Minimal Spanning Tree application to determine market correlation structure

Bui Thanh Khoa[1*], Tran Trong Huynh[2], Vo Dinh Nhat Truong[2], Le Vu Truong[2], Do Bui Xuan Cuong[1], Tran Khanh[1]

[1]Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam
[2]FPT University, Hanoi, Vietnam
[*]Corresponding author: buithanhkhoa@iuh.edu.vn

ARTICLE INFO

DOI:10.46223/HCMCOUJS.tech.en.13.1.2669.2023

Received: March 01st, 2023

Revised: March 10th, 2023

Accepted: March 21st, 2023

Keywords:

Covid-19 pandemic; market correlation structure; mean of distances; Minimal Spanning Tree; T-Test

ABSTRACT

Determining the structure of market correlation is an important topic in theory and experiments. Under the impact of the Covid-19 pandemic, the market structure may be deformed. Therefore, this study examines the pandemic's impact on the market structure. This study considered the correlation structure of the VN30 portfolio (including 30 stocks with the largest market capitalization); the collecting period is from July 28, 2000, to July 30, 2021. The data was divided into 02 phases before and after the pandemic. The Kruskal algorithm is implemented to determine the Minimal Spanning Tree (MST) structure to define the structure of market correlation. This study compared the change in the structure before and after the Covid-19 pandemic by structures' mean of distances comparison. T-test results show that there are structural differences before and after the pandemic. Based on the research result, investors should change their risk management strategy to suit the market context because the previous structure has been changed.

1. Introduction

Substantial research on the structure and topological features of complex networks has been conducted by physicists during the last two decades (Dorogovtsev, Goltsev, & Mendes, 2008). They learned that tiny and finite loops in graphs are uncommon and unimportant. Therefore, it is reasonable to assume that trees play a crucial role in their architectural designs. All of these functions have seen heavy usage. That this assumption holds true even in complex networks with many loops and nodes is amazing. Therefore, the Minimum Spanning Tree (MST) technique, a correlation-based connected network with no loops, was chosen as a valuable, classic graph theory tool (Bollobás, 1998). The businesses are represented as vertices (nodes) in the graph, and the distances between them are calculated using their correlation coefficients (Tumminello, Di Matteo, Aste, & Mantegna, 2007). The required transformation of correlation coefficients into distances was done using a straightforward procedure.

This study investigates the temporal dynamics of a complex empirical network of companies listed on the Warsaw Stock Exchange (WSE). It is possible for the number of businesses (vertices) and their distances to vary. These numbers are consistent at times and inconsistent at others. As the network evolves, some edges may go, while others may arise. Consequently, neither the number nor the borders of enterprises are maintained. Consequently, their properties, including mean length and mean occupation layer, have altered (Brida & Risso, 2010), constantly changing over time, as explained below.

Through the application of the MST technique, which pointed out how the structure of the stock market changed from a hierarchical (power law) tree before the current global financial crisis to a superstar-like tree (superhub) ornamented by a hierarchy of trees (hubs) during the collapse (Sornette, 2017). After the 2008 financial crisis, this study found that the market structure and topology reflected by the thorny tree gave way to a power law tree adorned by a hierarchy of local star-like trees or hubs (the wealthiest of these hubs may be a contender for another superhub).

Although there have been many studies on the use of MST to determine the structure of the financial market, there has not been a consensus (Brida & Risso, 2010; Sornette, 2017); nonetheless, research on the Vietnamese stock market remains sparse. Our issue, particularly structural alterations caused by the Covid-19 pandemic, is the impetus for this study. This research provides a phenomenological basis for modeling dynamic structural and topological phase transitions and critical events in financial markets.

2. Literature review

One of the most consequential market events in history was the global financial crisis of 2008 (henceforth crisis). No one can really know how this disaster will unfold. Markets in the financial sector may be understood in a number of different ways (He & Deem, 2010). Network analysis has proven beneficial for assessing both its static and dynamic properties (Nobi & Lee, 2016). When it comes to cross-correlation networks, only stock price data and quasi-synchronously recorded global indexes from across the world are allowed to be analyzed (Nobi, Maeng, Ha, & Lee, 2015). A key field of research in today's globally interconnected economy is the dynamics between various financial institutions. Recently, structural changes in the global financial network have been evaluated using correlation-based network analysis of global financial indices with great success (Wang & Xie, 2015). By setting the threshold, the MST, and the Planar Maximally Filtered Graph (PMFG), the cross-correlation network may be used to investigate the properties of eigenvalues, eigenvectors, and the threshold network (TN).

Onnela, Saramäki, Kertész, and Kaski (2005) employed correlation network approaches to analyzing the structural changes in a regional financial system in the wake of the crisis. Based on the average cross-correlation threshold, Qiu, Zheng, and Chen (2010) created both static and dynamic financial networks. Network analysis using an absolute threshold for cross-correlation coefficients was used by Kumar and Deo (2012) to look at the international banking industry. All during the crisis, they saw that the European indexes were significantly linked at a high threshold. When studying the Tehran Stock Exchange and the Dow Jones Industrial Average, Namaki, Shirazi, Raei, and Jafari (2011) uncovered the threshold network. They discovered scale-free threshold networks within a small region around the threshold. Threshold network analysis was also used by Wang et al. (2016) to examine the Chinese stock market. According to Maqsood et al. (2020), publicly traded Turkish companies were unaffected by the MST scandal. The French FSBF 250 was named by Zhang, Zhuang, Tang, and Han (2022) as a major stock exchange reporting center for both MST and PMFG. Kumar, Kumar, and Kumar (2020) discovered Multi-Stage Trading (MST) and Hierarchical Networks (HN) in international currency and commodity exchanges.

This study explores the noticeable topological alterations in the MST before and after the Covid-19 pandemic. In addition, the research examined the link between geometrical and network distances and found a scaling relationship.

3. Research method

Research information comprises 30 businesses from the VN30 portfolio (price unadjusted), cycled in days. The amount and labeling of stock are shown in Table 1. The information is derived from the label and observations in the VN30 portfolio from Bui and Huynh (2022)'s study.

Since some businesses were created on July 28, 2000, and others on July 30, 2021, and some days aren't traded, the total number of company observations will be uneven. The Ho Chi Minh City Stock Exchange (Ho Chi Minh Stock Exchange, n.d.) is the source for this information. Time (date), ticker, closing price, open price, high price, low price, and trading volume are all recorded for each observation (vol). Stocks are only analyzed based on their final pricing.

The data was then separated into two parts, before and after the pandemic at the time of the outbreak, December 2019. Returns on stocks are approximated by Equation (1):

$$r_{it} = ln\frac{p_{it}}{p_{it-1}} \times 100\% \quad (1)$$

Where, p_{it} Is the closing price of stock i at time t.

Next, this study calculated the correlation matrix $\rho = \left[\rho_{ij}\right]_{i,j=1,2,\ldots 30}$, where: ρ_{ij} is the coefficient of linear correlation between stocks i and j. Then this study calculated the distance matrix using Equation 2:

$$d = \sqrt{2 - 2\rho} \quad (2)$$

Additionally, the MST may be calculated by determining the index distance matrix. We may express the separation between two nations, I and j, as $d_{ij} = \sqrt{2 - 2\rho_{ij}}$, if indices i and j are perfectly linked, $d_{ij} = 0$, and if they're perfectly anti-correlated, $d_{ij} = 2$. The N (N - 1)/2 potential index pairings are narrowed down to the N (N - 1)/2 most relevant pairs using Kruskal's method as the algorithm below:

```
algorithm Kruskal(G) is
    F:= ∅
    for each v ∈ G.V do
        MAKE-SET(v)
    for each (u, v) in G.E ordered by weight(u, v), increasing do
        if FIND-SET(u) ≠ FIND-SET(v) then
            F:= F ∪ {(u, v)}s ∪ {(v, u)}
            UNION(FIND-SET(u), FIND-SET(v))
    return F
```

Finally, this study compared the difference in MST profiles before and after the pandemic by using the T-test for the difference in distance between these two MST structures.

4. Result

4.1. Descriptive statistic

The results of the combined descriptive statistics of all variables are summarized in Table 1. The price fluctuates from 5.3 (thousand VND/share) to 665 (thousand VND/share), and the average price is 51.54 (thousand VND/share). Some stocks have unusually negative returns, such

as -74.257%, because these stocks immediately pay dividends, so the stock price, including dividends, immediately decreases significantly.

Table 1

Descriptive statistic

statistics	close	return
min	5.3	-74.257
median	37.2	0
max	665	9.742
mean	51.54	0.004
standard deviation	43	0.009

Figure 1 points out that the average return of stocks in the VN30 portfolio is approximately 0.004% (excluding dividends) and has a relatively low standard deviation of 0.009. The reason is that these companies have high market capitalization, so the overall risk is relatively low. Equities such as FPT, REE, SSI, and STB shifted to the right and closed at extremely high levels in some trading sessions. Some stocks, such as GAS and VIC, have a reasonably steady and symmetric distribution.

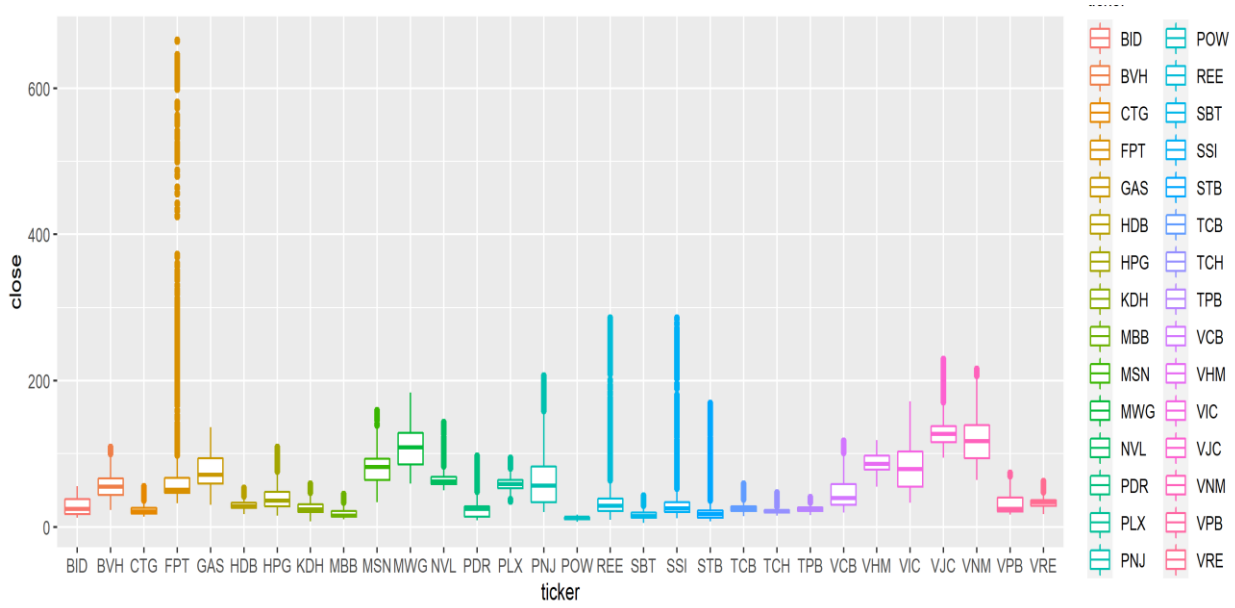

Figure 1. Movement of closing price across securities codes

4.2. Correlation between stocks, distance, and MST

During the Covid-19 pandemic, the correlation between stock pairings was more variable than before the outbreak. During a pandemic, the market is more volatile and unpredictable than during periods of calm. The standard deviation is a convincing illustration of this claim. Before the pandemic, the standard deviation of the correlation pairs in Table 2 is around 0.255, but this number is greater after the pandemic (0.276).

Figure 2 shows the results of using the Kruskal method to calculate MST before and after the pandemic. Before and after the pandemic, there has been an enormous shift. Significant changes in the MST positions suggest that the correlation structure of the VN30 portfolio has also changed considerably.

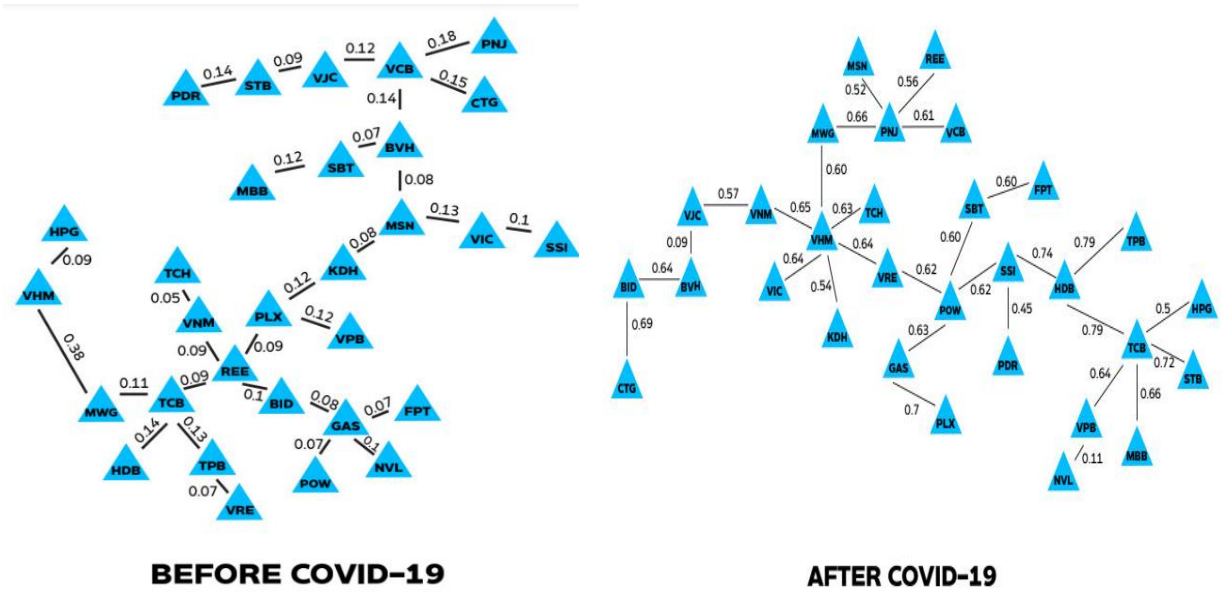

Figure 2. MST before and after the pandemic

4.3. T-test for the difference in distance

The correlation structure of the VN30 portfolio before and after the pandemic is seen in Figure 2. However, this hypothesis must first be tested. This research does a T-test on the assumption that the distances are normally distributed, have homogenous variance, and are independent. For the mean test, it is crucial to assume that the variances of the two populations are identical. Specifically, the F test was utilized to evaluate this assumption. The findings are given in Table 2.

Table 2

F-test result

F -Test: Comparing the variances of two independent populations		
The null hypothesis (H0) is that variances are equal		
Period	Before	After
Variance	0.003	0.025
Observations	29	
df	28	28
F Stat	0.12	
P value	2.69E-07	

The P-value is close to 0, indicating sufficient data to suggest a difference in variance between before and after the pandemic. Also, for this reason, the T-test assumes uneven variance. The findings of the T-test are presented in Table 3.

Table 3

T-test result

t-Test: Two-Sample Assuming Unequal Variances		
	Before	*After*
Mean	0.118	0.599
Variance	0.003	0.025
Observations	29	
df		35
t Stat	-15.471	
P(T< = t) two-tail	0.000	
t Critical two-tail	2.030	

It is assumed in this T-test that there is no difference in the means of distances or the null hypothesis H0. According to the results of the T-test, the p-value is very low, making it clear that H0 cannot be accepted. The findings point to a distance difference that is statistically significant. This evidence implies a difference between the VN30 portfolio's structure before and after the pandemic.

5. Conclusion

The study of the market correlation structure has important experimental and theoretical significance. Knowing the structure, investors can implement several investment strategies that suit their risk appetite and minimize risk for a given level of expected return.

This study used MST to determine the shortened structure of the VN30 portfolio (including 30 stocks with the largest capitalization HOSE). This research compared the structural change before and after the Covid-19 pandemic. The results of the T-test show a difference in distance before and after the pandemic. This result implies a change in the correlation structure of the VN30 portfolio.

One limitation of this study is that it only considers the correlation structure of stocks in the VN30 portfolio. Hence, this study proposes extending to all stocks in the HOSE market. Furthermore, further research should be directed toward applying MST in portfolio diversification.

References

Bollobás, B. (1998). *Modern graph theory* (Vol. 184). New York, NY: Springer Science & Business Media.

Brida, J. G., & Risso, W. A. (2010). Hierarchical structure of the German stock market. *Expert Systems with Applications, 37*(5), 3846-3852.

Bui, K. T., & Huynh, T. T. (2022). Forecasting stock price movement direction by machine learning algorithm. *International Journal of Electrical and Computer Engineering (IJECE), 12*(6), 6625-6634. doi:10.11591/ijece.v12i6.pp6625-6634

Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Reviews of Modern Physics, 80*(4), 1275-1335. doi:10.1103/RevModPhys.80.1275

He, J., & Deem, M. W. (2010). Structure and response in the world trade network. *Physical Review Letters, 105*(19), 1-4.

Ho Chi Minh Stock Exchange. (n.d.). Retrieved Januaray 10, 2023, from https://www.hsx.vn/

Ho Chi Minh Stock Exchange. (2021). *Trading data 2021*. Retrieved Januaray 10, 2023, from https://www.hsx.vn/Modules/Rsde/Report/ReportTradingView?fid=a2220adea05242a29e6130d9e3d19ae7

Kumar, S., & Deo, N. (2012). Correlation and network analysis of global financial indices. *Physical Review E, 86*(2), 1-8.

Kumar, S., Kumar, S., & Kumar, P. (2020). Diffusion entropy analysis and random matrix analysis of the Indian stock market. *Physica A: Statistical Mechanics and its Applications, 560*(2020), 1-24.

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., … Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management, 50*(2020), 432-451.

Namaki, A., Shirazi, A. H., Raei, R., & Jafari, G. (2011). Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and Its Applications, 390*(21/22), 3835-3841.

Nobi, A., & Lee, J. W. (2016). State and group dynamics of world stock market by principal component analysis. *Physica A: Statistical Mechanics and its Applications, 450*(2016), 85-94.

Nobi, A., Maeng, S. E., Ha, G. G., & Lee, J. W. (2015). Structural changes in the minimal spanning tree and the hierarchical network in the Korean stock market around the global financial crisis. *Journal of the Korean Physical Society, 66*(2015), 1153-1159.

Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E, 71*(6), 1-5.

Qiu, T., Zheng, B., & Chen, G. (2010). Financial networks with static and dynamic thresholds. *New Journal of Physics, 12*(4), 1-16.

Sornette, D. (2017). Why stock markets crash. In *Why stock markets crash.* Princeton, NJ: Princeton University Press.

Tumminello, M., Di Matteo, T., Aste, T., & Mantegna, R. N. (2007). Correlation based networks of equity returns sampled at different time horizons. *The European Physical Journal B, 55*(2017), 209-217.

Wang, C.-H., Zhong, Y., Zhang, Y., Liu, J.-P., Wang, Y.-F., Jia, W.-N., . . . Gao, X.-M. (2016). A network analysis of the Chinese medicine Lianhua-Qingwen formula to identify its main effective components. *Molecular BioSystems, 12*(2), 606-613.

Wang, G.-J., & Xie, C. (2015). Correlation structure and dynamics of international real estate securities markets: A network perspective. *Physica A: Statistical Mechanics and Its Applications, 424*(2015), 176-193.

Zhang, D., Zhuang, Y., Tang, P., & Han, Q. (2022). The evolution of foreign exchange market: A network view. *Physica A: Statistical Mechanics and Its Applications, 608*(2022), 1-25.