IDENTIFYING MARKET CRASHES USING ELASTIC NETS: A CHANGE POINT ANALYSIS

Vishwas Kukreti

School of Computational Integrative Sciences, Jawaharlal Nehru University, New Delhi, India.
*Correspondence : vishwa22_sit@jnu.ac.in

Abstract. *Change point detection (CPD) is a widely used statisti- cal technique in financial studies to identify abrupt shifts or transitions in the behavior of financial time series data. In the context of market crashes, CPD becomes crucial for detecting critical points in the evolution of financial sys- tems, characterized by sudden and severe declines in market prices and investor confidence. This research paper presents a novel approach to CPD using Elastic Net imputation for the detection of market crashes. The methodology involves split- ting the return time series of different stocks into equidistant windows and constructing networks based on these windows. Structural properties of the resulting graphs are computed, as these properties are known to change with variations in market conditions. To identify potential market crashes, we employ Elastic Net imputation by introducing missing values in the dataset of structural properties and predicting their values. The time points with the highest errors between the predicted and original values correspond to instances of market crashes. The proposed method is applied to the S&P 500 stocks dataset, demonstrating its effectiveness in detect- ing market crashes. The findings highlight the potential of Elastic Net imputation and structural properties of graphs for accurate and timely identification of market crashes, con- tributing to improved risk management and decision-making in financial markets.*

Keywords: *Change Point Detection, Elastic Nets, Market Crash, Network Theory.*

PACS numbers: 89.65.Gh, 89.75.Fb, 05.45.Tp

1. INTRODUCTION

Change point analysis, also known as change point detection, is a statistical technique used to identify points or periods in a dataset where there is a significant change in the underlying behavior, characteristics, or parameters of the data. It aims to detect and quantify abrupt shifts, discontinuities, or transitions in a time series or multivariate data [4]. The concept behind change point analysis is that data often exhibit different patterns or regimes over time, and identifying these changes is crucial for understanding the underlying processes or phenomena. By detecting change points, we can investigate factors causing the shifts, assess the impact of interventions or external events, and model the data more accurately. Change point analysis can be applied to various types of data, including time series data, spatial data, image sequences, and biological sequences. The changes detected can be related to mean shifts, variance changes, changes in the distributional shape, or changes in the relationships between variables.

For a sequence of independent random variables rep- resented by x_1, x_2, ..., x_n , with respective probability distribution functions denoted by F_1, F_2, ..., F_n. The objective of the change point problem is to assess the validity of the null hypothesis, which can be formulated as follows:

$$H_0 = F_1 = F_2 = ... = F_n. \tag{1}$$

The alternate hypothesis is:

$$H_1\text{: } F_1 = F_{k1} \neq F_{k2} = ... = F_{k2} \neq F_{k1} \tag{2}$$

where $1 < k_1 < k_2 < ... < k_q < n$ and q is an unknown quantity representing the number of change points, and k_1, k_2, ... k_q are the unknown positions that need to be estimated.

In the context of financial markets, a market crash represents a sudden and severe decline in asset prices, often accompanied by a significant increase in market volatility. A market crash is typically characterized by a departure from the previous market regime. It signifies a shift from a period of relative stability or upward trend to a sudden downward movement. This change in market behavior can manifest as a structural break, where the underlying dynamics governing the market fundamentally change [1]. Change point analysis helps identify these structural breaks by detecting changes in the statistical properties of financial data. It focuses on detecting points where key parameters such as mean, trend, volatility, or correlation significantly deviate from their previous values. These deviations indicate a departure from the previous market regime and the emergence of a new regime associated with the market crash. By applying change point analysis to financial time series data, we can identify the specific point or period when the market crash occurs. This provides valuable insights into the timing and magnitude of the crash, helping to understand the underlying factors and dynamics leading up to it. Additionally, change point analysis can help economists and analysts understand the dynamics leading up to market crashes and potentially develop early warning systems. By integrating change point analysis into predictive models, models can provide early warning signals or probabilities of market downturns.

Researchers have utilized various regression and machine learning models for change point detection and identifying market crashes.Logistic regression, support vector machines, random forests, recurrent neural networks, and hidden Markov models are commonly employed in this context. Logistic regression predicts market crashes based on historical data and relevant indicators. Support vector machines can discern boundaries between different market regimes. Random forests aggregate decision trees to identify significant market shifts. Recurrent neural networks capture temporal dynamics, while hidden Markov models model transitions between market regimes. The choice of model depends on data characteristics and the problem complexity, and ensemble methods can yield improved results. With the increased interest in network theory, researchers have tried to infer change points from the evolution of the structural properties of the network. Several models utilize network theory to detect change points in various fields, including finance. One approach is network connectivity analysis, where a financial network is constructed with variables or assets as nodes, and edges represent their relationships or correlations. Change points are identified by examining alterations in the network's structure or strength of connections. Community detection algorithms are another model that identifies change points by analyzing shifts in the community structure within the network. Changes in community assignments indicate shifts in relationships or dependencies between variables. Centrality measures, such as degree centrality or betweenness centrality, quantify node importance and can be used to detect change points when there are significant shifts in the influence of specific nodes. Dynamic network analysis models explicitly capture the temporal evolution of a network, allowing for the detection of change points when there are substantial variations in the network's structure or dynamics over time. Similarly, various regression and machine learning models can be utilized for change point detection and identifying market crashes. Logistic regression, support vector machines, random forests, recurrent neural networks, and hidden Markov models are commonly employed in this context. Logistic regression can predict market crashes based on historical data and relevant indicators. Support vector machines can discern boundaries between different market regimes. Random forests aggregate decision trees to identify significant market shifts. Recurrent neural networks capture temporal dynamics, while hidden Markov models model transitions between market regimes.

The choice of model depends on data characteristics and the problem complexity, and ensemble methods can yield improved results.

In this research paper we introduce a novel change point method for detecting market crashes by leveraging Elastic Net imputation. The proposed method follows a stepwise process involving the segmentation of return time series of different stocks into equidistant windows. For each window, a network is constructed, and various structural properties of the resulting graph are calculated. These structural properties are known to exhibit changes corresponding to shifts in market conditions. Subsequently, for each time point in the dataset of structural properties, a row is intentionally set as NA, and its value is predicted using Elastic Net imputation. Remarkably, it is observed that time points with the highest errors between the predicted and original values coincide with periods of market crashes. To demonstrate the efficacy of the proposed method, an application is conducted using S&P 500 stocks, wherein the results confirm the successful detection of market crashes.

2. METHODOLOGY

To detect market crashes, we propose a novel Change point method utilizing Elastic Net imputation. To validate the effectiveness of our method, we apply it to the S&P500 stocks as an application. By analyzing the identified time points with high prediction errors, we demonstrate that these correspond to the actual occurrences of market crashes in the S&P 500 market. The methodology involves the following steps:

- ***Data Preparation:***
  We gather the return time series data for various stocks and divide them into equidistant windows. This ensures a comprehensive analysis of local market dynamics.

- ***Network Construction:***
  For each window, we create a network by representing the stocks as nodes and establishing edges that represent their relationships. This network captures the interdependencies among the stocks.

- ***Calculation of Structural Properties:***
  From each net- work, we calculate the structural properties, which provide insights into the characteristics of the graph. These

properties are known to change with variations in mar- ket conditions and reflect important aspects of the market dynamics.

- ***Elastic Net Imputation***:
  Next, we employ Elastic Net imputation to predict the missing values in the dataset of structural properties. Elastic Net imputation is a regression technique that combines L1 and L2 regularization, balancing feature selection and handling multicollinearity.

- ***Error Analysis:***
  We compare the predicted values with the original values in the dataset of structural properties. By calculating the prediction error for each time point, we identify the time points with the highest errors. These time points correspond to periods where market crashes occur.

2.1. Data

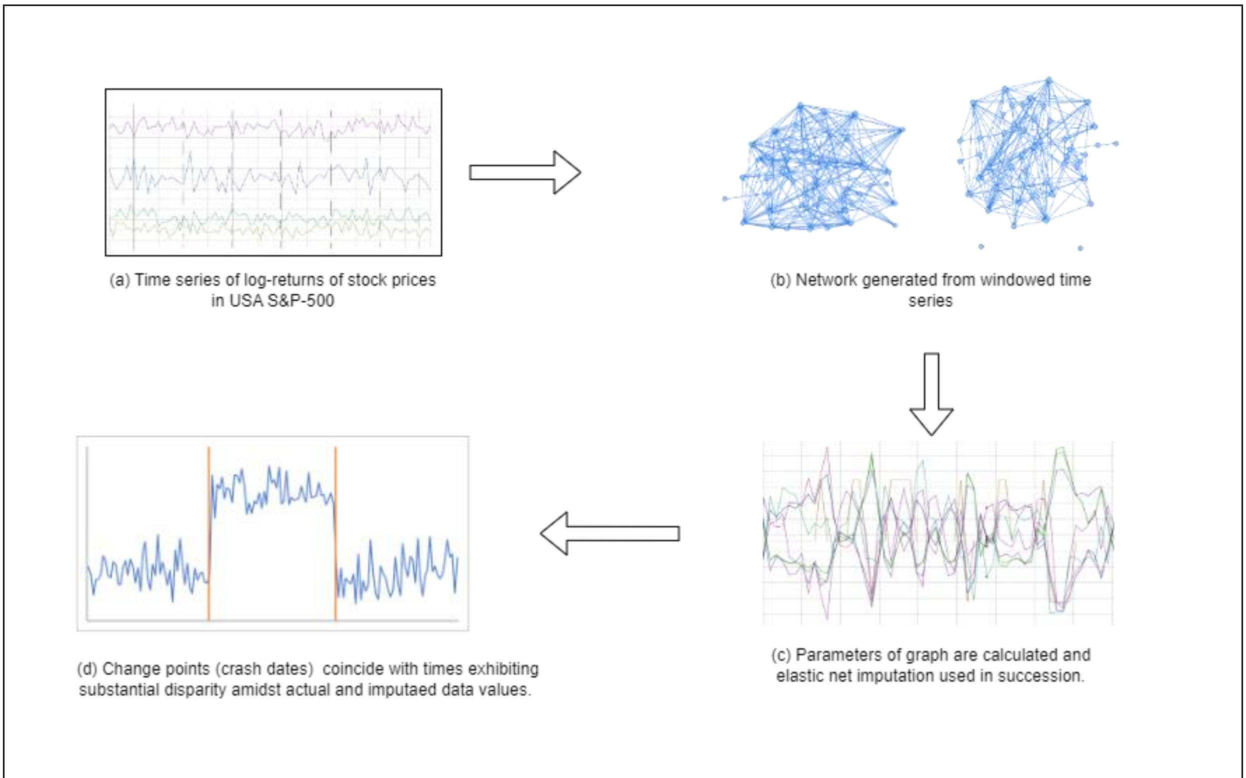
The research focuses on the utilization of historical daily adjusted logarithmic returns derived from the widely recognized Standard & Poor 500 index (S&P 500) during the period spanning from January 3rd, 2018 to March 30th, 2021. This dataset encompasses 1,183 days and is obtained from Yahoo Finance API, using daily closing prices.

The S&P 500 index is a market-cap-weighted benchmark index that reflects the performance of the US stock market, with larger companies exerting a greater influence due to their market capitalization. The dataset accounts for weekends by filling in the gaps with the previous day's values.

It includes all 500 constituent companies from various sectors such as Consumer Goods, Energy, Healthcare, Information Technology, Materials, Tele-communication Services, Utilities, and Industrials. The daily return time series is constructed as

$$r_i(t) = \log P_i(t)/\log P_i(t-1) \quad (3)$$

where $P_i(t)$ is the adjusted closing price of the i_{th} stock at time t.



Source: Realized by author

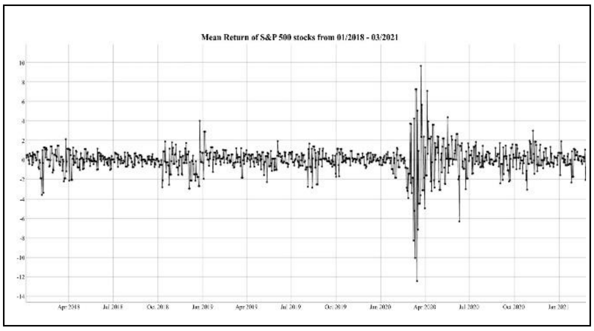
Fig. 1: A novel method to detect market crashes by utilizing Elastic Net imputation.

The process involves splitting return time series into equidistant windows, creating networks and calculating structural properties ((Fig. 1).

Changes in these properties reflect market conditions. By setting rows as NA and predicting their values, we identify the time points with the highest prediction errors as market crash occurrences.

We validate our approach using S&P 500 stocks. Notably, the dataset captures the market crash date triggered by the Corona pandemic shock in March 2020, which serves as a significant change point. On March 18th, 2020, the S&P 500 index experienced a substantial crash, losing more than 12% of its value in a single day. This crash was primarily attributed to the economic uncertainty and fear induced by the COVID - 19 pandemic.

Also it is worth mentioning that the S&P 500 index swiftly recovered from the crash, regaining all losses within a few months (Fig. 2).



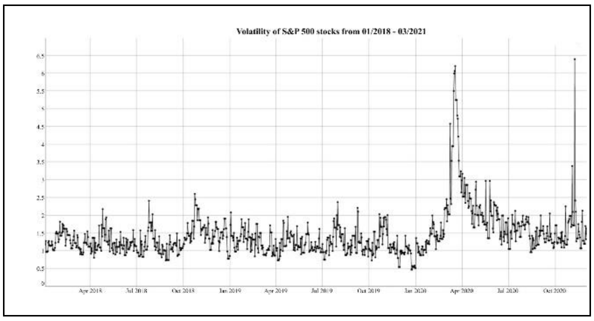
Source: Realized by author

Fig. 2: The mean return of S&P 500 stocks spanning the time period from January 2018 to March 2021.

Notably, the main market crash, attributed to the COVID-19 pandemic, occurred during the corona scare period, leading to a significant loss of approximately 12% of the market capitalization.

2.2. Network Construction

Graphical Gaussian models provide a flexible and widely- used methodology for generating networks that capture the relationships between variables. In this modeling ap- proach, each variable is represented as a node in the resulting graph, and the connections between variables are represented by edges. One common technique for constructing these networks involves utilizing the estimated precision matrix, which characterizes the partial correlations between variables in the model. The precision matrix, which is in verse of covariance matrix, plays a crucial role in network generation, as it provides valuable insights into the strength and nature of relationships among variables.



Source: Realized by author

Fig. 3: The mean volatility of S&P 500 stocks throughout the period from January 2018 to March 2021.

The visualization reveals that the volatility reaches its highest level during this critical period, reflecting the heightened uncertainty and turbulence in the market (Fig. 3).

By examining the non-zero entries in the precision matrix, researchers can identify the significant connections between variables and construct an informative network representation. Additionally, the inverse of the precision matrix, known as the covariance matrix, offers complementary information regarding the dependencies and associations among variables.

To facilitate the generation of networks from price data, we utilize the ’huge’ package in the R programming language [3]. The ’huge’ package offers efficient and scalable algorithms for estimating high-dimensional graphical mod- els and provides the necessary tools for constructing net- works based on the estimated precision matrix.

For a 30 days’ window with 15-day shift period, we capture the complex dependencies and inter-relationships present in the return data and construct a comprehensive network representation that highlights the significant connections among variables.

2.3. Structural Properties

The structural properties of a graph constructed from the return time series capture the state of the market by reflecting the underlying relationships and dependencies among the stocks. These properties provide insights into the overall structure and dynamics of the market, enabling us to assess its current state and anticipate potential changes. One crucial aspect is the network connectivity, which describes how the stocks are connected to each other in the graph. The presence or absence of connections and the strength of those connections reflect the degree of inter-dependence and co-movement among the stocks. A well-connected graph indicates a highly integrated market, where information and trends are easily transmitted among the stocks. On the other hand, a fragmented or loosely connected graph may suggest isolated sectors or less synchronized movements, potentially indicating diverse market conditions. Additionally, network centrality measures, such as degree centrality, betweenness centrality, and eigenvector centrality, provide information about the importance and influence of individual stocks within the network. Stocks with high centrality measures are considered influential or pivotal in the market. Changes in the centrality of specific stocks can indicate shifts in market dynamics or the emergence of new market leaders.

Moreover, community structure analysis helps identify groups or clusters of stocks that exhibit stronger internal connections compared to connections with stocks outside the cluster. These communities can represent sectors, industries, or groups of related stocks. Changes in the community structure can indicate shifts in sectorial relationships, market correlations, or the emergence of new market regimes.

The properties of network which were calculated are the following:

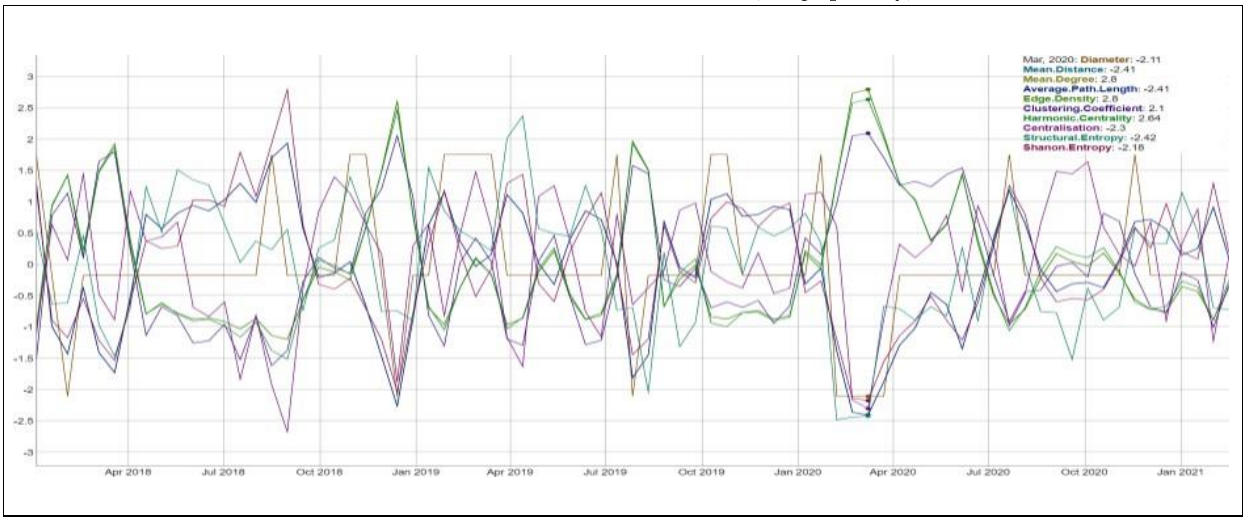
1. ***Entropy:***
In the context of a network made of stock price series as nodes, entropy can provide insights into the level of uncertainty or randomness in the connectivity patterns and relationships among the stocks. During a market crash, the network may exhibit higher entropy, indicating increased unpredictability and disorder in the interactions between stocks. This suggests that the relationships between stocks become less predictable, potentially reflecting the turbulent and volatile nature of the market during a crash.
2. ***Structural Entropy:***
Structural entropy measures the randomness or uncertainty in the overall structure of the network. During a market crash, the structural entropy of the network may increase as the relation- ships and connections between stocks undergo signifi- cant changes. The higher structural entropy suggests a greater level of complexity and disorder in the network structure during periods of market turmoil.
3. ***Mean Distance:***
The mean distance in the network mea- sures the average number of edges or steps required to move from one stock to another. In the context of a market crash, the mean distance may decrease, indi- cating shorter paths or stronger connections between stocks. This suggests a higher level of interconnectedness and co-movement among stocks during a crash, as they respond collectively to market-wide shocks.
4. ***Mean Degree:***
The mean degree of the network reflects the average number of connections or relationships that each stock has. During a market crash, the mean degree may increase, indicating a higher level of connectivity among stocks. This signifies that more stocks are influenced by or respond to market conditions, leading to a more synchronized behavior during the crash.
5. ***Average Path Length:***
The average path length mea- sures the average number of edges or steps needed to traverse from one stock to another across the network. In the context of a market crash, the average path length may decrease, indicating a shorter average distance be- tween stocks. This suggests a higher level of information flow and transmission of market shocks across the network during a crash.
6. ***Edge Density:***
Edge density represents the proportion of actual connections present in the network compared to the total possible connections. During a market crash, the edge density of the network may increase, indicating a higher level of connectivity and interdependencies among stocks. This reflects the closer relationships and stronger correlations between stocks as they respond collectively to market conditions.
7. ***Clustering Coefficient:***
The clustering coefficient measures the tendency of nodes (stocks) to form clusters or tightly connected groups. During a market crash, the clustering coefficient may increase, indicating the formation of cohesive substructures within the network. This suggests that stocks within the same cluster may exhibit similar price movements or share common re- sponses to market shocks during a crash.
8. ***Mean Degree Centrality:***
Mean degree centrality reflects the average level of influence or importance of stocks based on their connectivity. During a market crash, the mean degree centrality of the network may increase, indicating a higher average level of influence or impact of stocks. This suggests that more stocks play a crucial role in transmitting market information or driving the collective behavior of the network during a crash.
9. ***Mean Harmonic Centrality:***
Mean harmonic centrality measures the influence or importance of stocks based on their ability to efficiently reach other stocks within the network. During a market crash, the mean harmonic centrality may increase, indicating that certain stocks have a greater ability to disseminate information or control the flow of market dynamics. These stocks may act as key players in transmitting market shocks and driving the overall behavior of the network during a crash.

By examining these structural properties of the graph constructed from the return time series, we gain insights into the interdependencies, connectivity patterns, and overall organization of the market. These properties capture the state of the market by reflecting the collective behavior and relationships among the stocks, enabling us to understand the prevailing market conditions, detect changes in market regimes, and potentially anticipate or identify market crashes.

2.4. Elastic Net Imputation

Elastic Net regression is a statistical method used for re- gression analysis, particularly in scenarios where there are many predictor variables and potential multicollinearity (correlation) among them. It combines the properties of two popular regularization methods: Lasso regression and Ridge regression [5]. In Elastic Net regression, the objective is to find the best-fitting regression model that minimizes the sum of squared errors while simultaneously incorporating a penalty term that promotes sparsity (shrinking coefficients to zero) and handles multicollinearity.

The method achieves this by adding two penalty terms to the ordinary least squares (OLS) loss function: the L1 norm (Lasso penalty) and the L2 norm (Ridge penalty).



Source: Realized by author

Fig. 4: The evolution of network properties in the S&P 500 during market crashes, specifically focusing on the notable crash that occurred in March 2020

Figure 4 showcases key network properties, including entropy, structural entropy, diameter, mean distance, mean degree, average path length, edge density, clustering coefficient, mean degree centrality, and mean harmonic centrality. The vertical dashed line marks the oc- currence of the market crash in March 2020, indicating a significant deviation in the network properties during this period. The time series highlights the dynamic nature of these properties and their association with market crashes, shedding light on the structural changes and interconnected- ness within the S&P 500 network during times of financial turbulence.

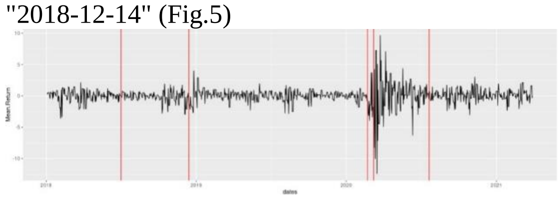
The L1 penalty encourages variable selection by shrink- ing less relevant variables to exactly zero, effectively per- forming feature selection. On the other hand, the L2 penalty encourages shrinkage of all variables while maintaining their non-zero values, reducing the impact of multicollinearity. Imputing values using Elastic Net can be achieved by employing the trained Elastic Net model to predict missing or incomplete values in a dataset. This process is particularly useful when dealing with datasets that contain missing or unreliable data. To impute missing values, the Elastic Net model utilizes the observed values of other variables to predict the missing values based on the learned relationships between the predictors target variable. We use 'mlim' package to impute values. [2]

The Elastic Net imputation process involves the following steps:

- ***Train an Elastic Net regression model*** using a dataset without missing values, where the target variable is known or complete.
- ***Identify the rows or instances in the dataset with missing values that need to be imputed***. For each row with missing values, use the trained Elastic Net model to predict the missing values based on available observed values of other variables in that row.

3. RESULTS AND DISCUSSION

We predict every values of every row, given the entire structural indicator data set. The mean difference of every imputed/predicted from the original value is calculated. It is seen that the mean difference is highest at points of market crash. The methodology successfully locates the change points as dates of market crash. For our dataset, the dates with highest difference were "2020-03-08", "2020-07-21", "2020-02-22", "2018-07-02", "2018-12-14" (Fig.5)



Source: Realized by author

Fig. 5: The red lines indicate dates of market "crash" **"2018-07-02", "2018-12-14","2020-02-22", 2020-03-08", "2020-07-21"**

4. CONCLUSION

In conclusion, change point detection plays a crucial role in understanding the critical points in the evolution of finan- cial systems, particularly in identifying market crashes. By utilizing various methods, such as change point analysis and network theory, researchers have made significant progress in detecting market crashes and understanding their underly- ing dynamics. Change point analysis provides a systematic approach for identifying abrupt changes or shifts in time series data, which can be applied to financial and economic indicators to detect market crashes. Network theory, on the other hand, offers a powerful framework for analyzing the interconnectedness and dependencies among financial variables, enabling the identification of change points associated with market crashes. The relevance of change point detection in the context of market crashes cannot be overstated. Market crashes represent critical moments in the financial system, characterized by sharp declines in asset prices, increased volatility, and widespread economic distress. Detecting market crashes in a timely manner is crucial for risk management, portfolio optimization, and decision-making in financial markets.

Through the application of various regression and machine learning models, such as logistic regression, support vector machines, random forests, recurrent neural networks, and hidden Markov models, researchers have successfully utilized change point detection to identify market crashes. Additionally, the incorporation of network theory has fur ther enhanced the understanding of market dynamics and the detection of change points through the analysis of network properties, such as connectivity, community structure, and centrality measures.

Furthermore, the use of Elastic Net imputation in change point detection has provided a novel approach to fill missing values and predict the occurrence of market crashes. By im- puting values based on the relationships between variables captured by Elastic Net regression, researchers have success- fully identified time points with the highest prediction errors as indications of market crashes.

The findings and methodologies discussed in this paper underscore the importance of change point detection in the context of market crashes. The ability to identify critical points in the evolution of financial systems not only con- tributes to a deeper understanding of market dynamics but also facilitates risk assessment, decision-making, and the development of effective strategies for managing and mitigating the impact of market crashes.

Continued research and advancements in change point detection techniques, coupled with the integration of network theory and machine learning approaches, hold great potential for improving market crash detection and enhancing financial analysis in the future.

5. REFERENCES

[1] Sayantan Banerjee and Kousik Guhathakurta. 2020. “Change- point analysis in financial networks.” In: *Stat* 9.1, e269.

[2] EF Haghish. 2021. *mlim: Multiple Imputation with Automated Machine Learning*. 2022.

[3] Haoming Jiang et al. 2021. *High-dimensional undirected graph estimation*.

[4] Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2020. “Selective review of offline change point de- tection methods.” In: *Signal Processing* 167, p. 107299.

[5] Hui Zou and Trevor Hastie. 2005. “Regularization and variable selection via the elastic net.” In: *Journal of the royal statistical society: series B (statistical methodology)* vol 67.2 (2005), pp. 301–320.