SFB 649 Discussion Paper 2017-012

Industry Interdependency Dynamics in a Network Context

Ya Qian*
Wolfgang Karl Härdle*
Cathy Yi-Hsuan Chen*

*Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Industry Interdependency Dynamics in a Network Context *

Ya Qian[†], Wolfgang Karl Härdle[‡], Cathy Yi-Hsuan Chen [§]

April, 2017

Abstract

This paper contributes to model the industry interconnecting structure in a network context. General predictive model (Rapach et al. 2016) is extended to quantile LASSO regression so as to incorporate tail risks in the construction of industry interdependency networks. Empirical results show a denser network with heterogeneous central industries in tail cases. Network dynamics demonstrate the variety of interdependency across time. Lower tail interdependency structure gives the most accurate out-of-sample forecast of portfolio returns and network centrality-based trading strategies seem to outperform market portfolios, leading to the possible 'too central to fail' argument.

Keywords: dynamic network, interdependency, general predictive model, quantile LASSO, connectedness, centrality, prediction accuracy, network-based trading strategy

JEL classification: C32, C55, C58, G11, G17

*The financial support from the Deutsche Forschungsgemeinschaft, Humboldt-Universität zu Berlin and IRTG 1972 'High Dimensional Non Stationary Time Series' is gratefully acknowledged.

[†]Corresponding author. Humboldt-Universität zu Berlin, C.A.S.E.-Center of Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. Email: qianya@hu-berlin.de

[‡]Humboldt-Universität zu Berlin, C.A.S.E.-Center of Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. Research fellow in Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, 6th Level, School of Economics, Singapore 178903. Research fellow in Wang Yanan Institute for Studies in Economics(WISE), A307,Economics Building, Xiamen University, Xiamen, 361005 China. Research fellow in National Chiao Tung University, No. 1001, Daxue Rd, East District, Hsinchu City, Taiwan 300.

[§]Humboldt-Universität zu Berlin, C.A.S.E.-Center of Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany.

1 Introduction

Interdependency among different assets is always the key topic of portfolio management. From the very beginning of portfolio theory (Markowitz 1952) , correlation between every two assets is considered to be one of the most important factors in portfolio construction. Ever since the recent financial crisis, studies in interdependencies in the context of risk management have increased rapidly with most of them showing a great interest in the dependency structure within financial sector, i.e. financial contagion (Rodriguez 2007, May & Arinaminpathy 2010, Hasman 2013, Georg 2013, Acemoglu et al. 2015). However, as broad asset allocation including industry assets becomes more and more popular, interdependency among industries started to attract more attention as well. Some research takes the perspective of the interdependency among financial sector and other real economy sectors (Baur 2012, Chiu et al. 2015, Claessens et al. 2012). With no exception, work in this direction concentrates on the effect of financial sector on other real economy sectors, not the way around. Whilst if we consider an easy example containing three corporations, being them oil (X), car manufactory (Y) and autos dealer (Z) from the perspective of supply chain, we have to say that interconnectivity across different industries (not limited to financial to others) is pretty common, as the work done by Rapach et al. (2016). They use the one-period predictive model to establish return predictability among different industries as a depiction of industry interdependency among various sectors and claim the interdependency is pretty widespread among each other. Nevetheless, the industry interdependency in an extreme or stress situation hasn't been addressed intensively. One may imagine that the interdependency may not necessarily show a monotonic linearity w.r.t the quantile level being considered. We therefore contribute to the extant literature by aiming this extreme interdependency which can be referable in the industry portfolio in a market downturn. Tail event based quantile regression with LASSO regularization is implemented here, which is cast into a dynamic network context.

We study the industry interdependency from the network point of view for mainly three important reasons: First, it has been proved in literature as an excellent tool to depict interconnectivities. Real network analysis includes the work of Schweitzer et al. (2009), by taking a socioeconomic perspective, they argue a network architecture built upon trade, $R\&D$ alliances, ownership or credit-debt relationships can vividly study the strategic behavior of the interacting agents. Gençay et al. (2015) use North American supplier-customer network data of public companies to assess counterparty risk and detect counterparty network effects as significant determinants of credit spreads. In the empirical part of Zhu, Pan, Li, Liu & Wang (2016), they test the 'Chinese Twitter' - Xinlang Weibo social network and observe a significant network effect in Chinese social activities. Zhu, Wang, Wang & Härdle (2016) extend Zhu, Pan, Li, Liu & Wang (2016) into the quantile regression framework to consider tail risks. They then exert the quantile network autoregressive model to describe Chinese stocks' interconnecting behivors on the basis of common shared ownership information. For artificial networks, statistical methods need to be used to construct linkages. Based on vector autoregressive (VAR) model, Diebold & Yilmaz (2014) propose a generalized variance decomposition to define a weighted directed network. They apply their method to US financial institutions and it turns out to be coordinated pretty well with the 2008 financial crisis. Similarly, Billio et al. (2011) use linear as well as nonlinear Granger-Causality tests to construct pairwise connections in the network and apply it to monthly returns of different sectors of finance department stocks. Their empirical results show the advantage of network models in measuring the systematic risk levels. Chan-Lau et al. (2016) adopt a default correlation model to construct the forward-looking partial default correlations, which turn out to be the network element. Depending on their network construction, they study the systemic risk of over 1000 exchange-traded banks in the global network framework and argued that connectivity hasn't been paid enough emphasis in Financial Stability Board. Härdle et al. (2016) propose a nonlinear semiparametric quantile regression method on CoVaR to construct a tail-event driven network in order to study the systemic risk among different financial sectors and conclude that the interconnectedness is growing during financial crisis period with largest systemic risk receivers and emitters being the most systemically important.

The second reason for adopting network methodology due to industry characteristics themselves. As it is not just limited to the dyadic relationship, industry interconnectivity has more complicated dependency structure. It focuses on concentration. In a bunch of various companies from different industries, some companies are more important in the sense that they are connected to more of the others with others just being less important locating in the periphery. An obvious example concerning concentration is the banking industry acts actively in the group of companies which need funding. None of these funding-needed companies can thrive without the financing from the banking sector. In addition, Baxamusa et al. (2015) provided the empirical evidence that in the customer-suplier network, the more central the firm, the lower its returns from the acquasition activities. And this concentration can be easily established in the network framework.

Thirdly, it is not hard to properly specify the node set and edge set in our research question, which are the basic elements of a network structure. Fama-French industry portfolios are the nodes, while for edges establishment, we have to use appropriate statistics methods. Former popular calibration methods on interdependency analysis include correlation analysis (see Chiang et al. (2016), for instance), vector autoregressive models (e.g. Diebold & Yilmaz (2014)) and copula based methedology (Poshakwale & Mandal (2016)). However, to include the information of return predictability and further use the network to do forecasts, we construct the edges as the one-month ahead return predictive model parameters. As Rapach et al. (2016) argue, there is a significant relationship between their general predictive model and the US production network. We therefore claim the setup of using predictive model parameters as edges is reasonable.

The motivation to incorporate tail risks are the consideration of the parallelity between industry portfolios and financial stocks. Lots of research has investigated the importance of downside risk in financial stocks and comparatively, beyond just financial sector, we conjecture that industry interdependencies are also affected by extreme situations. Therefore, we refer to Tibshirani (1996) and Li & Zhu (2008) to introduce quantile LASSO techniques upon the industry return general predictive model of Rapach et al. (2016). We have the penalizing techniques come into play here to solve high dimensionality problem. For different quantiles, we can model median level as well as tail level interconnectivities to fulfill the purpose of comparison. In a nutshell: Tail event based quantile regression with LASSO regularization is cast into a dynamic network context. Our main contribution is to extend the general predictive regression framework into the tail case using quantile LASSO regularization to construct networks under different quantile levels and compare median and tail-centered return data of industry portfolios in order to show the increase of the connection during extreme periods. Based on the differences between normal and extreme markets, we would also like to compare the prediction accuracy of the one-month forward return. For utilizing the network information into financial markets, we will construct network-based trading strategies of industry portfolios as well in order to see whether the markets can be beaten. Lastly, by studying the dynamic structure of industry portfolio network across time, we would like to discern some evolution pattern of this industry network.

The remaining of our work is organized as follows: Section 2 describes the econometric model that we are using to construct the industry network; Some basic concepts of network structure and its key parameters are given in Section 3; Section 4 shows the empirical network analysis of 49 industry portfolios obtained from Kenneth French's data library. Analysis in this part includes the construction of whole network as well as dynamic networks, predicition accuracy computation and performance of network-based trading strategy; Section 5 concludes and summarizes. Tables and figures are organized in Appendices at the end.

2 Econometric Modelling of Industry Interdependency

Rapach et al. (2016) propose a general predictive model of industry returns to study the interdependency among 30 industries. They compare their model to American production network, concluding that their model could represent a good construction of industry network. However, only the sector return is considered in their work, while we all agree that in stress situations, tail events carry information on the network infrastructrue. This motivates to extend the general predictive model to different tails, e.g. quantile levels, to investigate the role tail risks playing in industry portfolios.

The general predictive model proposed by Rapach et al. (2016) is given as follows:

$$r_{i,t+1} = \beta_{0,i} + \sum_{j=1}^{N} \beta_{i,j} r_{j,t} + \varepsilon_{i,t+1}, t = 1, \ldots, T-1 \tag{1}$$

where $r_{i,t}$ is the monthly return of industry portfolio i at time t; N is the total number of industries and $\varepsilon_{i,t}$ is the white noise error term.

The generalized quantile regression is described as: $\{X, Y\} = \{x_i, y_i\}_{i=1}^{n}$, $x_i = (x_{i1}, \ldots, x_{ip})^\top \in \mathbb{R}^p$, $\tau \in (0,1)$.

$$Y = X\beta + \varepsilon \tag{2}$$

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau (y_i - x_i^\top \beta) \tag{3}$$

where $\rho_\tau(\cdot)$ is an asymmetric loss function:

$$\rho_\tau(u) = |u|^\alpha |\, \mathbf{I}(u \leq 0) - \tau|, \alpha \geq 1 \tag{4}$$

with $\alpha = 1$ and $\alpha = 2$ corresponding to a quantile and expectile regression respectively, see Breckling & Chambers (1988). The aforementioned general predictive model (1) is a special case of the generalized quantile regression (2), if we set $y_i = r_{i,t+1}, x_i = (1, r_t), \alpha = 1, \tau = 0.5$, where $r_t = \{r_{j,t}\}_{j=1}^{N}$.

For large dimension p one runs into singularity problems and a plethora of too many small coefficients. Prediction accuracy and model interpretability become so big problems with large p that standard ordinary least squares (OLS) turns out to be valid no longer. Standard techniques for improving the OLS estimates contain subset selection and ridge regression while neither of which solves the two problem simultaneously. The way to go here is the implementation of Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996):

$$\hat{\beta}(LASSO) = \arg\ \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_i| \tag{5}$$

where λ is a nonnegative regularization parameter, p is the total number of possible covariates that explains Y and the second term $\sum_{i=1}^{p} |\beta_i| = \|\beta\|_1$ is the l_1 norm.

Combining the idea of tail event QR with LASSO leads us to

$$\hat{\beta}(qLASSO) = \arg\ \min_{\beta_0,\beta} \sum_{t=1}^{T-1} \rho_\tau(r_{i,t+1} - \beta_0 - r_t^\top \beta) + \lambda\|\beta\|_1 \tag{6}$$

where r_t denotes the return vector of all industries at time t; β the vector of coefficients of the regression and β_0 the intercept. The l_1-norm quantile LASSO model can be refered to Li & Zhu (2008).

As is known from Härdle & Simar (2015), the solution to (6) yields a finite subset of nonzero elements of the $\hat{\beta}(qLASSO)$ vector. The coefficients in this 'active set' may be called 'prominent' since all other coefficients are actually zero.

Later in Section 4 we will use quantile LASSO regression method to build the network across different industry portfolios. Before going to that, we would like to give a brief introduction of network structure, the main graphic tool in our analysis.

3 Network Structure

A binary set $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the network structure of a system where $\mathcal{V}$ denotes the collection of vertices (also called nodes) in the system and $\mathcal{E}$ stands for the collection of edges (or linkages, etc) between each pair of the vertices. In our application, vertices $\mathcal{V}$ are the industries and the edges $\mathcal{E}$ are constructed as the prominent quantile LASSO coefficients. A network is always corresponding to an adjacency matrix which specifies the edges between each pair of the nodes. At the very beginning, adjacency matrix is simply a symmetric binary matrix but it is later extended to weighted and asymmetrix ones, which is exactly the case in our later empirical analysis.

Given a network $\mathcal{G}$, two important and interesting questions are always asked: first, how to measure the graph level connectivity of the network? Second, which parameter gives us insight into the relative importance of each vertice?

3.1 Revisit *Fagiolo (2007)*

The answer to the first question is the concept of 'connectedness'. Connectedness is a measure specified in network analysis depicting the degree of interdependency among all nodes and the whole network connectedness is achieved by averaging all the node-specific connectedness. In the context of graph theory, connectedness is usually refered to as clustering coefficient, which measures the inherent tendency of nodes clustering together. The global verison of clustering coefficient gives the measure of the connectedness of the whole network. The most common definition is designed for undirected and binary adjacency matrices, following Fagiolo (2007), we use four patterns (cycle, middleman, in and out) to depict directed networks.

- *cycle*: there is a cyclical relation among i and any two of its neighbors ($i \to j \to h \to i$ or viceversa);
- *middleman*: when one of i's neighbors reach a third neighbor directly with an outward edge or indirectly with i as a medium;
- *in*: i has two inward edges;
- *out*: i has two outward edges.

For an asymmetrix binary adjacency matrix, *in-degree*, *out-degree total-degree* and *bilateral-degree* of node i are defined as:

$$
\begin{aligned}
d_i^{in} &= \sum_{j \neq i} a_{ji} = (A^\top)_i \mathbf{1} \\
d_i^{out} &= \sum_{j \neq i} a_{ij} = A_i \mathbf{1} \\
d_i^{tot} &= d_i^{in} + d_i^{out} = (A^\top + A)_i \mathbf{1} \\
d_i^{\leftrightarrow} &= \sum_{j \neq i} a_{ij} a_{ji} = A_{ii}^2
\end{aligned}
$$

where $A^\top$ is the transpose of A, A_i the i_{th} row of A, A_{ii} the i_{th} diagonal element of A, $\mathbf{1}$ is the N-dimensional column vector $(1, 1, \cdots, 1)^\top$.

Based on above notations, the number of all possible triangles that node i could form (T_i^D):

$$
T_i^D = d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}
$$

For weighted adjacency matrix W, the four patterns of clustering coefficient are defined as:

$$
\begin{aligned}
C_i^{cyc} &= \frac{(W^{[1/3]^3})_{ii}}{d_i^{in} d_i^{out} - d_i^{\leftrightarrow}} \\
C_i^{mid} &= \frac{(W^{[1/3]} W^{[1/3]^\top} W^{[1/3]})_{ii}}{d_i^{in} d_i^{out} - d_i^{\leftrightarrow}} \\
C_i^{in} &= \frac{(W^{[1/3]^\top} W^{[1/3]^2})_{ii}}{d_i^{in}(d_i^{in} - 1)} \\
C_i^{out} &= \frac{(W^{[1/3]^2} W^{[1/3]^\top})_{ii}}{d_i^{out}(d_i^{out} - 1)}
\end{aligned}
$$

where $W^{[1/3]}$ denotes the matrix with each element generated as the cubic roots of W's elements. To get overall connectedness, we just average these C_i^* via:

$$C^* = N^{-1} \sum_{i=1}^{N} C_i^*$$

where * stands for elements in *cyc*, *mid*, *in*, *out*.

3.2 Centrality Measures

For the second, the answer is centrality. Centrality basically answers the question 'what characterizes the important vertices?' There are various kinds of centrality definitions. In the simplest cases, *Degree centrality* measures how many ties each node has and assigns the biggest value of importance to the node which has the largest number of ties. A more complex extension of degree centrality is to consider the directions of linkages in directed networks. Therefore we have 'in' as well as 'out' degree centralities. Though simple and easily to exert, degree centrality assigns equal values to all the edges. Baveias (1950) defined the *closeness centrality* of a node as the average length of the shortest path between the node and all other nodes. Freeman (1977) introduced the *betweenness centrality* that measures the number of times that the node plays as a bridge along the shortest path between any other two nodes. A more appropriate version of these two centrality measures is to incorporate the concept of 'cost' in which case we define the shortest path in the sense of actual lenghth instead of number of nodes. Since the version with 'cost' pays attention to the actual distance between each pair of node, it is more appliable to real world networks, for instance, transportation networks. As for differentiating the relative importance of different nodes, these two measures contributes little. A good measure to incorporate relative importance of different nodes is *Eigenvector centrality*. For a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, *eigenvector centrality* of node v - $C_E(v)$ equals

$$C_E(v) = \frac{1}{\lambda} \sum_{t \in M(v)} C_E(t) = \frac{1}{\lambda} \sum_{t \in \mathcal{G}} a_{v,t} C_E(t) \tag{7}$$

where λ is the maximum eigenvalue of the adjacency matrix A; $M(v)$ the set of neighbors of v and $a_{v,t}$ the element of A in row v and column t. According to this measure, a node in a network is important if it is linked to other important nodes. Hence one does not consider the edges between every pair of the nodes as equally important, one assigns different importance value through the first eigenvector. Similar to degree centrality, when in directed graphs, we can have 'in' and 'out' eigenvector centrality measures to discriminate the 'receiving' and 'emitting' effects respectively. Since industries need to be treated differently and receving and emitting effects have to be set apart, we are going to adopt eigenvector centrality for the weighted, directed graphs in the empirical part.

4 Empirical Results

4.1 Data

Monthly return of the 49 industry portfolios constructed by Kenneth R. French is used as our data sample. The data is available from Kenneth French's webpage [1]. As monthly data is mostly often used in industry portfolio analysis, we select it from January 1970 to January 2017 with 565 observations in total.

As quantile-quantile (QQ) plots (Figure 1 to Figure 5) of the 49 industry portfolios show, compared to normal distributions, tail behaviors exist in most industries, which justifies our analysis of focusing on industry network structure at different tail levels.

4.2 Whole sample network

We now come to the network construction based on Equation (6). The edge between node i and node j exists if and only if the lagged return of industry j (i) is selected by LASSO as the significant predictor of the return of industry i (j). The edges are constructed as directional: if i (j) helps predicting j (i) , then the edge goes from i to j (j to i); if i

[1] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

helps predicting j and j helps predicting i as well, then the edge between i and j has both arrows. Furthermore, the edges in our industry network are also weighted. And the weight assigned to each edge depends on the absolute value of the beta coefficients. To sum up, the adjacency matrix corresponding to our industry interdependency network is the absolute value of beta coefficients.

To discern the probable varieties between median network and tail-event driven network, we set $\tau = (0.05, 0.5, 0.95)$, denoting the crisis, stable and boom situations respectively. Figure 6 depicts the whole sample (1970.01-2017.01) industry networks under these three situations. In every subfigure, we locate the 49 industries in a circle and fix their positions in favor of convenient comparison among different τ levels. The nodes in the network have different sizes according to their 'in' and 'out' eigenvector centrality scores respectively, as introduced aforementioned. We arrange the network plots with eigenvector centrality from the 'in' direction in the left panel and 'out' the right panel of each subfigure in Figure 6. Specifically, the leading industries possess larger sizes in our industry network. The grey arrows with directions within the network circle suggest the intensity of the interconnectivity of the industry network. Comparing the density of edges in these three figures, we can clearly reach the conclusion that in extreme cases, the industry network connection increases a lot when comparing to stable situation, which means the whole economy becomes more connected in extreme cases, as is listed in Table 2. Meanwhile, the leading industries change as well. To see this perspective more precisely, we list the top leading industries which has eigenvector centrality score larger than 0.200 (both 'in' and 'out') under each situation in Table 3. Comparing the leading industries in different cases, we have several interesting findings:

a. More central industries are identified under extreme cases with more even centrality score distribution. It is a signal of intercorrelation rise among various industries in extremes.

b. Financial-related industries (banking, insurance and trading) play important roles as risk emitters under whatever market situations (banking has rank 6 as risk emitters when $\tau = 0.95$ with centrality score 0.186). It is determined by the nature of financial-related industries.

(a) Banking evolves as a leading risk emitter when extreme events happen. It complies with our common knowledge that when markets go to extremes, banking affects other industries more.

(b) Insurance arises as the NO.1 risk emitter when market goes down. It is a reflection of the market sentiment.

(c) However, when market goes down, financial trading also becomes a crucial risk receiver which to some degree, signifies the hard time that traders have to endure during financial crisis.

c. Coal remains leading risk emitter in various markets. It may be accounted to the relative position of coal industry in the supply chain. As an upperstream industry, coal price can actually affect the return of many others which need it as raw material. Furthermore, no matter what the market looks like, basic production still has to be done.

d. Gold always stays as top risk receiver for which we may argue from the role gold plays as a financial hedging instrument.

e. The leading industries detected has little to do with industry size according to our empirical analysis.

4.3 Dynamic networks

In the last section we see the differences in whole sample networks under different τ_s. However the information is limited since it is a static picture. One step further, we can gain more insight by investigating the dynamics of networks. Through this, we expect to discover some potential patterns in industry networks. As introduced in Sec 3.1, we plot the 4 clustering coefficients under three τ_s in a moving window framework. We compute the network structures using quantile LASSO for samples of every three years' data. The data sample starts from 1970.01, then we move the window forward every three months to compute the next one, that is, the first sample is 1970.01-1972.12, the second one 1970.04-1973.03, so on so forth. Finally, we get 177 data points of each clustering coefficients and their plots are shown in Figure 7 (arranged by different τ_s) and Figure8 (arranged by different clustering patterns in a directed network).

The dynamic networks vary across time are shown in Fig 7 and Fig 8. Under each τ levels, the 'cycle' and 'out' clustering are lower than 'middleman' and 'in' clustering in our industry network. Besides, 'cycle' and 'out' also possess smaller volatilities. However, the four have similar pattern under each τ across time. Around 1997-2002, 2007-2010 and 2012-2013 there are sharp ups and downs in all connectedness measures no matter whether we take tail risks into account. These time slots also approximately match the economic crisis in history. When comparing each of the four connectedness measures under various τ_s, we once again validate that median level network is less connected than those in tail cases, from the perspective of network dynamics. When zooming in those special time periods, we have an interesting finding that lowertail connectedness behave oppositely with uppertail and median ones in periods 2000-2002, 2011-2012, 2016 thereafter but move more simultaneously during the recent financial crisis during 2007-2008. This, probably, can be explained by the relative important positions financial-related industries possess in the industry network. When the financial-related industries triggered the crisis in 2007 and 2008, lots of industries are affected which leads to the comovement of the connectedness under different τ_s. This argument also complies with centrality analysis in section 4.2.

4.4 Three additions based on specific network construction

4.4.1 Prediction Directions

As far as we discussed, we use the abslute value of the beta coefficients as the adjacency matrix inasmuch as to comply with the cannonical definition of adjacency matrix. However, this setting has the drawback of ignoring the signs of coefficients, which is very useful in return prediction. As is contained in the quantile LASSO regression (6), not only do we know the magnitude of industry return prediction network (the absolute values of the betas), but also the information on directions of industry return predictability (the signs of the beta coefficients). Therefore we add the information of whole sample interdependency in the image plots in Fig 9 and in Table 4 and of rolling window predictability in Table 5, which, in our opinion, are valuable additions to the cananical network analysis. In order for later return prediction, the rolling window in constructed as follows: we first use the data from 1970.01-1995.12 to construct the first 'network' composed of beta coefficients(here with signs). After the first step, we enlarge our estimation window with one more month every time, i.e. in the second step we compute the beta coefficients network using data from 1970.01-1996.01, and so on so forth. Finally we get 253 beta networks of the industry interdependency.

In all three image plots (Figure 9), the horizontal represents the predictive power one industry getting from all others (receiving) and vertical stands for the predictive power that industry to others (emitting). Comparing these three image plots in Table 4, we can see that in common, the LASSO method selects quite a small set of significant predictors out of the 2401 cells in total. Besides, the differences are quite obvious. The difference between normal market and extreme markets is that when market changes from stable to extremes, a larger quantity of significant connections are detected (either in crisis or in booms). In network language, we say the entire industries become more interconnected in stress situations. Furthermore, there are still some differences between these two extremes. First of all, more negative connections are detected in boom than that in crisis while more positive ones in crisis than in boom, indicating a higher and non-diversifed tail risk and a difficulty w.r.t industry diversification in the market crisis. The benefit of industry diversification is diminished in this situation, implying an inevitable tail risk. Second of all, the average connectedness is positive when $\tau = 0.05$ and negative when $\tau = 0.95$. It sends out the signal that bad market conditons tend to affect most industries in the same direction while good ones is more favorable for portfolio management since the average connectedness is negative. The averaged dynamic directions of the beta coefficients under rolling window framework dipict a similar pattern in Table 5 as that in Table 4. Combining them together, we claim it's hard to diversify industry portfolios in crisis, indicating a higher and non-diversified tail risk, which, when translates into risk-return relationship, tells us that we can expect higher returns with the tolerance of higher risks, if we construct appropriate trading strategies. This point is going to be shown in 4.4.3.

4.4.2 Prediction Performance

To make use of the industry network structure that we constructed before and the prediction directions information, a dynamic network structure within a rolling window framework is used to predict the one-month ahead industry returns. We here compare the performance of the interdependencies under different τ levels in the use of predicting future returns. Based on the predicted and actual monthly industry returns, we calculate the root mean squared error (RMSE) of the three models, i.e. with different τ_s, as a measure of the prediction accuarcy. Specifically, we first use the data from 1970.01-1995.12 to construct the first 'network' composed of beta coefficients(here with signs). We multiply this beta matrix with the industry monthly returns in 1995.12 as the prediction of the industry monthly returns in 1996.01. After the first step, we enlarge our estimation window with one more month every time, i.e. in the second step we compute the beta coefficients network using data from 1970.01-1996.01 and multiply the new beta matrix with the returns in 1996.01 as a prediction of those in 1996.02, and so on so forth. Finally we get 253 predictions of the monthly returns for each industry. Lastly, we compute square of the differences between the predicted returns and the real ones and average them over the length of time series, i.e., 253. Figure 10 shows the average RMSEs of the 49 industries in our data sample for the three τ_s.

As expected, beta 'network' dynamics under extreme cases have better prediction performance (smaller out-of-sample forecast error) than that in normality. In general, for these three cases, we claim that lower tail predition achieves the highest accuracy (the smallest out-of-sample forecast error). Therefore, incorporating tail risks contributes to industry return prediction which justifies the necessity of extending extant studies which focus only on median level to different quantiles.

4.4.3 Network centrality-based trading strategies

As we can see from the aforementioned discussion, some vital characteristics about the industry interdependency network have been investigated. However, we still lack the tactic to make use of the network information to profit from financial markets. We are going into this direction here. Network centrality-based trading strategies are considered due to the importance of the concept of centrality as well as our specific construction of industry network. As discussed earlier, centrality measures the relative importance of different industries in the network structure, which, in our framework, is based on one-month-ahead return predictability. Therefore the most central nodes demonstrate the relatively important roles of influencing or being influenced by others more than the rest in the sense of return predictability. That is, industries with higher 'out' ('in') eigenvector centrality scores can also be called 'prediction emittors' ('prediction receivers') which affect more of others (are affected more by others) in the sense of return prediction.

We conjecture that highly centralized industries are more likely to ourperform market portfiolios due to the similar effect as 'too connected to fail', we call it 'too central to fail'. Centralized industries are more connected to other central industries and hence possess more complicated risk structures, which in turn, lead to higher excess returns. However, we would also like to conjecture that the centrality-based trading portfolios have no abnormal return,i.e., no mispricing about the industry portfolios. To verify our assumption, we do the same rolling window computation as described in Subsection 4.4.2 to generate a time series of 253 networks under each τ level and find the first and least leading industries of each network. Then we construct the 4 centrality-based trading strategies and balance them every month with the updated beta network. Finally, we calculate the average of annualized cumulative log-returns of these strategies and of the MKT and report their *t-stat* to decide whether the excess returns are significant. Also, we regress the monthly excess returns of these strategies to Fama-French three risk factors so as to check the existence of abnormal returns. Details are reported in Table 6 and Table 7. The 4 trading strategies we constructed are as follows:

- long the top 10 leading 'in' centralized industries (HI strategy)
- long the top 10 leading 'out' centralized industries (HO strategy)
- long the bottom 10 leading 'in' centralized industries (TI strategy)
- long the bottom 10 leading 'out' centralized industries (TO strategy)

The empirical results mostly authenticate the assumption under various market situations: except for HI in median case, all other more centralized industry portfolios outperforms less centralized ones and outperform market portfolios as well. Furthermore, they have no significant abnormal return when regressing on Fama-French three risk factors. Even in market with huge downside risk, the more centralized portfolios gain a sizeable excess return. Therefore, our empirical analysis stands up for the assumption of 'too central to fail' of the industry networks, which rests on complicated risk structures of more centralized industry portfolios.

5 Conclusion

This study extends the general predictive model of industry portfolios to different quantile levels so as to incorporate tail risks in interdependency measurement and construct network analysis under different market situations. By comparing median level ($\tau = 0.50$) with upper ($\tau = 0.95$) and lower ($\tau = 0.05$) tail networks, we find out that interdependency across USA industries increases a lot during extreme market situations in the whole period from 1970.01 to 2017.01. Similar results are achieved under dynamic network framework - connectedness in stress situations is always higher than that in normality. The time series of four connectedness measures corresponding to weighted directed networks show significant varieties of the interdependency structure dynamics. In addition, leading industries vary as well when market switches from stable to highly volatile. An obvious finding is that financial-related industries evolve as leading ones under stress situations which highlights their role in bad times. At last, three more additions to network analysis are summarized in the last subsection. First, prediction directions differ under different τ_s: lower-tail case involves more positive coefficients while upper-tail circumstance has more negative ones, which reflects the movement in the same direction of different assets in crisis, i.e. asset returns are highly affected in the same direction when market goes down. While in promising situations, they tend to be negatively related therefore it is more effective to do risk management during these periods since we can easily find negative-related assets. Second, when quantifying the one-month-ahead industry return prediction accuracy using dynamic coefficients networks under different quantile levels, as expected, the lower tail case gives the best prediction performance in the sense of RMSE. Third, four trading strategies based on network centrality dynamics are constructed and compared with market portfolio. Empirical results report significant sizeable excess returns for more centralized industry portfolios, which outperform less centralized ones and market portfolio, even in bad market situations. With a risk adjustment for Fama-French factors, these strategies do not possess significant risk-adjusted abnormal returns. Therefore, the higher returns come from the more comlicated risk structures central industries endow. To conclude, we argue for the possibility of the effect 'too central to fail'.

References

Acemoglu, D., Ozdaglar, A. & Tahbaz-Salehi, A. (2015), 'Systemic risk and stability in financial networks', *The American Economic Review* **105**(2), 564–608.

Baur, D. G. (2012), 'Financial contagion and the real economy', *Journal of Banking & Finance* **36**, 2680–2692.

Baveias, A. (1950), 'Communication patterns in task-oriented groups', *Journal of the Acoustical Society of America* **22**, 725–730.

Baxamusa, M., Javaid, S. & Harery, K. (2015), 'Network centrality and mergers', *Review of Quantitative Finance and Accounting* **44**, 393–423.

Billio, M., Getmansky, M., Lo, A. & Pelizzon, L. (2011), 'Econometric measures of connectedness and systemic risk in the finance and insurance sectors', *Journal of Financial Economics* **17**(1), 1–23.

Breckling, J. & Chambers, R. (1988), 'M-quantiles', *Biometrika* **75**, 761–771.

Chan-Lau, J. A., Chuang, C., Duan, J.-c. & Sun, W. (2016), Banking network and systematic risk via forward-looking partial default correlations, Technical report, Working papers//International Monetary Fund and National Singapore University.

Chiang, T. C., Lao, L. & Xue, Q. (2016), 'Comovement between chinese and golobal stock markets: Evidence from aggregate and sectoral data', *Review of Quantitative Finance and Accounting* **47**, 1003–1042.

Chiu, W.-C., Peña, J. I. & Wang, C.-W. (2015), 'Industry characteristics and financial risk contagion', *Journal of Banking & Finance* **75**, 411–427.

Claessens, S., Tong, H. & Wei, S.-J. (2012), 'From the financial crisis to the real economy: Using firm-level data to identify transmission channels', *Journal of International Economics* **88**, 375–387.

Diebold, F. X. & Yilmaz, K. (2014), 'On the network topology of variance decompositions: Measuring the connectedness of financial firms', *Journal of Econometrics* **182**, 119–134.

Fagiolo, G. (2007), 'Clustering in complex directed networks', *Physical Review E* **76**, 026107–1–8.

Freeman, L. C. (1977), 'A set of measures of centrality based on betweenness', *Sociometry* **40**(1), 35–41.

Gençay, R., Signori, D., Xue, Y., Yu, X. & Zhang, K. (2015), 'Economic links and credit spreads', *Journal of Banking & Finance* **55**, 157–169.

Georg, C.-P. (2013), 'The effect of the interbank network structure on contagion and common shocks', *Journal of Banking & Finance* **37**, 2216–2228.

Härdle, W. K. & Simar, L. (2015), *Applied Multivariate Statistical Analysis*, Springer.

Härdle, W. K., Wang, W. & Yu, L. (2016), 'Tenet: Tail-event driven network risk', *Journal of Econometrics* **192**, 499–513.

Hasman, A. (2013), 'A critical review of contagion risk in banking', *Journal of Economic Surveys* **27**, 978–995.

Li, Y. & Zhu, J. (2008), 'l_1-norm quantile regression', *Journal of Computational and Graphical Statistics* **17**(1), 1–23.

Markowitz, H. (1952), 'Portfolio selection', *Journal of Finance* **7**(1), 77–91.

May, R. M. & Arinaminpathy, N. (2010), 'Systemic risk: the dynamics of model banking systems', *Journal of The Royal Society Interface* **7**, 823–838.

Poshakwale, S. S. & Mandal, A. (2016), 'Sources of time varying return comovements during different economic regimes: Evidence from the emerging indian equity market', *Review of Quantitative Finance and Accounting* pp. 1–34.

Rapach, D. E., Strauss, J. K., Tu, J. & Zhou, G. (2016), Industry interdependencies and cross-industry return predictability, Technical report, Working papers//Washington University in St. Louis, Olin School of Business.

Rodriguez, J. C. (2007), 'Measuring financial contagion: A copula approach', *Journal of Empirical Finance* **14**(3), 401–423.

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A. & White, D. R. (2009), 'Economic networks: The new challenges', *Science* **325**, 422–425.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.

Zhu, X., Pan, R., Li, G., Liu, Y. & Wang, H. (2016), 'Network vector autoregression', *Annals of Statistics* .

Zhu, X., Wang, W., Wang, H. & Härdle, W. K. (2016), Network quantile autoregression, SFE 649 Discussion Paper 2016-050, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. available at `http:\sfb649.wiwi.hu-berlin.de/papers/SFE649DP2016-050.pdf`.

6 Appendices

6.1 Tables

Code	Full name	Code	Full name
Agric	Agriculture	Drugs	Pharmaceutical Products
Food	Food Products	Chems	Chemicals
Soda	Candy & Soda	Rubbr	Rubber & Plastic Products
Beer	Beer & Liquor	Txtls	Textiles
Smoke	Tobacco Products	BldMt	Construction Materials
Toys	Recreation	Cnstr	Construction
Fun	Entertainment	Steel	Steel Works Etc
Books	Printing and Publishing	FabPr	Fabricated Products
Hshld	Consumer Goods	Mach	Machinery
Clths	Apparel	ElcEq	Electrical Equipment
Hlth	Healthcare	Autos	Automobiles and Trucks
MedEq	Medical Equipment	Aero	Aircraft
Ships	Shipbuilding, Railroad Equipment	LabEq	Measuring and Control Equipment
Guns	Defense	Paper	Business Supplies
Gold	Precious Metals	Boxes	Shipping Containers
Mines	Non-Metallic & Industrial Metal	Mining Trans	Transportation
Coal	Coal	Whlsl	Wholesale
Oil	Petroleum and Natural Gas	Meals	Restaurants, Hotels, Motels
Util	Utilities	Banks	Banking
Telcm	Communication	Insur	Insurance
PerSv	Personal Services	RlEst	Real Estate
BusSv	Business Services	Fin	Trading
Hardw	Computers	Other	Almost Nothing
Softw	Computer Software	Rtail	Retail
Chips	Electronic Equipment		

Table 1: 49 industry portfolios from French's data library

τ	*cycle*	*middleman*	*in*	*out*
0.05	0.013	0.016	0.016	0.013
0.50	0.003	0.006	0.006	0.003
0.95	0.015	0.019	0.020	0.014

Table 2: Four connectedness measures of whole sample industry network

INET-connectedness

τ	Leading industries	
	receivers	emitters
0.05	Gold(44)(0.293), Softw(3)(0.284), Smoke(26)(0.279), Hardw(10)(0.262), Fin(12)(0.248)	Insur(8)(0.378), Books(32)(0.355), Autos(21)(0.287), Bank(1)(0.252), Toys(42)(0.241), Coal(48)(0.228), Agric(41)(0.209), Aero(24)(0.209), Hlth(34)(0.203)
0.50	Gold(44)(0.612), Smoke(26)(0.238), Meals(25)(0.234), Coal(48)(0.216)	Coal(48)(0.624), Fin(12)(0.342), Ships(46)(0.289), Clths(31)(0.246), Hlth(34)(0.246)
0.95	Softw(3)(0.356), Guns(39)(0.281), Gold(44)(0.249), Steel(33)(0.224), Agric(41)(0.222), Hlth(34)(0.219)	Ships(46)(0.506), Drugs(2)(0.229), Other(11)(0.225), Coal(48)(0.214), Boxes(40)(0.213)

Table 3: Top central industries under various stress situations (the number in the first (second) parentheses is the rank of industry size (eigenvector centrality score))

INET-eigcentr

τ	No. of Coefficients	No. of Nonzeros	No. of Negatives	No. of Positives	Max	Min	Average
0.05	2401	498	158	340	0.517	-0.477	0.007
0.50	2401	218	86	132	0.228	-0.108	0.001
0.95	2401	658	348	310	0.367	-0.729	-0.003

Table 4: Summary predictability magnitude and directions - whole sample

INET-totaldire

τ	No. of Coefficients	No. of Nonzeros	No. of Negatives	No. of Positives	Max	Min	Average
0.05	2401	579	218	361	0.535	-0.618	0.007
0.50	2401	255	99	156	0.199	-0.150	0.001
0.95	2401	718	396	321	0.575	-0.590	-0.003

Table 5: Summary predictability magnitude and directions - rolling window

INET-rollingdire

Strategies	$\tau = 0.05$				$\tau = 0.50$			
	HI	HO	TI	TO	HI	HO	TI	TO
Excess Returns	0.094**	0.088**	0.074*	0.079**	0.059	0.089***	0.084*	0.065*
t-stat	2.157	2.050	1.678	2.068	0.929	2.761	1.855	1.756
Strategies	$\tau = 0.95$				MKT			
	HI	HO	TI	TO				
Excess Returns	0.088*	0.095***	0.065*	0.072	0.073			
t-stat	1.911	2.742	1.853	1.577	1.328			

Table 6: Excess returns of centrality-based trading strategy and market portfolio. *,**, *** denoting the 10%, 5% and 1% significance level respectively.

INET-nettrading

Strategies	intercept	Mkt-Rf	SMB	HML	R^2	Adjusted R^2
HI-lowertail	0.002	0.978***	0.125***	0.147***	0.791	0.788
HO-lowertail	-0.000	1.020***	0.085**	0.467***	0.835	0.833
TI-lowertail	-0.002	1.045***	0.228**	0.564***	0.875	0.874
TO-lowertail	0.000	0.970***	0.095***	0.331***	0.890	0.888
HI-median	-0.002*	1.148***	0.155***	0.153***	0.886	0.885
HO-median	0.000	1.048***	0.173***	0.359***	0.810	0.808
TI-median	0.001	0.983***	0.147***	0.321***	0.797	0.795
TO-median	0.000	0.968***	0.126***	0.279***	0.885	0.883
HI-uppertail	0.001	0.968***	0.251***	0.285***	0.768	0.766
HO-uppertail	0.001	1.001***	0.169***	0.339***	0.853	0.851
TI-uppertail	0.001	0.870***	-0.042	0.334***	0.842	0.840
TO-uppertail	0.000	0.979***	0.070**	0.284***	0.888	0.886

Table 7: Coefficients of excess portfolio returns regressing on Fama-French risk factors. *,**, *** denoting the 10%, 5% and 1% significance level respectively.

INET-regression

6.2 Figures

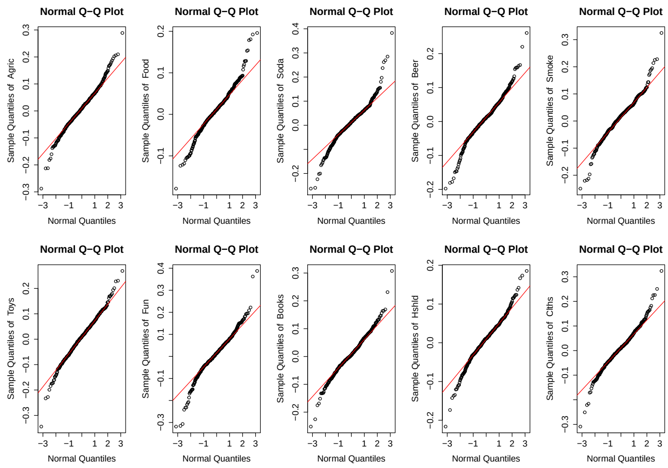

Figure 1: QQ plots of the industries

INET-qqplot

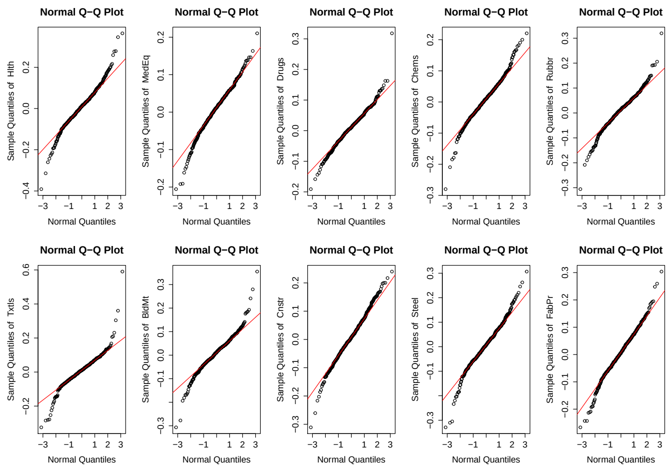

Figure 2: QQ plots of the industries

INET-qqplot

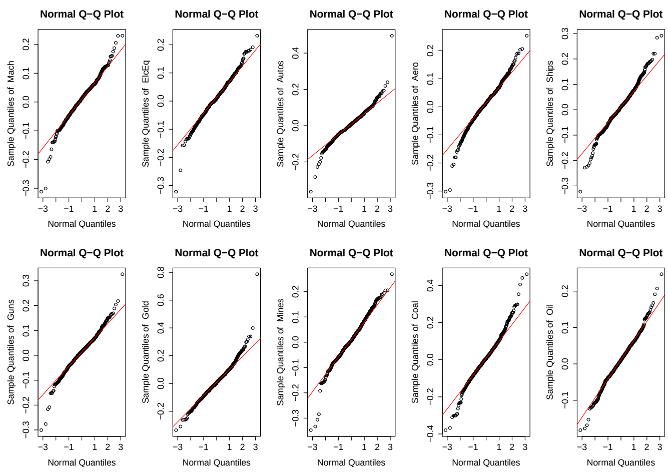

Figure 3: QQ plots of the industries

INET-qqplot

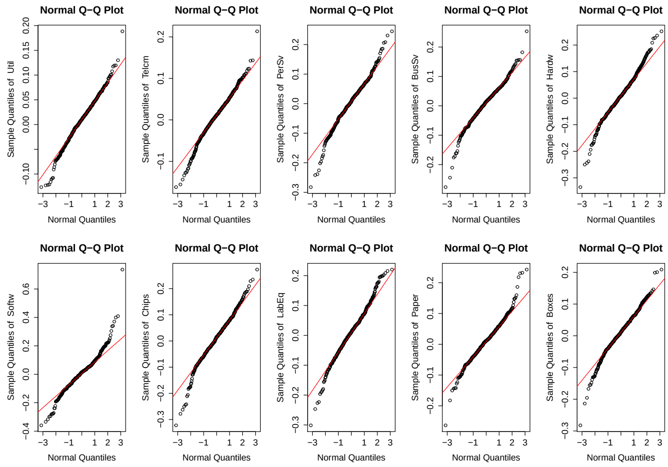

Figure 4: QQ plots of the industries

INET-qqplot

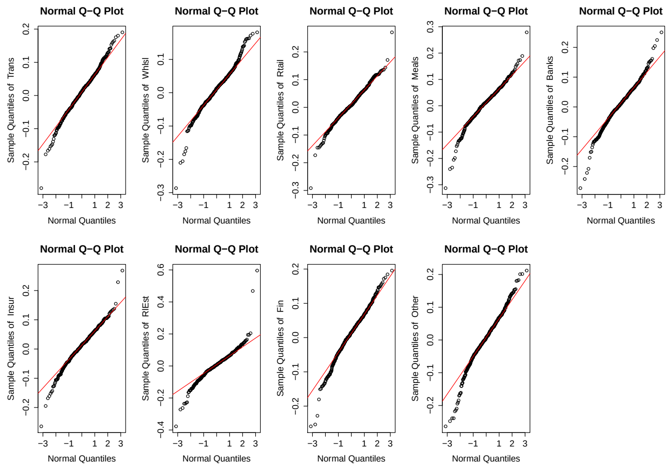

Figure 5: QQ plots of the industries

INET-qqplot

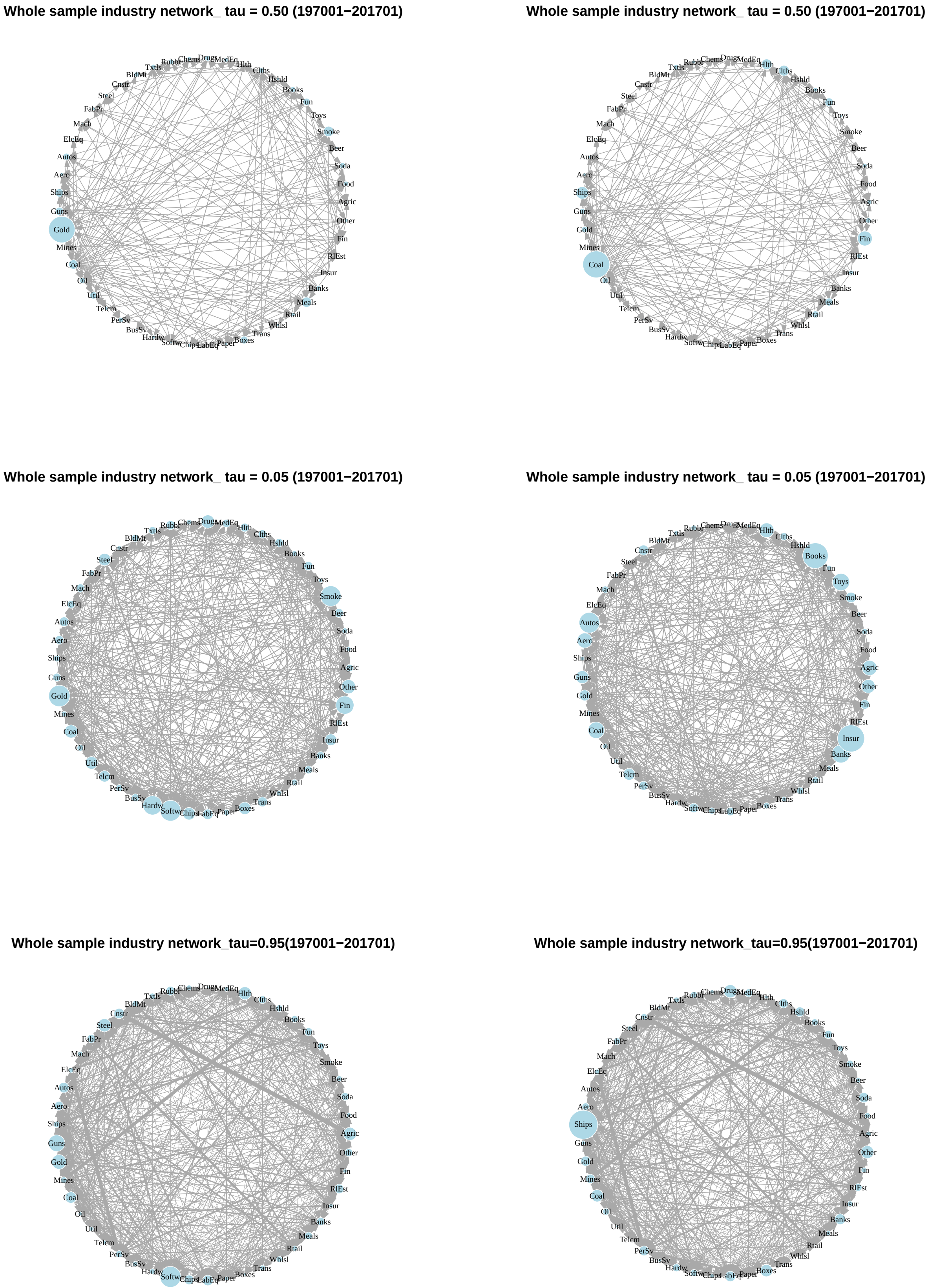

Figure 6: Whole sample network of industry portfolios with larger size denoting the eigenvector centrality('in'-left, 'out'-right) - $\tau = 0.50$ (top), $\tau = 0.05$ (middle), $\tau = 0.95$ (bottom)

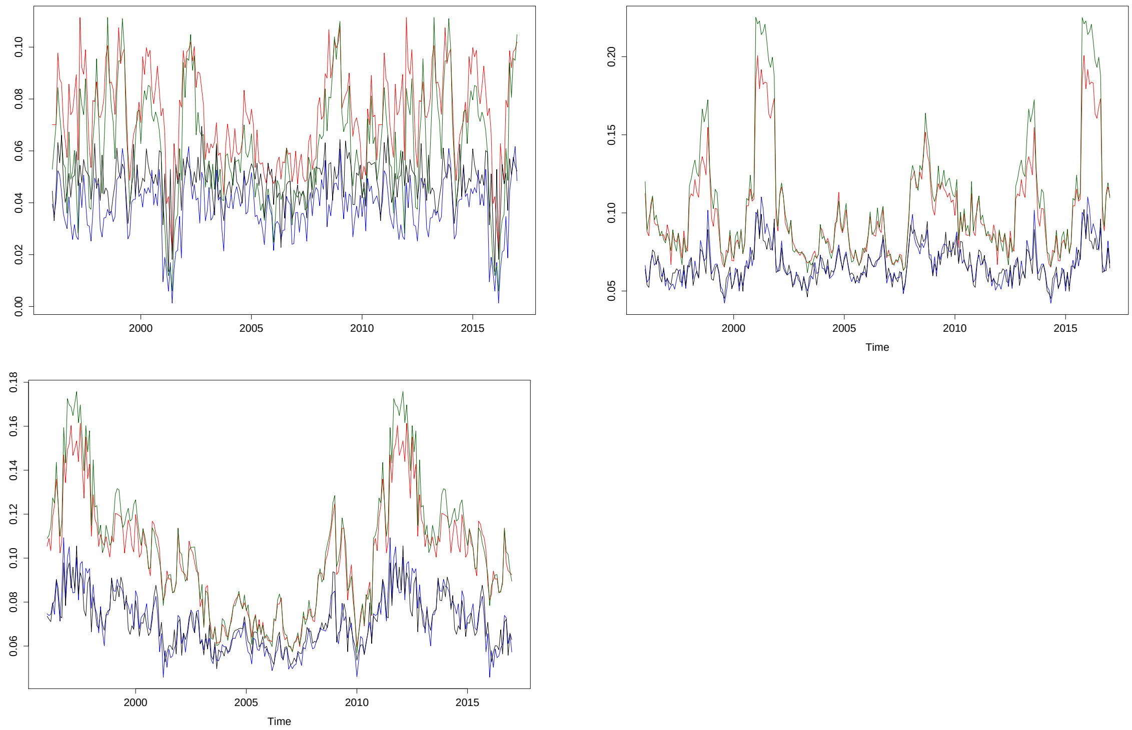

Figure 7: The cycle, middleman, in and out connectedness of industry networks: upperleft($\tau = 0.50$), upperright($\tau = 0.05$), lower($\tau = 0.95$)

INET-connectedness

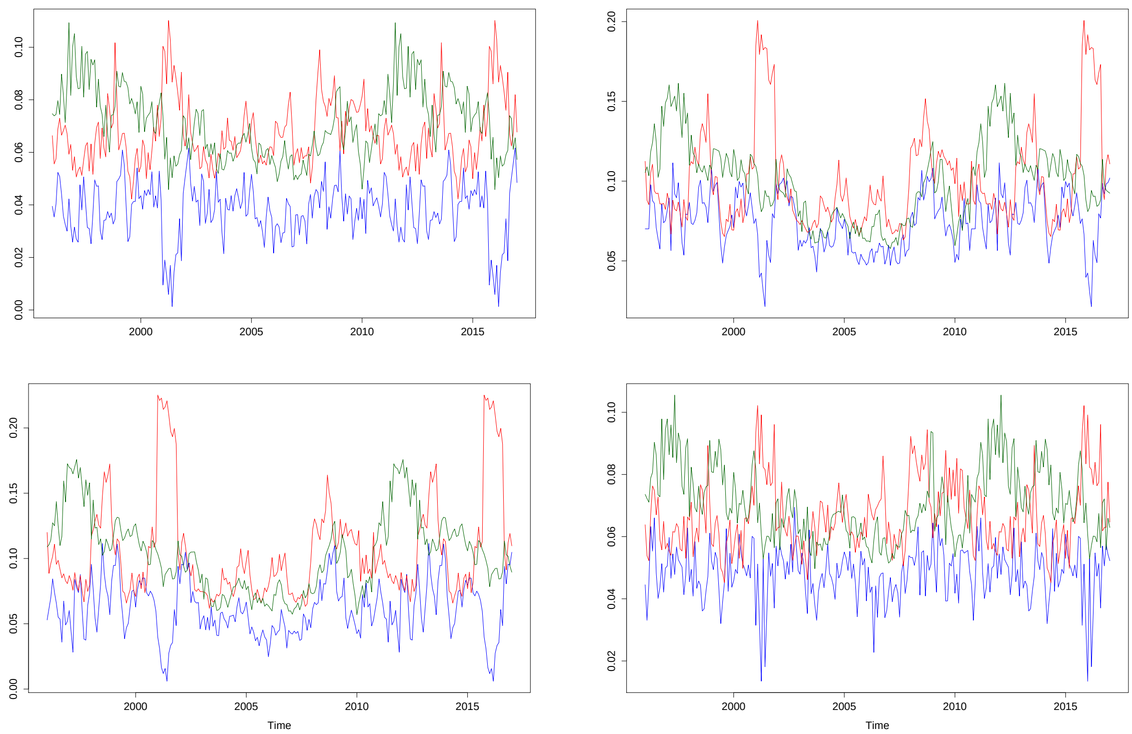

Figure 8: The comparison of industry network connectedness under $\tau = 0.50$, $\tau = 0.05$ and $\tau = 0.95$: upperleft(cycle), upperright(middleman), lowerleft(in), lowerright(out)

INET-connectedness

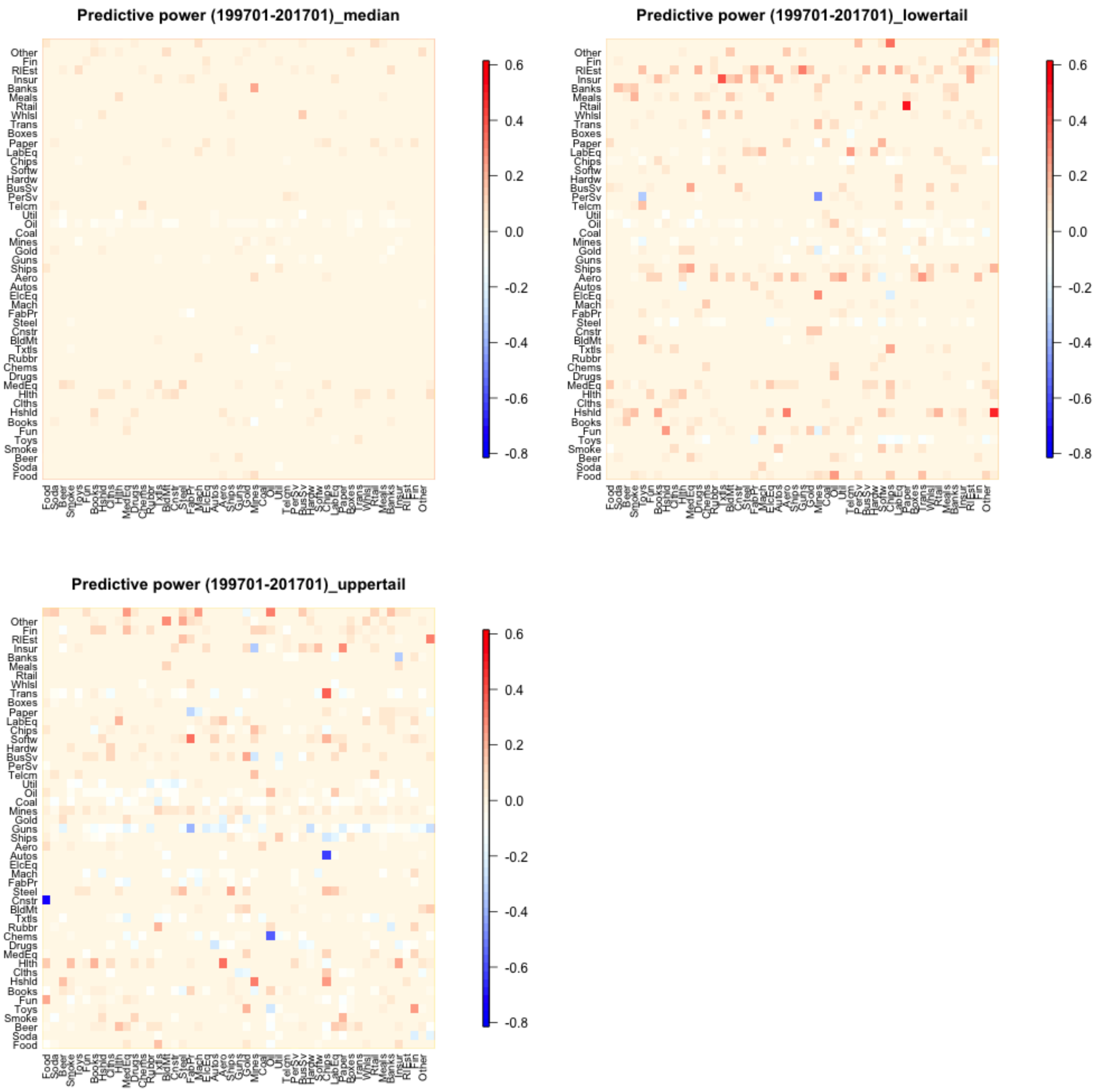

Figure 9: Connectedness across 49 industries - $\tau = 0.50$ (upperleft), $\tau = 0.05$ (upper-right), $\tau = 0.95$ (lower)

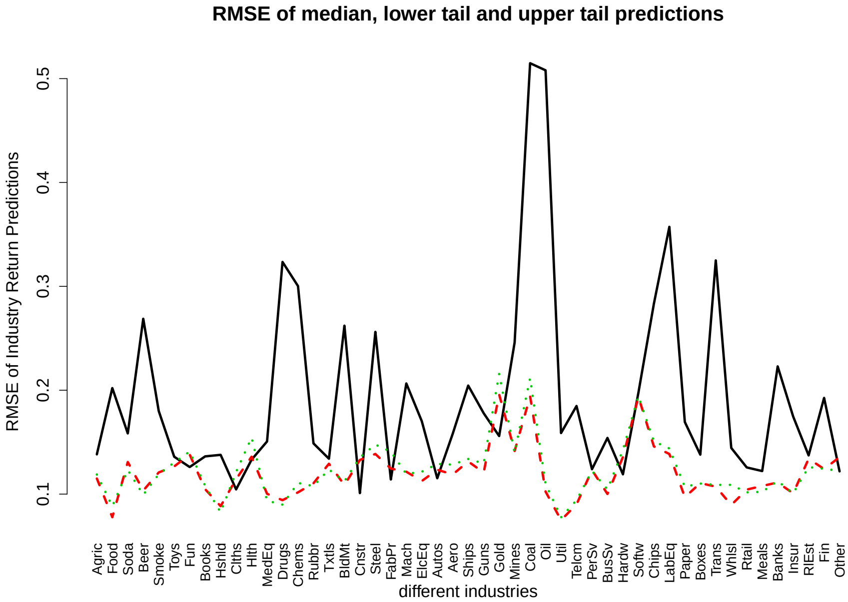

Figure 10: RMSE of industry return prediction over 253 months for $\tau = 0.05$, $\tau = 0.50$ and $\tau = 0.95$

INET-RMSE

SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger and Marliese Uhrig-Homburg, January 2017.
002 "Estimating location values of agricultural land" by Georg Helbing, Zhiwei Shen, Martin Odening and Matthias Ritter, January 2017.
003 "FRM: a Financial Risk Meter based on penalizing tail events occurrence" by Lining Yu, Wolfgang Karl Härdle, Lukas Borke and Thijs Benschop, January 2017.
004 "Tail event driven networks of SIFIs" by Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle and Yarema Okhrin, January 2017.
005 "Dynamic Valuation of Weather Derivatives under Default Risk" by Wolfgang Karl Härdle and Maria Osipenko, February 2017.
006 "RiskAnalytics: an R package for real time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods" by Lukas Borke, February 2017.
007 "Testing Missing at Random using Instrumental Variables" by Christoph Breunig, February 2017.
008 "GitHub API based QuantNet Mining infrastructure in R" by Lukas Borke and Wolfgang K. Härdle, February 2017.
009 "The Economics of German Unification after Twenty-five Years: Lessons for Korea" by Michael C. Burda and Mark Weder, April 2017.
010 "DATA SCIENCE & DIGITAL SOCIETY" by Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, May 2017.
011 "The impact of news on US household inflation expectations" by Shih-Kang Chao, Wolfgang Karl Härdle, Jeffrey Sheen, Stefan Trück and Ben Zhe Wang, May 2017.
012 "Industry Interdependency Dynamics in a Network Context" by Ya Qian, Wolfgang Karl Härdle and Cathy Yi-Hsuan Chen, May 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".