DSSR

Discussion Paper No. 34

Examining Investors Trading Performance Through Brokers Network

Hongwei Chuang

February, 2015

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Examining Investors Trading Performance Through Brokers Network

Hongwei Chuang

Graduate School of Economics and Management, Tohoku University
27-1 Kawauchi, Aoba Ward, Sendai 980-8576, Japan
hongwei@econ.tohoku.edu.tw

Abstract

Institutional investors play an important role in determining the stock returns. However, not so many research conduct the empirical research to investigate their behaviors because of the lankness of dataset. This paper, by using the unique dataset which contains all daily buy and sell information in each broker, I explore the applications of multivariate vector autoregression (MVAR) models in constructing the financial networks of stock market. In the spirit of Granger causality, the proposed methodology provides a direct way with statistical significance and economic meaning to describe the mutual interconnectedness of brokers. A financial network I analyzed illustrates as the structure properties for the underlying interactions among institutional investors.

Key words: Financial Network; Vector Autoregression Model; Multivariate Time Series Analysis.
JEL classification: C3; G2

---

1. Introduction

Networks have long appeared in almost every aspect of science and technology. Researchers in economics, sociology, physics and computer science [1] have developed theoretical models for extracting the insights from networks and have proposed many practical applications based on their empirical studies. The PageRank algorithm of Google is an example. However, the impact on market quality of investors' trading behaviors within a financial network remains less developed. Hence, the purpose of this paper is to study the underlying interactions among brokers in order to investigate the relationship between investors' trading profits and the brokers' financial network.

[1] For economics, more detail can be found in Schweitzer, Fagiolo, Sornette, Vega-Redondo, Vespignani, and White (2009); For sociology, more details can be found in Wasserman (1994); For physics and computer science, more details can be found in Easley and Kleinberg (2010).

To construct a financial network, it is necessary to define the node (vertex) and link (edge). In this study, I analyze the dynamic interactions among brokers, therefore, it is natural to represent the brokers as the nodes. As such, the definition of links would be problematic since the definitions of links in financial network literature usually describes the links from the trading motives for each pair of investors. That is, if two investors trade the same stock in the same direction (buy or sell) in a short time period, the two investors are treated as being connected. However, this definition is not applicable to this study because a broker can be represented as a group of investors. In this case, it is much easier to link a pair of brokers if they trade the same stocks, than it is to compare a pair of investors. Comparing the brokers usually results in a fully connected financial network. In other words, the financial network is a complete graph [2] .

Hence, I propose to use a statistical methodology to counter this problem– instead of defining the links by using the trading motives, such that the definition of links can catch up to the trading behaviors of the brokers. The principle is that if we consider the portfolio holdings of one broker during a period of time and denote it by $\vec{v}_1$. If $\vec{v}_1$ is Granger caused by other brokers' portfolio holdings, say $\vec{v}_2, ..., \vec{v}_k$ where $k \leq N$ (N is the total number of brokers) in a multivariate time series sense, then brokers $v_2, ..., v_k$ are going to connect with v_1. Based on this principle, I adopt the multivariate vector autoregression (MVAR) models in constructing the brokers' financial network to find the Granger causality among the brokers' portfolio holdings. In fact, the proposed methodology works not only for the brokers' financial network, but also for those networks with a sparse trading matrix.

The first contribution of this paper is the discovery that the links are directional and the interconnectedness is a system-wide concept. The measurement of interconnectedness concerns the multivariate time series relationships which reflect a generous structure financial network, not only a pairwise mutual connection, as in Billo, Getmansky, Lo, and Pelizzon (2012). The directional linkage is considered because the brokers trading network in which also contains the institutional investors trading behaviors in the market. Usually, these institutional investors are more informative than other individual investors. Their past trading information can affect the institutional/individual investors in other broker firms. It is necessary to identify who are leaders and who are followers in the market. Moreover, MVAR models can not only implement the principle and measure the daily time-varying interconnectedness among brokers, but also can implement the Granger-Causality test to determine the magnitude of interconnectedness.

[2] A complete graph is a simple undirected graph in which every pair of distinct vertices is connected by a unique edge.

The second contribution of this study is the unique dataset I have collected from the Taiwan Stock Exchange (TSE), which consists of all brokers' daily trading information on all trading days from November 6, 2001 to March 2, 2011. There are a total 95 brokers and 1,330 stocks being traded during the sample period. The dataset contains order types (buy or sell), daily total trading amounts, and average prices for stocks that are traded by each broker. The financial network analysis is based on brokers' data instead of account-level data, because most stock market investors are illiquidity traders. Tracking their trading behaviors would become confusing and possibly be irrelevant to constructing a financial network. Even worse, it could result in an estimation bias.[3] In this paper, I show that analyzing the brokers' financial network could prevent this issue.

Finally, my empirical findings provide substantial support for a positive relationship between investors' gains and losses and brokers' financial network centrality. When a broker has more connections with other brokers, which means the lagging investors are more central in the network, these investors are more profitable in the market. This is also shown in the conjecture of the theoretical model in Walden (2013). This empirical finding can also be supported by recent studies on the effect of social networks on decisions made by investors, which indicate that the central investors in the network tend to access more information and trade earlier in the right direction than the peripheral investors. In my regression analysis, I find that one standard deviation increase in rescaled centrality of a group of investors leads to a 0.1 to 2.7% increase in returns over the short-term. For the long-term, there is a 0.1 to 2.0% increase in returns .

The remainder of the paper is organized as follows: Related literature review is in Section 2. Section 3 describes the details of the data and my methodology to construct the Granger-Causality network. Section 4 illustrates my main empirical results. Section 5 concludes the paper.

2. Related Literature

My study follows up on mainstream interests in investigating the equilibrium analysis of information diffusion within a population of investors. For example, Shiller and Pound (1989) survey 131 NYSE investors and ask what prompted their initial interest in their most recent stock purchase or sale. The survey reveals that for the majority of investors, discussion with peers prompted the action. Ivkovi and Weisbenner (2007) find similar empirical evidence for households. Hong, Kubik, and Stein (2005) document that mutual fund managers in a given city tend to have trading behavior that correlate more strongly with other managers in the same city, as opposed to with managers in dif-

[3] As shown in Gomez-Rodriguez, Leskovec, and Krause (2012), the exact maximum likelihood estimation is not feasible for large networks.

ferent cities. Cohen, Frazzini and Malloy (2008, 2010) demonstrate that past educational connections facilitate information transmission from managers to security analysts, and examine educational connections between mutual fund managers and corporate board members to identify information transfer via social networks. Ozsoylev, Walden, Yavuz, and Bildik (2013) study the trading behavior of investors by using an account-level dataset of all trades on the Istanbul Stock Exchange in 2005. They find the central investors earn higher returns and trade earlier than peripheral investors with respect to information events. Walden (2013) introduces a dynamic noisy rational expectations model to explore agents' trading behavior and finds that agents who are more closely connected will have more period-by-period trades. Their profits can be determined by the measure of centrality of the network. Han and Yang (2013) also analyze a rational expectations equilibrium model to explore the implications of information networks for the financial market.

This study also links to other financial network literature which focuses on the allocation purpose of investors. The main concept is that the two nodes in a financial network can be connected because they implement similar investing strategies and thus hold the same financial assets in their portfolios. For example, Ibragimov, Jaffee, and Walden (2011) show that the recent financial crisis has revealed significant externalities and systemic risks arising from the interconnectedness of financial intermediaries' risk portfolios. Negative externalities arise because of intermediaries' actions taken to diversify their risks. Billo, Getmansky, Lo, and Pelizzon (2012) propose econometric measures to investigate the monthly returns of the 25 largest companies in the four sectors (hedge funds, banks, broker/dealers, and insurance) in the U.S. stock market. They find that all four sectors have become highly interrelated over the past decade, likely increasing the level of systemic risk in the financial and insurance industries through a complex network of relationships. The two nodes are more likely to share information and implement similar investing strategies so that they can also hold the same financial assets and diversify their risk in the market.

Haldane (2013) argues that financial innovation, in the form of structured products, increases further network dimensionality, complexity, and uncertainty. Certain financial instruments, including RMBS, ABS, CDO, and CDS, have created a series of gross claims among financial entities which far exceed their capital bases. For example, Lehman had gross CDS exposure approximately eight times its balance sheet in August 2007. This occurred because financial innovation in the form of structured credit also had the consequence of creating a network structure, which was non-hierarchical. Financial engineers created products in which elements of a loan portfolio were reassigned to a higher order subassembly. Thus, an automatic dependence was created among almost every substructure.

The financial network analysis is not only used in examining the stock market, but also in other financial assets markets. Allen and Gale (2000) studied how the banking system responds to contagion when banks are connected under different network structures. Their study showed that incomplete networks are more prone to contagion than complete ones. Better connected networks are more resilient because the proportion of losses in one bank's portfolio is distributed among more banks via interbank agreements. May, Levin, and Sugihara (2008) discuss the similarities in analyzing financial networks and ecosystems, especially as regards their vulnerabilities. The system includes over 8,000 banks, but 75% of the value is distributed among 66 banks. Flows between large money center banks are seen as links with higher weights.

3. Data and Methodology

Taiwan Stock Market

The dataset for my empirical study is collected from the TSE, which is the world's 12^{th} largest financial market. During my sample period, the market operated from 9:00 AM to 1:30 PM. Buy and sell orders interacted to determine the executed price, subject to applicable auto-matching rules, every 90 seconds. Orders were executed in strict price and time priority. A daily price limit of 7% held in each direction and a trade-by-trade intraday price limit of two ticks from the previous trade price was in effect. The commission fee for TSE was 0.1425% of the trading value and there was a transaction tax on stock sales of 0.3%. Capital gains were not taxed, whereas cash dividends were taxed at ordinary income tax rates for domestic investors and at 20% for foreign investors. Corporate income was taxed at a maximum rate of 25%, whereas personal income was taxed at a maximum rate of 40%. The accumulated investor account number at securities companies was approximately 16 million in 2011, almost 76% of the total population.

Brokers' Daily Trading Information

Instead of using account-level data, I use all daily trading information of brokers from November 6, 2001 to March 2, 2011 to carry out my empirical study covering the financial crisis in 2007. My reason for using the brokers' trading data set is that the majority of investors in a market trade infrequently. Tracking infrequent investors' trading behavior and identifying their mutual interactions could easily create distraction and would be irrelevant to a study of investors' trading behavior via networks. Most important, without this irrelevant information, my analysis of the size of the brokers' financial network will be more stable over time. The sample used records order types (buy or sell), trading total amounts, and average prices for stocks traded by each broker every day. Therefore, I can investigate the trading behavior of groups of investors within a financial network in a market. Further, these brokers can be categorized into these finance types: Banking (13 brokers), Bills Finance Corporations (1 broker), Specialized Brokerage Firms (32 brokers), Integrated Securities Firms (31 brokers), and Foreign Financial Institutions (18 brokers). During the sample period, a total of 95 brokers and 1,330 stocks had complete trading information. The average trading amount of these brokers is approximately 97% of the entire market's total daily trading amount.

Short- and Long-term Trading Gains and Losses

Based on the unique dataset, the first step of my analysis is to calculate each broker's daily buy and sell portfolios to mimic a group of investors' net daily purchases and sales over a period of time. I use a method similar to Barber, Lee, Liu, and Odean (2009), but focus on brokers' trades over the short- and long-term in order to evaluate groups of investors' trading gains and losses, which are denoted by the vectors $G_t = (g_{1,t}, ..., g_{n,t})'$ and $\text{Ł}_t = (\ell_{1,t}, ..., \ell_{n,t})'$ at time t respectively. For example, consider one of the brokers "Yuanta Securities." On March 29, 2002, it buys 900 shares of HTC and sells 700 shares. It makes 200 net shares of HTC, adding to the buy portfolio, whereas no HTC shares are added to the sell portfolio. The purchase price is then recorded as the total value of buying 900 shares minus the total value of selling 700 shares divided by the net shares of 200. Moreover, I consider the shares being included in the mimicking portfolios for a fixed horizon, z, where the short-term period is $z = 5$ (one week) and the long-term period is $z = 20$ (one month) trading days.

The volume-weighted (VW)-realized trading gains, $g^z_{i,t}$, can be calculated as:

$$g^z_{i,t} = \sum_{j=1}^{n_i} \frac{\#\text{net shares of stock } j \text{ purchased}}{\#\text{net total shares purchased by broker } i} \times r_j^{t+1}, \tag{1}$$

where r_j^{t+1} is the one-day return of the stock j after it was traded at time t and n_i denotes the number of stocks that the broker i holds within z-days holding period. An analogous calculation occurs for the VW-realized trading losses, $\ell^z_{i,t}$, which are defined as

$$\ell^z_{i,t} = \sum_{j=1}^{n_i} \frac{\#\text{net shares of stock } j \text{ sold}}{\#\text{net total shares sold by broker } i} \times r_j^{t+1}. \tag{2}$$

The net profit, $p^z_{i,t}$ is then defined as the difference between $g^z_{i,t}$ and $\ell^z_{i,t}$, that is,

$$p^z_{i,t} = g^z_{i,t} - \ell^z_{i,t}. \tag{3}$$

In the following study, I use the net profit time series data for brokers to represent the profits of groups of investors to fit the MVAR models.

Multivariate Vector Autoregression (MVAR)

To construct the financial networks of 95 brokers, I use the MVAR models to empirically measure their mutual interconnectedness from a system-wide perspective so that the network structure can be inferred. The Granger-Causality test be used to investigate whether one time series can provide forecasting power to another, and determine the magnitude of interconnectedness among these brokers. Hence, the interconnectedness I define can reflect both statistical correlations and economic connections for the multivariate time series of groups of investors' net profit. The definition of "link" in this study is different from Ozsoylev, Walden, Yavuz, and Bildik (2013), who define the connection between each pair of investors as if the two agents traded the same stock in the same direction in a short-term period and also different from Pareek (2012), who defined it as two fund managers who allocate 5% or more of their portfolio to the same stock being connected to each other.

MVAR model is defined as $Y_t = (y_{1t}, ..., y_{nt})', t = 1, 2, ..., T$, which denotes a weak stationary multivariate time series of dimension n defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A p-th order vector autoregressive can be represented in a simple form as

$$Y_t = c + \sum_{k=1}^{p} \Phi_k Y_{t-k} + \epsilon_t, \tag{4}$$

where c denotes an $(n \times 1)$ vector of constants and Φ_k denotes an $(n \times n)$ matrix of autoregressive coefficients for $k = 1, 2, ..., p$. The ϵ_t is a vector with Σ symmetric definite matrix. The stationary condition is satisfied if all roots of $|\Phi(Z)| = 0$ lie outside the unit circle.

For a large VAR model $(n > 2)$, the Granger-Causality test can be used to test whether one variable is influenced by other variables in the model system. Y_t can be arranged and partitioned in subgroups Y_{1t} and Y_{2t} with dimensions n_1 and n_2 $(n = n_1 + n_2)$, respectively. VAR model can be represented as a matrix form:

$$\begin{bmatrix} \Phi_{11}(\beta) & \Phi_{12}(\beta) \\ \Phi_{21}(\beta) & \Phi_{22}(\beta) \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} + \begin{bmatrix} \Sigma_1 \\ \Sigma_2 \end{bmatrix}. \tag{5}$$

The Wald statistic to test $H_0 : C\beta = c$, where C is a $s \times (n^2 p + n)$ matrix of rank s and c is an s-dimensional vector can be obtained from

$$\sqrt{T}(C\hat{\beta} - c)[C(\hat{\Gamma}_p^{-1} \otimes \Sigma)C']^{-1}(C\hat{\beta} - c) \xrightarrow{d} \chi^2(s). \tag{6}$$

In this study, I set the significance level of Wald test as 1%, i.e., $\alpha = 0.01$.

Granger-Causality Network

Based on the MVAR model, the financial network I construct followsa concept similar to the Grange-Causality graph in Eichler (2007), and I denote it as a "Granger-Causality" network. I define a network $G = (N, E)$, where

N is a set of elements called nodes ($N = 95$ in this study) and E is a set of directed edges which belong to the class

$$\{j \to i | i, j \in V, i \neq j\} \notin E \Leftrightarrow \Phi_{ij}(k) = 0 \; \forall k, \tag{7}$$

where Φ is the autoregressive coefficients matrix in Equation (4). A financial network can be represented as an adjacency matrix $A \in \{0, 1\}^{N \times N}$, with

$$A_{ij} = \begin{cases} 1 \text{ if broker } j \text{ is directly linked to } i, \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

Moreover, I assume that a broker is connected with himself; that is, $A_{ii} = 1$ for all i. This class represents the directed edges corresponding to direct causal relations among the components of Y_t which can be identified by the Granger Causality test as shown in Equation (6). For simplicity, I consider the case of $k = 1$ and use a 180-day rolling window to estimate Φ in Equation (4).

An Illustration

As an example, if I take the following four-dimensional VAR(1) process, with Equation (4):

$$Y_t = \Phi Y_{t-1} + \epsilon_t \tag{9}$$

with statistically significant parameters

$$\Phi = \begin{bmatrix} \Phi_{11} & 0 & \Phi_{13} & 0 \\ 0 & \Phi_{22} & 0 & \Phi_{24} \\ \Phi_{31} & 0 & \Phi_{33} & 0 \\ 0 & 0 & \Phi_{43} & \Phi_{44} \end{bmatrix}. \tag{10}$$

Therefore, the adjacency matrix A of Equation (8) can be represented as

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{11}$$

Based on the above adjacency matrix, the autoregressive structure can be visualized by the associated path diagram shown in Figure 1.

<Insert Figure 1 here>

Eigenvector Centrality

Once the network is constructed, to measure the relative importance within a network of a node, I calculate its centrality. Following the analysis of Ozsoylev, Walden, Yavuz, and Bildik (2013), I adopt the eigenvector centrality which is defined as:

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j \in G} A_{ij} x_j, \tag{12}$$

where $M(i)$ is a set of the neighbors of i and λ is a constant and A_{ij} is defined in Equation (8).

Eigenvectors of adjacency matrices are useful as measures of centrality or of status. However, they are misapplied to asymmetric networks in which some positions are unchosen (see Bonacich and Lloyd (2001)) and there will be many different eigenvalues λ for which an eigenvector solution exists (see Spizzirri (2011)). The additional requirement that all the entries in the eigenvector be positive, implies (by the Perron-Frobenius theorem) that only the greatest eigenvalue results in the desired centrality measure. The i-th component of the related eigenvector then gives the centrality score of the vertex i in the network. Power iteration is one of many eigenvalue algorithms that may be used to find this dominant eigenvector. I also use a more robust measure, rescaled centrality, to capture the information advantage than pure centrality in an empirically estimated investor network. The rescaled centrality of node i is as follows:

$$\frac{\text{Centrality}_i}{\text{In Degree}_i + \text{Out Degree}_i} \tag{13}$$

where In Degree_i measures how many links from other nodes to i and Out Degree_i measures how many links from i to other nodes.

4. Empirical Results

Power Law Distribution

The distribution of centrality is intimately related to asset pricing dynamics, and therefore may be informative about the aggregate behavior of a stock market as stated in Gabaix, Gopikrishnan, Plerou, and Stanley (2003). One of aggregate behaviors is that the degree distribution follows a power law in the form of

$$\mathbb{P}(X > x) \sim x^{-\alpha} \tag{14}$$

in the tail of the distribution. This is also called preferential attachment of investors. In examining the phenomenon, I plot the centrality distribution of a different z in Figure 2.

<Insert Figure 2 here>

In Figure 2, the black line is for $z = 5$ and the red dashed line is for $z = 20$.

I find that the network centrality is followed by the power law distribution.

Trading Profits v.s. Financial Networks

The summary statistics and the main empirical results are represented in Table 1 and Table 2. Table 1 and Table 2 show short-term trading behavior ($z = 5$) and long-term trading behavior ($z = 20$) of groups of investors, respectively. In both, I report the VW-realized trading gains, VW-realized trading losses and net profits from all trades under the five types of brokers. I also illustrate their centralities and rescaled centralities.

<Insert Table 1 here>

Table 1 illustrates the short-term trading behavior of groups of investors. For the rescaled centrality, Table 1 shows the investors belonging to the broker category Bankers are most connected with others, while investors belonging to the broker category Financial Corporations are the least connected. Investors trading in the broker category Integrated Securities firms are the most profitable, with a mean net profit of 2.4% during a one week period, while the investors trading in category Bills Finance Corporations are the least profitable with a mean net profit of -1.7%. Table 1 also indicates that the VW-realized trading losses are positively correlated to the rescaled centrality. When the rescaled centralities increase, the VW-realized trading losses increase as well.

<Insert Table 2 here>

Table 2 illustrates the relationships of long-term trading behavior among groups of investor. For the rescaled centrality, Table 2 shows the investors belonging to the broker category Integrated Securities Firms are the most connected with others, whereas the investors belonging to the broker category Bills Finance Corporations are the least connected. Investors trading in the category Integrated Securities Firms are the most profitable, with a mean net profit of 1.0%, while investors trading in the category Specialized Brokerage Firms are the least profitable with mean net profit of -0.6%. Table 2 also shows that the VW-realized trading gains are positively correlated to the rescaled centralities. When the rescaled centralities increase, the VW-realized trading gains also increase. We argue that the rescaled centralities of brokers can affect the investors trading gains and losses could because one possibility is that these brokerage firms provide better research reports to their clients, the research they provide could be correlated, which affect their client's trades.

Regression Analysis

I further investigate the relationship between the rescaled centrality and net profit via regression analysis. The models are

$$NPROF_{i,t}^{c} = \alpha^{c} + \beta^{c} RCTRAL_{i,t}^{c} + \epsilon_{i,t}, \tag{15}$$

where c denotes the three levels profitability: low, medium, and high. $NPROF_{i,t}^{c}$ and $RCTRAL_{i,t}^{c}$ indicate net profit and rescaled centralities for a group of investors in different categories. The regression results are shown in Table 3 and Table 4.

<Insert Table 3 here>

Table 3 is for $z = 5$. In Table 3, we can find that rescaled centrality is important in determining the net profit of a group of investors. Rescaled centrality can explain the high profitability of investors in the Integrated Securities Firms. We see one-standard deviation in the increase in the rescaled centrality of the high profitability of investors belonging to the category Integrated Securities Firms, leading to a 14.3% in return. For brokers in Bills Finance Corporations and Banking, rescaled centrality is also important in determining profitability of investors. I find a one-standard deviation increase in rescaled centrality of low profitability for investors belonging to Bills Finance Corporations and Banking, leading to 276.9% and 81.1% in returns, respectively.

<Insert Table 4 here>

The regression results in Table 4 are for $z = 20$. Rescaled centrality can explain the high-level profitability of investors belonging to category Foreign Financial Institutions. We find one-standard deviation increase in rescaled centrality for the high profitability of investors belonging to the category Foreign Financial Institutions, leading to 18.3% in return. In addition, we see one-standard deviation increase in rescaled centrality of medium profitability for investors belonging to the category Bills Finance Corporations and Specialized Brokerage Firms, leading to 206.8% and 22.8% in returns, respectively. In sum, rescaled centrality can explain investors' profitability.

5. Conclusions

Investors trading behaviors are more complicated than they appear. Therefore, the mechanisms needed in order to form links in a population in the market within a financial network structure are not readily apparent. In this study, I use the multivariate time series models and perform a statistical test to directly identify the links. According to the property of the dynamic similarities in portfolio holdings among a group of investors, I construct the Granger-Causality network to measure their relative importance within a brokers' financial network. My empirical results show the net profitability of groups of investors are positively correlated to their relative importance in the brokers' financial network.

My findings provide another viewpoint from which to study the trading behaviors of investors from the perspective of brokers' financial networks. The results demonstrate the existence of empirical evidence to support the hypothesis of information diffusion via investors' social networks which can affect their trading behaviors. Understanding the mechanism for forming links among investors will be useful in future study. It is also interesting to apply the network analysis to different financial assets.

References

Allen, F., and Gale, D., 2000, Financial contagion, *Journal of Political Economy* 108, 1–33

Billio, Monica, Getmansky, Mila, Lo, Andrew W. and Pelizzon, Loriana, 2012, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, *Journal of Financial Economics* 104, 535–559.

Barber, Brad M and Lee, Yi-Tsung and Liu, Yu-Jane and Odean, Terrance, 2009, Just how much do individual investors lose by trading?, *Review of Financial Studies* 22, 609–632.

Bonacich, Phillip and Lloyd, Paulette, 2001, Eigenvector-like measures of centrality for asymmetric relations, *Social Networks* 23, 191–201.

Cohen, L., Frazzini, A., and Malloy, C., 2008, The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116, 951–979.

Cohen, L., Frazzini, A., and Malloy, C., 2010, Sell-side school ties, *Journal of Finance* 65, 1409–1437.

Easley, D., and Kleinberg, J., 2010, Networks, crowds, and markets, *Cambridge University Press*

Eichler, Michael, 2007, Granger causality and path diagrams for multivariate time series , *Journal of Econometrics* 137, 334–353.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, E., 2003, A theory of power-law distributions in financial market fluctuations, *Nature* 423, 267–270.

Gomez-Rodriguez, M., Leskovec, J., and Krause, A., 2012, Inferring networks of diffusion and influence, *ACM Transactions on Knowledge Discovery from Data* 5:21, 1–37.

Haldane, A. G., 2013, Rethinking the financial network, *Springer Fachmedien Wiesbaden*, 243–278.

Han, B., and Yang, L., 2013, Social networks, information acquisition, and asset prices, *Management Science* 59(6), 1444–1457.

Hong, H., Kubik, J. D., and Stein, J. C., 2005, Thy neighbor's portfolio: wordofmouth effects in the holdings and trades of money managers, *Journal of Finance* 60, 2801–2824.

Ibragimov, R., Jaffee, D., and Walden, J., 2011, Diversification disasters, *Journal of Financial Economics* 99, 333–348.

Ivkovi, Z., and Weisbenner, S., 2007, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, *Review of Financial Studies* 20, 1327–1357.

May, R. M., Levin, S. A., and Sugihara, G. 2008. Complex systems: Ecology for bankers. *Nature* 451(7181), 893–895.

Ozsoylev, Han N and Walden, Johan, 2011, Asset pricing in large information networks, *Journal of Economic Theory* 146, 2252–2280

Ozsoylev, Han and Walden, Johan and Yavuz, M Deniz and Bildik, Recep, 2013, Investor networks in the stock market, *Review of Financial Studies*

forthcoming.
Pareek, Ankur 2012, Information networks: Implications for mutual fund trading behavior and stock returns, *Working Paper*
Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., and White, D. R., 2009, Economic networks: The new challenges, *Science* 325, 442.
Shiller, R. J., and Pound, J., 1989, Survey evidence on diffusion of interest and information among investors, *Journal of Economic Behavior & Organization* 12, 47–66.
Spizzirri, Leo, 2011, Justification and application of eigenvector centrality, *Working Paper*
Wasserman, S., 1994, Social network analysis: Methods and applications, *Cambridge University Press*
Walden, Johan, 2013, Trading, profits, and volatility in a dynamic information network model, *Working Paper*

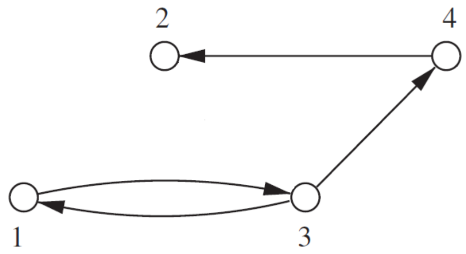

Figure 1: An Illustration of Granger-Causality Network
The figure illustrates an example of Grange-Causality network according to the four-dimensional VAR(1) model with the adjacency matrix shown in Equation (11).

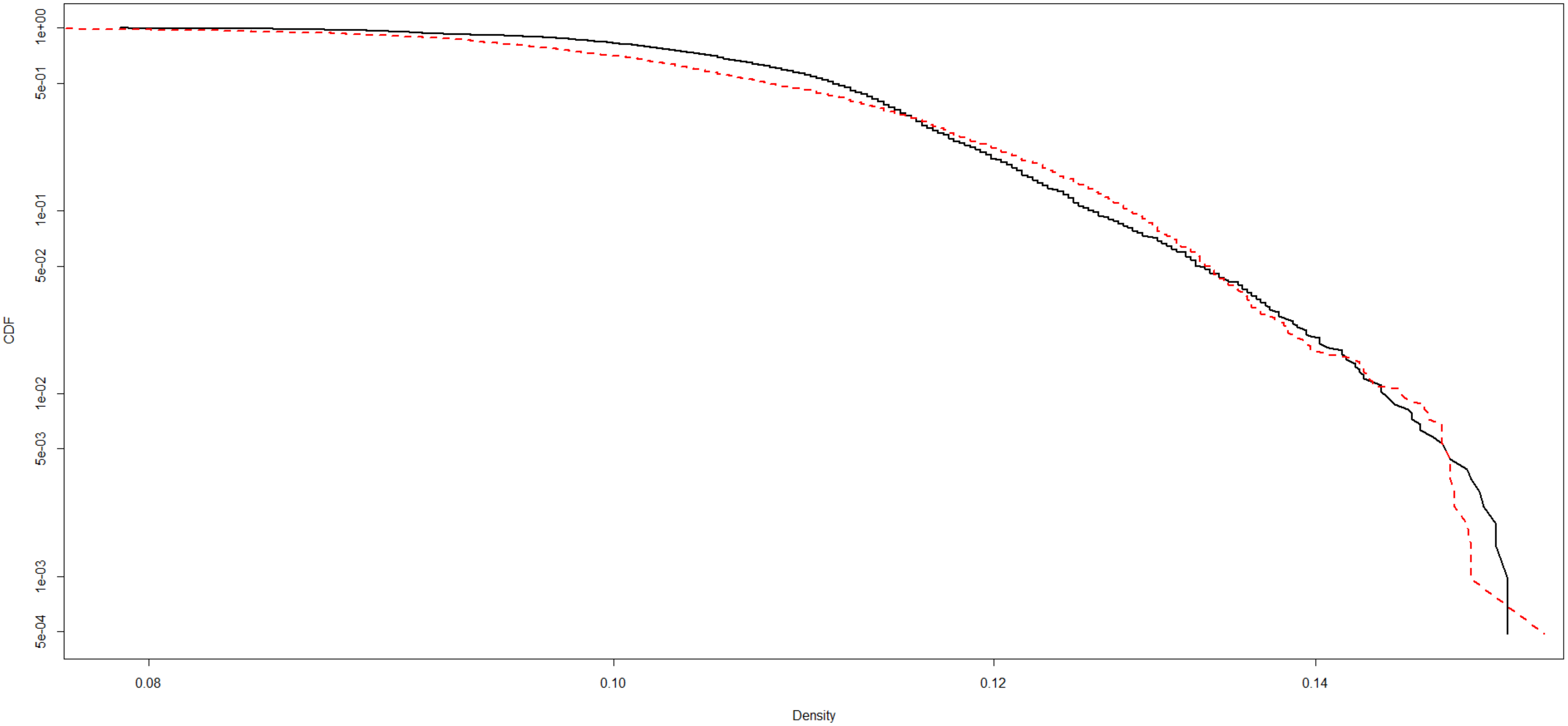

Figure 2: Preference Attachment Examination
The figure plots the centrality distribution of brokers' financial network to examine the preference attachment theory as shown in Equation (14). The black line is for z = 5 and the red dashed line is for z = 20.

Table 1: Summary Statistics (z=5)

The table reports the summary statistics of z=5 for five types of brokers. "Std" and "VW" mean standard deviation and volume-weighted, respectively.

z=5	Statistics	VW realized gain	VW realized loss	Net profit	Centrality	Rescaled Centrality
I. Foreign Financial Institutions (N=20750)	Mean	0.066	**0.068**	-0.002	0.091	**0.674**
	Std	1.588	1.599	0.703	0.061	0.455
	Skewness	-0.157	-0.170	0.069	0.885	0.963
	Kurtosis	2.070	2.023	2.807	0.721	1.025
II. Integrated Securities Firms (N=56025)	Mean	0.087	**0.064**	0.024	0.090	**0.668**
	Std	1.643	1.645	0.411	0.061	0.458
	Skewness	-0.426	-0.445	0.094	1.140	1.216
	Kurtosis	1.936	1.950	8.951	2.142	2.510
III. Specialized Brokerage Firms (N=66400)	Mean	0.057	**0.065**	-0.009	0.090	**0.672**
	Std	1.708	1.697	0.424	0.063	0.480
	Skewness	-0.428	-0.453	-0.025	1.166	1.232
	Kurtosis	1.840	1.862	9.462	2.176	2.445
IV. Bills Finance Corporations (N=2075)	Mean	0.037	**0.054**	-0.017	0.085	**0.635**
	Std	1.772	1.766	0.594	0.061	0.467
	Skewness	-0.432	-0.398	-0.362	0.822	1.020
	Kurtosis	1.597	1.620	3.876	0.392	1.213
V. Banking (N=24900)	Mean	0.064	**0.070**	-0.006	0.096	**0.711**
	Std	1.638	1.645	0.458	0.065	0.486
	Skewness	-0.415	-0.427	-0.120	1.273	1.259
	Kurtosis	1.974	2.008	5.261	3.056	2.835

Table 2: Summary Statistics (z=20)

The table reports the summary statistics of z=20 for five types of brokers. "Std" and "VW" mean standard deviation and volume-weighted, respectively.

z=20	Statistics	VW realized gain	VW realized loss	Net profit	Centrality	Rescaled Centrality
I. Foreign Financial Institutions (N=20750)	Mean	**0.059**	0.062	-0.003	0.089	**0.663**
	Std	1.559	1.558	0.424	0.067	0.500
	Skewness	-0.192	-0.174	0.003	1.202	1.261
	Kurtosis	2.249	2.142	3.128	1.965	2.287
II. Integrated Securities Firms (N=56025)	Mean	**0.069**	0.059	0.010	0.090	**0.667**
	Std	1.625	1.626	0.247	0.071	0.530
	Skewness	-0.444	-0.461	-0.023	1.495	1.489
	Kurtosis	2.066	2.103	15.272	3.533	3.314
III. Specialized Brokerage Firms (N=66400)	Mean	**0.053**	0.060	-0.006	0.085	**0.632**
	Std	1.680	1.672	0.214	0.065	0.489
	Skewness	-0.450	-0.473	0.034	1.189	1.228
	Kurtosis	2.004	2.038	11.417	1.954	2.100
IV. Bills Finance Corporations (N=2075)	Mean	**0.037**	0.042	-0.005	0.078	**0.584**
	Std	1.729	1.733	0.287	0.064	0.494
	Skewness	-0.455	-0.421	-0.122	1.035	1.155
	Kurtosis	1.800	1.803	4.546	0.573	1.000
V. Banking (N=24900)	Mean	**0.060**	0.062	-0.002	0.085	**0.632**
	Std Deviation	1.617	1.625	0.286	0.070	0.528
	Skewness	-0.428	-0.450	0.069	1.778	1.829
	Kurtosis	2.136	2.144	15.152	5.729	5.955

Table 3: Short-Term Profitability of Investors and Rescaled Centrality of Brokers

The table reports regression results of Equation (15) under three-level profitability of low, middle, and high for five types of brokers for z=5. It investigates the relationship between the rescaled centrality and net profit via regression analysis. The regression models are

$$NPROF^{c} = \alpha^{c} + \beta^{c}\ RCTRAL^{c} + \varepsilon$$

where c denotes the three level profitability of low, medium and high. NPROF and RCTRAL indicate net profits and rescaled centralities for a group of investors in different types of brokers, respectively. *, **, and *** are represented by the statistical significance of 0.1, 0.05, and 0.01. The standard errors are in parentheses.

z=5	NPROF	Low		Middle		High	
I. Foreign Financial Institutions	Intercept	0.000		-0.001		-0.011	
		(0.018)		(0.048)		(0.031)	
	RCTRAL	0.232		0.060		-0.041	
		(0.506)		(0.581)		(0.186)	
II. Integrated Securities Firms	Intercept	0.027	***	0.005		0.003	
		(0.006)		(0.017)		(0.010)	
	RCTRAL	-0.116		0.219		0.143	**
		(0.181)		(0.205)		(0.061)	
III. Specialized Brokerage Firms	Intercept	-0.019	***	0.008		-0.010	
		(0.006)		(0.016)		(0.010)	
	RCTRAL	0.168		-0.229		0.048	
		(0.164)		(0.197)		(0.057)	
IV. Bills Finance Corporations	Intercept	-0.107		-0.116		-0.100	
		(0.044)		(0.120)		(0.082)	
	RCTRAL	2.769	**	1.227		0.598	
		(1.270)		(1.457)		(0.504)	
V. Banking	Intercept	-0.037	***	0.007		-0.023	
		(0.011)		(0.029)		(0.016)	
	RCTRAL	0.811	***	-0.057		0.079	
		(0.299)		(0.347)		(0.092)	

Table 4: Long-Term Profitability of Investors and Rescaled Centrality of Brokers

The table reports regression results of Equation (15) under three-level profitability of low, middle, and high for five types of brokers for z=20. It investigates the relationship between the rescaled centrality and net profit via regression analysis. The regression models are

$$NPROF^{c} = \alpha^{c} + \beta^{c}\ RCTRAL^{c} + \varepsilon$$

where c denotes the three level profitability of low, medium and high. NPROF and RCTRAL indicate net profits and rescaled centralities for a group of investors in different types of brokers, respectively. *, **, and *** are represented by the statistical significance of 0.1, 0.05, and 0.01. The standard errors are in parentheses.

z=20	NPROF	Low		Middle		High	
I. Foreign Financial Institutions	Intercept	-0.009		0.000		-0.030	**
		(0.010)		(0.026)		(0.015)	
	RCTRAL	0.254		-0.094		0.183	**
		(0.353)		(0.338)		(0.088)	
II. Integrated Securities Firms	Intercept	0.013	***	0.010		0.009	
		(0.003)		(0.009)		(0.005)	
	RCTRAL	-0.163		0.038		0.000	
		(0.122)		(0.116)		(0.028)	
III. Specialized Brokerage Firms	Intercept	-0.006	**	-0.023	***	-0.006	
		(0.003)		(0.007)		(0.005)	
	RCTRAL	-0.051		0.228	**	0.004	
		(0.092)		(0.095)		(0.027)	
IV. Bills Finance Corporations	Intercept	0.001		-0.165		0.049	
		(0.017)		(0.059)		(0.040)	
	RCTRAL	0.217		2.068	***	-0.354	
		(0.608)		(0.784)		(0.237)	
V. Banking	Intercept	-0.003		0.024		-0.010	
		(0.005)		(0.016)		(0.009)	
	RCTRAL	-0.087		-0.294		0.037	
		(0.187)		(0.211)		(0.050)	