A Decadal Analysis of the Lead-Lag Effect in the NYSE

Aarush Sheth , Jonah Weinbaum, Kevin Zvonarek

1   Introduction

As is widely known, the stock market is a complex system in which a multitude of factors influence the performance of individual stocks and the market as a whole. One method for comprehending—and potentially predicting—stock market behavior is through network analysis, which can offer insights into the relationships between stocks and the overall market structure. In this paper, we seek to address the question: Can network analysis of the stock market, specifically in observation of the lead-lag effect, provide valuable insights for investors and market analysts? This inquiry is both interesting and pertinent for several reasons.

Firstly, grasping the relationships between stocks and the overall market structure can aid investors in making more informed, and potentially more profitable, decisions regarding their investments. Additionally, network analysis may offer new tools for monitoring the stock market and identifying trends or potential risks. Through this investors will be able to, hypothetically, observe the bad returns of a leading stock and therefore draw conclusions about the effects this will have on other closely tied stocks.

To tackle this question, we will build upon two previous studies—of which both observed a power-law distribution in stock returns. The first, performed by researchers at HIT China, constructed a network for the US stock market based on which two stocks were said to be connected if their returns fell in the respective returns of each other over some period to some threshold of accuracy. The study found that through the use of this technique, they were able to select a reasonably performing portfolio that beat the S&P500 [1]. The second study, performed by a group of researchers at Iran's Tehran University, employed community detection techniques to construct a correlation network. They discovered that the resulting communities were consistent with market sectors classified using the Standard Industrial Classification code. They also utilized network analysis and visualization software to generate visualizations of the return correlations among various public stocks, which offered an intuitive way to examine the overall correlation structure of different public stocks and identify key market segments [2].

While these prior works provide valuable insights into the network structure of the stock market, there remains much to expand on this approach. In our research, we concentrate on the lead-lag effect, which refers to the phenomenon where the returns

of one stock lead, over some period—referred to as the "lag"—the returns of another stock. By analyzing the lead-lag effect within the network of the stock market, we aim to offer insights into the dynamics of market behavior and potentially inform investment strategies.

To accomplish this objective, we will first construct a network of the market using the stock return data of the S&P500. We will then replicate the algorithm, and connection thresholds, used by the HIT paper. The algorithm work as follows: using Python's "numba" package—for increased speed we—imported the price data of the S&P500 over the past 10 years and constructed a 500x500 matrix of the closing prices of the stocks. If, over consecutive time periods, the return of stock "i" stays within a given interval of stock "j" then we consider stock "j" to lead to "i," where the time period is called the lag. If a stock leads another stock we place a "1" in our created pseudo-adjacency matrix; pseudo, as although it resembles an adjacency matrix it is not symmetric nor does it have 1's completely along the diagonal. Noticeably, stocks tend to lead themselves, but because of volatility, certain stocks do not.

With our ten years of data input into daily matrices, we plan to do a multitude of things. Foremost, we realize that the lead-lag relationship is a "strategy-enhancer," but not a strategy in itself. Thus, we plan to combine it with a known, semi-reliable, fundamental strategy: CAPM. Simply put, CAPM is an effective way to compute the expected return of a given security based on the market risk premium and the systematic risk of the security. With the coupling of these two strategies, we will, through backtesting, select a portfolio and compare its returns to the S&P500. We hypothesize that our portfolio should outperform the S&P500 over the ten years we backtest.

Next, we will explore additional methods of data analysis, including community detection techniques and principal component analysis. Using our matrices, we will construct daily and yearly lead-lag networks and visualize them. We will also calculate key network measures, such as diameter, density, average path length, average node degree, and clustering coefficient, to gain insight into the relationships between long and short-term lead-lag effects and their relative strength and reliability.

Furthermore, we will conduct PCA on our data set to identify any underlying factors. While we do not anticipate obtaining definitive results from this analysis, it may still provide valuable insights.

In conclusion, our research aims to determine whether network analysis of the stock market, particularly analyzing the lead-lag effect, can provide valuable insights for investors and market analysts. By building on previous work and applying network analysis techniques such as degree centrality, eigenvector centrality, hubs and authorities, and modularity to the stock market network, we hope to contribute to the understanding of the stock market's complex dynamics and potentially inform investment strategies and market monitoring efforts.

References

[1]: Li, Y., Wang, T., Sun, B., & Liu. Detecting the lead–lag effect in stock markets: Definition, patterns, and investment strategies. Financial Innovation, 2022.

[2]: A. Namaki, A.H. Shirazi, R. Raei, and G.R. Jafari. Network analysis of a financial market based on genuine correlation and threshold method. Physica A: Statistical Mechanics and its Applications, 2011.

[3]: R. D. Smith. The Spread of the Credit Crisis: View from a Stock Correlation Network. Journal of Korean Physical Society, 54:2460, June 2009.