Contributed Talk

Detecting Extreme Events in the Stock Market using Topological Data Analysis

Anish Rai[1], **Buddha Nath Sharma**[1], Md Nurujjaman[1], Sushovan[2]

[1]*National Institute of Technology, Sikkim, India*
[2]*George Washington University, USA*

The inherent unpredictability and fluctuations in financial markets have attracted the interest of researchers to examine their behaviors. Particular emphasis is given to understanding market behavior during Extreme Events (sudden extraordinary occurrences arising from normal conditions, such as stock market crashes and spikes) of past to enhance profitability and mitigate risks in the future. Topological Data Analysis (TDA) is emerging as an important method for analyzing financial time series data. It is motivated to infer robust qualitative and quantitative information about the structure of the data through its topology and geometry. Within TDA, Persistent Homology (PH) is a technique that has been particularly used to study stock market crashes. Its ability to study multidimensional time-series in one go makes it an apt technique to study crashes that occur over an entire market, sector, or continent. This technique converts the multidimensional data set to point clouds in Euclidean space. The topology of the point cloud is studied for various resolutions (ε) of the points. With the increase of ε, various topological features of different dimensions (connected components, loops, voids correspond to 0-, 1-, and 2-dimensional features respectively) are born and die. Their birth and death are summarized in *Persistence Diagrams* (PD). The difference between two PDs is estimated by *Wasserstein Distance* (WD). The evolution of topological dynamics can be studied by calculating the (WD) between subsequent PDs using sliding window technique. For better statistical analysis, PDs can be converted to *Persistence*

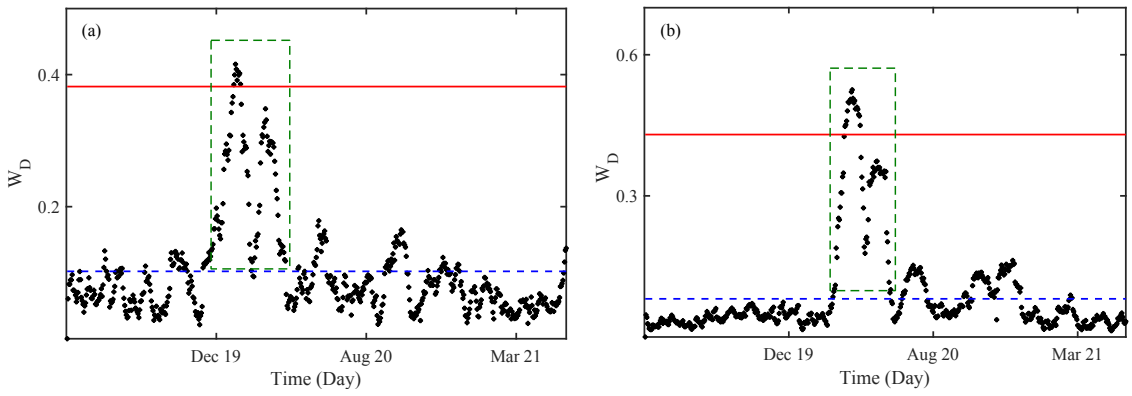

Figure 7: Plot (a) and (b) represent the Wasserstein distance (WD) of Asia and Europe continents, respectively, from 2019 to 2022. The spike in WD above the threshold $\mu + 4\sigma$ value coincides with the COVID-19 pandemic.

Landscapes (PL). In this work, we consider four major stock market indices from each continent. For each continent, we calculate the daily log returns of each index from 01-01-2006 to 01-01-2010 and from 01-01-2017 to 31-12-2022. We create a point-cloud dataset from the daily log-returns of the four indices. The point-cloud

data set is the input for the TDA. We then use a sliding window (ω) to the point-cloud dataset, and for each instant, we estimate the PD and PL for the particular window. For each PL, we calculate the L^1 and L^2 norm. Finally, we have calculated the *Wasserstein distance* (WD) between the successive PDs. We quantify the crashes as EEs if these measures cross a threshold of $\mu + 4\sigma$.