Analysis of Empirical Investor Networks and Information Events in Stock Market

Shiraz Farouq
Abstract

**Analysis of Empirical Investor Networks and Information Events**

Shiraz Farouq

We did further analysis to understand the dynamics of information diffusion in an Empirical Investor Network (EIN) [8]. We find that the timings of the trades are of crucial importance for central investors. We find further evidence that central traders have information advantage and information diffusion plays a central role in their profitability. We verify through our robustness tests that all results hold up when profits are calculated using actual realized returns instead of a fixed holding period.
## Contents

1 Introduction ................................................................. 2

2 The Theory behind Empirical Investor Networks (EIN) ............ 3
   2.1 Centrality Measures .................................................. 4
      2.1.1 Degree Centrality ................................................. 4
      2.1.2 Eigenvector Centrality ......................................... 5
      2.1.3 Rescaled Centrality .............................................. 6

3 Profitability Measures .................................................. 6
   3.1 Forward Pricing ....................................................... 7
   3.2 Realized Returns ..................................................... 7

4 Data Description .......................................................... 8

5 Analysis of Data .......................................................... 9

6 Computational Issues .................................................... 9
   6.1 Calculating realized profits ......................................... 9
   6.2 Finding information events .......................................... 10

7 Investor centrality, information diffusion and timing of the trades 11

8 Information events and information diffusion ......................... 11
   8.1 Frequency of information events in the stock market .......... 11
   8.2 Study of high and low information periods of the year ....... 12

9 Why some investors earn extreme returns than others? .............. 13

10 Robustness tests and market microstructure issues ................. 14
   10.1 Market microstructure ............................................... 14
   10.2 Realized returns as a measure to profitability ................. 14

11 Robustness of alternative profitability measures .................... 17

12 Limitations of the current work ...................................... 17

13 Concluding remarks ..................................................... 17

14 Acknowledgements ...................................................... 19
1 Introduction

Over the years, various theories have been put forward to understand the fundamental determinants of success for investors in stock market trading. Like for instance

1. Sophisticated Investors
2. Variations in liquidity
3. Investment Styles
4. Differential Information

Walden et al. in their research [8] used network theory to understand investor behavior in order to seek further insight into the fourth determinant. The underlying hypothesis is that heterogeneous information diffuses across social networks as a function of space and time, and that trading behavior and profitability of investors depend on their corresponding position in the network. In order to proxy real investor networks, an empirical investor network or simply EIN has been proposed. Estimating an EIN, roughly, is based on the idea that investors may be linked together if they happen to fall on the same side of a trade of the same stock within a certain interval in time. However, to stand any middling chance of being considered a fundamental determinant of profitability for investors, an EIN firstly must not be random but rather trading agents must form some sort of strategic network that is fairly consistent and stable over time. Secondly, EIN should be able to capture the dynamics of information diffusion. Thirdly, the position of investor in the EIN i.e. his centrality must be able to exhibit a positive relation to returns. Finally, the discretion to act early on both the observable and unobservable information events before their neighbors should be a defining feature of central investors. Any other alternate explanations or theory that may attempt to nullify the information diffusion story must then also be able to satisfy all these four factual assertions about EIN. Indeed, however, several explanations fail to satisfy at least one of these assertions.

An EIN satisfies all the fours assertions, thus lending support to the idea that information diffusion is an important factor in determining investors trading behavior and profitability. Another important question that has been answered in [1] to some extent is that the dynamics of information diffusion are decentralized. EIN was found consistent with decentralized network structure. Further there was evidence that most trading activity occurs in a stock before news on it appears over on the media.

In the current study, the focus of our attention is to answer the following questions:
1. What happens to the profits of central traders if they delay their trades by a day?
2. What is the driving force behind extreme returns of certain investors?
3. What happens to the profits of central traders during months in which there are large numbers of earning announcements?
4. Are there more information events than those covered in conventional media?
5. Does fixing the holding period to 1 month induce spurious correlations between centrality and returns?
6. Is it important to define the profitability measure in a specific way, or are the results robust to alternative specifications?
We find that delaying trades by a day erodes as much as around 30% of the positive relationship between centrality and returns of central investors. Thus, the more central an investor is in its network, the more likely it is that he will earn higher profits compared to his neighbors. Also, the magnitude of centrality is the main driving force behind extreme returns of certain investors. We find that central traders indeed have an information advantage during the months of high numbers of earning announcements. We find evidence that along with special information events, there are also many unobserved information events in the market that the central investors can take advantage of. To substantiate this assertion we identified 3,291 information events in the year 2005 under the assumption that any return above 15% over consecutive days (3-week period in our case) corresponds to an information event in the market. Our robustness tests mitigate the concern that spurious correlation and other microstructure issues bias our results related to the relationship between centrality and returns. They also confirm that the specific way our profitability measure is constructed does not affect the results of the analysis.

2 The Theory behind Empirical Investor Networks (EIN)

A general social network comprising \( N \) agents, in a stock market, can be represented as neighborhood (adjacency) Matrix \( A \in \{0, 1\}^{N \times N} \)

\[
A_{ij} = \begin{cases} 
1 & \text{if } i \text{ and } j \text{ are connected} \\
0 & \text{otherwise} \end{cases} \quad (1.a)
\]

Also since the connections are bidirectional i.e. \(^1\)

\[A_{ij} = A_{ji}, \quad \forall i, j. \quad (1.b)\]

This implies that \( A \) is symmetric.

The motivating idea is that the stock market comprises of informed investors, say \( N_I \), and a large and a large number of uninformed investors, say \( N_U \), such that

\[N = N_I + N_U. \quad (2)\]

One of the major assumptions in defining the governing dynamics of EIN is that the informed trading agents only trade in the presence of a signal which corresponds to a profit \( \pi_0 > 0 \), in a setting that requires uninformed trading agents to be willing to accept the opposite side of the transaction with a profit \(-\pi_0\). Also, \( \pi_0 > 0 \) is above expected market return. In general, the trading behavior and profitability depends on the position the trading agent takes in the network.

Consequently, in an EIN defined over each discrete point in time, say \( t = 0, 1, 2, \ldots \), each agent \( n_i \in N_I \) receives a signal \( s_t^i \). If \( \psi_t^i \) represents the set of signals that agent \( i \) received at time \( t \),

\(^1\)Symmetric information sharing is important, otherwise an agent will not have any incentive to form a relationship.
then a simplifying assumption is that only one such signal is relevant. A trading agent $n_i$ may then go on to share its signal with other trading agents in its network between times $t, t + 1$ with a probability $q_1 > 0$. Similarly, the set of agents in the network of $n_i$, marked as $n^1_i$, may in turn decide to trade at time $t + 1$, and then share their signal with the agents in their corresponding networks, marked $n^2_i$, between time $t + 1, t + 2$ with probability $q_2 > q_1$. Eventually, $n^2_i$ trade on the signal received at time $t + 2$ and by time $t + 3$, the signal is fully incorporated into the stock market and then no further profits above market return are possible. With information slowly diffusing into the market, not only by agents themselves but also through other media between time $t, t + 3$, the profits follow the sequence $\pi_2 < \pi_1 < \pi_0$. This is inline with the assumption that as time passes, expected profits on a particular signal reach some asymptotic limit. This discrete version is also easily extendable to a generalized model in continuous time.

Now that we have a theory to define the Empirical Investor Network (EIN), we can define it formally.

**Definition 1:** An Empirical Investor Network (EIN), comes into existence in the stock market over a period of time when at least two investors, say $i, j \in N$, such that $i \neq j$ trade in the same stock, say $k$, and in the same direction (either buy or sell) at least $M$ times within some time interval $\Delta t$ in $T$.

### 2.1 Centrality Measures

Once we have our EIN, we need to understand some of the macro and micro level properties of the network. When we speak about centrality of an investor, it implies a microstructure measure to understand how important a particular node is within a network. We discuss three such measures to define the concept of centrality in the EIN. A detailed overview of these measures can be found in texts related to social networks such as [5].

#### 2.1.1 Degree Centrality

This is a measure that tells how connected the node is i.e. how many neighbors does the node have

$$D_i = \sum_{j=1}^{n} A_{ij}. \quad (3.a)$$

The normalized version of this measure is

$$d_i = \frac{D_i}{n-1}. \quad (3.b)$$

---

It is assumed that the probability of sharing information is greater in the second stage, given that information loses its importance with time.
2.1.2 Eigenvector Centrality

The problem with degree centrality is that it gives no weight to well connectedness of its neighbors. A node connected to the nodes that are well connected themselves is more important than the nodes that are not. The second important concern is that the more further the nodes are, the more time it will take for the information signal to reach the investor. A micro measure that takes into account these two fundamental opposing forces is therefore needed. Fortunately, eigenvector centrality seems to efficiently counterbalance the interplay between the connectedness and information delay. How? Eigenvector centrality is basically a sum of powers of the degree matrix or simply a sum of degrees of different orders. This implies that with higher order degree i.e. the further the nodes are, the more signals reach an investor, but the more delayed these signals are.
Specifically, if $C_i$ is the centrality of the $i^{th}$ node then

$$C_i = \frac{1}{\lambda} \sum_{j \in A(i)} C_j = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} C_j,$$

(4.a)

where $j \in A(i)$ is the set of nodes connected to the $i^{th}$ node.

The vector form of the above equation is simply

$$C = \frac{1}{\lambda} AC,$$

(4.b)

where $\lambda$ is an eigenvalue of $A$ corresponding to the eigenvector $C$.

### 2.1.3 Rescaled Centrality

Our formal definition of EIN suffers from one major coincidental error. In an EIN, two investors may appear to be linked because they happen to trade in the same stock within the same interval just by random chance. This issue becomes more difficult to disentangle with noise traders who trade in random directions for many different reasons. One way to manage the issue is to empirically set the value of $M$ to a higher threshold. But this might also penalize genuine connections. Another way is to define a centrality measure that somehow mitigates this concern in an effective way. Rescaled Centrality, which is basically defined as the ratio of eigenvector centrality and degree centrality mitigates the concern by penalizing noise traders for their increased measure of degree centrality.

### 3 Profitability Measures

Profitability measures are an important tool to quantify information content embedded in a signal. Continuing in this spirit, we say that if each investor $i$ executes a trade $k$ consisting of $N_{ik}$ shares with price $N_{ik}$, then the trading quantity is defined as

$$Q_{ik} = N_{ik} \times P_{ik}.$$  

(5.a)

The vector $Q_i$ of all trades executed by the investor $i$ during the whole year and is given by

$$Q_i = \sum_k Q_{ik}.$$  

(5.b)

Measuring returns follows from Barber et al (2009). Thus, for each trade $k$ over a time span of $\Delta T$, the return is defined by

$$\mu_{ik} = sign \left( \frac{P_{t+\Delta T} - P_t}{P_t} \right),$$

(6.a)

where $sign$ depends on the direction of the trade, $P_t$ is the price of the stock at time $t$ and $P_{t+\Delta T}$ is the closing price of the stock $\Delta T$ days later.

The total return to the investor $i$ is the value weighted average return of all the trades executed
during the year.

\[ \mu_i = \text{sign} \sum_k \mu_{ik} Q_{ik} / \sum_k Q_{ik}. \]  \hspace{1cm} (6.b)

However it may happen that the investor is just lucky to trade in the direction of the market without any valuable information. In order to take this effect into consideration we define \( \mu_{ik}^e \), as the excess return above market for trade \( k \)

\[ \mu_{ik}^e = \text{sign} \frac{p_{t+\Delta T} - p_t}{\frac{p_{M}}{t+\Delta T}} = \frac{p_{t+\Delta T} - p_t}{p_t}, \]  \hspace{1cm} (6.c)

where \( P_M \) is the market value of the ISE-100 index.

Redefining, (6.b) to account for market movement, we have the total excess return to the investor \( i \)

\[ \mu_i^e = \text{sign} \sum_k \mu_{ik}^e Q_{ik} / \sum_k Q_{ik}. \]  \hspace{1cm} (6.d)

3.1 Forward Pricing

To analyze the scenario in which the traders were assumed to delay their trades by \( \Delta \tau = 1 \) day, we modified 6.a and 6.c. Thus, for each trade \( k \) over a time span of \( \Delta T + \Delta \tau \), the return is defined by

\[ \mu_{ik} = \text{sign} \frac{p_{t+\Delta T+\Delta \tau} - p_{t+\Delta \tau}}{p_{t+\Delta \tau}}, \]  \hspace{1cm} (7.a)

and the excess return is given by

\[ \mu_{ik}^e = \text{sign} \frac{p_{t+\Delta T+\Delta \tau} - p_{t+\Delta \tau}}{p_{t+\Delta \tau}}. \]  \hspace{1cm} (7.b)

3.2 Realized Returns

Since we had access to individual transaction records, we constructed a measure based on actual realized returns. The main reason to build such a measure was to test the robustness of our results. We kept track of inventory of each stock maintained by investor \( i \). So if each investor \( i \) executes a trade \( k \) consisting of \( N_{ik} \) shares with price \( P_{ik} \), then from the buying perspective, the trading quantity can be defined as

\[ Q_{ik} = N_{ik} \times P_{ik}. \]  \hspace{1cm} (8)

Now from the selling perspective, the weighted average cost \( \eta \) of trade \( k \) is then given by

\[ \eta_{ik} = \frac{\sum_k Q_{ik}}{N_{ik}}. \]  \hspace{1cm} (9)
The realized return for each trade \( k \) from the sellers perspective is then
\[
\mu_{ik} = \frac{P^t - \eta_{ik}}{P^t},
\] (10.a)
and the total realized return to investor \( i \) is the value weighted average realized return of all the trades executed during the year is
\[
\mu_i = \frac{\Sigma_k \mu_{ik} Q_{ik}}{\Sigma_k Q_{ik}}.
\] (10.b)

Now the market adjusted cost of inventory at time \( t \) is the excess cost
\[
\zeta_t = \frac{\Sigma_k Q_{ik} P^t}{P^t M} - \frac{\Sigma_k Q_{ik}}{P^t M - 1}.
\] (11)

The excess realized return over market for trade \( k \) is
\[
\mu_{ik}^e = \frac{P^t - \zeta^t_{ik}}{\zeta^t_{ik}},
\] (12.a)
and the total excess realized return to the investor \( i \) is
\[
\mu_i^e = \frac{\Sigma_k \mu_{ik}^e Q_{ik}}{\Sigma_k Q_{ik}}.
\] (12.b)

## 4 Data Description

The Istanbul Stock Exchange (ISE) is an autonomous professional organization founded in 1986. The exchange deals with trading in a wide variety of financial instruments like stocks, bonds, certificates etc. The ISE is an order driven, multiple-price, and continuous auction market without specific market makers. The stock trading activities are carried out in two separate sessions i.e. 0930-1200 and 1400-1630 hours. The market capitalization of 201 billion USD ranked ISE 19th across the world in the year 2005. The vast majority (94.7%) of the institutional investors at the start of 2005 were foreigners. The trading system at the ISE is completely automated.

The dataset used in the analysis consists of over 43 million transactions that were executed from January 2005 to December 2005. There were 313 stocks that were traded during this period by 580,142 active account holders. Of these active accounts, 489 belonged to institutional investors and the remaining, 579,673 belonged to individual investors. Each trader is identified by its unique identification number and each transaction constituted of the following information:

a. Timing of the trade.
b. Stock Ticker.
c. Price per share of the stock.
d. Number of shares traded.
e. Account ID of the buyer and the seller.
f. Account type of the buyer and the seller (individual or institutional).
g. Type of Trade (Long or Short).

Furthermore, prices for transactions were adjusted for stocks splits and dividends payments in our analysis.

5 Analysis of Data

In order to analyze the data, a preliminary code for the project was written in C++ by Gil Shallom at the University of California, Berkley. It was later extended by subsequent research assistance work by Niclas Eriksson and Ludvig Larruy[4] and then later by Andreas Kieri and Joakim Saltin[6] here at the Division of Scientific Computing of the Department of Information Technology at Uppsala University. The analysis in the current project further builds on the work done by the previous participants. The current project was run on Tintin, a high performance computing cluster available at the UPPMAX facility of Uppsala University.

6 Computational Issues

6.1 Calculating realized profits

Since we have to deal with around 43 million trade transactions, it is important to design an efficient algorithm to compute realized returns. The main problem here is to deal with how to maintain the inventory of each stock held by the investors. One way is to implement LIFO or FIFO lists. In this case we would need to keep track of different prices. This would have made computational and memory requirements a bit on the higher side. A simpler approach to calculate cost of inventory is the weighted average cost method. The method is simple in the sense that instead of maintaining a list of transactions for each stock maintained by each investor, we only need to take care of the total trading quantity of each stock and its total monetary value. So we have matrices $I_{is}$ and $N_{is}$ to represent the total trading quantity and the total monetary value of each stock $s$ held by investor $i$. Short sales are ignored here. The matrices are adjusted for stock splits and dividend payments.

For each trade record, we match if the seller has sufficient inventory. If not, the extra inventory is ignored. For instance, if the current trade being read consists of $k$ shares of stock $s$, sold by seller $u$ to the buyer $v$ at price $P^t$, then the number of shares to take into consideration is given

$$\lambda = \min(k, I(u, s)),$$

and the inventory is updated as

$$I(u, s) = I(u, s) - \lambda,$$  \hspace{1cm} (14.a)  

$$I(v, s) = I(v, s) + k.$$  \hspace{1cm} (14.b)
The cost is measured by
\[ c_{uks} = \frac{N(u, s)}{I(u, s)}. \] (15)

Finally, the profit \( \pi \) for the seller \( u \) is recorded as
\[ \pi(u, k) = \pi(u, k) + \lambda(P^t - c_{uks}). \] (16)

The implementation of the scenario where we needed to check what happened when the traders traded with a delay of one day is a pretty straightforward implementation of equations 7.a and 7.b.

### 6.2 Finding information events

The basic idea is to find the number of times a stock return exceeds some threshold in a consecutive order. This is done through a greedy type algorithm. We now outline the algorithm that we worked with for the purpose.

Let \( n \) be the total number of trading days in the year.

Let \( r \) be a vector containing returns for the year.

Let \( \alpha \) be some threshold return that is considered an information event.

Let \( p \) be a scalar defined over the interval \( I = [r_k, r_{k+1}, ..., r_{k+n}] \) such that
\[ p = r_k + r_{k+1} + ... + r_{k+n} > \alpha \]

and \( k \) is the minimum day in the interval \( I \).

Let \( m \) be a vector containing the minimum day in the interval \( I \) i.e. each \( k \) in each \( I \) if \( |p| > \alpha \).

#### Algorithm I

for \( j = 1 \) to \( (n-1) \)
\[ p=0; \]
for \( k = j \) to \( n \)
\[ p=p+r(k) \]
if \( |p| > \alpha \)
\[ m(j)=k \]
break
end
end

The proof of feasibility of this algorithm is by contradiction; any other choice of the next interval than the minimum \( k \) would lead to fewer points to generate new intervals over.
7 Investor centrality, information diffusion and timing of the trades

We verify that the timely processing of information signals is a characteristic attribute of central investors. To further analyze this point, we sort investors according to their centrality and then divide them into high and low centrality groups using median centrality as the cutoff. We then study the impact of delayed trades around information events. Using excess market return as a measure of performance our regression results show that delaying trades of high centrality investors by one day leads to a decrease in their excess return by 0.21%. Comparatively, low centrality investors show increase in their excess return by 0.26% when their trades were delayed by a day. Furthermore, the t-stats of these tests came out to be highly significant. Clearly, high centrality investors lose their performance to low centrality investors when trades are delayed around information events.

Further, we study the relationship between centrality and returns with delayed trades. Our regression results show that delaying trades by a day decreases the centrality coefficient from 0.0027 to 0.0019. We also observe that the economic significance of one standard deviation increase in centrality reduced to 0.42% from 0.57%. In fact, delaying trades by a day erodes around 30% of positive relationship between centrality and returns. Clearly, the relationship between centrality and returns becomes weaker when trades are delayed.

8 Information events and information diffusion

We perform additional analysis to study behavior of central investors around information events. We find that there are many unobserved information events in the market which can potentially provide significant performance boost to central investors. Furthermore, we observe that central investors have information advantage before information diffuses into the market.

8.1 Frequency of information events in the stock market

In the previous two instances of the projects related to the current paper, it has been verified that centrality of an investor is determined both by his ability to trade earlier than his neighbors in general, and to special information events in particular. However, it can be argued on certain grounds that comparing returns obtained from these two studies may not be appropriate despite being consistent. For instance, a point can be made that the study of special events constituted of such news that there were high chances of big price change in the underlying stock. Another point that can be made is that special events invariably find their way to a greater audience and as a result central investors may not necessarily always have information leverage, thereby reducing their advantage to make profits on such events. To counter these assertions, we identified 3,291 events in 2005, where the absolute return of the stock was greater than 15% within a 3-week period. This finding further strengthens our argument that there are many unobserved information events
in the market and central investors are well positioned to take advantage of them and increase their potential profits.

8.2 Study of high and low information periods of the year

Central investors have an information advantage and studies of known information events confirm this. We also now know that while many information events may be unobservable, they are indeed not rare. As an additional test, we use earnings announcements as proxy to unobservable information events. In the table below we list the number of earnings announcements each month during the year 2005.

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>41</td>
<td>178</td>
<td>109</td>
<td>182</td>
<td>87</td>
<td>19</td>
<td>176</td>
<td>120</td>
<td>55</td>
<td>179</td>
<td>80</td>
</tr>
</tbody>
</table>

We see that the earnings announcements during the months of March, May, August and November are higher than in the other months of the year. So in our analysis, we split the months over the year 2005 into two lists; one consisting of months in which there were many earnings announcements (March, May, August, Nov) and the other one consisting of months in which there were few such announcements (January, February, April, June, July, October, December). A natural hypothesis to test here is that during the months of increased information flow, central traders should have clear information advantage over their neighbors. We compare regression results based on profits obtained during the months of increased information flow with the remaining months.

Table 1 displays results from regressions of value-weighted returns (Panel A) and value-weighted excess returns (Panel B). The first row displays coefficients while the second row displays the t-statistics. Columns 2-4 display results in high-information months while columns 5-7 display the regressions in the low-information months. We do not assume error terms to be normally distributed, so we therefore also perform OLS regression that is robust to heavy tailed error terms (columns 3, 5). In addition, we also perform an iteratively re-weighted least square (using Ramseys E-function) for multivariate regressions (columns 4, 8). The variable $\mu$ is the value-weighted return for all trades of an investor for the entire year assuming a $\Delta T=30$-day holding period for each trade and $\mu_e$ is the excess return which is simply value-weighted return adjusted for market return (ISE 100 index return). Centrality and degree is calculated using $\Delta t = 30$ minute window. The variables $\Delta \mu$ and $\Delta \mu_e$ give the economic significance of the results by showing the change in results returns (excess returns) given one standard deviation increase of the variable in uni-variate regressions and centrality or rescaled centrality in multivariate regressions, all else being equal. Further data of investors in the bottom and top two percentiles of connectedness is discarded to avoid influence of outliers on the results.

Looking at Panel (A) of Table 1, we observe that centrality coefficient during the months of high information flow is 0.019 but decreases to 0.0008 during the months of low information flow. Furthermore, the economic significance of centrality during the periods of high information flow is 3.6% compared to just 0.2% during the months of low information flow. The results are similar in case of excess returns (Panel B). This further strengthens our claim that central investors have
Table 1: High- and low-information periods.

### A. Returns

<table>
<thead>
<tr>
<th></th>
<th>High-information periods</th>
<th>Low-information periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS $t$-error Ramsey</td>
<td>OLS $t$-error Ramsey</td>
</tr>
<tr>
<td>Centrality ($c$)</td>
<td>0.019 0.015 0.019</td>
<td>0.0008 0.0025 0.0014</td>
</tr>
<tr>
<td></td>
<td>$&gt; 20$ 12.5 $&gt; 20$</td>
<td>1.7 2.6 2.9</td>
</tr>
<tr>
<td>Degree ($d$)</td>
<td>-0.016 -0.013 -0.017</td>
<td>0.0015 -0.0018 0.0007</td>
</tr>
<tr>
<td></td>
<td>$&lt; -20$ -10.7 $&lt; -20$</td>
<td>3.0 -1.9 1.3</td>
</tr>
<tr>
<td># of trades ($n$)</td>
<td>0.0064 0.0059 0.0064</td>
<td>0.0007 0.0006 0.0007</td>
</tr>
<tr>
<td></td>
<td>$&gt; 20$ $&gt; 20$ $&gt; 20$</td>
<td>6.8 3.0 7.2</td>
</tr>
<tr>
<td>Quantity ($q$)</td>
<td>-0.0054 -0.0049 -0.0054</td>
<td>-0.0010 -0.0007 -0.0010</td>
</tr>
<tr>
<td></td>
<td>$&lt; -20$ $&lt; -20$ $&lt; -20$</td>
<td>-14.5 -5.0 -14.4</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.016</td>
<td>0.0031</td>
</tr>
<tr>
<td>$\Delta \mu$</td>
<td>3.6% 2.8% 3.6%</td>
<td>0.2% 0.5% 0.3%</td>
</tr>
</tbody>
</table>

### B. Excess returns

<table>
<thead>
<tr>
<th></th>
<th>High-information periods</th>
<th>Low-information periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS $t$-error Ramsey</td>
<td>OLS $t$-error Ramsey</td>
</tr>
<tr>
<td>Centrality ($c$)</td>
<td>0.010 0.0080 0.010</td>
<td>0.0025 0.0026 0.0029</td>
</tr>
<tr>
<td></td>
<td>19.6 7.9 19.8</td>
<td>5.9 3.0 6.8</td>
</tr>
<tr>
<td>Degree ($d$)</td>
<td>-0.010 -0.0088 -0.011</td>
<td>-0.0023 -0.0033 -0.0028</td>
</tr>
<tr>
<td></td>
<td>$&lt; -20$ -8.6 $&lt; -20$</td>
<td>-5.2 -3.8 -6.5</td>
</tr>
<tr>
<td># of trades ($n$)</td>
<td>0.0023 0.0017 0.0022</td>
<td>-0.0001 0.000003 -0.0004</td>
</tr>
<tr>
<td></td>
<td>19.3 7.0 17.8</td>
<td>-0.7 -2.3 0.05</td>
</tr>
<tr>
<td>Quantity ($q$)</td>
<td>-0.0019 -0.0013 -0.0018</td>
<td>-0.0001 0.00003 -0.0002</td>
</tr>
<tr>
<td></td>
<td>$&lt; -20$ -7.9 $&lt; -20$</td>
<td>-2.4 -2.3 -2.8</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0032</td>
<td>0.00004</td>
</tr>
<tr>
<td>$\Delta \mu$</td>
<td>1.9% 1.5% 1.9%</td>
<td>0.5% 0.5% 0.6%</td>
</tr>
</tbody>
</table>

information advantage and that information diffusion plays a significant role in the profitability of central investors.

### 9 Why some investors earn extreme returns than others?

Our analysis of earning announcements as proxy to unobserved events revealed that the profitability of central investors was higher during periods of high information diffusion. Further, in another analysis, we split the investors among high and low centrality groups and delayed the trades of investors belonging to high centrality group by a day around information events. We observed that central investors belonging to high centrality group lost their information advantage against the investors belonging to the low centrality group. The impact of this delay was that excess market return of high centrality investors decreased by about 0.21%. These observations help us to determine one more key to the information diffusion story i.e. why are certain investors more likely to earn higher profits while others are not. It is now clear that high centrality investors are more likely to earn higher profits compared to those with lower centrality, especially during the period of increased information diffusion.
Robustness tests and market microstructure issues

Several additional analysis were carried out in previous projects to support the information diffusion argument over alternate explanations of investors profitability such as algorithmic trading or other trading strategies, momentum and price impact from illiquidity, market microstructure effects etc. Here we employ a robustness test to further support information diffusion argument and to rule out potential of spurious correlation arising out of fixing the holding period to 1 month for each trade.

10.1 Market microstructure

Market microstructure concerns how institutional rules and trading behavior of market agents affect the prices and returns of assets in the market. There is a wide range of literature available on microstructure issues [3,7]. Microstructure effects, whether due to illiquidity, bid-ask spread, price impact, or non-synchronous trading work on smaller time scales, and hence can be ignored in cases of long investment horizon. However, in cases where market collusion between trading agents is a possibility, ignoring these effects may not be a good approach. Microstructure effects per se do not cause spuriousness of stock returns, as discussed in [1]. Spurious correlation across agents has been discussed in [2]. Thus we did an additional test to be sure that spurious correlation whether across time or across agents is not inflating our results.

10.2 Realized returns as a measure to profitability

One of the main driving forces behind trading decisions is information. But a trade position may not necessarily be closed even after the information is completely diffused into the price without other compelling reasons, e.g. liquidity needs. However, in order to measure information content embedded in trading decisions over short time horizon, we fixed the holding period to one month. But such an assumption has the potential of inducing spurious correlation in measured return of investors thereby inflating the significance of our results. To address this concern, we designed a robustness test using a new profitability measure based on actual realized returns. However, since the data spans only over one year, not all trades may be closed and therefore this new specification of the profitability measure reduced the number of trades that the measure is based upon. Also, the number of traders reduced to 322,766 since many investors did not close their trade at all during the period. The EIN is created using 30 minute window with $M = 3$. Furthermore for our analysis, we discard the data of investors in the bottom and top two percentiles in terms of connectedness to avoid the impact of outliers on our results.

First we regress returns and excess market returns based on fixed holding period of 1 month on degree, centrality, rescaled centrality, number of trades and trading quantity. Table 2 shows the results of the regression which we will refer to as our base test case. We then perform the same regression, but using actual realized returns and actual realized excess market returns instead. Table 3 shows the results of the second regression which we will refer to as our robustness case. The variables $\Delta \mu$ and $\Delta \mu^c$ give the economic significance of the results by showing the change in
### Table 2: Centrality and returns.

The table displays results from regressions of value-weighted returns (Panel A) and value-weighted excess returns (Panel B) on log centrality, log degree, log rescaled centrality, log number of trades, and log volume.

Results (excess returns) given one standard deviation increase of the variable in uni-variate regressions and centrality or rescaled centrality in multivariate regressions, all else being equal. Columns 1-5 in Table 2 and Table 3 are the results of uni-variate regressions while columns 6-7 represent OLS multivariate regressions. As in our base test case, we do not assume error terms to be normally distributed, so we, therefore, also perform OLS regression that is robust to heavy tailed error terms (columns 8-9). In addition, we also perform an iteratively re-weighted least square (using Ramseys E-function) for multivariate regressions (columns 10-11). The univariate regression results of both the tables are similar except that the sign of centrality in our robustness test goes the wrong way. In the multivariate case of our robustness test, rescaled centrality and centrality come up with the positive sign and are statistically significant while degree shows up the wrong sign. This is similar to our base test case results of Table 2. The additional robustness test using actual realized returns thus further confirms our base test case that centrality is indeed more important than degree in determining returns. Indeed, the results of both the tests do not contradict each other.

Furthermore, we observe that both the measures give similar economic significance of returns.
Table 3: Realized returns. The table displays results from regressions of value-weighted realized returns (Panel A) and value-weighted excess realized returns (Panel B) on log centrality, log degree, log rescaled centrality, log number of trades, and log volume.

In case of returns based on one month holding period, a one standard deviation increase in centrality translated into 0.7%-1.8% return depending on the type of regression used. Similarly, returns based on actual realized returns have a slightly wider range of 0.1%-2.9%. Thus, the new tests further confirm the importance of centrality in determining returns.

Given the strong correlations between degree and centrality, our new test further works as a second robustness test to rule out the effect of multicollinearity on the results. After analyzing the results obtained by two different profitability measures, we can say that they do not contradict each other, thus lending support to our main tests described by Table 2. It may be noted that while not all the variables in the two tests are significant or have the right sign, this is not a major issue here since while doing a number of tests, rarely do all the coefficients come to be significant or point in the right direction.

The assumption of fixed holding period may cause spurious mechanical relationship between centrality and returns to inflate the significance of coefficients in our base test case. Our robustness test alleviates this concern as well. Furthermore, since this specification is based on realized returns, it also addresses the concern that the positive relationship between centrality and returns can be a direct consequence of price impact or some other microstructure effects. Also, since
microstructure effects, whether due to illiquidity or adverse selection work on smaller time scales, our assumption of one month holding period is our base test case is large enough to mitigate such concerns. Furthermore, we have no reason to believe that the specification based on actual realized returns is affected by microstructure effects.

11 Robustness of alternative profitability measures

In our robustness tests, we found that the profitability measure based on fixed holding period and the one based on actual realized returns do not contradict each other. Indeed, we can now conclude here that no matter how we define the profitability measure, conclusions based on them do not change i.e. profitability measures are robust to any intuitive and logical alternative construction.

12 Limitations of the current work

EIN is just another way to understand the role of differential information on the governing dynamics of investor behavior and their subsequent success and failures in stock market. However, there can still be alternative explanations, mechanisms and motivations that could in theory generate an EIN type network. Notwithstanding, such explanation or motivations must satisfy the four main assertions as discussed before and several such explanations fail to satisfy at least one of them. Also, EIN does not directly identify information channels such as word of mouth communications or exchange of opinions through different media or ability of certain investors to create consistent information mosaic, but rather tries to substantiate the mechanism indirectly. Also, EIN cannot differentiate between type of information event i.e. if it is based on a valid fundamental reasoning or driven by some rational or irrational sentiments or biases.

13 Concluding remarks

Our robustness analysis of the previous work shows that the main conclusion that central investors outperform peripheral investors is robust to alternative specification and measures of agents profitability. This is an important result because we do not directly observe agents network positions and therefore aside from the spurious correlation issues across time the question of spurious correlation across agents was also a concern. According to market micro-structure literature, cases where agents are expected to collude with each other, it was not a good idea to ignore microstructure effects. And in an EIN, agents are in a way colluding with each other through private signals. Choosing 30 day holding period may be considered a long enough time to ignore spurious correlation across time. Furthermore, by forcing traders to liquidate after a fixed time horizon, we embedded a quasi-mechanical relationship in their returns. Hence, this might lead up to artificial agent-agent correlation. Our purpose for using actual realized returns is that they are based on actual reason/information rather than a mechanical construct. Therefore, with actual realized returns our aim was to alleviate the concern of spurious correlations both across agents and across
time. Indeed the results hold up when actual realized profits are measured instead of using a fixed holding period of one month.

Our study of the effects of delaying the trades gave us strong evidence that central agents in an EIN have information advantage and they act on it in a timely manner. As much as 30% of the positive relationship between centrality and returns of central investors is lost as a result of delaying the trades by a day.

Since the amount of information diffusing into the market at any given point time is unobservable, we used earnings announcements for each month as a reasonable proxy. Our study found strong evidence that profitability is higher in months with much information diffusion. We also found evidence that there are many unobservable events in the market which the central trades may take advantage of. In fact, 3,291 information events were identified for the year 2005 under the assumption that any return above 15% over consecutive days (3-week period in our case) corresponds to an information event in the market.

All our results support that information diffusion is an important determinant of profitability of central investors in a market network.
14 Acknowledgements

I would like to thank my supervisor Johan Walden for guiding me throughout this very interesting project and giving me his regular feedbacks on my questions. The UPPMAX support team has been crucial in providing me answers to questions related to the Tintin cluster. So a big thanks to them. I would also like to thank Ayesha Suboor for proof reading my initial report and giving her valueable comments. I am extremely grateful to Lina von Sydow for her help in this project and for reading my report throughly and suggesting changes to make it better. Finally, I would like to thank Per Lötstedt, Bertil Gustafsson, Lina von Sydow, Victor Shcherbakov, Slobodan Milovanovic, Josef Höök for being in the audience of this thesis presentation.
References


