Risk Transmission Mechanism across Industries in China based on VAR-Lasso

Zhaoyang Chen

May 27, 2020

Abstract

It is crucial to learn how the risk has been transmitted between industries in China before and after some big events. To conduct the empirical research on risk transmission mechanism in China based on daily data of the SWS secondary industry index, I utilize the LASSO algorithm to compress, select and estimate variables and build a high dimensional VAR model to calculate the pairwise risk connectedness between different industries. With the help of network analysis, I visualize the outcome of the VAR-Lasso model. Then, both full sample estimation and rolling window estimation are applied to make a static and dynamic study of the risk transmission network. As is revealed in dynamic analysis, clustering characteristics can be easily seen in risk transmission in the market, especially between elements in the same industrial chain or between industries that are closely connected. Particularly, Oil exploitation, Insurance, Banking, and Railway Transportation are functioning as the efficient intermediary node in the whole transmission process. As for dynamic analysis, the overall risk connectedness reaches the summit of the great stock damage in 2015 and the shock of Covid-19. Comparisons are made on the risk transmission network before and after those big events.

1 Data

I obtain the data of 104 industries from the SWS secondary industry index, from October 12, 2010 to May 23, 2020. The data is in classical OHLC format, which contains high price, low price, open price and close price during the day. However, I find that 8 industries in the data contain missing values and most of them are 100% missing. Thus, I delete them from the raw data and the number of industries becomes 96. Before constructing the model, I have to transform the OHLC data to another format. Following [5], I calculate daily stock return as

$$\widetilde{\theta}_{it}^2 = 0.511(H_{it} - L_{it})^2 - 0.019[(C_{it} - O_{it})(H_{it} + L_{it} - 2O_{it}) - 2(H_{it} - O_{it})] - 0.383(C_{it} - O_{it})^2, \quad (1)$$

where $H_{it}, L_{it}, O_{it}, C_{it}$ are, respectively, the logs of daily low, opening, high, and closing prices for $industry_i$ on day_t.

2 Methods

2.1 Diebold–Yilmaz volatility connectedness framework

Here, I will introduce the Diebold–Yilmaz volatility connectedness framework and measures, which serve as the base of the full model. The framework is fundamentally about connectedness measures based on variance decompositions, which is proposed and developed by [6] and [3].

2.1.1 Variance Decompositions for Risk Connectedness Measurement

Connectedness measures based on variance decompositions are practical and simple. First, they are directly linked to network model, and they are also related to systemic risk, such as marginal CoVaR [2] and ES [1]. Secondly, different connectedness at different horizons is allowed, leading to selection of a preferred horizon and the examination of a variety of horizons if desired[3]. Finally, they make obvious intuitive sense, answering a key question, which at the most granular pairwise level is "How much of entity i's future uncertainty (at horizon H) is due to shocks arising not with entity i, but rather with entity j?".

2.1.2 Vector Autoregression Models

I conduct the variance decomposition on the VAR model. In time series analysis, the auto-regression representation is $x_t = \sum_{i=1}^{p} \Phi_i x_{t-i} + \varepsilon_t$, where $\varepsilon_t \sim (0, \Sigma)$. And the moving average representation is $x_t = \sum_{i=0}^{\infty} A_i \varepsilon_{t-i}$, where the $N \times N$ coefficient matrices A_i obey the recursion $A_i = \Phi_1 A_{i-1} + \Phi_2 A_{i-2} + \ldots + \Phi_p A_{i-p}$, with A_0 an $N \times N$ identity matrix and $A_i = 0$ for $i < 0$.

Identification of the lagging parameter becomes challenging in the high-dimensional situations, which is the key question I want to answer in the paper. Here, I follow [4] in using the "generalized identification" framework of [7], which produces variance decompositions invariant to rank. Instead of attempting to orthogonalize shocks, the generalized approach allows for correlated shocks but accounts appropriately for the correlation[4] .

2.1.3 Risk Connectedness Measures

In this part I want to introduce the risk connectedness measures. Let's begin with the pairwise directional connectedness, and proceed with total directional connectedness to calculate the systematic connectedness.

Suppose we are in a huge market network, node $j's$ contribution to node $i's$ H-step-ahead generalized forecast error variance, $\Theta_{ij}^{g}(H)$, is

$$\theta_{ij}^{n}(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' A_h \sum e_j)^2}{\sum_{h=0}^{H-1} (e_i' A_h \sum A_h' e_i)}, H = 1, 2, \ldots \tag{2}$$

where θ_{jj} is the standard deviation of the disturbance of the jth equation Σ is the covariance matrix of the disturbance vector ε, and e_i is the selection vector with one as the ith element and zeros otherwise.

Next I normalize each entry of the generalized variance decomposition matrix (Equation 1) by the row sum to obtain pairwise directional connectedness from node j to node i:

$$\widetilde{\theta}_{ij}^{g}(H) = \frac{\theta_{ij}^{g}(H)}{\sum_{j=1}^{N} \theta_{ij}^{g}(H)} \tag{3}$$

As a matter of notation, I now convert from $\widetilde{\theta}_{ij}^{g}(H)$ to $C_{i\leftarrow j}^{N}$ (C is, of course, for risk connectedness), which is more directly informative.

After obtaining the pairwise directional risk connectedness measure $C_{i\leftarrow j}^{N}$, I can move to total directional risk connectedness measures. Total directional risk connectedness to node i from all other nodes j is

$$C_{i\leftarrow \bullet} = \frac{\sum_{j=1,j\neq i}^{N} \widetilde{\theta}_{ij}^{g}(H)}{\sum_{i,j=1}^{N} \widetilde{\theta}_{ij}^{g}(H)} = \frac{\sum_{j=1,j\neq i}^{N} \widetilde{\theta}_{ij}^{g}(H)}{N}. \tag{4}$$

Similarly, total directional risk connectedness from node i to all other nodes j is

$$C_{\bullet\leftarrow i} = \frac{\sum_{j=1,j\neq i}^{N} \widetilde{\theta}_{ji}^{g}(H)}{\sum_{i,j=1}^{N} \widetilde{\theta}_{ji}^{g}(H)} = \frac{\sum_{j=1,j\neq i}^{N} \widetilde{\theta}_{ji}^{g}(H)}{N}. \tag{5}$$

Finally, I obtain the systematic connectedness measure. Using the normalized entries of the generalized variance decomposition matrix (Equation 2), I measure total directional risk connectedness as

$$C^{H} = \frac{\sum_{i,j=1,j\neq i}^{N} \widetilde{\theta}_{ji}^{g}(H)}{\sum_{i,j=1}^{N} \widetilde{\theta}_{ji}^{g}(H)} = \frac{\sum_{i,j=1,j\neq i}^{N} \widetilde{\theta}_{ji}^{g}(H)}{N}. \tag{6}$$

I call this total risk connectedness *systematic* connectedness. It is simply the sum of total directional risk connectedness whether “to” or “from.”

2.2 Volatility Connectedness Estimation with Lasso

So far I have discussed Diebold–Yilmaz volatility connectedness measurement[3]. Now I want to discuss the estimation with LASSO, which can be useful in the high-dimension network analysis. I conduct the connectedness assessment on an estimated VAR approximating model. Due to the high-dimension property of the data, I need the VAR to be estimable in high dimensions, somehow recovering degrees of freedom. People can do so by pure shrinkage (as with ridge regression) or pure selection (as with traditional criteria like AIC, BIC or SC). However, combining shrinkage and selection with variants of the LASSO is particularly crucial. To understand the LASSO, consider the OLS estimation:

$$\widetilde{\beta} = \arg\min_{\beta} \sum_{i=1}^{T} (y_t - \sum_{i} \beta_i x_{it})^2 \tag{7}$$

subject to the constraint:

$$\sum_{i=1}^{K} |\beta_i|^q \le c. \tag{8}$$

Equivalently, consider the penalized estimation problem:

$$\widetilde{\beta} = \arg\min_{\beta} [\sum_{i=1}^{T} (y_t - \sum_{i} \beta_i x_{it})^2 + \lambda \sum_{i=1}^{K} |\beta_i|^q]. \tag{9}$$

Concave penalty functions non-differentiable at the origin produce selection, whereas smooth convex penalties (e.g., q = 2,the ridge regression estimator) produce shrinkage. Hence penalized estimation nests and can blend selection and shrinkage. The LASSO algorithm introduced in the seminal work of [8], solves the penalized regression problem with q = 1. Hence it shrinks and selects. Moreover, it requires only one minimization, and it uses the smallest q for which the minimization problem is convex[3].

An extension of the LASSO, the so-called adaptive elastic net (Zou & Zhang, 2009), not only shrinks and selects, but also has the oracle property. In the implementation of the adaptive elastic net, I solve

$$\widetilde{\beta}_{AEnet} = \arg\min_{\beta} [\sum_{i=1}^{T} (y_t - \sum_{i} \beta_i x_{it})^2 + \lambda \sum_{i=1}^{K} \omega_i (\frac{1}{2}|\beta_i| + \frac{1}{2}\beta_i^2], \tag{10}$$

where $\omega_i = 1/|\widetilde{\beta}_{i,OLS}|$, and λ is selected equation by equation by 10-fold cross-validation. Note that the adaptive elastic net penalty averages the “LASSO penalty” with a “ridge penalty,” and, moreover, that it weights the average by inverse ordinary least squares (OLS) parameter estimates, thereby shrinking the “smallest” OLS-estimated coefficients most heavily toward zero.

2.3 Network Model

The issue of how to display results takes on great importance in high-dimensional network analysis. In our subsequent work, I will construct a network connectedness model with 97 nodes, but presenting and examining $97 \times 97 = 9409$ estimated pairwise variance decompositions would be thoroughly uninformative. Therefore, I characterize the estimated network graphically using five instruments: side color, node size, node color, node location and link arrow sizes. Throughout, I use the open-source Gephi software (https://gephi.github.io/) for network visualization.

Next, I want to explain the detail information of each graphical representation of the network model.

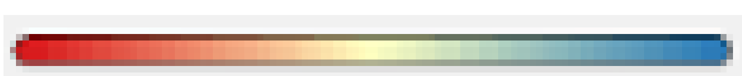

Figure 1: Color in the Network Plot

Side color indicates the relationship between two industries In order to show the degree of correlation between industries, I convert the impact of each industry in the variance decomposition results into weights, which makes the comparison criteria more consistent. The color shown in Figure 1 shows that the weights are getting larger and larger from left to right, that is, as the weight of the share increases, the color of the edge changes from red to blue.

Node size indicates the impact to others First calculate the total risk transmission effect of each industry to all other industries, and then assign weights to the nodes. Generally speaking, the larger the node, the greater the impact on the outside world, and the smaller the node, the smaller the impact on the outside world.

Node location indicates average pairwise directional connectedness I use the ForceAtlas2 algorithm in Gephi to determine the node position. This algorithm can help to find a stable state and accurately balance the repulsive force and attractive force. This stable state makes the two nodes repel each other like the same magnetic pole, and the link between the nodes attracts each other like an elongated spring. It is worth noting that the attractive force between the links is proportional to the pair-wise average conduction effect. In addition, the steady-state node position depends on the initial node position, so the final result is not unique. Based on the research purpose of this article, I am interested in the relative rather than absolute equilibrium node position between the nodes, so the difference of the initial nodes can be ignored.

Node color indicates the volatility of each industry To calculate the logarithmic daily volatility of a single industry to measure the risk of a single industry. The color shown in Figure 1 represents the increase in weight from left to right. Here, the industry risk is used as the weight to the color of the node. That is, the color of the node changes from red to blue, indicating that the risk is increasing. Based on this, it is easier to discover the risk industries of the network transmission graph.

Link arrow sizes indicates pairwise directional connectedness 'to' and 'from' Since all arrow sizes show all pairwise directed risk conduction effects, from which all other graphs can be derived, most additional graphs (especially node shadows and positions) are redundant in principle. But these additional graphics are very useful for describing large network topologies.

3 Result

3.1 Static Analysis of Risk Connectedness

I estimate logarithmic volatility VARs using the adaptive elastic net as described above. Then I compute variance decompositions and corresponding connectedness measures at horizon $H = 10$, using the estimated VAR parameters.

3.1.1 Clustering analysis of industries

The clustering feature can help us explore the potential transmission chains between industries. Generally, the clustering feature will become more obvious as the relationship among industries becomes closer, thus leading to stronger risk transmission. The result shows that the clustering characteristics of many industries in similar business areas and close cooperations are obvious. Specifically, this conclusion can be divided into the following seven categories and I will focus on the first two categories.

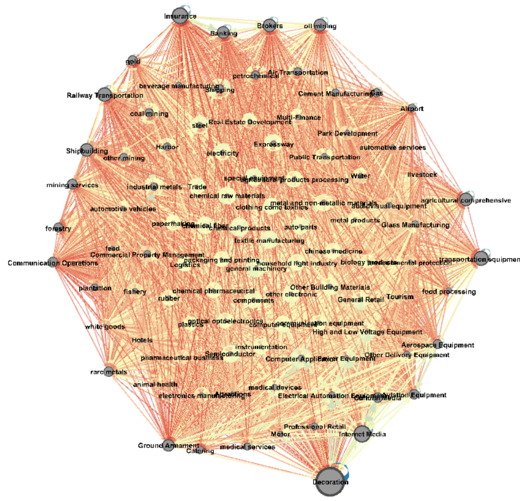

Figure 2: Risk Connectedness (2010-2020)

The first one is finance and real estate. In *fig 2*, the clustering features between real estate development, diversified finance, insurance, banking, and broker are significant. Real estate is a capital-intensive industry. Its prosperity and development are inseparable from the financial support of the financial market. Moreover, the prosperity of the real estate market is an important guarantee for investors to establish confidence, and therefore feeds back the development of financial markets. The close investment relationship between finance and real estate determines the fact that they prosper mutually, which indicates that if one of them become risky, the other one is less likely to survive from the risk.

The second one is transportation, including transportation equipment, air transportation, railway transportation, ports, airports, and shipbuilding industries. The various components in transportation are mutually constrained and interdependent. Thus, the industry is closely connected and the transmission characteristics are remarkable.

The third category includes industrial materials, fuels and heavy industry, such as oil exploration, petrochemicals, chemical materials, gas, steel, metal and non-metal materials, industrial metals, gold, rare metals and other industries; the fourth category contains entertainment and consumption, including hotels, tourism, attractions, restaurants, etc.; the fifth category includes agriculture, plantation, livestock and poultry breeding, agricultural integration, fisheries, agricultural products processing.The sixth category contains electronic devices. For example, computer applications, computer equipment, electronics manufacturing, electronic appliances, etc.; the seventh category includes medical care, medical services, medical devices, Chinese medicine, pharmaceutical industry, chemical pharmaceuticals, animal health, etc. There are clustering phenomena between industries in each category.

3.1.2 Intermediary Analysis of Risk Connectedness

The four industries with the greatest risk connectedness effects are oil exploration, banking, insurance, and railway transportation. It should be noticed that the indirect risk transmission effects in the risk connectedness process are very common. Taking the oil exploitation industry as an example, the oil transmission industry has a significant risk transmission effect on the petrochemical, industrial metal and insurance industries, but the direct transmission effects on other industries are not very significant. For one thing, when risks occur within the oil extraction industry, it directly transmits risks; For another, it transmits risks to the entire market through the interlinkages of petrochemical, industrial metals and insurance. It is clear that indirect volatility connectedness is more influential than direct risk conduction.

The risk of a single industry can be measured through the return volatility of the industry index. Thus, risk characteristics of a single industry are conducive to finding the source of risk. The four industries with the highest volatility are: transportation equipment, ground equipment, aviation equipment and mining services. Conversely, the average volatility of the four industries with the highest transmission effects (insurance, banking, oil exploration, and rail transportation) are particularly small, indicating that the four industries themselves have little risk. Therefore, the risks transmitted by these industries are more likely from other industries. For example, in the transportation equipment, although the transportation equipment industry itself has no significant effect on the total volatility connectedness to other industries, its own volatility is large, and it will indirectly export risks to the entire market through the important industry of railway transportation.

Some important industries can act as risk-transfer mediators, channeling the risks of the risk-source industry to the entire market. For governments, on the one hand, I must focus on the above-mentioned important industries, strengthen the internal control mechanism and external supervision of the industry, and mitigate risks in a timely manner. On the other hand, when there are internal risks in these industries, risk prevention and control measures should be taken in time for vulnerable industries, then the indirect volatility connectedness process is blocked.

3.2 Dynamic analysis of risk connectedness

In this section, firstly, I will use rolling window algorithm to process the data and generate the overall risk connectedness in the market. Then, I can estimate, respectively, the connectedness network before and after Covid-19. By comparing two network models, I can roughly learn the influence of the Covid-19. Finally, I will estimate the eigenvector centrality degree and find the industry at the central point.

3.2.1 Overall Risk Connectedness in China from 2015 to 2020

I estimate the overall risk connectedness through *equation 6*, which is also called the systematic risk. To calculate the systematic risk, I choose 60 days as the rolling sample period and get the total volatility connectedness, which continuously changes with time. Then, I show the time series plot and decompose it to get the tendency.

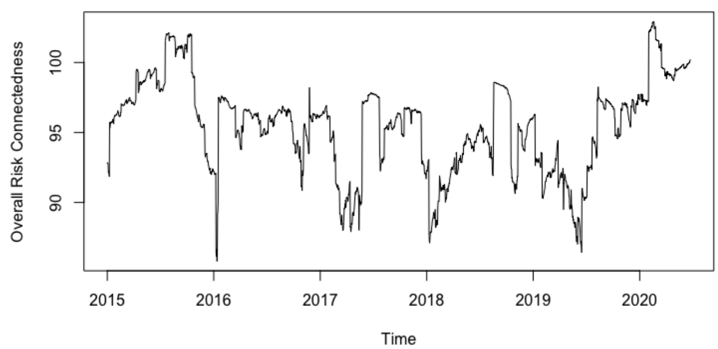

Figure 3: Overall Risk Connectedness from 2015 to 2020

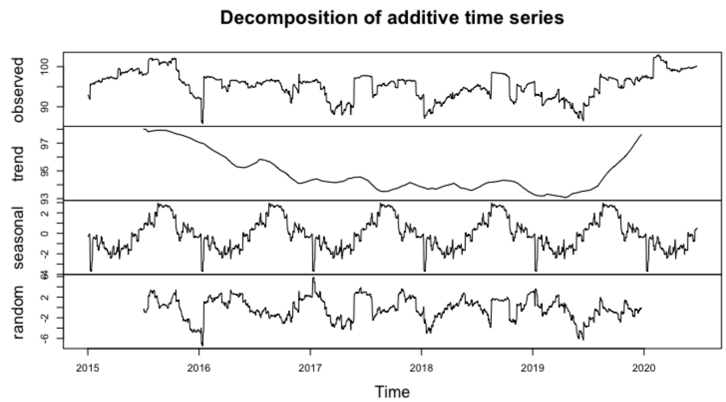

Figure 4: Decomposition of Overall Risk Connectedness from 2015 to 2020

In *fig 3* and *fig 4*, two peaks are obvious, in August, 2015 and February, 2020, which correspond to the stock disaster in China and the outbreak of Covid-19. I can conclude that the emergence of crisis boosts the risk connectedness effects. Also, I can compare the connectedness network before and after the crisis to make our conclusion clear.

3.2.2 Comparison before and after the Covid-19

From the beginning of January 2020 to March 2020, Covid-19 shocked the whole of China. Start from Wuhan, Hubei, the disease killed around two thousand people in China and greatly depress every aspect of the country. Here, I compare the industry risk connectedness networks before and after the disease in order to analyze the influence of risk transmission mechanism. From the comparison of the graphs before and after the stock crash, I can find that both financial institutions and transportation departments play an important role before and after the disease. Also, the epidemic has boosted the emergence of risk in other areas such as forestry, agriculture, and the internet.

For one thing, the volatility connectedness effects become stronger among most industries after the

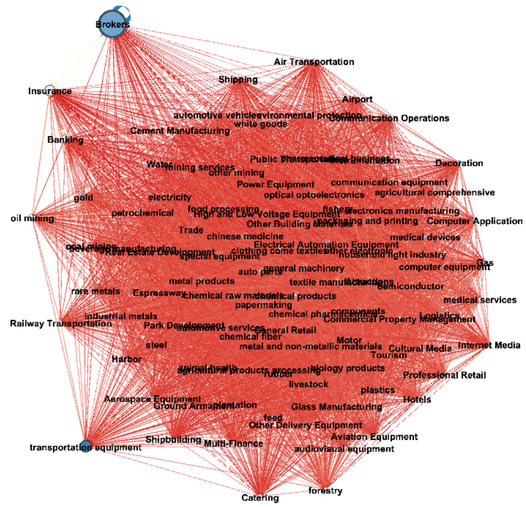

Figure 5: Risk Connectedness before Covid-19

Figure 6: Risk Connectedness after Covid-19

Covid-19 since the overall color has changed, compared with the graph before the Covid-19. The Internet industry, for example, has much stronger transmission effects to others after the Covid-19. Plus, industry clusters have enhanced internal risk transmission effects. For example, ground military equipment, aerospace equipment, and aviation equipment present obvious clustering characteristics. Additionally, broker, oil exploration, insurance, and banking are closely related through remarkable connectedness relationship. That is to say, the volatility connectedness between individual industry and industry clusters is significantly enhanced, and the systemic risks of the entire industry are greatly invigorated after Covid-19 in 2020.

References

[1] Viral V Acharya, Lasse H Pedersen, Thomas Philippon, and Matthew Richardson. Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47, 2017.

[2] Tobias Adrian and Markus K Brunnermeier. Covar. Technical report, National Bureau of Economic Research, 2011.

[3] Mert Demirer, Francis X Diebold, Laura Liu, and Kamil Yilmaz. Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1):1–15, 2018.

[4] Francis X Diebold and Kamil Yilmaz. Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66, 2012.

[5] Mark B Garman and Michael J Klass. On the estimation of security price volatilities from historical data. *Journal of business*, pages 67–78, 1980.

[6] Stefan Klößner and Sven Wagner. Exploring all var orderings for calculating spillovers? yes, we can!—a note on diebold and yilmaz (2009). *Journal of Applied Econometrics*, 29(1):172–179, 2014.

[7] H Hashem Pesaran and Yongcheol Shin. Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1):17–29, 1998.

[8] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.