Independent study report

Implementation of Stock Network Portfolio allocation algorithm on the multi-layer stock network

Shreekant Gokhale (gokhale6)

Industrial & Enterprise Systems Engineering, UIUC

Abstract

Daily stock price data can be converted to the stock networks based on the correlation between the stocks. These stock networks are useful to get an idea about the interactions of stock price behavior. The structural changes in the stock networks according to the state of the market can be exploited to select the stocks to invest in, for maximizing the returns along with minimizing the risk using diversification. In this study, different aspects of stock correlations are handled using different correlation methods such as Pearson, Spearman, and Mutual Information correlations. The network created using these correlation properties can be combined to form multi-layer stock networks. These multi-layer networks are more robust than single-layer networks as they consider the correlation information from multiple aspects. Moreover, a random walk algorithm can be used on these networks to select diversified stocks for equity portfolios. The performance of these portfolios is further compared using the Sharpe ratio. Stocks selected using multi-layer network outperform other portfolios selected using single layer networks.

1. Introduction

Complex network analysis has been widely used to gain different perspectives on various real-life problems. Stock market data analysis has been one of these fields where creating a network based on the daily stock price data gives an additional advantage. The major advantage of the stock network is that it can easily tell you how different stocks are related to each other depending on the period for which the network is formed. This relation factor cannot be accessed this easily just from the daily prices data.

Moreover, these stock networks can be used for asset management problems like stock portfolio allocation. The stock portfolio allocation using the stock networks problem can be mainly divided into 3 parts.

i) Finding correlation coefficients for stock pairs: Pearson, Spearman, Mutual Information, etc.
ii) Network representation: Threshold, MST,
iii) Stock allocation strategy: Deterministic vs Stochastic

The aim of the Stock portfolio allocation is that we select the proportions in which we invest in each stock to get the maximum return in the future. This selection problem is very complex as stock prices depend on many factors which cannot be anticipated. Also, it is very difficult to forecast stock prices through time series models as stock prices do not follow a particular pattern and cannot be regressed. Hence, we are allocating the stocks based on the study of the correlations of stocks in historical data, assuming these correlations affect future prices.

The work related to all these 3 parts is discussed in detail in the Literature review section of this report.

For the scope of this study, Stochastic stock allocation using the random walk algorithm method is selected. In this study, different kinds of correlation methods are used for the formation of various single-layer networks whose results are compared based on the return they have given in the future. Also, a combination of different correlation techniques is used to construct the multilayer network which allows us to consider different aspects of stock prices relation in the stock network.

The rest of the report explains the data collection, correlation calculation, network formation, and random walk algorithm steps implemented in the code. Following that results that are achieved in the comparison of different portfolio performances are presented.

2. Literature Review

2.1 Correlation Techniques

Stock price data mainly comprises of daily prices of each stock on a particular day. This data does not help us to comprehend the behavior of the market and thus needs to be reformed into useful data. The correlation coefficient between the two stocks is one of the useful parameters which gives us an idea about the change in the price of one stock if we have information about the other stock. If this correlation coefficient is positive, then both stocks will behave similarly meaning, if the price of one stock increases the other stock price will increase as well. The extent to which other stock price increases will be given by the magnitude of the correlation coefficient. In case of a negative coefficient, other stock price decreases with an increase in the price of the first stock. All the correlations that are covered in this study are normalized and hence range between [-1,1].

1) Pearson correlation:
   This is the most widely used correlation for the stock network data. It takes into consideration only the linear correlation between two stocks. Pearson correlation is defined as follows:

$$c_{xy} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2 \sum(y_t - \bar{y})^2}}$$

   Where,
   x_t = log return closing prices of first stock at time t
   y_t = log return closing prices of second stock at time t
   $\bar{x}$ = mean of x values for the given time period
   $\bar{y}$ = mean of y values for the given time period

   Log return values are calculated as follows [1]:

$$r_{i,t} = \frac{p_{i,t} - p_{i,t-1}}{p_{i,t}}$$

   Where, $p_{i,t}$ indicates the price of stock 'i' on time t.

The drawback of the Pearson correlation is that, as it only considers linear correlation, it is possible that two stocks are correlated in some sense, but their Pearson correlation coefficient is negligible. To avoid this, we will be looking at other correlation coefficients and their specific advantages.

2) Spearman correlation:

Spearman correlation is found out by first finding the rank of the values between which correlation is to be found. Spearman correlation is basically the Pearson correlation coefficient of the ranks of the values instead of those values themselves. Spearman correlation finds the monotonic correlation between the stocks. Spearman correlation is given as [2]:

$$c_{xy} = \frac{\sum(R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\sum(R(x_i) - \overline{R(x)})^2 \sum(R(y_i) - \overline{R(y)})^2}}$$

Where, $R(x_i)$ is the rank of the log return value of first stock price at time t.

3) Kendall Tau correlation:

This is also a measure of rank correlation. But instead of using Pearson correlation for rank data, it is calculated by finding the number of concordant and discordant pairs. A pair is called concordant if the sort order of the log return prices of stocks x and y is same for two different dates t_1 and t_2. In that case, pair (t_1,t_2) is called concordant pair. If that is not the case, if the sort order is not same, then that pair is discordant pair.

Kendall tau correlation between two stocks is given as:

$$c_{xy} = \frac{(No. of\ concordant\ pairs) - (No. of\ discordant\ pairs)}{(Total\ No. of\ pairs)}$$

Kendall Tau correlation also finds the monotonic relation between the pair of stocks and is more robust than Spearman correlation.

4) Grey relational correlation:

Deng [3] proposed this method to find the geometric correlation of the time series. In this correlation technique, the closing price data is normalized for the chosen period using the minimum and maximum price in that period. Grey relational correlation showcases better prediction capabilities in the financial data [4].

The grey correlation between two stocks is calculated as [5]:

$$c_{xy} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_{x,y}(t)$$

Where $\varepsilon_{i,j}(t)$ is given as:

$$\varepsilon_{x,y}(t) = \frac{min_y min_t \Delta_{y,t} + \rho \times\ max_y max_t \Delta_{y,t}}{\Delta_{y,t} +\ \rho \times\ max_y max_t \Delta_{y,t}}$$

ρ is any value in the interval (0,1] which is generally set at 0.5 [6].

$\Delta_{y,t}$ is given by $|x'(t)-y'(t)|$ where, $x'(t)$ is the min max normalized closing price at time t.

5) Mutual Information

Mutual information correlation is the entropy theory concept. It is a more robust way of calculating the dependence between two stocks than rank correlations and grey relational coefficients. Normalized version of mutual information is calculated as [7]:

$$NMI(x, y) = \frac{2I(x, y)}{H(x) + H(y)}$$

$$H(x) = -\sum_{t} p(x_t) log_2 p(x_t)$$

$$H(x, y) = -\sum_{t}\sum_{t} p(x_t, y_t)\, log_2 p(x_t, y_t)$$

$$I(x, y) = H(x) + H(y) - H(x, y)$$

Where, p(x) is probability distribution of x and p(x,y) is joint probability distribution of (x,y)

2.2 Network formation techniques

The correlation matrix formed by the collection of the correlation coefficients of all the possible stock pairs is further used to convert into the network. This network has all the stocks as the nodes. Edges between the stocks are decided depending on the type of network to be formed.

The correlation threshold network is the most widely used network formation method. In this method, two stocks are connected by an edge if the correlation between them is higher than the threshold value. The weight assigned to these edges is equal to the correlation between these networks. Threshold networks are used where clustering or stochastic techniques are implemented.

Moreover, the correlation matrix can be converted to simpler network formations using different filtration methods. Minimum Spanning Tree (MST), Asset graph and Planar Maximally Filtered Graph (PMFG) are such examples. In these networks only the most significant edges are considered in the network.

In the case of MST, correlation between the stock pairs is converted to the distance attribute. Distance between two nodes representing stocks x and y is calculated as [8]:

$$d_{x,y} = \sqrt{2(1 - c_{x,y})}$$

Hence, higher the correlation between the stocks lower will be the distance between them. All the stocks represented as nodes are connected by the edges assigned with the weights equal to the corresponding distance attribute. Then minimum spanning tree is found out which gives us most significant correlations covering all the stocks. MST networks are useful while studying the effect of the state of the market on the network topology. MST gives the overall view of the market for the chosen time period, which otherwise is not possible with the daily stock price data.

Asset graph is another such network suggested by Onnela et al.[9] where n-1 smallest distance pairs are chosen from the ordered list which give us the highest correlated stock pairs. As both MST and AG have n-1 edges comparison between these two networks is easily possible. Based on the distribution of the stocks of particular economic sector in the different clusters formed in MST and AG we can predict the state of the market. A crash in the market affects all the sectors in a similar manner and hence stocks from the particular sector are spread all over the network. In other words, clusters in the MST networks are not defined by sectors and each cluster will have mixture of multiple sectors when there is downfall in the market. But in case of the bubble in specific sector, correlations between the stocks of this sector are most significant and hence, clusters can easily be seen in the MST networks which are majorly defined by the economic sector of the stocks which are causing bubble in the market.

Tumminellow et al. [10] proposed another network structure called PMFG where the most significant edges (with smaller distance) in the network are sequentially chosen such that the network can be drawn on the planar surface without its edges crossing each other. Such structure can have maximum 3(n-2) edges in the network. Main advantage of this network structure is that the level of filtering can be chosen by changing the genus of the surface. Also, PMFG network contains negatively correlated edges as well which can be useful in problems like portfolio management where the diversification of the stocks is required.

Chen W. et al. [5] suggest the formation of multi-layer networks to combine different aspects of relations between the stocks in the single network. The correlation matrix for this network is calculated using the combination of Spearman, Grey relational and MI correlation matrices using the union function. Such multi-layer networks are observed to be better representatives of the market and are more robust than the single layer networks.

2.3 Stock allocation strategies for portfolio management

In the stock allocation problem, we want to maximize the return on investment and minimize the loss risk by allocating the money to different assets. The solution also includes the proportion in which these investments must be made in the available assets. In this study, only stock allocation strategies where stock network is used are discussed.

1) Deterministic techniques:
   In deterministic techniques, topological study of the networks based on different states of market is exploited to select the stocks to invest in. For studying the effects of market behavior on the network structure, various centrality measures such as Degree centrality, Betweenness centrality, Eccentricity, Closeness, Eigenvector centrality are used. A combination of these measures is used to find the peripherality attribute of the node in the network. For topological analysis PMFG networks are preferred. Li et al. [11] suggest that peripherality in the network can be used as an indicator to select the stocks for portfolio management.
   Ren et al. [12] indicate that the central portfolios perform better in declining market conditions. On the other hand, they indicate that peripheral portfolios outperform central portfolio when the market is stable. These observations can be used to decide the strategy to select the assets in the portfolio.

2) Stochastic techniques:
   In these strategies, the threshold stock network is covered using the random walk. Chen et al. [13] show that highest visited nodes after the random walk can be used as a stock selection strategy in the portfolio. This method is referred to as Stock Network Portfolio Allocation (SNPA). The network used in SNPA is the threshold network where two thresholds are used to include positive as well as negative correlation values. Each available stock is represented as a node and edges are formed between these nodes if the correlation between the pair of stocks is either lesser than the negative threshold or if it is higher than positive threshold. This ensures diversification in the portfolio selected.

In the random walk algorithm [14], probability of jumping from node x to y (transition probability) is defined by the softmax function on the correlation coefficients between stocks x and y.

$$P_{xy} = \frac{e^{c_{xy}}}{\sum_i e^{c_{xi}}}$$

$\forall$ i $\in$ nodes of the stock network

With the calculated transition probabilities, a next instance in the random walk is chosen. This process is carried out k many times. k is the number of iterations of random walk to carry out. The q many most visited nodes in the network are then selected in portfolio. Weights assigned to these stocks are defined as:

$$w_i = \frac{v_i}{\sum_{i=1}^{q} v_i}$$

Where v_i is the number of visits to the node 'i' during the random walk. In this case, threshold values and the number of iterations are the hyperparameters. Positive and negative threshold values are selected in such a way that the resulting stock network is completely connected.

Freitas et al. [14] propose hybrid SNPA algorithm which includes the forecast modeling like XGboost along with the random walk algorithm to select the stocks in the portfolio. A major advantage inclusion of forecasting is that it is no longer assumed that the future prices of the stocks are solely dependent on the historical behavior.

3. Data and methods used

For the scope of this study, daily stock price data of the top 100 stocks in US is chosen which is extracted from Yahoo Finance website. The daily closing price of these stocks for the time period 1st April 2021 to 31st March 2022 is selected for the correlation analysis.

Log return values of these closing prices are calculated for further calculating Pearson, Spearman and Kendall Tau correlations between between all pairs of stocks. A correlation matrix for each of these correlations is found out.For the given time period, Grey relational correlation and NMI correlation matrices are also calculated. Note that all these correlation matrices are normalized. Single layer networks for Grey relational and NMI correlation are also formed.

According to Chen W. et al. [5] multilayer correlation matrix is calculated by the matrix multiplications of Spearman, Grey relational and NMI correlation matrices.

The best possible upper and lower threshold limits are chosen for all the correlation matrices so that the resulting stock network does not have any disconnected component. After finding these threshold values, 5 single-layer and one multi-layer stock threshold networks are formed.

Moreover, Random walk algorithm is applied on these networks to find out the stock portfolios and weights corresponding to each network. q value for these portfolios is chosen to be 5 ie. 5 most visited stocks are selected as the portfolio.

4. Results

Results for each of these portfolios are shown in the following table.

Network	Stocks in the portfolio	Corresponding economic Sectors
Pearson	ATVI: Activision Blizzard	Communication Services
	KHC: Kraft Heinz	Consumer Goods
	XEL: Xcel Energy	Utilities
	AEP: American Electric Power	Utilities
	GILD: Gilead Sciences	Healthcare
Spearman	XEL: Xcel Energy	Utilities
	AEP: American Electric Power	Utilities
	KHC: Kraft Heinz	Consumer Goods
	GILD: Gilead Sciences	Healthcare
	EXC: Exelon	Utilities
Kendall Tau	AEP: American Electric Power	Utilities
	GILD: Gilead Sciences	Healthcare
	XEL: Xcel Energy	Utilities
	KHC: Kraft Heinz	Consumer Goods
	EXC: Exelon	Utilities
Grey relational	TMUS: T Mobile	Communication Services
	BIDU: Baidu	Communication Services
	CDW: CDW Corp.	Technology
	SPLK: Splunk	Technology
	AMGN: Amgen	Healthcare
NMI	CPRT: Copart	Consumer Goods
	AVGO: Broadcom	Technology
	INTU: Intuit	Technology
	PEP: Pepsico	Consumer Goods
	ASML: ASML Holding	Technology
Multi-layer	BKNG: Booking Holdings	Consumer Goods
	XEL: Xcel Energy	Utilities
	KHC: Kraft Heinz	Consumer Goods
	AEP: American Electric Power	Utilities
	REGN	Healthcare

Table 1: Stocks are arranged according to the decreasing weightage assigned in the portfolio.

Performance of these portfolios is compared using the Sharpe ratio. Sharpe ratio is the portfolio performance metric which is calculated as the ratio of average return to the standard deviation of each portfolio. Sharpe ratio gives us the risk adjusted performance of the portfolio.

Based on the Stocks selected, and weight assigned in each portfolio, Sharpe ratio is found. Results for this analysis are shown in the following table.

Network Portfolio	Annualized Sharpe Ratio
Pearson	0.024
Spearman	0.844

Kendall Tau	0.807
Grey relational	-0.003
NMI	0.928
Multi-layer	1.012

Table 2: Annualized Sharpe ratio for each of the portfolio

5. Conclusion

In this study, SNPA algorithm is implemented on various single layer and a multilayer stock networks to see if portfolio performance changes depending on the correlation criteria. From the table 1, we can see that stocks are selected from the diverse sectors. Recommendation of stocks changes according to the correlation used. The same kind of correlations select similar stocks but the weightage assigned to these stocks changes. It can be seen that, Pearson, Spearman, and Kendall Tau networks select almost the same stocks in their portfolio. But from the Sharpe ratio it can be seen that merely considering linear correlations between the stocks is not sufficient for the portfolio to perform well. As monotonicity factor is included using either Spearman or Kendall Tau correlation, correction in the weightage of the stocks selected is done. These portfolios are near the acceptable Sharpe ratio range.

Stock selections made by the NMI network are far different that the previous portfolio selections. This can be a confirmation of the fact that different correlation methods consider different aspects of stock relations in the market. From the Sharpe ratio analysis, it can be seen that multi-layer network portfolio outperforms all the other portfolio selection.

6. Future Scope

The experiments carried out in this study are not sufficient enough to conclude the effectiveness of any particular correlation method or any particular portfolio allocation method. We can carry out hyperparameter analysis in the SNPA algorithm for better results. Also, these experiments need to be performed on different time periods to see if the portfolio allocation method can be generalized. Dynamic stock networks can be useful to consider the changing topological properties of the stock networks along with the time.

7. References

1] Masmoudi M., & Abdelaziz F. (2018) Portfolio selection problem: a review of deterministic and stochastic multiple objective programming models, Annals of Operations Research, 267, 335-352.

2] A. Kumar, S. Abirami, Aspect-based opinion ranking framework for product reviews using a spearman's rank correlation coefficient method, Inform. Sci. 460 (2018) 23–41.

3] D. Julong, Introduction to grey system theory, J. Grey Syst. 1 (1) (1989) 1–24.

4] S.L. Lin, S.J. Wu (2011). Is grey relational analysis superior to the conventional techniques in predicting financial crisis? Expert Syst. Appl. 38 (5) 5119–5124.

5] Chen W, Qu S., Jian M., & Jiang C. (2021) The construction of multilayer stock network model, Physica A 565, 125608.

6] F. Xu, S. Zhang, K. Wu, Z. Dong, Multi-response optimization design of tailor-welded blank (TWB) thin-walled structures using taguchi-based gray relational analysis, Thin-Walled Struct. 131 (2018) 286–296.

7] Guo X., Zhang H & Tian T. (2018) Development of stock correlation networks using mutual information and financial big data. PLoS ONE 13(4): e0195941

8] Birch J., Pantelous A., & Soramaki K. (2015) Analysis of correlation based networks representing DAX 30 stock price returns, Computational Economics, 47(4), pp. 501-525

9] Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, K., & Kanto, A. (2003). Asset Trees and Asset Graphs in Financial Markets. Physic Scripta, T106: 48-54.

10] Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R.N. (2005). A Tool for Filtering In-formation in Complex Systems. *Proceedings of the National Academy of Sciences, 102*(30): 10421-10426.

11] Li Y., Jiang X., Tian Y., Li S., & Zheng B. (2019). Portfolio optimization based on network topology. Physica A. 515, pp 671-681.

12] Ren F., Lu Y., Li S., Jiang X., Zhong L., & Qiu T. (2017) Dynamic portfolio strategy using clustering approach. PLoS ONE 12(1): e0169299.

13] Chen, W., Jiang, M., & Jiang, C. (2019). Constructing a multilayer network for stock market. Soft Computing, 24, 1–17.

14] Freitas W., & Bertini J. (2023) Random walk through a stock network and predictive analysis for portfolio optimization. Expert Systems With Applications 218, 119597.