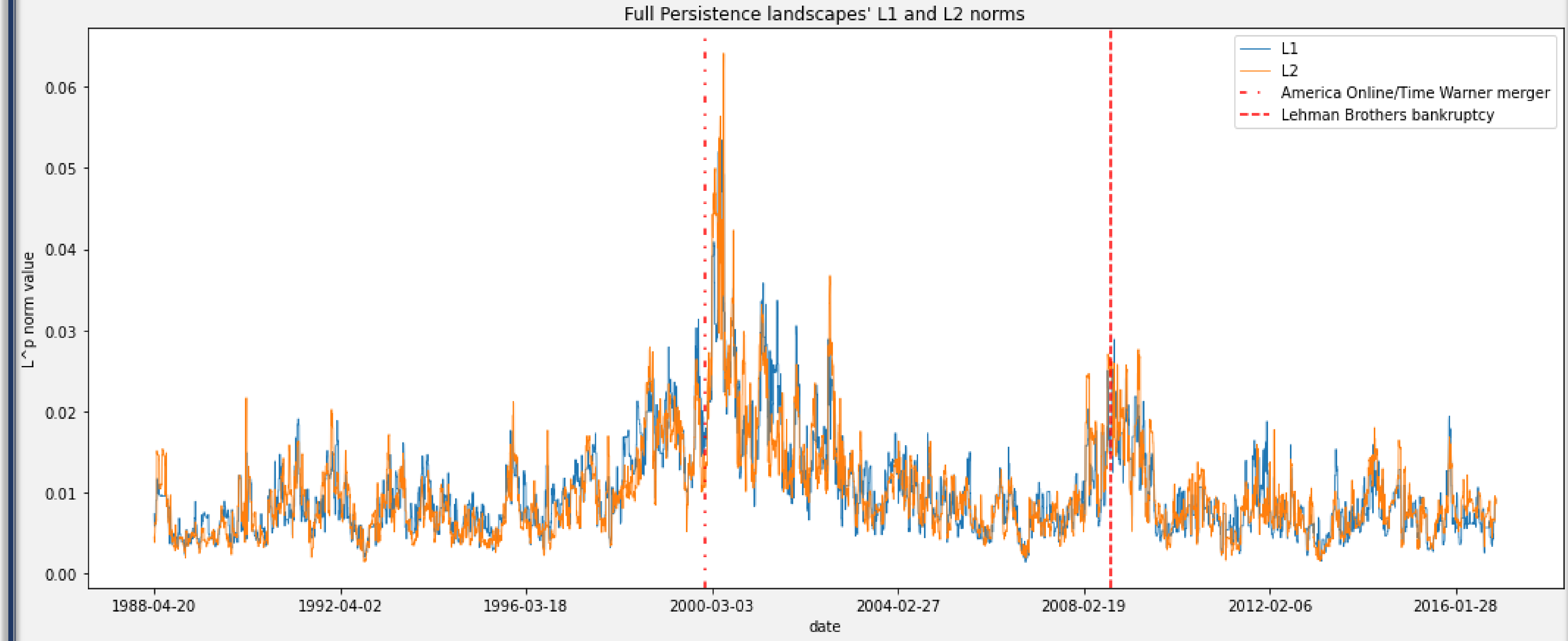
5 – Reproducing the paper

6 –Explorations & Conclusion



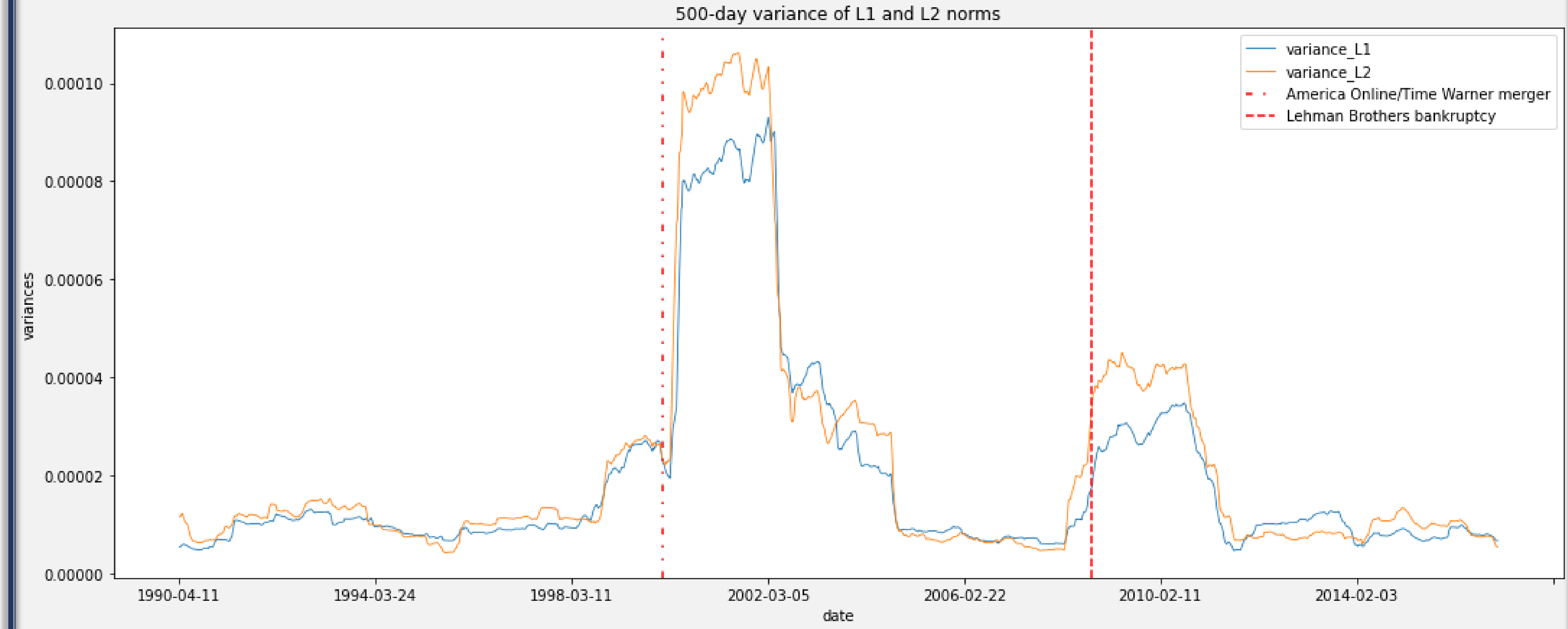
5- Reproducing the paper

Example results with window size 80



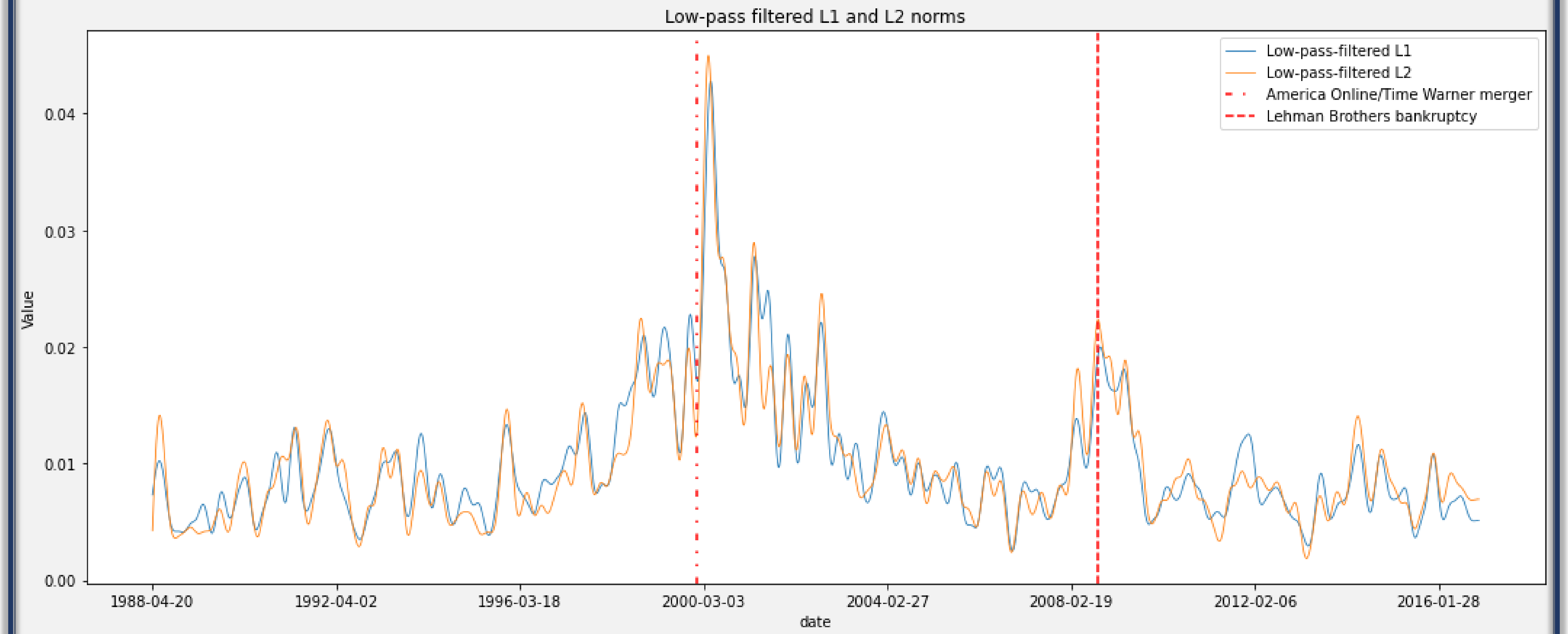
5- Reproducing the paper

Example statistics of results with window size 80 - 1/4



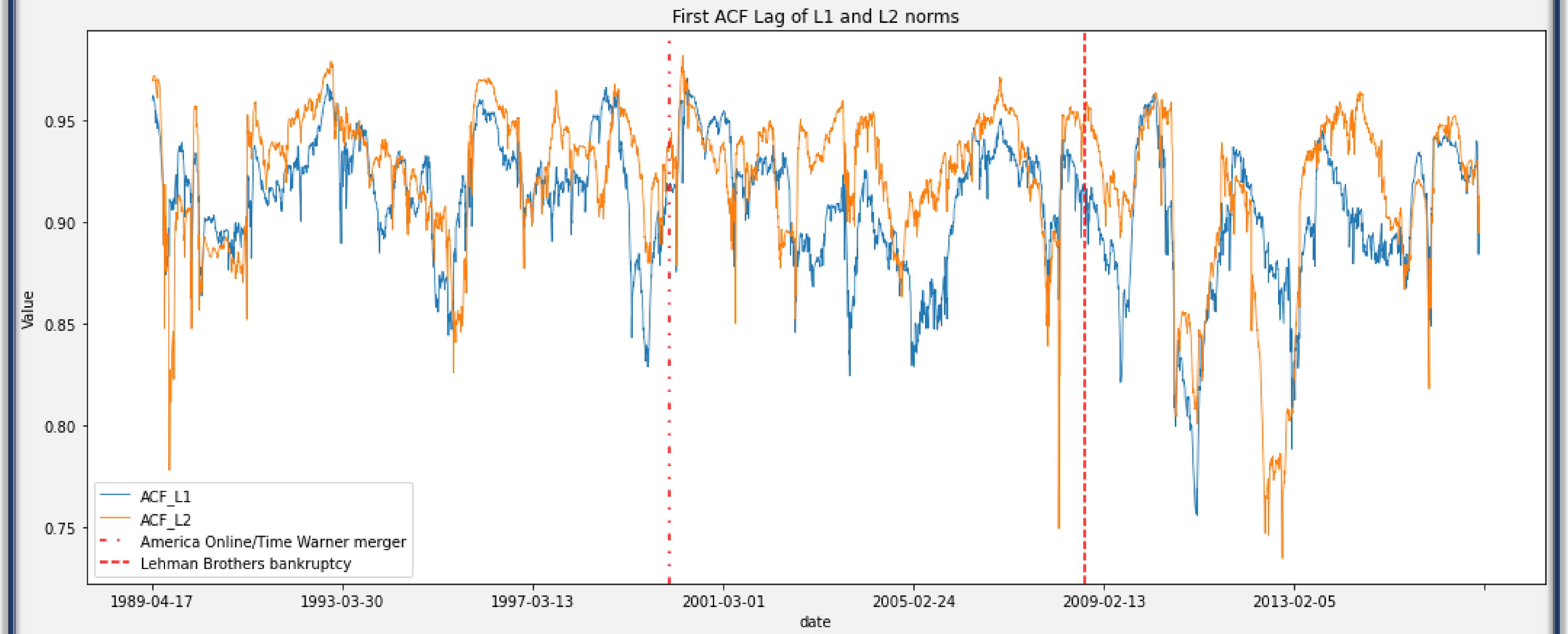
5- Reproducing the paper

Example statistics of results with window size 80 - 2/4



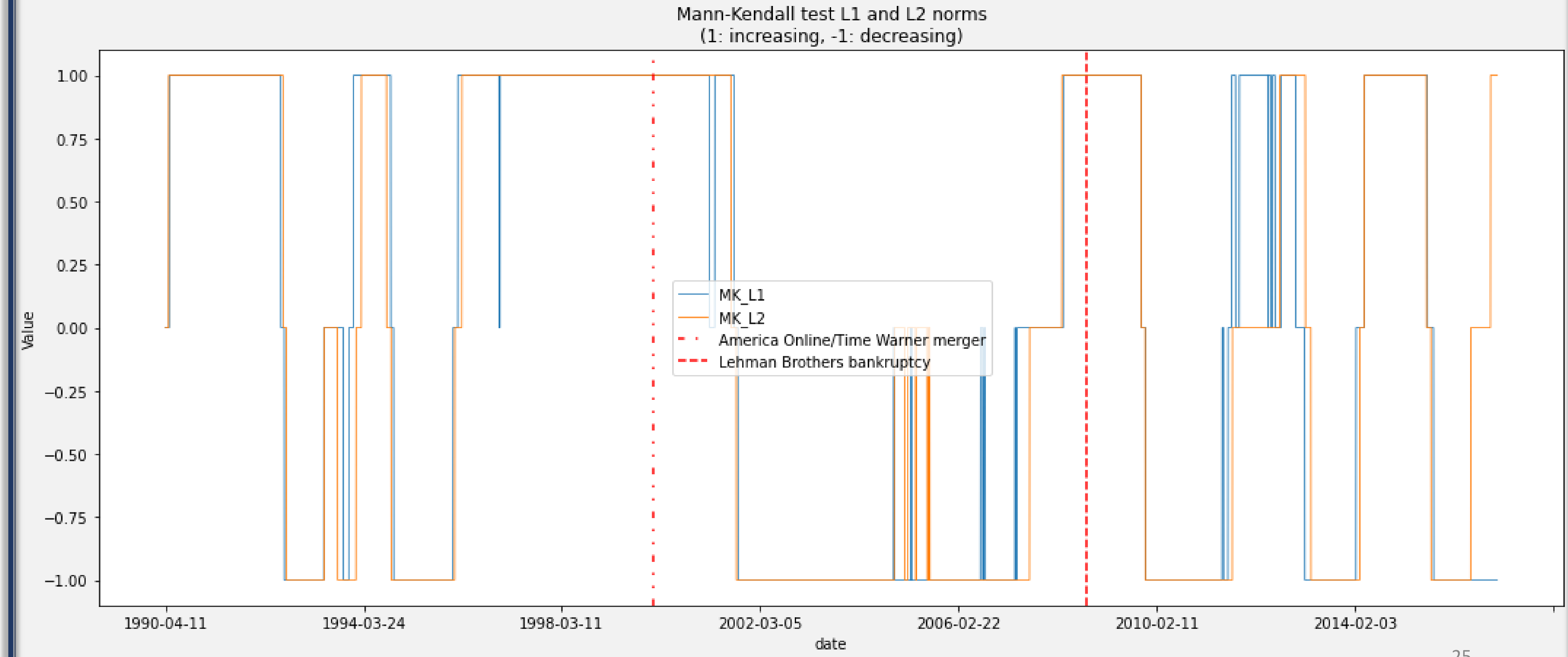
5- Reproducing the paper

Example statistics of results with window size 80 - 3/4

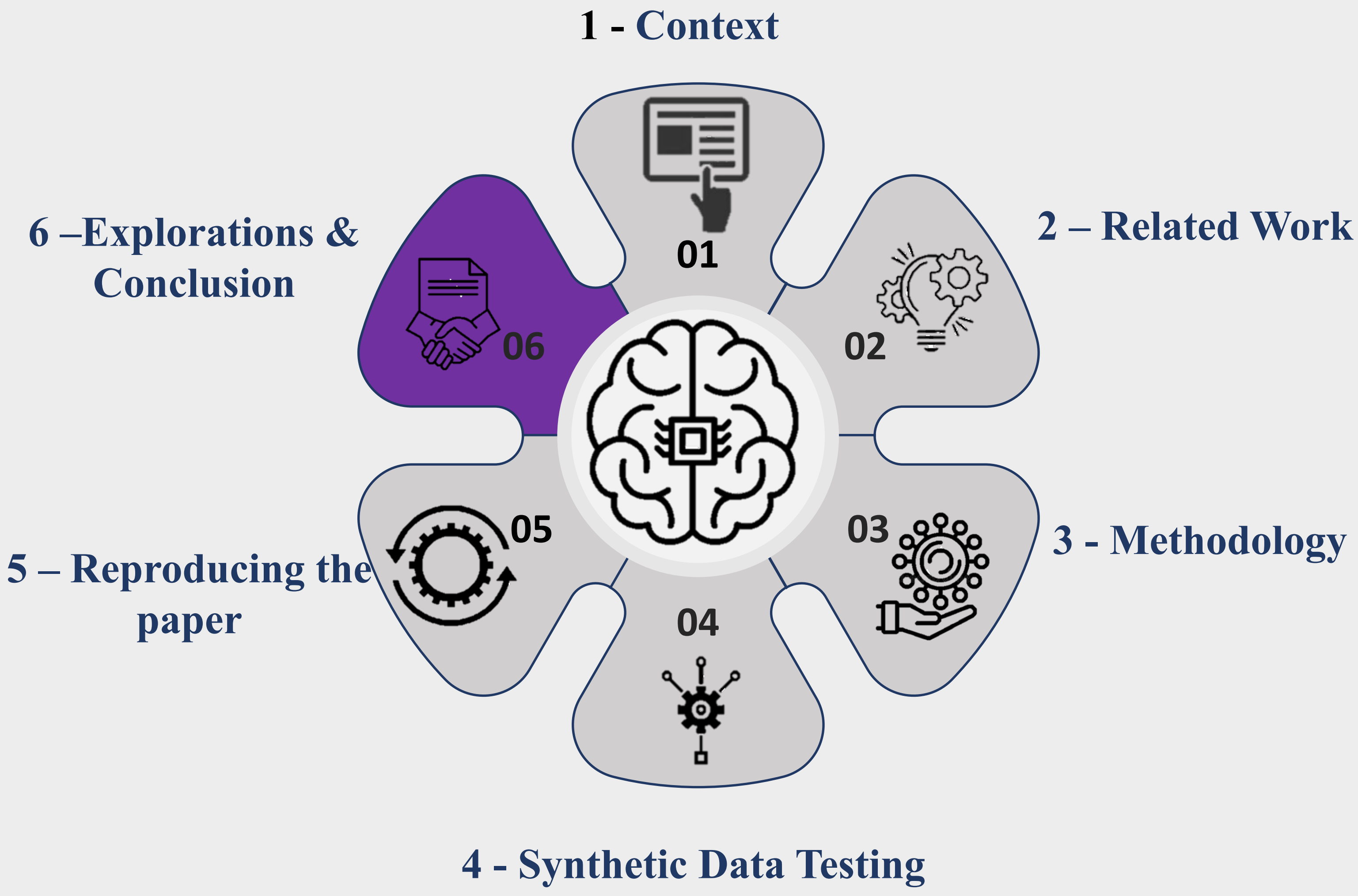


5- Reproducing the paper

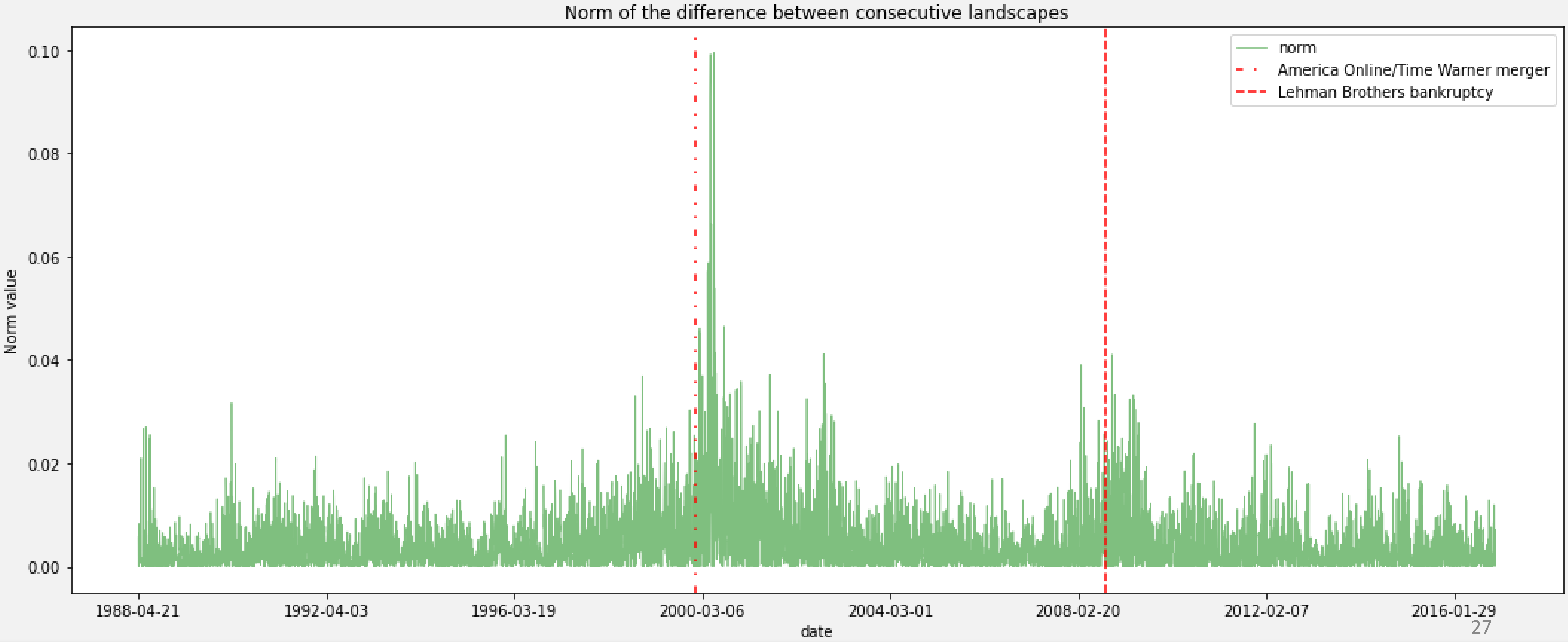
Example statistics of results with window size 80 - 4/4



Explorations & Conclusion



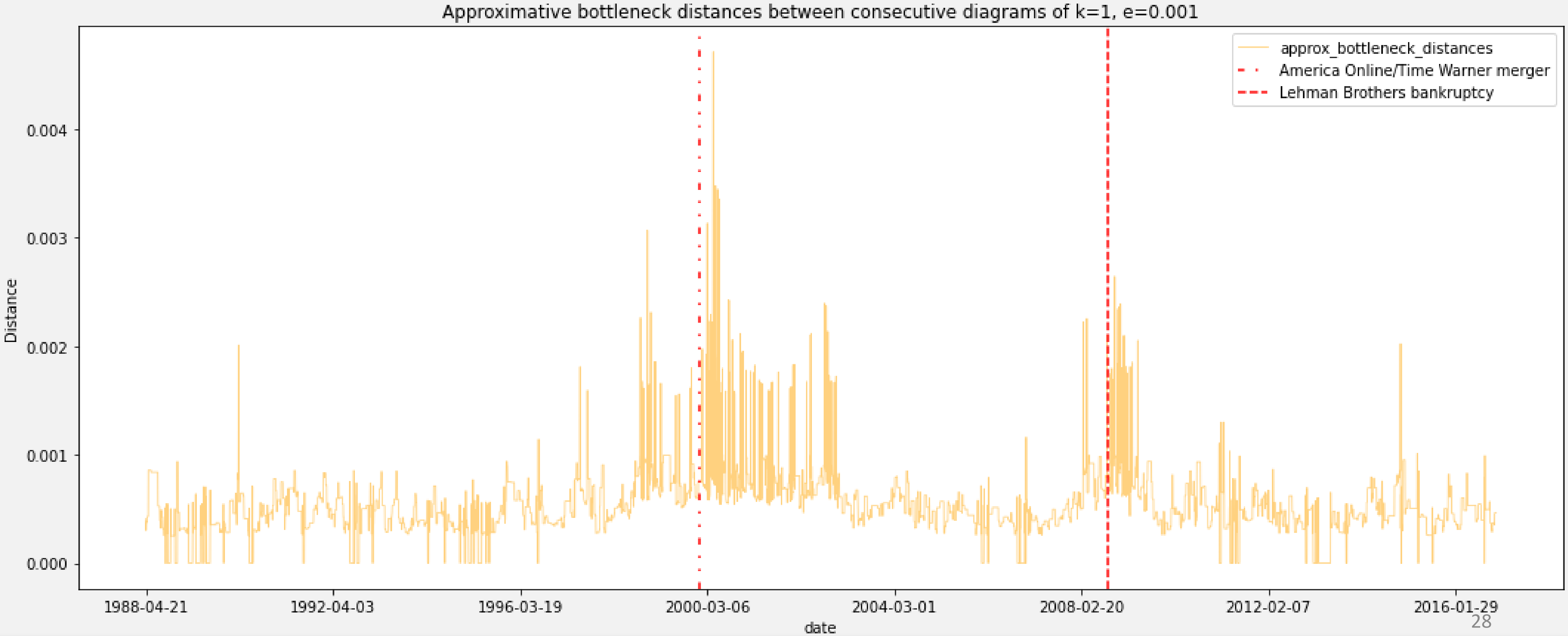
Norm of the difference between persistence landscapes



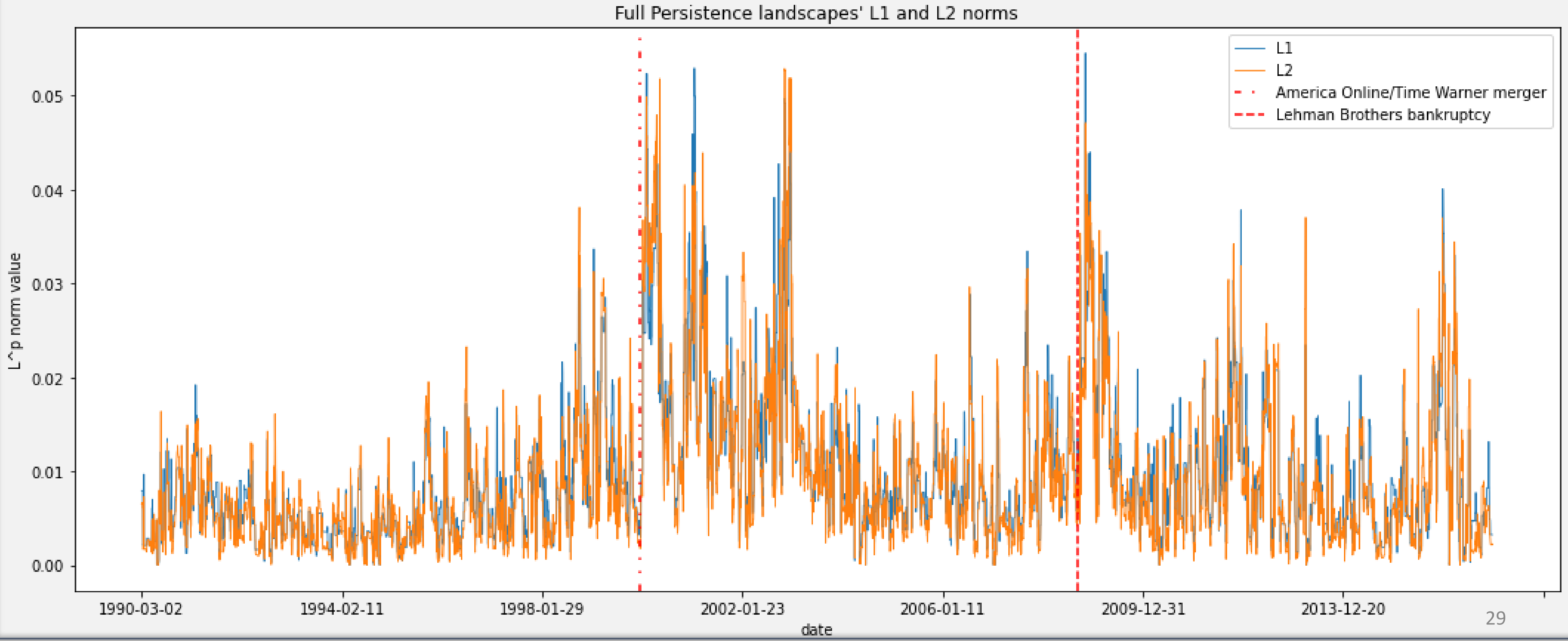
6 –Explorations & Conclusion

Other possible analyses | **Point of Contention** | **Conclusion**

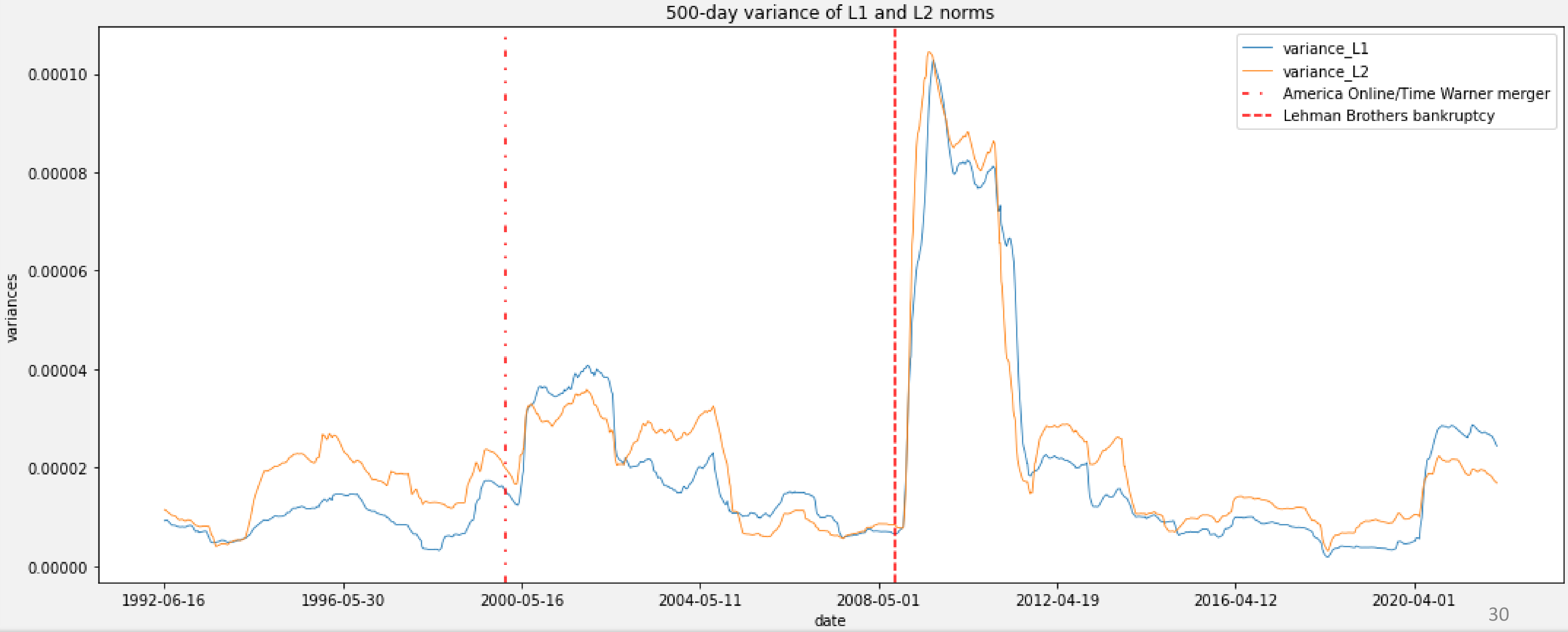
Bottleneck distance between persistence diagrams



Replacing the S&P500 with the VIX

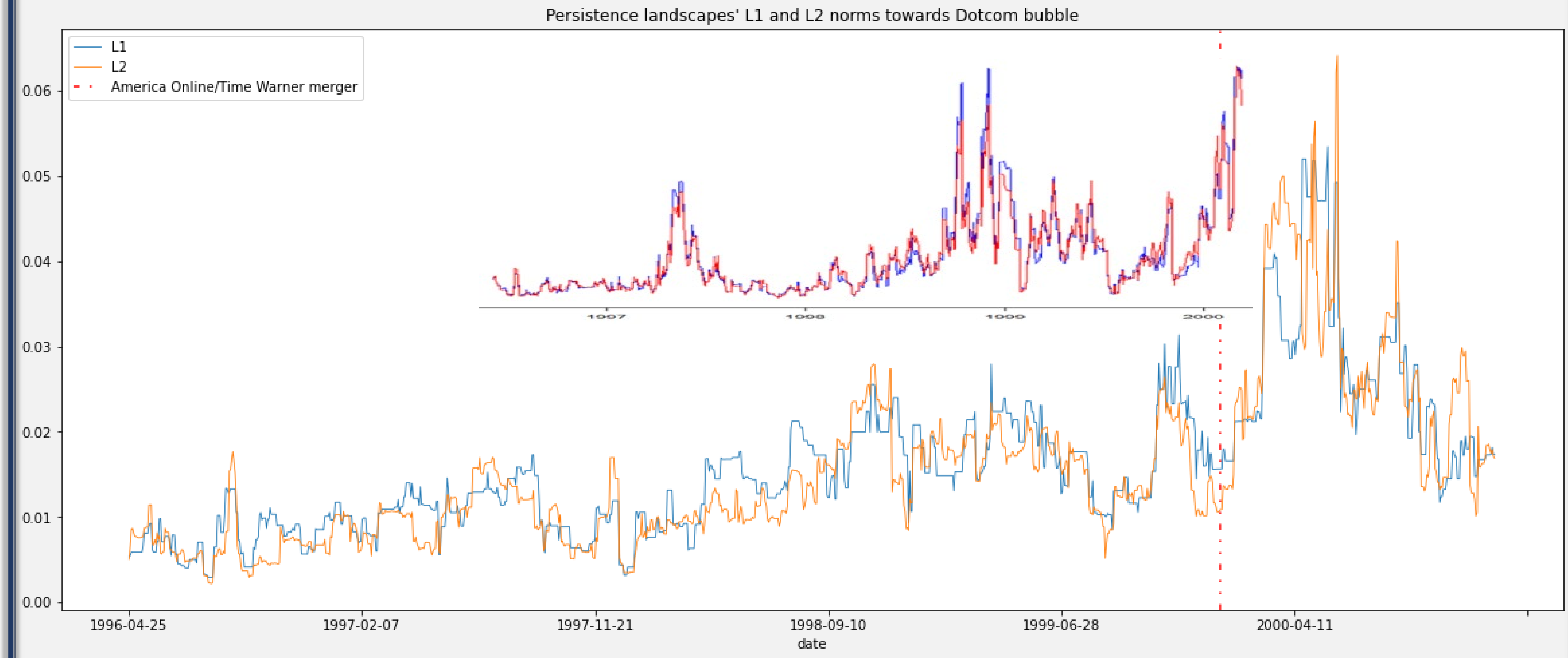


Looking at the Covid period (with data available: NASDAQ, RUSSELL 2000, VIX)



6 –Explorations & Conclusion

Other possible analyses | **Point of Contention** | Conclusion



6 –Explorations & Conclusion

Other possible analyses	Point of Contention	Conclusion

+ Novel way to extract information from noisy market data (robust)

+ Effectively a volatility measure with some future trend evidence

- The paper might oversell the method

- Computationally expensive compared to other econometrics methods like VaR or ES

<u>Other areas of exploration:</u> application to intra-day volatility