Complex Systems and Networks HW 1

Rachael Judy, Connor Klein, Josh Smith

February 21, 2024

1 Q1: Annotated Bibliography

Economic markets are extremely volatile and dynamic and present an active area of research. Previous models of the market treat it as a linear system, which does not fully capture the actual behavior of the market. Developing an accurate model of the market would result not only in a significant economic advantage but provide the possibility of analyzing global financial systems, market sentiment, inter-dependencies, and macroeconomic indicators. Due to the readily available detailed historical data of the stock market and the aforementioned advantages gained by correctly modeling the market, this field has been studied by numerous research groups. Within this field, specific areas of research have included correctly modeling the markets as economic systems, forecasting stock prices in various global markets based on feature selection and through various means, detecting anomalies in the market, and information diffusion in the market. By reviewing previous and active research in the field, new parameters and methods for network models can be developed.

Modeling the stock market as a complex network involves viewing entities, stocks, or market groups as interconnected nodes with edges representing their relationships, whether these are transactions performed or correlations. This approach provides new perspectives on market dynamics, allowing analysts to discover previously unknown relationships and dependencies not apparent in traditional methods that fail to capture the influence of nodes on one another. This provides key advantages of capturing emerging trends and recognizing feedback loops. Leveraging tools such as power law, centrality measures, and neural networks, researchers can identify critical entities and groups and better understand the forces at play.

The papers summarized here discuss the general modeling approaches taken for representing financial markets and impacts of shock events in those markets in terms of complex systems. This includes examining broker and stock networks and highlighting means for determining edges in the graphs such as correlation and mutual information. They also discuss the results of several models in different global markets and in selecting better representative indices of the stock market and detecting anomalies.

1.1 Explaining Financial Markets in Terms of Complex Systems

M. Kuhlmann, ”Explaining financial markets in terms of complex systems,” *Philosphy of Science*, vol. 81, no. 5, Dec., pp. 1117-1130, 2014.

The researcher theorizes that the US financial market can be modeled as a nonlinear complex system. His belief stems from the study of econophysics, which explains economic phenomena with methods and models used to describe physical phenomena. He compares the market to that of a ferromagnet, where microprocesses accumulate to create macro effects. These microprocesses are not random and linear but are instead a result of the state of their neighbors. Additionally, modeling the market as a complex system allows some of the microprocesses to be filtered out, resulting in the larger macro functions being derived without all the information. Kuhlmann postulates that understanding structural mechanisms is where the research needs to be focused to get closer to understanding complex systems in general. These structural mechanisms are broken into two classes: a structural start with boundary conditions and emerging dynamical systems. In this way, he relates the US financial market to ferromagnetism. That is, the underlying structure is very similar, however, the function is entirely different. This article describes how the financial market could be developed as a complex network, but no work is done; instead, Kuhlmann states that the models fail because human behavior is not sufficiently realistic in these models. This complexity is derived through the structural mechanisms of interacting agents. This paper describes a thought experiment without much obvious proof backing his claims. If work were done to advance the modeling of the complexity of human interactions through these structural networks, a significant step would be made in modeling the US financial market as a whole.

1.2 The Relevance of Broker Networks for Information Diffusion in the Stock Market

M. Di Maggio, F. Franzoni, A. Kermani, and C. Sommavilla, ”The relevance of broker networks for information diffusion in the stock market,” *Journal of Financial Economics*, vol. 134, no. 2, Nov., pp 419-446, 2019.

This report was compiled by the Harvard Business School to determine the effect of centrality of brokers in the US stock market. Specifically, the researchers postulate and test a multitude of theories validating that brokers play a key role in diffusing information from originators to their institutional investors. They theorize why this might be the case, with examples such as brokers wanting more clients, clients wanting to make more money, and the more clients a broker has (centrality), the more insider information they will know. They validate these theories by analyzing the trades of brokers and their connections after a period using private 13D forms. They found that not only were activist traders more likely to follow large trades made by hedge fund investors with the same broker, but also central brokers—those with the most connections—had clients who, on average, performed 40 points better than peripheral brokers. They followed up, validating that these central brokers were not significantly better than peripheral ones with smarter clients, instead showing that on other trades, the clients are equivalent. The researchers performed many tests to prove that in the US market, it is not skill that matters most but who you know. Powered by this information, they developed an equation to estimate the performance of an active investor based on their broker's centrality and relationship with that broker. This data was taken from 1999 to 2014 and included information from 30 different brokers. The researchers appeared to cover all their bases in terms of the scope in which they were attempting to qualify. More work can be done with this information, such as the application of modeling the market as a set of centralized and peripheral brokers to see if their work can reasonably predict the market.

1.3 Contagion in financial networks

P. Gai and S. Kapadia, ”Contagion in financial networks,” *Proceedings of the Royal Society*, vol. 466, no. 2120, Aug., pp 2401–2423, 2010.

This paper presents a model of contagion for unexpected shocks in arbitrary network structures in finance and uses numerical simulations with edges initialized by a Poisson random graph to illustrate the results. They authors discussed both the benefit of being able to absorb loss of one institution among other entities, reducing the probability of failure, and the danger of more linkages increasing the potential for spread and damage from the second round of defaults. Their directed network of interconnected balance sheets examines the number of vulnerable nth neighbors for each entity and uses recursive equations to consider possible contagion and identify phase transitions representing thresholds for outbreaks. The paper also briefly addressed how illiquid assets that are sold on default can be reduce prices generally in the liquid market, introducing a second potential source of contagion upon default. Their model has the advantage of representing arbitrary network structures not dependent on the real world model. However, the paper focused on a single disruptive event, namely the default of a single bank, and other events cannot be assumed to have the same effects. Additionally, the analytical method treated the network of a failing bank as a tree with no cycles, reducing the value of the numerical simulation results.

1.4 Article 5

1.5 Article 6

1.6 Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction

G. Li, A. Zhang, Q. Zhang, D. Wu, and C. Zhan, “Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 5, May, pp. 2413–2417, 2022.

In this paper, the Pearson Correlation Coefficient (PCC), a measure of strength and direction of the relationship between two variables, is used with a Broad Learning System (BLS) for feature selection to perform time series prediction of stocks selected from the

Shanghai Stock Exchange. The authors compared the results of their method with ten machine learning methods including Adaboost, Gradient Boosting Decision Tree, Convolutional Neural Networks, and others. The BLS system is based on a random vector functional link neural network that transforms input features to feature nodes that connect to the output layer. Weights are initially randomly generated and then improved through pseudoinversion. This provides a less time intensive alternative to a deep learning framework. The paper provides a detailed explanation of the features such as closing price and rates of change and equations used and provides metrics showing superior performance over other prediction methods. The main deficiencies in the paper are the assumptions that the small sample size of selected stocks represent the system's performance overall; these selections are regional and do not necessarily represent the overall market. The paper does state that future work will involve evaluating the model through real investment returns.

1.7 Networks in financial markets based on the mutual information rate

P. Fiedor, "Networks in financial markets based on the mutual information rate," *Physical Review E*, vol. 89, no. 5, May, 2014.

In this paper, Fiedor argues that correlation, which deals in linear dependencies and is commonly quantified with the Pearson Correlation Coefficient (PCC) within complex system modeling of the market, does not capture the nonlinear behavior of economic markets. He instead proposes using mutual information and the mutual information rate with Lempel-Ziv complexity as a similarity measure between closing prices on consecutive days. Defining minimal spanning trees and planar maximally filtered graphs, he represents the market entities and their relationships, demonstrating greater clustering by sector based on the mutual information metric. He also notes that mutual information rate demonstrates more about the system dynamics than the structure. The results corroborate previous papers using the PCC regarding the predominance of the financial sector in markets and suggest further study of differential and transfer entropy. While the paper highlights how mutual information and mutual information rate provide a different view of the market interconnectivities, it does not clearly demonstrate its claim that using mutual information as the metric is superior in modeling the true relationships of market entities, neither in theory or in simulation.

1.8 A perspective on complex networks in the stock market

J. Park, C. H. Cho, and J. W. Lee, "A perspective on complex networks in the stock market," *Frontiers in Physics*, vol. 10. Dec., 2022. [Online serial]

This paper sets out to visualize and analyze the connectedness of stocks in the Korean financial market at specific time windows. To accomplish this, the researchers connected their nodes (securities) using an absolute threshold calculated from the cross-correlation coefficients of their logarithmic returns within a small-time window. In other words, securities are linked based on their similarities in changes relative to one another. The findings indicate that stocks were much more connected during the 2008 global financial crisis than before it occurred (2006-07). Specifically, the largest cluster pre-crisis comprised 4.8% of securities, whereas post-crisis, the largest cluster increased to around 20%. The researchers focused on analyzing the overall structure of the Korean market, but it appears possible to analyze the variance of any stock and determine if there is a high cross-correlation with others after a set period, effectively allowing for a form of market prediction. Additionally, the researchers only considered a small time window, less than a year, for market analysis; a more comprehensive representation of the network might have been obtained over a longer time period.

1.9 A hierarchical view of a national stock market as a complex network"

Y. Y. Baydilli, S. Bayir, and I. Tuker, "A hierarchical view of a national stock market as a complex network," *Economic Computation & Economic Cybernetics Studies & Research*, vol. 51, no. 1, Jan., pp. 205–222, 2017.

In this paper, the authors construct a financial network using hierarchical methods for the Borsa Istanbul 100 Index (BIST-100) of 100 stocks bargained during 2011-2012 to investigate stock correlation and identify risky stocks in market. They used a mean spanning tree to determine the sub-dominant ultra-metric distances to analyze clustering and a hierarchical tree to extract factors affecting the dynamics of the complex system. They then calculated normalized tree length and degree distribution with a power law assumption. From the network, they determined that the correlation coefficients became approximately Gaussian and gaps in the MST shrank during crisis. Additionally, the network displayed small-world and scale-free characteristics with some nodes seemingly controlling information flow in a network. They also highlihgted clusters in the final graph representing different sectors of industry as well as clusters with high liquidity, with financial institutes and liquid stocks in central positions. Based on their observations, they claimed the MST concept is a good guide for risk management and highlighted useful features such as normalized tree length for recognizing risk. However, this study did not examine a long enough time period or enough stocks to strongly verify the authors' claims.

1.10 A network perspective of the stock market

C. K. Tse, J. Liu, and F. C. M. Lau, "A network perspective of the stock market," *Journal of Empirical Finance*, vol. 17, no. 4, Sep., pp. 659–667, 2010.

This paper presents network constructions of US stocks and proposes a new methodology for computing stock indices. The nodes of the networks consist of stocks, and the edges are established with a winner-take-all approach based on a threshold for cross correlation of daily stock prices, price returns, and trading volumes. The networks display a power-law distribution and suggest that the majority of stocks' time indices are strongly influenced by a small subset of stocks, particularly in the financial sector. They propose a new approach of using the networks for selection of stocks to represent the market indices. The paper highlights some flaws that may be present in the current methodology of determining stock indices and expands the importance of interconnectivity in examining the stock market.

However, their methods are not examined over enough data to generalize the approach; the two windows examined by the authors are each two years in length and centered around the financial crisis of 2007-2008 which may not represent the overall market over time. Further, it does not examine the impact of adjusting several somewhat arbitrary parameters such as the threshold for creating an edge.

1.11 System abnormality detection in stock market complex trading systems using machine learning techniques

P. A. Samarakoon and D. A. S. Athukorala, "System abnormality detection in stock market complex trading systems using machine learning techniques," in 2017 National Information Technology Conference (NITC), Sep., pp. 125–130, 2017.

Using several different supervised learning approaches, this paper explores detecting faults and anomalous behaviors in stock market systems. These methods can reduce the human domain knowledge necessary to detect and correct issues in a trading system. The system developed predicts the overall system state based on the individual states of the sequencing component, distribution component, and matching component. Key features were selected based on filter selection methods using statistics from the data. For the classification, the group used the C4.5 tree algorithm, the naive Bayesian classifier, and the Random Forest algorithm. From metrics such as accuracy, precision, recall, and ROC, the authors conclude the Random Forest approach, which builds decision trees during training, can best represent system state. While the paper provides a good basis for developing a warning system for anomalous behavior and suggests future work in specific areas such as false positive reductions, it does not clearly interpret its results or provide a control group of human anomaly detectors. While it discusses the challenge filtering the features, it is somewhat biased by the assumption that the classic statistical relationships and components selected by domain expertise will best represent the market.

2 Q2: ChatGPT Topic Description