ESSAYS ON TRADING AND FINANCIAL ECONOMETRICS

Jiangmin Xu

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
Adviser: Yacine Aït-Sahalia

June 2014
Abstract

This dissertation studies trading and investment in financial markets through the lens of financial econometrics. Chapter 1 develops a continuous-time model of the optimal strategies of high-frequency traders (HFTs) to rationalize their pinging activities – defined as rapid submissions and subsequent cancellations of limit orders inside the bid-ask spread. The current worry is that HFTs ping inside the spread to manipulate the market. In contrast, the HFT in my model uses pinging to control inventory or to chase short-term price momentum without any learning or manipulative motives. I use historical message data to reconstruct limit order books, and characterize the HFT’s optimal strategies under the viscosity solution to my model. By gauging the model’s implications against data, I show that pinging is not necessarily manipulative and is rationalizable as part of the dynamic trading strategies of HFTs.

In Chapter 2, joint with Harrison Hong, we use overdispersed Poisson regression models to study social networks in finance. We count an investor’s social connections in different cities as proportional to the number of stocks held by this investor that are headquartered in those cities. When connections are formed in an i.i.d. manner, the count of such connections in any city follows a Poisson distribution. Using data from institutional investors’ holdings, we find instead overdispersion for a number of cities like San Jose and San Diego, which suggests that investors have non-i.i.d. propensities to be connected to these cities. Overdispersed cities have a large number of graduates from local universities who work in the fund industry. Managers with relatively high non-i.i.d. propensities to have social contacts significantly outperform other managers.

In Chapter 3, I propose a continuous-time model for the joint stochastic process of asset price and trading volume to study the transmission mechanism from changes in trading volume to price movements at the high-frequency level. A GMM-based estimation procedure is developed based on the model’s closed-form moment conditions.
I estimate the model on real-world high frequency financial data and find that, jumps in volume have a strong cross-excitation effect on jumps in price. Other implications of the model are also discussed.
Acknowledgements

I would like to express my deepest gratitude to my advisor Yacine Aït-Sahalia for his advice, patience, encouragement and invaluable insights throughout all these years. It has been an honor and a privilege to be his student, and a wonderful experience to grow under his academic guidance.

I also want to express my sincere gratitude to Harrison Hong for his tremendous support, enlightenment and wisdom. It has been a great pleasure to work with and learn from him, and I am very thankful for all his time, attention and input.

Moreover, I would like to thank Christopher Sims, Wei Xiong, Valentin Haddad, David Sraer, Markus Brunnermeier for their valuable suggestions and helpful comments about my work. Thank you also to all my friends at the Economics Department for many useful conversations. Thank you to Cheng Sun, for her infinite patience and constant support. Thank you to all my non-econ friends, for making Princeton feel like home.

Special thanks to Karen Neukirchen for her extraordinary administrative support and contributions to the Bendheim Center for Finance. Thanks also go to Kathleen DeGennaro and Laura Hedden for their exceptional effort in administering the graduate program, and to Jennifer Bello for her administrative help.

In addition, I would like to thank NASDAQ for providing me with the access to their TotalView-ITCH database, as well as seminar participants at Guanghua School of Management of Peking University, Hong Kong University of Science and Technology, Princeton University, Shanghai Advanced Institute of Finance, the Princeton-QUT-SMU conference and Warwick Business School for help comments on my work.

Finally, thank you to my parents, to whom I owe everything.
To My Parents
# Contents

Abstract ................................................................. iii

Acknowledgements ....................................................... v

1 Optimal Strategies of High Frequency Traders 1

1.1 Introduction ......................................................... 1

1.2 Related Literature .................................................. 4

1.3 The Model .......................................................... 7

1.3.1 Processes of Limit Order Book Characteristics .............. 7

1.3.2 The HFT’s Trading Strategy .................................... 10

1.3.3 Order Execution Processes .................................... 12

1.3.4 Comments on the Model ....................................... 13

1.4 Equilibrium and Solution Method ................................. 16

1.4.1 The Objective of the HFT ...................................... 16

1.4.2 Definition of Equilibrium and Dynamic Programming Equations 20

1.4.3 Numerical Solution ............................................ 24

1.5 Data and Estimation ............................................... 27

1.5.1 Data .......................................................... 28

1.5.2 Estimation ..................................................... 29

1.6 Computation and Simulation Results ............................... 31

1.6.1 Optimal Strategy Profiles .................................... 31

1.6.2 Percentage Attributions of Optimal Strategies ............... 40
2 Count Models of Social Networks in Finance

2.1 Introduction ..................................................... 53
2.2 The Model ......................................................... 61
  2.2.1 Notation ..................................................... 62
  2.2.2 The Null Model ............................................. 63
  2.2.3 The Overdispersed Model .................. 63
  2.2.4 Likelihood Function .................................. 64
2.3 Data and Estimation ........................................... 66
  2.3.1 Data ....................................................... 66
  2.3.2 Rolling Estimation .................................. 68
2.4 Main Results ................................................... 69
  2.4.1 Transformation Parameters .................. 69
  2.4.2 Gregariousness Parameters ................ 70
  2.4.3 Group Size Parameters ......................... 70
  2.4.4 Overdispersion Parameters ................ 72
  2.4.5 Robustness Checks: Controlling for Local Bias and Fund Styles 73
2.5 Networks, Demographics and Performances .................... 74
  2.5.1 What Explains Overdispersions? ............ 74
  2.5.2 Managerial RPC Measure and Demographics 80
  2.5.3 Overdispersion in Counts vs. Portfolio Weights 84
  2.5.4 Managerial RPC and Fund Performance .......... 85
2.6 Extensions ...................................................... 88
## 2.6.1 A General Increasing Transform

88

## 2.6.2 A Censored Model

92

## 2.7 Conclusion

93

## 2.8 Appendix

95

### 2.8.1 Mutual Fund Results for Industry Groups

95

### 2.8.2 Retail Investor Results for City Groups

100

### 2.8.3 Retail Investor Results for Industry Groups

105

## 3 Asset Trading and Price Movement at High Frequency

109

### 3.1 Introduction

109

### 3.2 The Dynamics of Asset Price and Volume

113

#### 3.2.1 Mutually Exciting Jump Processes

113

#### 3.2.2 Joint Process of Price and Volume

114

#### 3.2.3 Further Specifications on Jumps

115

#### 3.2.4 Comments on the Model

117

### 3.3 Closed-form Moment Expressions

117

#### 3.3.1 Moment Functions of the Model

118

#### 3.3.2 Covariance Density Matrix of the Hawkes Process

119

### 3.4 Estimation

120

#### 3.4.1 The GMM Procedure

121

#### 3.4.2 Reduction of Parameter Space

123

### 3.5 Empirical Analysis

124

#### 3.5.1 Description of Data

125

#### 3.5.2 Empirical Results

125

### 3.6 Conclusion

127

### 3.7 Appendix

129

#### 3.7.1 Stationarity of the Bivariate Hawkes Process under Exponential Decay

129
3.7.2 Explicit Expression for the Covariance Density Matrix . . . . 130

Bibliography 133
Chapter 1

Optimal Strategies of High Frequency Traders

1.1 Introduction

A recent and ongoing heated debate concerns high-frequency traders and high-frequency trading activities (HFT stands for high-frequency trader/trading thereafter). The interest in this subject has grown significantly after the Flash Crash, because HFTs appear as a black-box mystery to the general public as well as to the academic world.\(^1\) One type of HFT activity that has attracted a great deal of attention due to HFTs’ speed advantage is the so-called pinging activity. Pinging, or the most aggressive fleeting order activity, is defined as the submission of limit orders inside the bid-ask spread that get cancelled very shortly.\(^2\) These activities

---

\(^1\)For a comprehensive review of the Flash Crash, see the study by Kirilenko, Kyle, Samadi, and Tuzun (2011). See also Easley, de Prado, and O’Hara (2011)’s examination on the same subject.

\(^2\)Existing studies (e.g. Hasbrouck and Saar (2009)) have used the term “fleeting orders” to denote submissions and subsequent quick cancellations of limit orders in general. Therefore, to be consistent with the literature, I use the term “pinging” throughout this paper to denote the most aggressive fleeting order activities, i.e. fleeting orders that take place inside the bid-ask spread.
occur in the scale of seconds or milliseconds with extremely low latency, which is “the hallmark of proprietary trading by HFTs” (Hasbrouck and Saar (2013)).

Regulators have expressed concerns over such pinging activities, the main one being that pinging used by HFTs might be manipulative. As pointed out in a concept release of the Securities and Exchange Commission (SEC), HFTs could use pinging orders to detect and learn about hidden orders inside the spread. Hidden orders are limit orders completely non-observable to other market participants. They have become increasingly popular in the past 5 to 10 years and are allowed by many stock exchanges around the world nowadays. This learning would enable HFTs to ascertain the existence of potential large trading interest in the market. Consequently, they would be able to trade ahead and capture a price movement in the direction of the large trading interest.

This chapter aims to rationalize the pinging activity levels observable in the data through a theoretical setup without manipulative elements. It develops a continuous-time model of the optimal trading strategies of HFTs absent any learning or strategic feedback effects. The model exploits two well-known forces in the existing literature: inventory control (e.g. Ho and Stoll (1981)) and trend chasing (e.g. Hirschey (2013)).

To achieve my purpose, I first incorporate the existence of hidden orders inside the bid-ask spread into a continuous-time model along the lines of Ho and Stoll (1981). As a result, the HFT in my model would ping for hidden orders inside the spread as a cheaper way to control his inventory compared to using market orders. This would produce pinging orders that execute against hidden orders. However, with inventory control as the sole motive for pinging, it is not possible to obtain a large

---

3 To give a illustration of the speed with which pinging takes place, Hautsch and Huang (2011) find that the median cancellation time is below one second for limit orders submitted inside the spread on NASDAQ.

4 According to Hautsch and Huang (2012), on average over 20% of trades on NASDAQ are executed against hidden orders in October 2010.

5 SEC concept release on equity market structure, January 21, 2010.
number of cancelled pinging orders at the same time. The reason is that when the HFT uses pinging to control his inventory, he intends to fill his pinging orders and is not incentivized to cancel many of them.

In order to resolve this problem, I then introduce short-term price momentum into the model as a channel for cancelled pinging orders to occur. The momentum is modelled through the predictability of depth imbalance on the direction of price movements. When the HFT sees large depth imbalance and anticipates a likely directional price move, he could use pinging orders as directional bets to chase the price momentum. Moreover, if there is a subsequent large change in depth imbalance, the HFT would cancel his pinging orders and adjust his strategy according to the variation in momentum. Therefore, the model will now give rise to pinging as well as cancellation due to the HFT’s momentum-chasing behaviors.

The model is solved numerically due to its complexity, and I use historical order book message feed data from NASDAQ to reconstruct limit order books and estimate model parameters. The optimal HFT strategies are characterized based on the viscosity solution to my model and the parameter estimates. There are two main findings. First, for stocks whose order books have high depths with relatively stable movements and whose spreads tend to be narrow, approximately 20% of the HFT’s optimal strategies are attributable to pinging. On the other hand, for stocks that tend to have low order-book depths, volatile order-book movements and wide spreads, pinging accounts for nearly 50% of the HFT’s optimal strategies. These pinging percentages from the model are proven to match most of the observable pinging activity levels from the data. Thus it implies that the pinging activities occurred in reality can be mostly rationalized by my model without any learning or manipulative components. Second, I demonstrate that the mechanism of momentum chasing plays a much more important role than that of inventory control in rationalizing pinging activity.

---

I am grateful to NASDAQ for providing me with the access to their historical TotalView-ITCH real-time limit order book message feed database.
activities for low-depth and wide-spread stocks. In contrast, the two mechanisms carry around similar weights in rationalizing pinging activities for high-depth and narrow-spread stocks. Furthermore, both mechanisms are shown to be necessary for the model to rationalize the pinging occurrences found in the data.

Beyond pinging rationalization, my model also generates a couple of additional auxiliary predictions. They concern about the directions of pinging activities and the frequencies of cancelled pinging activities with regard to the depth imbalance of order books. These predictions are also found to be largely consistent with the data, which further suggests that pinging is rationalizable as part of the dynamic trading strategies of HFTs.

In what follows, I build a continuous-time, partial equilibrium model that captures a wide range of HFT strategies and explore their empirical content and implications. I begin in Section 1.2 by discussing the relations between my model and the existing literature. Section 1.3 lays out the structure of my model. The model’s equilibrium and my numerical solution are presented in Section 1.4. Section 1.5 is devoted to parameter estimations. I then discuss the main findings of this paper in Section 1.6 and draw out the model’s auxiliary predictions and evaluate them on empirical data. Finally, Section 1.7 concludes. All proofs are relegated to the Appendix.

1.2 Related Literature

There is a long line of empirical studies (see, e.g., Brogaard, Hagströmer, Nordén, and Riodan (2013), Brogaard, Hendershott, and Riodan (2013), Hagströmer and Nordén (2013), Hasbrouck and Saar (2013), Hendershott and Riodan (2013), Hendershott, Jones, and Menkveld (2011) and Menkveld (2013)) that have tried to understand the effects of high-frequency and algorithmic trading activities on market quality. This
line of research is economically important as HFT firms account for over 50% of all US equity trading volume in 2012.  

On one hand, the majority of the empirical studies show that, on balance, HFTs are beneficial for market quality. For instance, Brogaard, Hendershott, and Riodan (2013) find that HFTs enhance price discovery and market efficiency on NASDAQ, with prices reflecting information more quickly. Hasbrouck and Saar (2013) also show that increased HFTs’ low-latency activities are associated with lower posted and effective spreads, lower short-term volatility and increased market depth. Additionally, Hendershott, Jones, and Menkveld (2011) illustrate that algorithmic tradings improve liquidity and make quotes more informative for NYSE stocks.

On the other hand, there are empirical analyses demonstrating that some of the HFT strategies are speculative/anticipatory in nature. For example, Hirschey (2013) finds that HFTs on NASDAQ tend to anticipate future order flow and trade ahead of it, through aggressively taking liquidity from the market. Moreover, Kirilenko, Kyle, Samadi, and Tuzun (2011) show that HFTs exacerbated market volatility during the Flash Crash by trading in the direction of the downward price spiral. In addition, Baron, Brogaard, and Kirilenko (2012) find that aggressive, liquidity-taking HFTs earn short term profits at the expense of other market participants.

This paper focuses on the optimal trading strategies of HFTs and their pinging activities in particular. It builds on the work of Ho and Stoll (1981) – HS81, Avellaneda and Stoikov (2008) – AS08, and Guilbaud and Pham (2013) – GP13. HS81 first introduced, in a continuous-time partial equilibrium model, a market maker who optimally chooses the bid and ask prices of his limit orders to maximize his expected utility of (terminal) wealth. The market maker uses limit orders only, so that inventory can only be managed through limit orders. AS08 recast HS81 in a modern HFT setting by introducing the trading environment for HFTs not present in HS81 – the

---

limit order book. However, an HFT in AS08 is still a pure market maker, since he uses only limit orders to maximize his utility and control his inventory.

GP13 brought in market orders to the framework of HS81 and AS08. In their model, an HFT can trade via both limit orders and market orders. The HFT can post limit orders at the best quotes or improve the quotes by one tick. Otherwise he can use market orders instead to change his inventory instantaneously. The purpose of introducing market orders as another control mechanism is to better address the execution risk and the inventory risk faced by HFTs when they use limit orders only.

My model extends the continuous-time configuration of GP13 for HFTs and makes two contributions. Firstly, hidden orders are introduced into the limit order book, which are not present in either AS08 or GP13. Hence the HFT can ping for hidden liquidity when his limit orders improve the best bid/ask prices and are submitted inside the spread. This captures the idea of pinging strategies as identified in Hasbrouck and Saar (2009)’s empirical study. Secondly, I model the order book’s depth imbalance at the best quotes as a stochastic process that has an effect on the movement of the mid-prices and on the existence of hidden orders. Thus the HFT can use the depth imbalance as a (imperfect) signal to anticipate the likely price movement. Therefore, the HFT will utilize pinging or market orders to conduct directional/momentum tradings when necessary. This captures the idea of anticipatory strategies as identified in Hirschey (2013), which is not modelled explicitly in GP13.

Due to market orders being impulse (jump) controls in continuous time, the value function of the HFT in my model is not necessarily smooth and differentiable everywhere. Consequently, I will apply the viscosity solution technique to my model, which is similar to the one used in GP13. The viscosity solution concept, originally

8HS81 and AS08 consider only limit orders that are regular (continuous) controls in continuous time. Therefore, the value functions of HFTs in their models are smooth and differentiable at every point, so that they have standard Hamilton-Jacobi-Bellman equations and classical solution techniques apply.
introduced by Crandall and Lions (1983), is a generalization of the classical solution concept to Hamilton-Jacobi-Bellman equations (or partial differential equations in general) where value functions might not be everywhere differentiable. I will discuss the setup of my model and its solution in detail in the next two sections.

1.3 The Model

The economy is defined in a continuous time, finite horizon $\mathcal{T} = [0, T]$, with a single risky stock that can be traded in a limit order book (LOB). There is a high-frequency trader (HFT) who trades this stock using either limit orders or market orders. The LOB has a tick size $\delta$, so that prices of all orders are in multiples of $\delta$. Characteristics of the LOB, such as spread and mid-price, evolve stochastically and are exogenous to the HFT.

The formalization of my model is further specified as follows.

1.3.1 Processes of Limit Order Book Characteristics

To introduce the main characteristics of the LOB, I will fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying the usual conditions. Therefore, all stochastic processes and random variables are defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

To start with, the depth imbalance $F$ at the best quotes is defined to be the difference of the log of the size of the depth at the best bid price from that at the best ask price, so that $F > (\leq) 0$ means bid (ask) side imbalance. It follows an

---

A limit order of size $q$ at price $p$ is an order to buy or sell $q$ units of the asset at the specified price $p$; its execution occurs only when it meets a counterpart market order. A market order of size $q$ is an order to buy or sell $q$ units of the asset being traded at the lowest (for buy) or the highest (for sell) available price in the market; its execution is immediate. Given an asset, the best bid (resp. ask) price is the highest (resp. lowest) price among all active buy (resp. to sell) limit orders in the order book. The spread is the difference between the best ask price and the best bid price, which is strictly positive. The mid-price is the mid-point between the best bid and the best ask price. For a more detailed explanation of limit order book variables, please refer to the non-technical survey of Gould, Porter, Williams, McDonald, Fenn, and Howison (2013).
Ornstein-Uhlenbeck process with mean zero:

\[ dF_t = -\alpha_F F_t \, dt + \sigma_F \, dW_t, \quad (1.3.1) \]

where \( \alpha_F \) measures the speed of mean-reversion and \( \sigma_F \) is a constant volatility parameter.

Next, let

\[ S = \{ S_t \}_{t \geq 0} \quad (1.3.2) \]

denote the bid-ask spread of the LOB. It follows a continuous time Markov chain on the state space \( S = \{ \delta, 2\delta, 3\delta \} \), with a constant jump intensity \( \lambda^S \) representing the times when orders of participants in the market affect the spread. The probability transition matrix of \( S \) is denoted by \( \rho = (\rho_{ij})_{1 \leq i,j \leq 3} \), with \( \rho_{ii} = 0 \).

Furthermore, the mid-price \( P \) of the stock is assumed to evolve according to a pure jump process:

\[ dP_t = dJ_{1t} + dJ_{2t}. \quad (1.3.3) \]

The first component, \( J_{1t} \), has a constant jump intensity \( \lambda_1^J \), and jump sizes equal to \( \delta/2 \) with probability \( \psi_1(F_t) \) and \( -\delta/2 \) with probability \( 1 - \psi_1(F_t) \); while the second component, \( J_{2t} \), has a constant jump intensity \( \lambda_2^J \), and jump sizes equal to \( \delta \) with probability \( \psi_2(F_t) \) and \( -\delta \) with probability \( 1 - \psi_2(F_t) \), where the functions \( \psi_i : \mathbb{R} \mapsto [0, 1] \) are assumed to have the form

\[ \psi_i(u) = 1/(1 + \exp(-\beta_i u)), \quad \text{for} \; i = 1, 2, \quad (1.3.4) \]

with \( \beta_i \) being positive constants. In addition, \( J_1 \) and \( J_2 \) are independent.I allow the depth imbalance \( F \) to have an impact on the directions of mid-price jumps, since HFTs mostly seek information from the LOB itself to forecast price movements in

\[ \text{The best bid and best ask prices are thus} \; P_t - S_t/2 \; \text{and} \; P_t + S_t/2 \; \text{respectively.} \]
the millisecond environment. Depth imbalances capture liquidity pressures within the LOB. Therefore, it is informative about at which side of the book the observable depth is likely to become depleted first, resulting in a change in the mid-price. Hence it is a natural signal (albeit imperfect) for the HFTs to infer the direction of future price movements, based on the current state of the LOB.

The influence of the depth imbalance $F$ on the directions of the mid-price jumps is interpreted as the short-term price momentum in the model. If the parameter $\sigma_F$ is large so that $F$ is volatile, it leads to more frequent appearances of stronger momentum (signals). Thus the HFT would be more likely to engage in trend-chasing actions. This effect is amplified if the jump intensities $\lambda_1^J$ and $\lambda_2^J$ of the mid-price are higher, i.e. there are more realizations of price momentum.

Besides these standard LOB characteristic variables, orders with limited pre-trade transparency have become increasingly popular on electronic trading platforms recently. Major US stock exchanges, such as NASDAQ, NYSE and BATS, permit submissions of hidden limit orders into their LOBs. The price, size and location of these orders are completely concealed from other market participants, and they can be placed inside of the observable bid-ask spread without affecting the visible best bid and ask quotes. This interesting mechanism of hidden orders has created a vast number of order activities in markets as HFTs try to “ping” for hidden liquidity inside of the spread by posting aggressive “fleeting orders” that are cancelled a few instants later if not executed.

Consequently, to account for this important phenomenon, I will specify the existence of hidden orders in the following fashion. If the spread $S_t$ equals $2\delta$, the probability of hidden bid orders sitting at the mid-price $P_t$ is $\varphi_1(F_t)$ and the prob-

\[11\] Hautsch and Huang (2012) give a detailed empirical analysis of hidden orders on Nasdaq stocks. They find that, during the month of October 2010, on average, 20.1% of all trades are executed against hidden orders. However, only a small proportion of the hidden depth get executed, which implies that the share of undetected hidden depth is much greater.

\[12\] This pinging phenomenon was first documented in Hasbrouck and Saar (2009).
ability of hidden ask orders is $\pi_1 - \varphi_1(F_t)$. Similarly, if $S_t = 3\delta$, the probabilities of hidden bid orders at $P_t - \delta/2$ and $P + \delta/2$ are $\varphi_1(F_t)$ and $\varphi_2(F_t)$ respectively, so that those of hidden ask orders at $P_t - \delta/2$ and $P_t + \delta/2$ are $\pi_1 - \varphi_1(F_t)$ and $\pi_2 - \varphi_2(F_t)$. Since existing literature suggests that the probability of hidden orders inside the spread is positively correlated with the own-side depth imbalance.\textsuperscript{[13]} I let the functions $\varphi_i$ take the form of

$$\varphi_i(u) = \pi_i / (1 + \exp(-\kappa u)), \quad \text{for } i = 1, 2,$$

(1.3.5)

where $\kappa$ and $\pi_i$ are positive constant and $\pi_i < 1$. When $\pi_1$ and $\pi_2$ are large, hidden orders are more likely to be found within the spread. Thus the HFT would have more opportunities to take advantage of these hidden orders through the use of pinging orders.

### 1.3.2 The HFT’s Trading Strategy

At any time $t$, the HFT in my model can submit limit buy and sell orders specifying the prices that he is willing to pay and receive, but they will be executed only when incoming market orders fill his limit orders. The quantity of the HFT’s limit orders is fixed at one lot (100 shares).\textsuperscript{[14]} On the other hand, instead of limit orders, the HFT can send out market buy or sell orders for immediate execution. The market orders will cross the spread, i.e. trading at the best ask (resp. best bid) price on the opposite side, and are thus less price-favorable.

\textsuperscript{[13]}Buti and Rindi (2013) provide a theoretical model that gives rise to this positive correlation, and Hautsch and Huang (2012) offer an empirical confirmation.

\textsuperscript{[14]}The reason that I fix the quantity of limit orders to be one lot is motivated by a common finding that appears in a vast number of empirical studies, whether they examine HFT activities (e.g. Hasbrouck and Saar (2013)) or analyze limit order books (e.g. Hautsch and Huang (2012)). The finding is that the average size of limit orders nowadays is just slightly bigger than 100 shares.
Make Strategy

When using limit orders to make the market, HFT can place his quotes at the best available bid and ask, hence joining the existing queues at these prices. This strategy simply amounts to traditional market making, which means that the HFT tries to passively capture the spread by posting limit orders at the best available quotes.

Furthermore, the HFT can also “ping” inside the spread by improving either the best bid or ask price by one tick whenever the spread $S_t$ is greater than $\delta$, i.e. submitting a buy (sell) limit order at $P_t - S_t/2 + \delta$ ($P_t + S_t/2 - \delta$). Such a pinging strategy of putting limit orders inside the spread is commonly used by HFTs in practice to capture the market order flow at the best quotes, because of the price-time priority associated with these limit orders. More importantly, they are used to ping for hidden orders inside the spread, as identified by Hasbrouck and Saar (2009). Nevertheless, despite being able to obtain faster executions, these pinging limit orders do receive worse execution prices than their queuing counterpart. Hence this is a trade-off faced by the HFT when he contemplates whether to improve the spread or not.

Because the submission, update or cancellation of limit orders entails no cost, it is natural to model the make strategy of the HFT as a continuous-time predictable control process:

$$\theta_{t}^{mk} = \{\theta_{t}^{mk,b}, \theta_{t}^{mk,a}\}, t \geq 0,$$  \hspace{1cm} (1.3.6)

where $\theta_{t}^{mk,b} \in \{0, 1\}$ and $\theta_{t}^{mk,a} \in \{0, 1\}$. $b, a$ stand for the bid and the ask side respectively. The predictable processes $\theta_{t}^{mk,b}$ and $\theta_{t}^{mk,a}$, with values equal to 0 or 1, represent the possible make regimes: 0 indicates that the limit order is joining the queue at the best price, whereas 1 indicates improving the best price by $\delta$. Note that

---

1 I define market making to be the strategy such that the HFT submit limit orders simultaneously at both sides of the LOB. Quoting at one side of the book only is not allowed, since it does not satisfy the standard definition of market making.
if the spread is at its minimum $\delta$, $\theta_{t}^{mk,b}$ and $\theta_{t}^{mk,b}$ both can only take the value of 0, since improving the best bid/ask will simply be considered as posting a market buy/sell order instead.

**Take Strategy**

Instead of limit orders, the HFT may also employ market orders (take strategy) to obtain instant execution. However, unlike limit orders, market orders take liquidity from the LOB and are subject to transaction fees. As a result, the costly nature of market orders implies that, if the take strategy is performed continuously, the HFT will go bankrupt in finite time. Hence I shall model the HFT’s take strategy as an impulse control in continuous time:

$$\theta^{tk} = \{\tau_{n}, \zeta_{n}\}_{n \in \mathbb{N}}. \quad (1.3.7)$$

Here, $\{\tau_{n}\}$ is an increasing sequence of stopping times denoting the moments that the HFT uses market orders. $\zeta_{n} \in [-\zeta_{\text{max}}, \zeta_{\text{max}}] \setminus \{0\}$ are $\mathcal{F}_{\tau_{n}}$-measurable random variables that represent the number of shares (in lot size) purchased if $\zeta_{n} > 0$ or sold if $\zeta_{n} < 0$, at these stopping times. I confine the size of market orders to be less than or equal to some small constant $\zeta_{\text{max}}$, so that the HFT remains small relative to the market and his market orders do not eat through the LOB.\footnote{Eating through the book means the size of an market order is larger than and thus exhausts the available depth at the best price(s), resulting in a price impact and a widening of the spread.}

**1.3.3 Order Execution Processes**

Limit orders of the HFT, when joining the queue at the best bid and ask, will be executed only if counterpart market orders arrive in the next instant and their sizes are large enough to fill the HFT’s limit orders completely. I model the arrivals of
exogenous buy and sell market orders by two independent Poisson processes, \( M^b \) and \( M^a \), with intensities given by \( \lambda^b \) and \( \lambda^a \) respectively. In addition, when a buy (resp. sell) market order arrives, the HFT’s sell (resp. buy) limit order in the queue will be filled with a probability given by the fill-rate function \( h(F_t) \) (resp. \( h(-F_t) \)), where

\[
h(u) = 1/(1 + \exp(\varsigma_0 + \varsigma_1 u)), \quad \varsigma_0, \varsigma_1 \text{ are positive constants.}
\] (1.3.8)

This fill-rate function depicts the idea that the probability of execution of a limit order becomes higher/lower if the own-side queue is relatively shorter/longer, which illustrates the time priority structure of the LOB.

However, if an limit order jumps the queue instead, it will receive an instantaneous execution as long as there is a matching hidden order on the opposite side that resides inside the spread. Otherwise, the limit order will be fully matched if a counterpart market order arrives at the next moment. Limit orders placed inside the spread by the HFT are not subject to the fill rate function, since they have price priority compared to those limit orders at the prevailing best quotes.

Turning to the HFT’s market orders, they will execute immediately by hitting either inside-spread hidden orders or limit orders at the best quote on the opposite side. Moreover, there is a per share fee of \( \epsilon \) associated with each market order that is sent by the HFT. The fee is fixed and paid directly to the market exchange operating the LOB. Thus the HFT cannot claim it back by any means.

### 1.3.4 Comments on the Model

The basic layout of the model resembles Guilband and Pham (2013) and is easy to justify given the goals of the paper. However, there are two important aspects of the model that are new compared to the current literature on high-frequency trading strategies and are essential for rationalizing HFTs’ pinging activities without ma-
Manipulation or learning. First, the opportunity of using limit orders to grasp hidden liquidity has not been properly addressed in existing continuous-time models of HFTs, where the limit order strategy only fulfills the role of the HFT as a traditional market maker or liquidity provider. In my model, limit orders can also be utilized by the HFT as a “pinging” strategy, because of the existence of inside-spread hidden orders. Compared to conventional market making, pinging is certainly more aggressive as it seeks a faster execution at a less favorable price. However, it is not as aggressive as market taking since it does not cross the spread. Given the large number of pinging activities already identified in empirical work, for example, Hasbrouck and Saar (2009, 2013), it is essential to include hidden orders and pinging strategies in a theoretical model of HFT behaviors.

Another aspect that has not been fully dealt with in the continuous-time literature concerns the prospect of directional or momentum tradings by the HFT. In a standard setup, the HFT would only use market orders to reduce its inventory whenever it gets out of control. However, in my model, the HFT would also implement market orders as a momentum strategy to aggressively take liquidity from the market and pursue directional trading when the signal of depth imbalance on price movement is strong enough. There are empirical studies that have identified some most profitable HFTs as predominantly liquidity takers on the market (e.g. Baron, Brogaard, and Kirilenko (2012)) and have recognized the anticipatory directional trading behaviors of HFTs (e.g. Hirschey (2013)). Therefore, it is crucial to incorporate momentum directional trading into any theoretical models on HFT strategies.\textsuperscript{17}

When both hidden orders and price momentum exist in the model, the HFT has two motives to carry out pinging strategies. One is to control inventory. The HFT could ping inside the spread to hit hidden orders when he needs to unwind his inventory.

\textsuperscript{17}Since market orders are costly, if they are used solely to control inventory when it is really necessary, it would be difficult to reconcile this with the findings that market taking can be (very) profitable.
inventory, which is cheaper than using market orders. Nonetheless, if hidden orders existed with high probabilities $\pi_1$ and $\pi_2$ and inventory control was the HFT’s only motive for pinging, the model would generate a large number of pinging orders filled by hidden orders. There would not be many cancelled pinging orders at the same time, contradicting with the numerous ones observed in reality.

The existence of price momentum, i.e. the predictability of depth imbalance on mid-price jumps, are then necessary to bring about cancelled pinging activities. This is because now the other motive for the HFT to ping is to chase price trends. The HFT could employ pinging orders to establish directional positions when anticipating likely price movements. However, if the volatility parameter $\sigma_F$ is large so that depth imbalance (momentum signal) varies turbulently, it would induce the HFT to cancel his pinging orders frequently and change his strategies based on the swings in momentum. This is escalated if the mid-price jump intensities $\lambda_{1J}$ and $\lambda_{2J}$ are also large. Large intensities prompt more trend-chasing pinging activities from the HFT since his directional bets have higher chances to materialize, yet the HFT often needs to cancel his pinging orders due to volatile changes in price momentum.

As a consequence of the possibility of pinging and momentum trading, the HFT in my model should not be considered exclusively as a turbo-charged market maker in the traditional sense, whereby it provides liquidity to the market with passive limit orders on the one side, and on the other side, liquidity-taking through market orders is only necessitated by inventory-control requirement.
1.4 Equilibrium and Solution Method

1.4.1 The Objective of the HFT

In order to derive an equilibrium of my model, I begin by stating the cash holding and the inventory processes of the HFT, and defining his optimization problem under stochastic evolutions of the LOB as laid out in the previous section.

Let $X$ and $Y$ denote the cash holdings and the inventory held by the HFT respectively. If a market making strategy $\theta_{mk}^t$ is used at $t$, the cash holding $X$ and the inventory $Y$ evolve according to:

\begin{align*}
    dY_t &= d\tilde{M}_t^a - d\tilde{M}_t^b \tag{1.4.1} \\
    dX_t &= -\left(P_t - \frac{S_t}{2} + \delta\theta_{mk,b}^t\right) d\tilde{M}_t^a + \left(P_t + \frac{S_t}{2} - \delta\theta_{mk,a}^t\right) d\tilde{M}_t^b, \tag{1.4.2}
\end{align*}

where

\begin{align*}
    d\tilde{M}_t^a &= \theta_{mk,b}^t \left(B_a^t dt + (1 - B_a^t) dM_t^a\right) + (1 - \theta_{mk,b}^t) h(F_t) dM_t^a \tag{1.4.3} \\
    d\tilde{M}_t^b &= \theta_{mk,a}^t \left(B_b^t dt + (1 - B_b^t) dM_t^b\right) + (1 - \theta_{mk,a}^t) h(-F_t) dM_t^b. \tag{1.4.4}
\end{align*}

Here, $B_a^t$ (resp. $B_b^t$) is an indicator that equals one if ask (resp. bid) hidden orders exist inside the spread. $B_a^t$ and $B_b^t$ have respective distributions $\eta^a(F_t, S_t)$ and $\eta^b(F_t, S_t)$. $\eta^a$, $\eta^b$ match the existence probabilities of ask/bid hidden orders that sit at the best bid/ask price plus $\delta$, given $F$ and $S$. Therefore, when the HFT’s buy (resp. sell) limit order makes the market at the prevailing best bid (resp. ask), its inventory increases (resp. decreases) by one lot if sell (resp. buy) market orders arrive at the next instant and fill the limit order of the HFT. Alternatively, if the HFT’s buy (resp. sell) limit order is pinging inside the spread, its inventory increases (resp. decreases) by one lot whenever the limit order hits an opposite-side hidden order. Otherwise, it rises (resp.
falls) by one lot if there is an arrival of sell (resp. buy) market orders within the next instant. Cash holdings thus increases (resp. decreases) by an amount equal to the quoted price of the sell (resp. buy) limit order multiplied by the order’s uncertain execution state.

On the other hand, the dynamics of $X$ and $Y$ jump at $t$ if the HFT exercises a take strategy $\theta_{tk}^t$ instead:

$$Y_t = Y_{t-} + \zeta_t,$$

$$X_t = X_{t-} - \left[ \zeta_t P_t + |\zeta_t| \left( \frac{S_t}{2} - H_t + \epsilon \right) \right]$$

(1.4.5)

(1.4.6)

where $\epsilon$ is the fixed fee per share paid to the market exchange, $H_t$ is an integer random variable with a probability mass function $G(\cdot | S_t, F_t)$ that takes the value

$$H_t = \begin{cases} 
0 & \text{if } S_t = \delta \\
\delta & \text{if } S_t > \delta \& \text{market order hits hidden order at } P_t + \text{sign}(\zeta_t)(S_t/2 - \delta) \\
2\delta & \text{if } S_t > 2\delta \& \text{market order hits hidden order at } P_t + \text{sign}(\zeta_t)(S_t/2 - 2\delta),
\end{cases}$$

(1.4.7)

and the probability distribution of $H_t$ matches the existence probabilities of inside-spread hidden orders given $S_t$ and $F_t$. As a result, when the HFT submits a market order, its inventory jumps up or down at $t$ by the size of the order (since take strategies are impulse control). Moreover, its cash holdings changes by the value of the order $\zeta_t P_t$ plus the cost associated with the market order. The cost consists of the uncertain part due to crossing the spread $(S_t/2 - H_t)$ and the constant, fixed transaction fee $\epsilon$.

---

18 In essence, the term $-H_t$ measures the reduction in the cost of market orders for the HFT given the possible existence of hidden orders inside the spread when $S_t > \delta$, i.e. it does not necessarily pay the full spread-crossing cost of $S_t/2$ (relative to the mid-price).
Given the processes of cash holdings $X$ and inventory $Y$, the objective of the HFT is the following. He wants to maximize over the finite horizon $[0, T]$ the profit (cash earnings) from his trades in the LOB, while at the same time keeping his inventory at bay. In addition, the HFT has to liquidate all his inventory at the terminal date $T$. Hence the HFT’s optimization problem is given by

$$\max_{\{\bar{\theta}^{mk}, \bar{\theta}^{tk}\}} \mathbb{E}_0 \left[ X_T - \gamma \int_0^T Y_t^2 \, d[P, P]_t \right], \quad \text{s.t.} \quad Y_T = 0,$$  

(1.4.8)

where the maximization is taken over all admissible strategies $\Theta$. The integral term $\gamma \int_0^T Y_t^2 \, d[P, P]_t$ is a quadratic-variation penalization term for holding a nonzero inventory in the risky stock, where $\gamma > 0$ is a penalization parameter and $[P, P]_t$ denotes the quadratic variation of the mid-price $P$.

Let me rewrite the above optimization problem in a more straightforward formulation where the terminal constraint $Y_T = 0$ is removed. To this end, I introduce the function

$$Q(x, y, p, f, s) = x + py - |y| \left( \frac{s}{2} - H + \epsilon \right),$$  

(1.4.9)

which represents the total cash obtained after an immediate liquidation of the inventory $y$ via a market order, given the cash holdings $x$, the mid-price $p$, the depth imbalance $f$ and the spread $s$.\footnote{H is the same integer random variable defined on the previous page.} I can now reformulate the problem (1.4.8) equivalently as

$$\max_{\{\bar{\theta}^{mk}, \bar{\theta}^{tk}\}} \mathbb{E}_0 \left[ X_T + P_T Y_T - |Y_T| \left( \frac{S_T}{2} - H_T + \epsilon \right) - \gamma \int_0^T Y_t^2 \, d[P, P]_t \right].$$  

(1.4.10)

The proof for the equivalence of the two formulations is shown in the appendix.

**Lemma 1.4.1.** (1.4.8) and (1.4.10) are equivalent.
Having defined the objective, the value function of problem (1.4.10) for the HFT is then:

\[
V(t, x, y, p, f, s) = \sup_{\theta \in \Theta} \mathbb{E}_t \left[ X_T + P_T Y_T - |Y_T| \left( \frac{S_T}{2} - H_T + \epsilon \right) - \gamma \int_t^T Y_u^2 - d[P, P_u] \right],
\]

(1.4.11)

for \( t \in [0, T], (x, y, p, f, s) \in \mathbb{R}^2 \times \mathcal{P} \times \mathbb{R} \times \mathbb{S} \). Here, given \( \theta = \{\theta^{mk}, \theta^{tk}\} \in \Theta \), \( \mathbb{E}_t \) stands for the expectation operator under which the solution \((X, Y, P, F, S)\) to the processes (1.3.1)-(1.3.3) and (1.4.1)-(1.4.6), with initial state \((X_t, Y_t, P_t, F_t, S_t) = (x, y, p, f, s)\), is taken.

Before giving a definition of the equilibrium, I states the following lemma (with its proof shown in the appendix), which provides some bounds on the value function (1.4.11) and demonstrates that the value function is finite and locally bounded.

**Lemma 1.4.2.** There exist constants \( C_0 \) and \( C_1 \) such that for all \((t, x, y, p, f, s) \in [0, T] \times \mathbb{R}^2 \times \mathcal{P} \times \mathbb{R} \times \mathbb{S},\)

\[
Q(x, y, p, f, s) \leq V(t, x, y, p, f, s) \leq x + py + C_1 + C_0.
\]

(1.4.12)

Both of the lower and the upper bound have a intuitive financial interpretation. The lower bound indicates the value of the particular strategy that eliminates all the current non-zero inventory through a market order, and then waits by doing nothing until the time reaches \( T \). The upper bound is made of three terms. The first term, \( x + py \), is the marked-to-market value of the portfolio at mid-price; the second constant \( C_1 \) denotes the upper boundary on profit from the fictitious market-making strategy that participates in every trade but with zero cost of controlling inventory; and the constant \( C_0 \) at last represents a bound on profit for any directional frictionless
market-taking strategy on a virtual asset that is always priced at the mid-price. This lemma is useful later on when I derive the solution to the model’s equilibrium.

1.4.2 Definition of Equilibrium and Dynamic Programming Equations

The problem of (1.4.11) is a mixed regular/impulse stochastic control problem in a jump-diffusion continuous time model, to which the methods of dynamic programming naturally lends itself. In order to characterize the equilibrium and the associated dynamic programming equations, I need to introduce two mathematical operators as follows. For any admissible strategy \( \theta_{mk} = \{ \theta_{mk,b}, \theta_{mk,a} \} \), I will define the second-order non-local operator \( \mathcal{L} \):

\[
\mathcal{L} \circ V(t, x, y, p, f, s) = (\mathcal{L}^P + \mathcal{L}^F + \mathcal{L}^S) \circ V(t, x, y, p, f, s) \\
+ g^a(f, s, \theta^m_{bk})(\eta^a(f, s) + (1 - \eta^a(f, s))\lambda^a) + (1 - \theta^m_{bk})\lambda^a h(f) \\
+ g^b(f, s, \theta^m_{ak})(\eta^b(f, s) + (1 - \eta^b(f, s))\lambda^b) + (1 - \theta^m_{ak})\lambda^b h(-f),
\]

(1.4.13)

where \( \mathcal{L}^P, \mathcal{L}^F, \mathcal{L}^S \) in the first term are the infinitesimal generators of the processes of the mid-price \( P \), the depth imbalance \( F \) and the spread \( S \) respectively, and the next two terms denote the non-local operator induced by the (expected) jumps of the cash process \( X \) and inventory process \( Y \) when the HFT applies an instantaneous make strategy \( \theta^m_{mk} \) at date \( t \). In addition,

\[
g^a(f, s, \theta^m_{bk}) = \theta^m_{bk}(\eta^a(f, s) + (1 - \eta^a(f, s))\lambda^a) + (1 - \theta^m_{bk})\lambda^a h(f) \\
g^b(f, s, \theta^m_{ak}) = \theta^m_{ak}(\eta^b(f, s) + (1 - \eta^b(f, s))\lambda^b) + (1 - \theta^m_{ak})\lambda^b h(-f),
\]

(1.4.14)

(1.4.15)
which correspond to the expected rate of execution for the HFT’s bid and ask limit order respectively.20

Besides the operator \( L \), the impulse control operator \( M \) for an admissible take strategy \( \theta^{tk} \) shall be given by

\[
M \circ V(t, x, y, p, f, s) = \sup_{\zeta} \int_H V(t, x - \zeta p - |\zeta|(s/2 - H + \epsilon), y + \zeta, p, f, s) \, dG(H \mid s, f)
\]

(1.4.16)

for \( \zeta \in [-\zeta_{max}, \zeta_{max}] \). It is generated by the jumps in \( X \) and \( Y \) due to \( \theta^{tk}_t \) used at \( t \).

With \( L \) and \( M \) defined, the dynamic programming equation associated with the value function (1.4.11) is the Hamilton-Jacobi-Bellman quasi-variational inequality (HJB-QVI):

\[
\max \left\{ \frac{\partial V}{\partial t} + \sup_{\{\theta^{mk}\}} \{L \circ V\} - \gamma y^2 \mathbb{E}_t[d[P, P]_t], \ M \circ V - V \right\} = 0, \text{ on } [0, T) \quad (1.4.17)
\]

together with the terminal condition

\[
V(T, x, y, p, f, s) = x + py - |y|(s/2 + \epsilon) + |y| \int_H H \, dG(H \mid s, f). \quad (1.4.18)
\]

There is an explicit expression for the HJB-QVI (1.4.17) shown in the appendix. In particular, it writes out the full expressions for the infinitesimal generators \( (L^P, L^F, L^S) \) and the value of \( \mathbb{E}_t[P, P]_t \).

**Lemma 1.4.3.** The HJB-QVI (1.4.17) admits an explicit expression.

Having (1.4.17) and (1.4.18) at hand, the equilibrium concept of my model is presented in the following definition.

**Definition 1.4.1.** In the above continuous-time economy where the HFT trades a risky stock in a LOB that is governed by the stochastic processes laid out in Sec-
tion 2.A-2.C, the partial equilibrium is defined by a value function \( v(t; \cdot) \) and policy functions (strategies) \( \{\theta^{mk}, \theta^{tk}\} \), \( t \in [0, T] \), such that

(a) The policies solve the HFT’s maximization problem (1.4.10);

(b) Given the policy functions, the value function \( v \) solves the HJB-QVI (1.4.17) for \( t \in [0, T] \) with the terminal condition (1.4.18) at \( T \).

The next proposition demonstrates that a solution to the partial equilibrium exists and is unique. The proof is relegated to the appendix.

**Proposition 1.4.1.** There is a unique solution to the partial equilibrium of the model. In particular, the value function defined in (1.4.11) is the unique viscosity solution to (1.4.17) and (1.4.18). \(^{21}\)

I shall devote the next subsection to providing a numerical solution to the equilibrium of my model, i.e. the value function (1.4.11) that solves (1.4.17) and (1.4.18) and the corresponding optimal trading strategies of the HFT. However, prior to proceeding further, I present a lemma below that simplifies the value function (1.4.11) and reduces its dimensionality.

**Lemma 1.4.4.** (Proof in the appendix) The value function (1.4.11) can be decomposed as \( V(t, x, y, p, f, s) = x + py + \nu(t, y, f, s) \). Moreover, the reduced-form value function \( \nu \) satisfies the quasi-variational inequality and the terminal condition shown below.

---

\(^{21}\) The viscosity solution concept is a generalization of the classical solution concept to a partial differential equation. For a classic reference on viscosity solutions, see Crandall, Ishii, and Lions (1992) or Fleming and Soner (2005).
which are simplified from (1.4.17)-(1.4.18) after decomposing $V$:

$$\max \left[ \frac{\partial \nu}{\partial t} + y \mathbb{E}_t dP_t + \mathcal{L}^F \circ \nu + \mathcal{L}^S \circ \nu - \gamma y^2 \mathbb{E}_t d[P, P] + \right. $$

$$\left. \sup_{\theta_{mk}^a} \left\{ g^a(f, s, \theta_{mk}^a) \cdot \left( \nu(t, y + 1, f, s) - \nu(t, y + 1, f, s) + s/2 - \delta \theta_{mk}^a \right) + \right. $$

$$g^b(f, s, \theta_{mk}^b) \cdot \left( \nu(t, y - 1, f, s) - \nu(t, y - 1, f, s) + s/2 - \delta \theta_{mk}^b \right) \right\},$$

$$\sup_{\zeta} \left\{ \nu(t, y + \zeta, f, s) - |\zeta| \left( \frac{s}{2} + \epsilon \right) + |y| \int H dG(H \mid s, f) \right\} = 0, \text{ on } [0, T)$$

(1.4.19)

with terminal condition:

$$\nu(T, y, f, s) = -|y|(s/2 + \epsilon) + |y| \int H dG(H \mid s, f). \quad (1.4.20)$$

Before digging into the numerical solution, I would like to provide some economic rationale behind the functions that the pinging strategy serves in the HFT’s maximization problem. On one hand, hidden orders exists inside the spread with probability $\eta^a$ or $\eta^b$ and the HFT can use pinging orders ($\theta_{mk}^a = 1$ or $\theta_{mk}^b = 1$) to control his inventory by trying to hit those hidden orders. This is cheaper than using market orders as market orders cross the spread and pay the transaction fee, which amounts to a cost of $S_t/2 + \epsilon$. However, if inventory control is the only motive for the HFT to utilize the pinging strategy, the HFT will always try to execute but not cancel his pinging orders. Thus the model could only produce pinging that executes against hidden orders without many cancellations.

This is why we need short-term price momentum on the other hand, i.e. the effect of depth imbalance $F$ on $\mathbb{E}_t dP_t$ through the functions $\varphi_1$ and $\varphi_2$. With the presence of momentum, the HFT will also employ the pinging strategy to chase the price trend
before it is gone. More importantly, the HFT will cancel his pinging orders when there
is an abrupt change in imbalance $F$. If the momentum becomes too strong to wait, i.e. $F$ becomes very large in absolute value, the HFT will cancel the pinging orders and place his directional bets via market orders. And if the momentum weakens substantially or even reverses, the HFT will also cancel the pinging orders since he does not want to be adversely hit. Hence such trend-chasing behaviors of the HFT enable the model to produce pinging activities as well as cancellations at the same time.

1.4.3 Numerical Solution

In this part, I focus on the numerical solution to the value function $V$ of (1.4.11), and the associated policy functions (optimal strategies). In particular, since $V(t, x, y, p, f, s) = x + py + \nu(t, y, f, s)$, I will provide a backward, finite-difference scheme that solves the quasi-variational inequality (1.4.19) and (1.4.20), which completely characterizes the reduced-form value function $\nu$. The numerical method is based on the finite-difference scheme developed by [Chen and Forsyth (2008)], as well as the scheme used in [Guilbaud and Pham (2013)].

To begin with, I consider a time discretization on the interval $[0, T]$ with time step $\Delta_T = T/N_T$ and a regular time grid

$$T_{N_T} = \{t_k = k\Delta_T, k = 0, \ldots, N_T\}. \quad (1.4.21)$$

Secondly, I need to discretize and localize the state spaces for $Y$ and $F$ on two finite regular grids, with bounds $M_Y, M_F$, and step sizes $\Delta_Y = M_Y/N_Y \equiv 1$, $\Delta_F = M_F/N_F$.
respectively, where \( N_Y, N_F \in \mathbb{N} \), so that

\[
\forall_{N_Y} = \{ y_i = i \Delta_Y, i = -N_Y, \ldots, N_Y \}, \quad \mathbb{F}_{N_F} = \{ f_j = j \Delta_F, j = -N_F, \ldots, N_F \}.
\]

(1.4.22)

Next, I define two finite-difference matrices, \( D_1 \) and \( D_2 \), for calculating first and second order derivatives against \( F \) on the \( F \)-grid \( \mathbb{F}_{N_F} \), where \( D_2 \) uses central difference and \( D_1 \) uses forward difference when \( f_j < 0 \) and backward difference when \( f_j \geq 0 \):

\[
D_2 \nu(t, y, f_j, s) = \frac{\nu(t, y, f_{j+1}, s) - \nu(t, y, f_j, s) + \nu(t, y, f_{j-1}, s)}{(\Delta F)^2}
\]

\[
D_1 \nu(t, y, f_j, s) = \begin{cases} 
\frac{\nu(t, y, f_{j+1}, s) - \nu(t, y, f_j, s)}{\Delta F} & \text{if } f_j < 0 \\
\frac{\nu(t, y, f_j, s) - \nu(t, y, f_{j-1}, s)}{\Delta F} & \text{if } f_j \geq 0
\end{cases}
\]

(1.4.23)

I now state the main part of the numerical scheme. To this end, I introduce

the explicit-implicit operator for the time-space discretization of the quasi-variational inequality (1.4.19), that is, for any \((t, p, f, s) \in [0, T] \times \mathcal{P} \times \mathbb{R} \times \mathcal{S}\) and any real-valued function \( \phi : \mapsto \phi(t, y, f, s) \), I define

\[
\mathcal{A}(t, y, f, s, \phi) = \max \left\{ \tilde{\mathcal{L}}(t, y, f, s, \phi), \tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t, y, f, s, \phi) \right\},
\]

(1.4.24)

\[\text{22}\]The two matrices \( D_1 \) and \( D_2 \) are defined in a similar fashion to the finite difference space derivatives in Section 5.1 of Cont and Voltchkova [2005]. The reason that I switch from forward to backward difference in \( D_1 \) when \( f_j \) becomes greater than 0 is detailed over there.
where

\[
\tilde{L}(t, y, s, \phi) = \left( I_{N_F \times N_F} - \Delta_T \sigma_T^2 D_2 - \Delta_T \alpha_F (P_{N_F 1}^T \cdot D_1) \right)^{-1} \times \\
\left( \dot{\phi}(t, y, s) + \Delta_T y \frac{E_t d P_t}{d t} + \Delta_T \mathcal{L}^S \left( \phi(t, y, s) \right) - \Delta_T \gamma y^2 \mathcal{E}_t d[P, P]_t + \right. \\
\Delta_T \sup_{\theta_{mk}} \left\{ g^a(\cdot, s, \theta_t^{mk,b}) \left( \phi(t, y + 1, s) - \phi(t, y, s) + \frac{s}{2} - \delta \theta_{mk,b} \right) + \\
\left. g^b(\cdot, s, \theta_t^{mk,a}) \left( \phi(t, y - 1, s) - \phi(t, y, s) + \frac{s}{2} - \delta \theta_{mk,a} \right) \right\} \right)
\]

(1.4.25)

and

\[
\tilde{M} \circ \tilde{L}(t, y, f, s, \phi) = \sup_{|\zeta| \leq \zeta_{max}} \left\{ \tilde{L}(t, y + \zeta, f, s, \phi) - |\zeta| (\frac{s}{2} + \epsilon) + |y| \int_H H dG(\cdot | s, f) \right\}.
\]

(1.4.26)

When inventory \( y \) is on the boundary of \( Y_N, \) i.e. \( y = -M_Y \) or \( y = M_Y, \) make and take strategies are confined to the buy side or the sell side only, so that \( y \) does not go off its grid. Then, I approximate the solution \( \nu \) to (1.4.19)-(1.4.20) by the numerical solution \( \varpi \) on \( T_N \times Y_N \times F_{N_F} \times S \) to the backward explicit-implicit finite difference scheme:

\[
\varpi(T, y, f, s) = -|y| (\frac{s}{2} + \epsilon) + |y| \int_H H dG(\cdot | s, f)
\]

(1.4.27)

\[
\varpi(t_k, y, f, s) = \mathcal{A}(t_{k+1}, y, f, s, \varpi), \; k = N_T - 1, N_T - 2, \ldots, 0,
\]

(1.4.28)

where \([1.4.27]\) and \([1.4.28]\) approximate \([1.4.20]\) and \([1.4.19]\) respectively.

\[23\) In the operator \( \tilde{L}, : \) denotes the column vector of \( F_{N_T \times N_F}, I_{N_F \times N_F} \) is an \( N_F \) by \( N_F \) identity matrix, \( E_t d P_t, \frac{E_t d d[P, P]_t}{d t} \) are vectors evaluated on \( F_{N_F}, 1^T_{N_F} \) is an \( N_F \times 1 \) vector of 1s, and \( \cdot \cdot \) denotes element-by-element product for vectors and matrices. \( \tilde{L} \) is expressed as a vector on the grid \( F_{N_T}, \) because of the implicit time step that I used when approximating the generator \( \mathcal{L}^F. \)
The complete solution algorithm for the value function and the policy functions is summarized in the backward induction steps below:

1. At the terminal date \( t_{N_T} = T \): for each combination of \((y, f, s)\), make
   \[
   \varpi(T, y, f, s) = -|y|\left(\frac{s}{2} + \epsilon\right) + |y| \int_H dG(H | s, f).
   \]

2. (Backward Induction) From time step \( t_{k+1} \) to \( t_k \) where \( k \) runs from \( N_T - 1 \) back to 0, for each combination of \((y, f, s)\):
   - Calculate \( \tilde{L}(t_{k+1}, y, f, s, \varpi) \) from (1.4.25) and obtain \( \theta^{mk,*} \).
   - Calculate \( \tilde{M} \circ \tilde{L}(t_{k+1}, y, f, s, \varpi) \) from (1.4.26) and obtain \( \theta^{tk,*} \).
   - If \( \tilde{L}(t_{k+1}, y, f, s, \varpi) \geq \tilde{M} \circ \tilde{L}(t_{k+1}, y, f, s, \varpi) \), set \( \varpi(t_k, y, f, s) \) to be equal to \( \tilde{L}(t_{k+1}, y, f, s, \varpi) \) and the policy at \( t_k \) is thus \( \theta^{mk,*} \), given \((y, f, s)\).
   - Otherwise, let \( \varpi(t_k, y, f, s) = \tilde{M} \circ \tilde{L}(t_{k+1}, y, f, s, \varpi) \), and \( \theta^{tk,*} \) is taken to be the policy at \( t_k \), given \((y, f, s)\).

Finally, I state the convergence theorem of my numerical solution \( \varpi \) to the reduced-form value function \( \nu \) (and hence the convergence of discretized policy functions), with its proof left to the appendix.

**Proposition 1.4.2.** The solution \( \varpi \) to the numerical scheme (1.4.27)-(1.4.28) and the corresponding discretized policies converge locally uniformly, respectively, to the reduced-form value function \( \nu \) and the optimal strategies on \([0, T] \times \mathbb{R} \times \mathbb{R}\) as

\[
(\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F) \to (0, 0, \infty, 0, \infty), \ \forall s \in S.
\]

### 1.5 Data and Estimation

An important aspect of my paper is to examine the optimal trading strategies of the HFT given exogenous evolutions of the LOB. In order to quantify the implications
of my model based on the numerical solution, I need to obtain values for the pa-
rameters that govern the stochastic processes of the LOB characteristics, and this
section concerns the estimation of these parameters. The table below summarizes the
parameters to be estimated.

Table 1.5.1: Parameters of order book characteristics

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Explanation</th>
<th>Estimation Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$ , $\rho$</td>
<td>jump intensity and transition matrix of $S$</td>
<td>Non-parametric</td>
</tr>
<tr>
<td>$\alpha_F$, $\sigma_F$</td>
<td>mean-reversion and volatility parameters of $F$</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>$\lambda_{1,2}$</td>
<td>jump intensities of $P$</td>
<td>Non-parametric</td>
</tr>
<tr>
<td>$\beta_{1,2}$</td>
<td>distribution parameters of $P$ jump directions</td>
<td>Logistic regressions</td>
</tr>
<tr>
<td>$\kappa$, $\pi_{1,2}$</td>
<td>distribution parameters of hidden orders</td>
<td>Logistic regressions</td>
</tr>
<tr>
<td>$\varsigma_{0,1}$</td>
<td>Parameters of limit order fill rates</td>
<td>Logistic regressions</td>
</tr>
<tr>
<td>$\lambda^{M,a}$, $\lambda^{M,b}$</td>
<td>market order arrival intensities</td>
<td>Non-parametric</td>
</tr>
</tbody>
</table>

1.5.1 Data

I use Nasdaq TotalView-ITCH 4.0 limit order book message feed data on three types
of stocks listed on Nasdaq during the month of June 2012 (21 trading days). The
three types consist of stocks with narrow spreads and high order-book depths, stocks
with medium spread and depth levels, and stocks with wide spreads and low order-
book depths. I focus on three representative stocks, with one from each of the types:
INTC (Intel, narrow spread and high depth), QCOM (Qualcomm, medium spread
and medium depth), and AMZN (Amazon, wide spread and low depth).

The TotalView data include all real-time Nasdaq limit order book messages of
these stocks, stamped to millisecond precision. The complete message feeds allow
me to reconstruct the whole limit order books and their complete evolutions for the
three stocks. Since the limit order book characteristics of my model concern only the
first level of the book, I track the evolutions at the top of the book for estimation
purposes. As in common practices, I use data between 9:45am and 15:45pm to avoid certain erratic market movements.\footnote{Please refer to the appendix for a complete description of the limit order book data.}

1.5.2 Estimation

I shall employ standard nonparametric estimators for all the intensity parameters as well as the transition matrix $\rho$.\footnote{For a standard reference, see, for example, Karr (1991).} The parameters of the depth imbalance $F$ are then estimated using maximum likelihood, since the transition density of an Ornstein-Uhlenbeck process is known in closed form.\footnote{For a reference, see Aït-Sahalia (1999) and Aït-Sahalia and Mykland (2003).} In addition, the parameters governing various probabilities are estimated by logistic regressions, as these distribution functions all have logistic forms. I conduct the estimations for all trading days in June 2012, and then calculate the averages of these daily estimates as my final estimated parameter values to be fed into my numerical solutions. The mean values of my daily parameter estimates are presented in Table 1.5.2 with Newey-West heteroskedasticity and autocorrelation consistent (HAC) standard errors in parentheses.

There are three aspects of the estimation results that are worth being pointed out. Firstly, compared to Amazon, Intel and Qualcomm have a less volatile imbalance process $F$ as well as a higher tendency to stay at lower-spread positions. Secondly, due to its lower book depth, Amazon’s mid-price jumps are more often, as measured by both $\lambda_1^J$ and $\lambda_2^J$. Thirdly, the estimates of hidden order parameters, market order arrival rates and limit order fill rates are more alike for those three stocks. We will see in the next section that these dissimilarities and similarities in parameter estimates will lead to a much different optimal HFT strategy profile for Amazon as opposed to Intel or Qualcomm.

Besides the estimated parameters, the fixed parameters in Table 1.5.3 are also used in my numerical quantification and simulation study of the optimal HFT strategies.
Table 1.5.2: Order Book Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>INTC</th>
<th>QCOM</th>
<th>AMZN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spread</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda^S$</td>
<td>0.161/s</td>
<td>0.312/s</td>
<td>0.578/s</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.017)</td>
<td>(0.027)</td>
<td></td>
</tr>
</tbody>
</table>
| $\rho$             | \[
\begin{pmatrix}
0 & 1 & 0 \\
0.99 & 0 & 0.01 \\
0 & 1 & 0
\end{pmatrix}
\] | \[
\begin{pmatrix}
0 & 0.97 & 0.03 \\
0.95 & 0 & 0.05 \\
0.08 & 0.92 & 0
\end{pmatrix}
\] | \[
\begin{pmatrix}
0 & 0.74 & 0.26 \\
0.22 & 0 & 0.78 \\
0.14 & 0.86 & 0
\end{pmatrix}
\] |
| **Imbalance**      |       |       |       |
| $\alpha_F$         | 0.308 | 0.547 | 0.734 |
| (0.016)            | (0.027) | (0.035) |       |
| $\sigma_F$         | 0.777 | 1.429 | 2.336 |
| (0.029)            | (0.053) | (0.074) |       |
| **Mid-price**      |       |       |       |
| $\lambda_1^J$      | 0.161/s | 0.307/s | 0.522/s |
| (0.010)            | (0.018) | (0.027) |       |
| $\beta_1$          | 2.744 | 2.651 | 2.610 |
| (0.077)            | (0.076) | (0.075) |       |
| $\lambda_2^J$      | 0.052/s | 0.075/s | 0.121/s |
| (0.003)            | (0.005) | (0.007) |       |
| $\beta_2$          | 4.766 | 2.921 | 1.881 |
| (0.244)            | (0.257) | (0.129) |       |
| **Hidden order**   |       |       |       |
| $\kappa$           | 1.196 | 1.036 | 0.992 |
| (0.034)            | (0.030) | (0.021) |       |
| $\pi_1$            | 0.238 | 0.233 | 0.230 |
| (0.010)            | (0.010) | (0.009) |       |
| $\pi_2$            | 0.125 | 0.117 | 0.114 |
| (0.006)            | (0.006) | (0.006) |       |
| **MO arrival**     |       |       |       |
| $\lambda^{M,a}$    | 0.110/s | 0.130/s | 0.181/s |
| (0.006)            | (0.006) | (0.012) |       |
| $\lambda^{M,b}$    | 0.110/s | 0.130/s | 0.179/s |
| (0.004)            | (0.006) | (0.010) |       |
| **Fill rate**      |       |       |       |
| $\varsigma_0$      | 1.320 | 1.454 | 1.648 |
| (0.039)            | (0.055) | (0.066) |       |
| $\varsigma_1$      | 0.399 | 0.462 | 0.673 |
| (0.019)            | (0.022) | (0.031) |       |

Note: The table shows the average values of the daily parameter estimates in the month of June 2012, using limit order book data reconstructed from Nasdaq TotalView-ITCH 4.0 real-time message feeds. Newey-West HAC standard errors are in parentheses. The standard errors of the transition matrix estimates are not shown since they are close to zero. Intensity parameters are all measured at per second frequency. In addition, I normalize the tick size $\delta$ to be $0.1$ for Amazon, since on average, the spread and limit order prices of Amazon tend to be in multiples of $0.1$ instead of the minimum tick size $0.01$ that is used for Intel and Microsoft. For definitions of all the parameters, please refer to Table 1.5.1.
Table 1.5.3: Fixed Parameters

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discr/loc parameters</strong></td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td>3600</td>
</tr>
<tr>
<td>$\Delta T$</td>
<td>0.5</td>
</tr>
<tr>
<td>$M_Y$</td>
<td>30</td>
</tr>
<tr>
<td>$\Delta Y$</td>
<td>1</td>
</tr>
<tr>
<td>$M_F$</td>
<td>10</td>
</tr>
<tr>
<td>$\Delta F$</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Model constants</strong></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.003</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2</td>
</tr>
<tr>
<td>$\zeta_{max}$</td>
<td>10</td>
</tr>
<tr>
<td><strong>Backtest parameters</strong></td>
<td></td>
</tr>
<tr>
<td>$N_{MC}$</td>
<td>10000</td>
</tr>
<tr>
<td>$X_0$</td>
<td>0</td>
</tr>
<tr>
<td>$Y_0$</td>
<td>0</td>
</tr>
<tr>
<td>$P_0$</td>
<td>10</td>
</tr>
</tbody>
</table>

Note: I choose the per share fee of market orders to be $0.003, which corresponds to the same transaction fee stated by Nasdaq for its stocks. Furthermore, the tick size $\delta$ is set to be $0.01$, which is the tick size for all Nasdaq stocks that have prices above $1$. However, I make the tick size $0.1$ for Amazon, for the same reason indicated in the note under Table 1.5.2.

1.6 Computation and Simulation Results

In this section, I provide numerical results obtained with the optimal HFT strategy computed via the implementation of my numerical scheme (1.4.27)-(1.4.28), for the three stocks – Intel (INTC), Qualcomm (QCOM) and Amazon (AMZN). I use as inputs the estimated parameter values shown in Table 1.5.2 together with the fixed parameters listed in Table 1.5.3.

1.6.1 Optimal Strategy Profiles

As seen from the reduced-form value function $\nu$, the optimal HFT strategy depends on time $t$, inventory $Y$, depth imbalance $F$ and spread level $S$. Therefore, I will characterize the strategy as a function of inventory and depth imbalance, for spread...
equal to $\delta$ and $2\delta$, near $t = 0$ and $t = T$. The strategy profiles are mostly time invariant if time $t$ is not very close to the terminal date $T$. In addition, the optimal strategy profiles for Qualcomm are not shown since they are in between the ones for Intel and the ones for Amazon.

Figure 1.6.1: Optimal Strategy, INTC, $t = 10$, spread = $\delta$ (left) / $2\delta$ (right)

I will start with Figure 1.6.1, which illustrates the optimal HFT strategies for Intel at $t = 10$, with inventory and depth imbalance level shown on the horizontal and the vertical axis respectively. Consider Figure 1.6.1(A) on the left first, where spread equals $\delta$. The orange-colored central region denotes market making through submitting a limit order at both the best bid and ask prices. The two blue regions stand for inventory management; buy (resp. sell) indicates using buy (resp. sell) market orders to increase (resp. decrease) inventory towards zero. Partial inventory control occurs when inventories are only partially unwound, whereas inventory control represents complete liquidation so that inventory jumps precisely back to zero. Furthermore,

27 The reason is that the key parameter estimates $(\lambda_S, \alpha_F, \sigma_F, \lambda_1^J, \lambda_2^J)$ of Qualcomm are between those of Intel and Amazon, as pointed out in Table 1.5.2.

28 For instance, if the inventory is $-5$ lots, the strategy will specify purchasing $z$ lots with $z < 5$ in the region of partial inventory control (buy), yet it will specify buying exactly 5 lots in the region of inventory control (buy) to make inventory become zero.
momentum represents utilizing market orders to change inventory and set up a directional position; (buy) and (sell) denote establishing positive and negative inventory holding positions respectively. Finally, there is no pinging since it is not possible when the spread is at its minimum $\delta$, and the majority of the graph is represented by traditional market-making / inventory-control behaviors.

The economic intuition behind Figure 1.6.1(A) is explained as follows. The situation where depth imbalance is less than zero will be focused on, since a similar but symmetrically reversed exposition can be applied to the opposite case where depth imbalance is greater than zero. To begin with, when depth imbalance is mildly positive, the probability of a positive mid-price jump is somewhat higher than that of a negative one. If the HFT has a negative inventory, he will face an inventory risk as the price would move against his holdings. Thus he will reduce such risk via a market order. Due to the cost of market-taking, the HFT either partially or completely disposes his inventory depending on the amount of the risk he has, i.e. how positive the depth imbalance is.

On the contrary, if the HFT holds a positive inventory, he would enjoy a possible gain from a positive mid-price jump and choose to make the market as a result. Since market taking is expensive and the signal effect from depth imbalance on the mid-price movement is not strong enough, the HFT would not use market orders to accumulate additional positive inventory, i.e. the expected return is not large enough.

Furthermore, when depth imbalance becomes considerably more positive, the probability of a positive mid-price jump is much higher than that of a negative one.

\footnote{For example, under momentum (buy), if the inventory is $-5$ lots, the strategy will dictate a purchase of $z$ lots with $z > 5$, and if the inventory is 1 lot, the strategy will dictate a purchase of $z$ lots with $z \geq 1$.}

\footnote{Note that market-making here would not achieve this risk-reduction purpose, for two reasons. First, market-making implies that the HFT would still post a sell limit order at the best ask. Second, the effect of a positive depth imbalance on the fill-rate function means that a buy limit order of the HFT has a smaller chance of being filled. Consequently, instead of decreasing, the two together exacerbate the inventory risk faced by the HFT.}
If the HFT has a negative inventory, he faces a substantial inventory risk, yet also a clear opportunity to chase the upward price momentum. Anticipating the likely price increase, the HFT aggressively takes the liquidity from the market through buy market orders to establish a directional (positive) inventory position, which gives rise to the momentum (buy) region in the graph. However, the HFT stops chasing the price momentum if his inventory is rather positive. As the mid-price jump intensities of Intel are not very large, being too aggressive and obtaining too much positive inventory would result in inventory risk on the opposite side if the mid-price does not jump up soon enough.\(^{31}\)

Next, consider Figure 1.6.1(B) on the right, where spread equals \(2\delta\). The green and yellow regions on each side of market making represent pinging strategies, which the majority of the graph consists of. Here, pinging on the bid (resp. ask) side denotes submitting a buy (resp. sell) limit order inside spread at the best bid plus \(\delta\) (resp. best ask minus \(\delta\)), while letting the sell (resp. buy) limit order joining the queue at the best ask (resp. best bid). To understand the intuition behind Figure 1.6.1(B) and compare it to the case of Figure 1.6.1(A), I will again concentrate on the scenario in which the depth imbalance is less than zero. A symmetrically reversed explanation can be constructed similarly for the opposite scenario.

When depth imbalance is modestly greater than zero, if the HFT’s inventory level is quite negative, he will remove the inventory risk through market orders. However, the inventory control is only partial since complete inventory controls are too costly under \(S = 2\delta\) as opposed to \(S = \delta\). Alternatively, if the HFT’s inventory level is closer to zero, he will instead use a pinging strategy from the bid side to reduce his inventory risk, for three reasons. Firstly, despite execution uncertainty, pinging is free and the risk is smaller with inventory level not far from zero, which decreases the HFT’s desire for immediacy. Secondly, a buy pinging order might hit a sell hidden

\(^{31}\)The depth imbalance is mean-reverting towards zero. Consequently, the anticipated price jump would become less likely if it does not occur very soon.
order inside the spread. Thirdly, a positive depth imbalance implies a lower fill rate for normal buy limit orders, so that it is optimal for the HFT to jump the queue. On the other hand, if the HFT holds a significant level of positive inventory, he is confronted with the inventory risk due to the possibility of no upward jump in mid-price. Hence he will ping on the ask side, which has a high chance of hitting a bid-side hidden order inside the spread (as the depth imbalance is positive) and entails no cost compared to using market orders.

When the positivity of depth imbalance becomes sizable, the HFT would reduce his inventory risk more aggressively through market orders if his inventory is below zero. However, provided that the inventory is positive but small, the HFT would instead pursue price momentum by pinging on the bid side, which increases the execution probability of his buy limit order. Moreover, chasing the momentum via market-taking is suboptimal in this case since the expected gain is less than the cost of using market orders \( S = 2\delta \). As a result, depending on the configuration of depth imbalance and inventory, pinging strategies would serve two different functionalities: unwinding inventory or pursuing price momentum.

Figure 1.6.2: Optimal Strategy, INTC, \( t = T - 3 \), spread = \( \delta \) (left) / \( 2\delta \) (right)
After that, let us examine Figure 1.6.2 which shows the optimal HFT strategies for Intel at $T - 3$. Consider Figure 1.6.2(A) on the left first, with spread equal to $\delta$. Since time $t$ is close to the terminal date, inventory management becomes a large concern for the HFT as he must liquidate all positions at $T$. Consequently, if the HFT is holding negative (resp. positive) inventory and the mid-price is more likely to jump up (resp. down) because of positive (resp. negative) depth imbalance, he will aggressively unload his inventory through market orders to reduce the risk associated with such mismatch of inventory against depth imbalance. This is the reason why approximately half of the graph is made of inventory controls. Conversely, when the HFT’s inventory corresponds to the likely price movement, he would still want to seize the expected gains from carrying a directional position. Hence the HFT would make the market instead if his inventory is positive (resp. negative) and the depth imbalance is the opposite. In this case, the HFT will not find it optimal to employ an aggressive momentum strategy to establish more directional positions in contrast to the situation of $t = 10$, since inventory control is the primary worry when time approaches $T$.

Consider Figure 1.6.2(B) next. Similar to Figure 1.6.2(A), inventory management accounts for a large part of the optimal strategy profile, but as in Figure 1.6.1(B), the inventory control is partial due to the costly nature of market orders under $S = 2\delta$. The main difference from Figure 1.6.1(B) is that pinging strategies now have one sole objective: reducing inventory risk. The HFT would only ping on the ask (resp. bid) side if his inventory and the depth imbalance are both positive (resp. negative). This is because pinging orders are free and have a larger probability of executing against an opposite-side hidden orders inside the spread compared to queuing limit orders.

Let us now turn our attention to Figure 1.6.3, which compares the optimal strategy for Amazon (right) to that for Intel (left) in the case of time $t = 10$ and spread $S = \delta$. It is clear that the major difference occurred to Amazon is that
Figure 1.6.3: Optimal Strategy, spread = \( \delta \), \( t = 10 \), INTC (left) v.s. AMZN (right)

(A) Intel

(B) Amazon

Then, let us look at Figure 1.6.4 that contrasts the optimal strategy for Amazon (right) to that for Intel (left) in the case of time \( t = 10 \) and spread \( S = 2\delta \). There are two main changes occurred to pinging strategies employed for Amazon. First, there is no pinging on the bid (resp. ask) side when the HFT’s inventory is negative (resp. positive) and the depth imbalance is also moderately negative (resp. positive). Second, pinging on the bid (resp. ask) side erodes everything when both the HFT’s inventory and the depth imbalance are above (resp. below) zero. These imply that when the HFT holds a directional position that matches the likely course of a mid-price change, pinging’s only aim is to chase the short-term price momentum. The
Figure 1.6.4: Optimal Strategy, spread = 2δ, t = 10, INTC (left) v.s. AMZN (right)

(A) Intel

(B) Amazon

reason is similar to the one given for Figure 1.6.3 and it is because Amazon has much higher mid-price jump intensities so that the HFT can afford to be more aggressive. Nevertheless, it is not optimal for the HFT to deploy momentum strategies via market-taking as it is too expensive and the price jump is already in favor of the HFT’s inventory. However, if the depth imbalance is extremely high, and the HFT carries inventory that is against the likely directional move of price, we observe that the HFT would actually use momentum strategies in the case of Amazon (purple areas). The large cost of market orders is absorbed by the reduction in inventory risk and the almost certain, short-term benefit from the anticipated price jump.

Finally, Figure 1.6.5 and Figure 1.6.6 are discussed together, which demonstrate the optimal HFT strategies for Amazon (right) towards the terminal date T and compare them to those for Intel (left). From these two figures, we notice that the optimal strategies for Amazon closely resemble those for Intel, despite the differences seen in Figure 1.6.3 and Figure 1.6.4 when t << T. Once more, this is the result of inventory management concern being a dominant force as time draws closer to the end, so that establishing directional positions to pursue short-term price momentum is no longer one of the objectives of the HFT.
Figure 1.6.5: Optimal Strategy, spread = $\delta$, $t = T - 3$, INTC (left) v.s. AMZN (right)

(A) Intel

(B) Amazon

Figure 1.6.6: Optimal Strategy, spread = $2\delta$, $t = T - 3$, INTC (left) v.s. AMZN (right)

(A) Intel

(B) Amazon
1.6.2 Percentage Attributions of Optimal Strategies

In this part, I conduct Monte Carlo simulation studies to quantify the properties of the HFT’s optimal strategies on the three stocks: Intel, Qualcomm and Amazon. The number of Monte Carlo runs is set to be $N_{MC} = 10000$, and I use a standard Euler scheme to simulate the paths of the state variables $(P, S, F, X, Y)$ as well as the exogenous market order arrivals $(M^b, M^a)$.

<table>
<thead>
<tr>
<th></th>
<th>INTC</th>
<th>QCOM</th>
<th>AMZN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market-making/Inventory-control</td>
<td>70.50%</td>
<td>58.30%</td>
<td>31.80%</td>
</tr>
<tr>
<td>Pinging</td>
<td>19.70%</td>
<td>28.10%</td>
<td>46.70%</td>
</tr>
<tr>
<td>Momentum (via market orders)</td>
<td>9.80%</td>
<td>13.60%</td>
<td>21.50%</td>
</tr>
</tbody>
</table>

Note: Please refer to Footnote 32 below on how these percentage breakdowns are obtained.

Table 1.6.1 summarizes the (Monte Carlo average) percentage attributions of the optimal HFT to three types of activity: traditional market-making/inventory-control, pinging, and momentum/directional trading via market orders.

There are two main implications that we can learn from Table 1.6.1. Firstly, for stocks with narrow spreads on average and abundant order book depths such as Intel, the HFT behaves like a market maker in the traditional sense. He is providing liquidity to the market most of the time. As a result, pinging activities account for only 20% of the optimal strategies and approximately 70% of the strategies are in the realm of liquidity provision (market-making) and inventory management. It corresponds to what we saw in the profiles of the optimal strategies for Intel early

---

Footnote 32: The percentage breakdowns are calculated as follows. For each one of the simulations, I compute the number of choices attributable to each type of the activity(strategy chosen optimally under the numerical solution given the simulated state variables. Then I divide these numbers by the total number of activities, which equals $T/\Delta T = 7200$, to arrive at the corresponding percentage numbers. Finally, I average these percentage numbers across all 10000 simulations to obtain the numbers shown in Table 1.6.1.
on, as Intel’s spread level is often equal to $\delta$. Secondly, for stocks like Amazon with wide spreads on average and an order book that has low depths and volatile movements, the HFT looks less like a market maker. Instead, he acts more like a short-term profit/momentum chaser, since pinging and momentum trading together account for about 70% of the strategies for Amazon. In particular, pinging constitutes nearly 50% of the optimal strategies. This also matches with the profiles of the optimal strategies for Amazon, where momentum trading and pinging constitute the majority of the strategy profiles under $S = \delta$ and $S = 2\delta$ respectively. Pinging can be considered as a strategy that demands liquidity from the market when its objective is to build directional inventory position. Hence for stocks with wide spreads and low book depths, the HFT is mainly a market taker and quite often he is trying to take liquidity in order to bet on the directional moves of price.\footnote{For each of the three stock types, I checked 10 stocks, including the representative ones (Intel, Qualcomm, Amazon) being focused here. The percentage attributions are similar, as well as the optimal strategy profiles, i.e. the HFT is more of a market maker (resp. taker) and the pinging percentage is lower (resp. higher) if the stock under consideration has higher (resp. lower) order book depths and narrower (resp. wider) spreads.}

Next, I compared the pinging percentages obtained from the model (as shown in Table 1.6.1) to the pinging percentages observable from the data to see how much pinging in the data can be rationalized by the model. To calculate pinging percentages in the data, I first compute the number of cancelled pinging activities for different stocks in a similar fashion to Hasbrouck and Saar (2009). This is defined as the number of limit orders submitted inside the spread and then cancelled in less than 2 seconds. Second, I need to compute the number of pinging orders that execute against hidden orders inside the spread. This is somewhat problematic since such pinging orders are identical to market orders that hit hidden orders. In order to deal with this issue, I treat non-consecutive orders executing on hidden orders as pinging orders. Therefore, the number of total pinging activities are calculated as the sum of the number of cancelled pinging activities and the number of pinging orders hitting
hidden orders. I then divide the number of total pinging activities by the total number of order book activities to arrive at the pinging percentages obtainable from the data:

\[
Pinging \% \text{ in the data} = \frac{\text{Number of total pinging activities}}{\text{Total number of order book activities}}.
\]

Finally, the model's pinging percentages are gauged against those from the data in Table 1.6.2.

<table>
<thead>
<tr>
<th></th>
<th>INTC</th>
<th>QCOM</th>
<th>AMZN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinging % observed in data</td>
<td>23%</td>
<td>39%</td>
<td>70%</td>
</tr>
<tr>
<td>Pinging % produced by model</td>
<td>20%</td>
<td>30%</td>
<td>50%</td>
</tr>
<tr>
<td>% of data’s pinging captured by model</td>
<td>85%</td>
<td>75%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Note: Pinging percentages produced by the model are approximations to the corresponding numbers shown in Table 1.6.1.

As clearly seen from Table 1.6.2, the pinging percentages produced by the model match quite closely to the percentages from the data. Moreover, at least over 70% of the pinging observable from the data can be captured by the model, and the number is higher for stocks with high depths and low spreads (like Intel). Hence the result indicates that most of the pinging activities observed in the data can be rationalized by the model with the two mechanisms of inventory control and trend chasing.

<table>
<thead>
<tr>
<th>Pinging’s % in Strategies</th>
<th>INTC</th>
<th>QCOM</th>
<th>AMZN</th>
</tr>
</thead>
<tbody>
<tr>
<td>No hidden orders</td>
<td>8%</td>
<td>16%</td>
<td>32%</td>
</tr>
<tr>
<td>No price momentum</td>
<td>12%</td>
<td>14%</td>
<td>18%</td>
</tr>
</tbody>
</table>
Table 1.6.3 further breaks down the roles that the mechanisms of inventory control and momentum chasing play in rationalizing pinging activities in the model. There are two noticeable features from Table 1.6.3. Firstly, for stocks with high depths and narrow spreads such as Intel, inventory control and momentum chasing contribute comparably to the rationalization of pinging. This can be seen from the similar resulting pinging percentages given by the model if either of the mechanism is shut down. Secondly, for stocks with low depths and wide spreads such as Amazon, momentum chasing carries more weight than inventory control in the rationalization of pinging. This is because the pinging percentage produced by the model decreases by a larger amount when the channel of price momentum is shut down than when the channel of hidden order is.

Additionally, Table 1.6.3 also implies that both inventory control and trend chasing are necessary for the purpose of pinging rationalization. Less than 50% of the pinging in the data is rationalized by the model if either mechanism is turned off, suggesting that both mechanisms are indispensable.

1.6.3 Auxiliary Predictions

Besides pinging rationalization, the model also yields a couple of other interesting auxiliary predictions regarding pinging activities with respect to depth imbalances. They are presented in this subsection.

The first auxiliary prediction of the model is related to the directions of pinging activities. The model implies that if the HFT sees positive (resp. negative) depth imbalance and hence positive (resp. negative) price momentum more often, he is more likely to pining from the buy (resp. sell) side and take positive (resp. negative) directional bets due to trend chasing motives. Consequently, it yields the following prediction:
**Prediction 1.6.1.** There is more pinging from the buy (resp. sell) side if depth imbalance is more frequently positive (resp. negative).

Now I check this prediction against data. To do this, I divide each trading day into 30-second intervals. Next, I calculate the number of buy and sell pinging activities in each of these intervals in a similar manner to the computation of the number of pinging activities shown in Section V(B). In addition, I also compute the durations (measured in seconds) of positive and negative imbalance in each interval. Then I use the following regression to measure and test the effect of imbalance durations on pinging activity directions:

\[ PO_t = \alpha + \beta DoI_t + x_t'\gamma + \epsilon_t. \]  \hspace{1cm} (1.6.1)

\(PO_t\) denotes the number of buy or sell pinging activities and \(DoI_t\) the duration (measured in seconds) of positive or negative depth imbalance in interval \(t\). \(x_t\) stands for other control variables which include \(DoI_{t-1}\), average spread and volatility of imbalance in interval \(t\), and \(\epsilon_t\) is an error term.

The regression is performed separately for buy pinging activities on positive imbalance and sell pinging activities on negative imbalance, and across all trading days in June 2012. The final parameter estimates are calculated in the same way as in Table 1.5.2 i.e. time series averages of daily parameter estimates. Overall, I find that the parameter \(\beta\) is statistically significant for all three types of stocks, with a value around 0.28 on average. This shows that the first auxiliary prediction is confirmed in the data.

The second auxiliary prediction of the model concerns the influence of depth imbalance volatilities on the frequencies of cancelled pinging activities. The model implies that if momentum strengthens (imbalance widens) by a large amount, contemporaneously...
neously the HFT would cancel his pinging orders and use market orders instead to chase momentum. And if momentum weakens by a large amount or even reverses (imbalance reduces by a large amount or reverses), contemporaneously the HFT would also cancel his pinging orders as his pinging orders risk being adversely hit. This implication thus yields the following prediction:

**Prediction 1.6.2.** *If depth imbalance is more volatile, there should be more cancelled pinging activities occurring at the same time.*

I employ a similar procedure as before to check this prediction, i.e. I divide each trading day into 30-second intervals and compute the number of cancelled pinging activities in each interval. Then I use the following regression to measure and test the effect of imbalance volatilities on cancelled pinging orders:

\[
CPO_t = \alpha + \beta \log VoD_t + x_t' \gamma + \epsilon_t. \tag{1.6.2}
\]

\(CPO_t\) denotes the number of cancelled pinging orders and \(\log VoD_t\) the log of depth imbalance volatility in interval \(t\). \(x_t\) again stands for other control variables which include \(\log VoD_{t-1}\), average spread and number of market order arrivals in interval \(t\), and \(\epsilon_t\) is an error term.

The regression is performed across all trading days in June 2012. The final parameter estimates are also calculated as the time series averages of daily parameter estimates. Overall, I find that the parameter \(\beta\) is statistically significant for all three types of stocks as well. Nevertheless, the magnitude of \(\beta\) is much higher for stocks with high depths and narrow spreads (with a value around 3.3) compared to stock with low depths and wide spreads (with a value less than 1). Therefore, the result suggests that the second auxiliary prediction is by and large confirmed in the data too.
1.7 Conclusion

In this paper, I build a continuous-time, partial equilibrium model on the optimal strategies of HFTs without any learning or manipulative ingredients to rationalize pinging activities observed in the data. The model improves on the works of Ho and Stoll (1981) as well as Guilbaud and Pham (2013) by introducing hidden orders inside the bid-ask spread and short-term price momentum. The HFT would use pinging orders inside the spread to either control inventory or chase price trends.

I demonstrate that for stocks with high order book depths and narrow spreads, pinging accounts for 20% of the optimal strategies in the model, whereas this number goes up to 50% for stocks with low order book depths and wide spreads. Then I compare these pinging percentages from the model to their corresponding counterparts in the data, and find that over 70% of the pinging activities in the data are captured by the model. The result thus suggests that most of the pinging in reality can be rationalized by my model. Furthermore, I show that for low-depth and wide-spread stocks, the majority of the pinging in the model is due to the momentum-chasing motive. However, for high-depth and low-spread stocks, the inventory control motive would play a similar role to the momentum-chasing motive in rationalizing pinging activities. In addition, I also develop a couple of other auxiliary predictions based on the model’s implications. They are both assessed on and found to be consistent with the data in general. Therefore, my model gives the overall message that pinging activities do not necessarily have to be manipulative and can be mostly rationalized as part of the standard dynamic trading strategies of HFTs.
1.8 Appendix

1.8.1 Data Description

In the first part of the appendix, I will give a detailed description of the limit order book data used in this paper. I utilize the Nasdaq TotalView-ITCH 4.0 database, which includes all real-time messages of limit order submissions, cancellations, executions and hidden order executions for every trading day since 7:00am EST when Nasdaq’s electronic limit order book (LOB) system starts accepting incoming limit orders, stamped to millisecond precision. The system is initialized by an empty order book where all overnight limit orders are resubmitted automatically at the beginning of each day. The TotalView-ITCH data record a unique identification for any limit order and I can identify the attribute of a limit order (cancelled, executed or neither) by tracking it through its order ID. Furthermore, trades are identified via the records of limit orders and hidden order executions. Since the trading direction of limit orders and hidden orders is recorded, I can exactly identify whether a trade is buyer-initiated or seller-initiated. Hence the LOB at any time can be reconstructed (up to millisecond precision) through continuously updating the book system according to all reported messages, which exactly represents the historical real-time-disseminated order book states of Nasdaq.

Furthermore, I consolidate all trade transactions into one single trade from a market order if they are logged at the same time-stamp in the data and have the same initiation types. In addition, I use the reconstructed LOB data between 9:45am and 15:45pm only, in order to avoid erratic effects that are likely to occur at market opening or closure.
1.8.2 Proofs for Various Lemmas and Propositions

The second part of the appendix is devoted to the proofs for several lemmas and propositions, which are omitted in the main paper.

**Proof of Lemma 1.4.1:** On one hand, I can show that the maximal value of (1.4.8) is smaller than that of (1.4.10). This is because, for any admissible strategy such that \( Y_T = 0 \), we immediately obtain \( Q(X_T, Y_T, P_T, F_T, S_T) = X_T \). On the other hand, given an arbitrary admissible strategy \( \theta \) and its associated state variable processes \( (X, Y, P, F, S) \), I can consider an alternative strategy \( \tilde{\theta} \), coinciding with \( \theta \) up to time \( T \) and employing a market order that liquidates all the inventory \( Y_T \) at the terminal date \( T \). The associated processes of the state variables \( (\tilde{X}, \tilde{Y}, P, F, S) \) under \( \tilde{\theta} \) satisfy \( (\tilde{X}_t, \tilde{Y}_t, P_t, F_t, S_t) = (X_t, Y_t, P_t, F_t, S_t) \) for all \( t < T \), and \( \tilde{X}_T = Q(X_T, Y_T, P_T, F_T, S_T) \), \( \tilde{Y}_T = 0 \). Therefore, this shows that the maximal value of (1.4.10) is smaller than that of (1.4.8), hence (1.4.8) is equivalent to (1.4.10).

**Proof of Lemma 1.4.2:** The lower bound is obvious, since \( V(t, x, y, p, f, s) = Q(x, y, p, f, s) \) by adopting the simple strategy that sets \( \theta_t^{tk} = y \) and terminates the problem immediately at time \( t \). On the other hand, we have

\[
V = (x + py) + \sup_{\theta} \mathbb{E}_t \left[(X_T - x) + (P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon\right) - \gamma \int_t^T Y_u^{2}d[P, P]_u\right] \\
\leq (x + py) + \sup_{\theta} \mathbb{E}_t \left[(X_T - x) + (P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon\right)\right].
\]

Since all the jump intensities \( (\lambda^S, \lambda^J_1, \lambda^J_2, \lambda^a, \lambda^b) \) are finite constants, the term

\[
\sup_{\theta} \mathbb{E}_t \left[(X_T - x) + (P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon\right)\right]
\]
cannot be greater than the finite, maximum profit achievable through a combination of a market-making strategy that participates in every trade when market orders arrive (an upper bound on $X_T - x$, denoted by $C_0$) and a directional, frictionless market-taking strategy that bets on the mid-price jumps (an upper bound on $(P_T Y_T - py) - |Y_T|\left(\frac{v_T}{2} - H_T + \epsilon\right)$, denoted by $C_1$).

**Proof of Lemma 1.4.3:** As stated under (1.4.18) in the main paper, only the infinitesimal generators ($L^P$, $L^F$, $L^S$) for the value function $V$ and the value of $E_t d[P, P]_t$ require explicit expressions. They are given below:

\[
E_t d[P, P]_t = \left(\frac{1}{4} \lambda_1^2 + \lambda_2^2\right) dt
\]

\[
L^P \circ V(t, x, y, p, f, s) = \left(V(t, x, y, p + \delta/2, f, s)\psi_1(f)
+ V(t, x, y, p - \delta/2, f, s)(1 - \psi_1(f))\right)\lambda_1^2 dt
+ \left(V(t, x, y, p + \delta, f, s)\psi_2(f)
+ V(t, x, y, p - \delta, f, s)(1 - \psi_2(f))\right)\lambda_2^2 dt
\]

\[
L^F \circ V(t, x, y, p, f, s) = V_f \cdot (\alpha_F f) dt + \frac{1}{2} V_{ff} \sigma_F^2 dt
\]

\[
L^S \circ V(t, x, y, p, f, s) = \left(\sum_{j=1}^3 \rho_{ij} [V(t, x, y, p, f, j\delta) - V(t, x, y, p, f, i\delta)]\right)\lambda^3 dt,
\]

where $V_f$ and $V_{ff}$ are the first and second order partial derivatives of $V$ against the state variable $F$. In addition, $E_t dP_t = \left(\lambda_1^2 \delta(2\psi_1(F_t) - 1) + \lambda_2^2 \delta(2\psi_2(F_t) - 1)\right) dt$.

**Proof of Proposition 1.4.1:** To prove Proposition 1.4.1 I need to show that the value function $V$ in (1.4.11) is the unique viscosity solution to (1.4.17) and (1.4.18). Since I have established the necessary growth (boundness) conditions on $V$ that are shown in Lemma 2, the proposition – the existence and the uniqueness of the viscosity solution $V$ (a.k.a. the value function) – is then a direct application of standard argu-
mments and results from stochastic control theory, e.g. Seydel (2009b,a), or Øksendal and Sulem (2007), Chapter 9.

**Proof of Lemma 1.4.4:** It is clear that the quasi-variational inequality (1.4.19) and the terminal condition (1.4.20) are simplified versions of (1.4.17) and (1.4.18), if I can decompose the value function \( V(t, x, y, p, f, s) \) as \( x + py + \nu(t, y, f, s) \). The required decomposition can be established by first noting that the mid-price \( P \) has constant jump intensities and jump size distributions depending only on the state variable \( F \), and then extending to my scenario the argument for a simpler case shown in Guilbaud and Pham (2013).

**Proof of Proposition 1.4.2:** I shall prove Proposition 1.4.2 by first establishing three properties for my numerical scheme and then applying a power theorem from Barles and Souganidis (1991).

**Lemma. (Monotonicity)**

For any \( \Delta_T > 0 \) such that \( \Delta_T \leq (\lambda^S + (\pi_1 + (1 - \pi_1)\lambda^a) + (\pi_1 + (1 - \pi_1)\lambda^b))^{-1} \), the operator \( A \) defined in (d) of the numerical scheme is non-decreasing in \( \phi \), i.e.

\[
\text{if } \phi < \tilde{\phi}, \text{ then } A(t, y, f, s, \phi) \leq A(t, y, f, s, \tilde{\phi}), \forall t, y, s \text{ and } f .
\]

**Proof.** From the expression in (e) of the numerical scheme, it is clear that \( g^a(f, s, \theta_t^{mk,b}) < \pi_1 + (1 - \pi_1)\lambda^a, \forall f, s, \theta_t^{mk,b} \) and \( g^b(f, s, \theta_t^{mk,a}) < \pi_1 + (1 - \pi_1)\lambda^b, \forall f, s, \theta_t^{mk,a} \). Thus \( 1 - \Delta_T \lambda^S - \Delta_T g^a(f, s, \theta_t^{mk,b}) - \Delta_T g^b(f, s, \theta_t^{mk,a}) > 0 \) as long as \( \Delta_T \leq (\lambda^S + (\pi_1 + (1 - \pi_1)\lambda^a) + (\pi_1 + (1 - \pi_1)\lambda^b))^{-1} < (\lambda^S + g^a(f, s, \theta_t^{mk,b}) + g^b(f, s, \theta_t^{mk,a}))^{-1} \), which implies that \( \tilde{L}(t, y, f, s, \phi) \) is monotone in \( \phi \) (as the sum of coefficients in front of \( \phi \) is positive) and so is \( A(t, y, f, s, \phi) \).
Lemma. *(Stability)*

For any $\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F > 0$, there exists a unique solution $\varpi$ to (1.4.27)-(1.4.28), and the sequence $\{\varpi\}$ is uniformly bounded.

**Proof.** By the definition of the backward scheme (1.4.27)-(1.4.28), the solution $\varpi$ exists and is unique. The uniform bound follows directly from the growth condition (lower and upper bounds) on the reduced-form value function $\nu$ (a modification on Lemma 1.4.2, i.e. the bounds on $V$).

Lemma. *(Consistency)*

The scheme (1.4.27)-(1.4.28) is consistent in the sense that, for all $(t, y, f) \in [0, T) \times \mathbb{R} \times \mathbb{R}$ and any smooth test function $\phi$, as $(\Delta_T, \Delta_Y, \Delta_F, M_F) \to (0, 0, \infty, 0, \infty)$, and $(t', y', f') \to (t, y, f)$, we have

$$
\lim 1/\Delta_T [\tilde{L}(t' + \Delta_T, y', f', s, \phi) - \phi(t', y', f', s)] = \frac{\partial \phi}{\partial t} + y \frac{\mathbb{E}_t dP_t}{dt} + \mathcal{L}^F \circ \phi + \mathcal{L}^S \circ \phi - \gamma y^2 \frac{\mathbb{E}_t d[P, P]|_t}{dt} + \\
\sup_{\theta^{mk}} \left\{ g^a(f, s, \theta^{mk,b}_t) \cdot \left( \phi(t, y + 1, f, s) - \phi(t, y, f, s) + s/2 - \delta \theta^{mk,b}_t \right) + \\
g^b(f, s, \theta^{mk,a}_t) \cdot \left( \phi(t, y - 1, f, s) - \phi(t, y, f, s) + s/2 - \delta \theta^{mk,a}_t \right) \right\}
$$

and

$$
\lim \tilde{M} \circ \tilde{L}(t', y', f', s, \phi) = \sup_{\zeta} \left\{ \phi(t, y + \zeta, f, s) - |\zeta|(s/2 + \epsilon) + |y| \int_H H dG(H | s, f) \right\}
$$

**Proof.** This follows from the result established in Section 6.1.2 of Chen and Forsyth (2008).

Proposition. *(Convergence)* The solution $\varpi$ to the numerical scheme (1.4.27)-(1.4.28) and the corresponding discretized policies converge locally uniformly, respec-
tively, to the reduced-form value function $\nu$ and the optimal strategies on $[0, T] \times \mathbb{R} \times \mathbb{R}$ as $(\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F) \to (0, 0, \infty, 0, \infty)$, $\forall s \in S$.

**Proof.** Given the properties of monotonicity, stability and consistency of the numerical scheme, this is a direct application of the result of Barles and Souganidis (1991).
Chapter 2

Count Models of Social Networks in Finance

2.1 Introduction

There is a growing use of social networks to model phenomena from all corners of financial economics. Models of social interactions, epidemics and network effects are thought to be the leading explanations for dramatic changes in stock market participation rates during the Internet Bubble Period or housing ownership rates during the Housing Bubble Period (see, e.g., Shiller and Pound (1989), Glaeser and Scheinkman (2001), Hong, Kubik, and Stein (2004), Glaeser, Gottlieb, and Tobio (2012), Han and Hirshleifer (2013)). Investor social networks, such as being a Boston money manager (Hong, Kubik, and Stein (2005)), being a Harvard Alumni (Cohen, Frazzini, and Malloy (2008)) or going to the same brokerage house in China (Feng and Seasholes (2008)), are associated with information sharing among group members which influence what stocks are held and how the stocks perform. Moreover, in the aftermath of the Financial Crisis of 2007, many have turned to the modeling of networks among banks and other financial intermediaries to explain financial contagion in the hopes
of discovering a more stable financial architecture (see, e.g., Allen and Gale (2007), Boyer, Kumagai, and Yuan (2006), Allen, Babus, and Carletti (2010)). Additionally, networks have also made their way to corporate finance as networks of CEOs, venture capitalists, entrepreneurs and banks are influential in allocating resources (see, e.g., Engelberg, Gao, and Parsons (2012), Lerner and Malmendier (2013), Shue (2013), Hochberg, Ljungqvist, and Lu (2010)).

A question fundamental and common to all these endeavors is how to measure the presence and value of social networks. Yet, no systematic approach has emerged. Instead, studies typically attack this challenge by being creative in utilizing special data and exploiting unique situations to identify network effects. This approach is largely necessitated by a lack of comprehensive information about social networks. While this approach has been highly effective in generating insights, the cost is that it is difficult to generalize results from one setting to another. And in many important settings, such detailed network data might simply not be available.

In this chapter, we show that overdispersed Poisson regression models, relying mostly on holdings or trade data that are typically available in most finance settings, can be used to study social networks in finance. These models were originally developed by statisticians Zheng, Salganik, and Gelman (2006) to analyze answers to survey questions from sociology (Killworth, Johnsen, McCarty, Shelley, and Bernard (1998), Killworth, McCarty, Bernard, Shelley, and Johnsen (1998), McCarty, Killworth, Bernard, Johnsen, and Shelley (2001)) about the count of friends a person has in different groups within the general population.

Importantly, they distinguish between being gregarious and being part of a network. Gregariousness is defined as people who differ in the expected number of social connections. However, their connections are formed randomly or independent and identically distributed (i.i.d.) as in the random networks model literature following Erdös and Rényi (1959). In contrast, being part of a network means that people
from certain groups have non-i.i.d. propensities to form ties with each other. These models make use of an important result from Erdős and Rényi (1959)—namely, if connections are formed randomly, then the count of the number of friends a person has in any group follows a Poisson distribution.

While the Poisson distribution fits well the count of friends in certain groups, like people named Nicole or postal workers, Zheng, Salganik, and Gelman (2006) find that it does not fit well the count of friends in other groups, like prisoners. For instance, the count of friends in the prison population is highly overdispersed, in that most people surveyed know zero but some know many prisoners. That is, the variance of this count distribution significantly exceeds the mean of the count distribution, in contrast to what one would find with a Poisson distribution in which the variance equals the mean. Overdispersion then captures social connections to the prison population that are formed in a non-i.i.d. manner as some people have a non-i.i.d. propensity to know prisoners. This is presumably because the prison population constitutes a network while people named Nicole do not.

Although such survey data are rare in financial markets, we show that these models can be extended to study social networks in finance by using plentiful data on the actions of agents in financial markets such as their investment holdings or trades. For concreteness, we study investors’ social networks by modeling the count of acquaintances in different cities, as proportional to the number of firms or stocks an investor holds that are headquartered in a given city. The idea is that since the stocks an investor chooses is a function of his network, we can infer that an individual who owns a “disproportionate” (in a sense that we will make precise shortly) number of stocks that are located in a certain city is more likely to have contacts in these cities.

Our extension of the network model in Zheng, Salganik, and Gelman (2006) can be easily applied to many other contexts in finance, such as banking networks where one
can count trades between a bank with other banks in different countries, or lending volume between banks and companies in different industries. In other words, while we do not have answers to survey questions about how many people investors know in different groups, we can proxy for answers to these questions by counting their investments across different categories.

Using panel data on the holdings of institutional investors in different cities, our dependent variable of interest is a monotonic transformation of the count of the number of stocks in those groups that are held by an investor. We estimate this model while allowing for heterogeneity in a number of important dimensions.

First, we allow for different gregariousness across managers—more gregarious managers have in expectation more stocks. We view this set of estimates as akin to investor fixed effects that allow some investors to hold more stocks than others. But it does not affect our inference of whether an investor belongs to a network. This inference is instead made controlling for this heterogeneity similar to the aforementioned statistics literature on social networks. Having a lot of friends is not the same as being part of a network. One could simply have equal numbers of friends and hold a lot of stocks in every group by chance.

Second, we allow different cities to have different numbers of potential connections based on how many stocks are headquartered in the city. A city like New York, which has many firms headquartered there, will have many connections attached to it. Again, this is a control or adjustment as we want to keep city or industry sizes roughly similar.

Controlling for these two factors, we can then use our model to estimate the degree of overdispersion of the cross-sectional distribution of the count of stocks in any given city held by investors. We allow the degree of overdispersion to vary across groups. That is, we can estimate a different overdispersion parameter for each city. If we have $N$ investors, $K$ groups, we end up estimating $N + 2K + J$ parameters with
$N \times K$ number of observations reflecting the number of stock holdings in different cities. In addition, $J$ is the degree of freedom needed to estimate semi-parametrically the transformation of the number of stocks into the number of social connections.

In our empirical analysis of the mutual fund holdings data from 1993 until now, we are careful to drop index and sector funds. Using the top 20 biggest Metropolitan Statistical Areas (MSAs) in terms of where stocks are headquartered as groupings, we find some overdispersion in most cities. Nevertheless, there is only pronounced overdispersion in San Jose, Los Angeles, New York, and San Diego, where the overdispersion parameter is around 2. Under the null Poisson random social connections setting, the overdispersion parameter for any given city should be 1. These results are robust to excluding managers located in a given city (and hence our results are not simply a manifestation of local bias) or controlling for fund style of the managers.

To get an intuition for our set-up, we plot in Figure 2.1.1 a histogram of the count of stocks headquartered in San Diego (left-panel) and Phoenix (right-panel) using the holdings of 1315 mutual fund managers reported in the fourth quarter of 2005. The x-axis is the number of stocks held by a manager. The y-axis is the frequency of managers. The counts are residuals after controlling for manager fixed effects and city size as described above. For San Diego, the mean of the count of stocks held by the mutual fund manager population is 1.83 with a variance of 12.7. For Phoenix, the mean is 1.43 and the variance is 2.4.

San Diego is highly overdispersed while Phoenix is near Poisson. This indicates that there are some managers who own many stocks in San Diego while most own few. We then make the inference that there is a San Diego network of investors while there is none or a small one for Phoenix. That is, managers who invest in San Diego are more likely to be part of a network that guides them toward San Diego stocks than Phoenix.
Our major concern is that overdispersion might simply capture data errors or some outliers but that are otherwise uninformative. Indeed, overdispersion is often treated as a nuisance rather than something fundamentally informative as pointed out by Zheng, Salganik, and Gelman (2006). However, in the context of social networks, the Poisson null model has a very natural interpretation a la Erdős and Rényi (1959), which is what makes this overdispersed Poisson regression model informative about networks.

We then relate our model’s outputs to demographic information about the investors as well as the performance of their investments. First, we try to understand what makes different cities overdispersed. We find that overdispersed cities tend to have a large number of graduates from local universities who work in the fund industry. We control for other city attributes such as the city’s real GDP, whether the city leans Democratic or Republican and the number of stocks headquartered in that city. Only the local university representation variable comes in significantly. This is consistent with the notion of university being an important source of networks in Cohen, Frazzini, and Malloy (2008).
Second, we use our model to calculate for each investor his relative propensity to have contacts in a city (RPC) and relate these managerial RPC scores to the manager’s demographic information. Our model gives a prediction for the expected number of stocks any investor should hold in a given city. An investor who holds a higher number of stocks than predicted is more likely to be part of that city (i.e. be part of that network). We sum up the investor’s scores across all the cities. We find that managers with an advanced degree, of the female gender, who younger and Republican-minded have higher RPC scores. Having an advanced degree and being younger play by far the biggest roles, consistent with our above finding that local university representation in the fund industry explains the overdispersion of different cities.

Third, we then regress fund performance on these managerial RPC scores, while controlling for a host of the usual explanatory variables for fund performance. We find that managers with higher RPC city scores outperform those with lower RPC scores by around 2.5% a year. Our findings here are reminiscent of the Industry Concentration Index of Kacperczyk, Sialm, and Zheng (2005). They find that managers who hold concentrated positions out-perform those that do not. Their interpretation is one of closet indexing as those with concentrated positions are less likely to be closet indexers. However, our measures and ICI are not very correlated and including ICI in the performance regression does not change the coefficient in front of our RPC score. Moreover, our findings are also not driven by familiarity bias (Pool, Stoffman, and Yonker (2012)). Namely, we show that managers do not hold concentrated positions in the city where they go to college. Our findings that social networks are valuable echoes prior studies, which document the value of investor and CEO networks such as Cohen, Frazzini, and Malloy (2008) and Engelberg, Gao, and Parsons (2012) using unique data on these networks.
In the Appendix, we consider a couple of additional sets of analyses. First, we replicate our findings using industry groupings as opposed to city groupings and find overdispersion for some industries. However, the performance results using industry groupings are weaker and the interpretation is potentially more problematic. Second, in addition to using institutional investor holdings data, we also use the Barber and Odean (2001) brokerage house retail investor holdings data and perform the same set of analyses. The problem with this dataset is that it is not representative of the retail investor universe in contrast to our mutual fund data which has the entire mutual fund universe. As a result, we generally find weaker social network effects for retail investors, but all the results are qualitatively similar in that we detect overdispersion in certain cities and industries and also find outperformance using our RPC scores. It is interesting to contrast the institutional and retail results as a statement about the prevalence of investor social networks across these two types of investors. The retail investor results also reassure us that our institutional investor results are not driven by unique mutual fund industry considerations. They seem more universal to investor networks.

Our contributions are two-fold. First, we introduce a new approach to the modeling of investors’ social networks by extending statistical models of surveys on acquaintances in groups. The existing approach in economics and finance in modeling social interactions focused on excessive correlation of investors’ actions due to them being part of the same group and sharing information (Glaeser and Scheinkman (2002), Hong, Kubik, and Stein (2005)). In other words, the null is that under non-social interaction there is no reason for the actions to be correlated after controlling for public signals. The challenge with the excess correlation approach is controlling for a rich enough set of public signals. This new approach differs in counting stock picks across different groups and making inferences on which groups are networks. Our new approach has a different null hypothesis premised on random social connections lead-
ing to a Poisson distribution of the count of investors’ holdings in any given group. The challenge here is to rule out that overdispersion is simply due to some outliers.

Second, and at the same time, our study contributes to the literature in statistics on the sociology of networks pioneered in Zheng, Salganik, and Gelman (2006). These papers have developed rich statistical models to study answers to surveys of questions on acquaintanceship networks. However, these studies are limited in terms of demographic or other information about the respondents. In applying these models to investor networks, we tap into a vast and rich database of such information about investor respondents including their investment performance. As a result, we can ask and answer many more questions about the determinants of the structure of social networks and the value of such networks for those who have them.

The rest of this chapter is organized as follows. We describe the model in Section 2.2 and the data and estimation procedures in Section 2.3. We collect the result for mutual fund investors in Section 2.4 and 2.5. We discuss extensions of our methodology in Section 2.6. We conclude in Section 2.7. The Appendix materials are in Section 2.8.

2.2 The Model

Our model follows Zheng, Salganik, and Gelman (2006)’s analysis of social networks. The key difference is that we do not observe answers to connections in different groups. We will instead use the number of holdings an investors has across different groups to proxy for their social connections. We will allow for a general flexible monotonic transformation of the number of connections in different groups to the number of holdings. This flexible transformation will be estimated along with the parameters describing the social networks.
2.2.1 Notation

Following [Zheng, Salganik, and Gelman (2006)](http://example.com), we use the following notation for the social networks between investors and their acquaintances in different cities. There is a total population of $N$ investors, with friends residing in a total of $K$ groups. Here, “group” is used interchangeably with “city”.

- $p_{ij}$: probability that investor $i$ knows person $j$,
- $a_i \equiv \sum_{j, j \neq i} p_{ij}$: gregariousness; the expected total number of social connections of investor $i$,
- $b_k \equiv \frac{\sum_{i=1}^{N} a_i}{\sum_{i \in S_k} a_i}$: proportion of total social connections that involves group $k$,
  where $S_k$ stands for “group $k$”
- $\lambda_{ik} \equiv \sum_{j \in S_k} p_{ij}$: investor $i$’s expected number of connections in group $k$,
- $g_{ik} \equiv \frac{\lambda_{ik}}{a_i b_k}$: investor $i$’s expected relative propensity to know people in group $k$,
- $y_{ik}$: number of friends from group $k$ made by person $i$,
- $z_{ik}$: number of stock picks that investor $i$ has in group $k$.

The parameters $\{a_i\}$ may be viewed as controls for investor fixed effects, while the parameters $\{b_k\}$ can be thought of as controls for group sizes.

We also model the count of acquaintances $y_{ik}$ in different groups as an increasing transformation of the number of stocks $z_{ik}$ an investor holds that belong to a given group\(^1\). Therefore, we have $y_{ik} = h(z_{ik})$, where $h$ is the increasing transformation. In our baseline setup, we assume $y_{ik}$ is proportional to $z_{ik}$, thus $y_{ik} = z_{ik}/c$ where $c$ is the transformation parameter.

---

\(^1\)By “belong to”, we mean a company is headquartered in a specific city (group).
2.2.2 The Null Model

If investors’ social connections are independently and identically formed as in the classical model of Erdős and Rényi (1959), the probability $p_{ij}$ of a link between an investor $i$ and a person $j$ from any particular group is the same for all pairs $(i, j)$. It then implies that $y_{ik}$ follows a Poisson distribution with its mean $\lambda_{ik} = ab_k$ equal to its variance. Furthermore, this model results in equal expected gregariousness $a_i$ for all investors and relative propensities $g_{ik}$ all equal to one. However, some investors may be more gregarious and have more social ties in expectation. To account for the variability in gregariousness, we let parameters $\{a_i\}$ vary across individual investors. Hence $y_{ik}$ follows a Poisson distribution with a mean $\lambda_{ik} = a_i b_k$, but relative propensities $g_{ik}$ are still all equal to one. We call this our null model.

2.2.3 The Overdispersed Model

An important departure from the null model is likely to occur if there are structured social networks formed in a non-i.i.d. fashion. To be more precise, we distinguish being part of a network from being merely gregarious. Being part of a network would mean that some investors have a non-i.i.d. relative propensity $\{g_{ik}\}$ to make connections to certain groups since the people in those groups constitute a structured network. As a result, we allow investors to differ not only in their gregariousness $\{a_i\}$, but also in their relative propensity $\{g_{ik}\}$ to accommodate for the effect of social influence. Consequently, $g_{ik} > 1$ if investor $i$ has a higher relative propensity to connect to people from group $k$ than an average investor in the population.

In the most general form where $\{g_{ik}\}$ varies for each $(i, k)$ pair, $y_{ik}$ is distributed as Poisson with a mean $\lambda_{ik} = a_i b_k g_{ik}$. Since it is not possible to identify each $g_{ik}$ later in the estimation if they are all different, for each group $k$, we let $g_{ik}$ follow a gamma distribution with a mean equal to 1 and a variance equal to $(\omega_k - 1)$ where
As a standard result, such a Poisson-gamma mixture leads to a (marginal) distribution/density for \( y_{ik} \) that is negative binomial (after integrating out \( g_{ik} \) and using an appropriate reparameterization)\(^2\)

\[
f(y_{ik}|a_i,b_k,\omega_k) = \frac{\Gamma(y_{ik} + \zeta_{ik})}{\Gamma(\zeta_{ik}) \Gamma(y_{ik} + 1)} \left( \frac{1}{\omega_k} \right)^{\zeta_{ik}} \left( \frac{\omega_k - 1}{\omega_k} \right)^{y_{ik}}, \tag{2.2.1}
\]

where \( \Gamma(\cdot) \) is the gamma function and \( \zeta_{ik} = a_i b_k / (\omega_k - 1) \). \( y_{ik} \) then has a mean equal to \( a_i b_k \) and a variance \( \omega_k a_i b_k \) that is greater than its mean \((\omega > 1)\). Therefore, we call this our overdispersed model. This is because variations in the relative propensities \( \{g_{ik}\} \) have resulted in overdispersions, i.e. \( y_{ik} \)'s variance exceeds its mean, in contrast to our Poisson null model with equal mean and variance \( a_i b_k \). Moreover, the \( \omega_k \)'s are called overdispersion parameters. They measure investors' non-identicalness in forming ties to certain groups and being part of structured social networks.

Our primary goal is to estimate the overdispersion parameters \( \{\omega_k\} \) from our overdispersed model and thus learn about diversities that exist in the formation of investors' social networks. As a byproduct, we also estimate the gregariousness parameters \( \{a_i\} \), representing the expected number of acquaintances know by investor \( i \), and the group size parameters \( \{b_k\} \) that gauge the proportion of social connections involving group \( k \).

### 2.2.4 Likelihood Function

Following from the density expression in (2.2.1), the likelihood function of \( y = \{y_{ik}\} \) in our overdispersed model is

---

\(^2\)The reason that it is not possible to identify all of the \( g_{ik} \)'s if each one of them is a different constant is because we only have \( N \times K \) number of observations of investors' stock picks. It is then not feasible to estimate \( N \times K \) number of \( g_{ik} \)'s with only \( N \times K \) number of data points.

\(^3\)For a reference on this type of Poisson-gamma mixture, see [Cameron and Trivedi (2005)](https://doi.org/10.1016/B978-0-7591-0327-0.50003-3), Chapter 20.
\[ p(y|a, b, \omega) = \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{\Gamma(y_{ik} + \zeta_{ik})}{\Gamma(\zeta_{ik}) \Gamma(y_{ik} + 1)} \left( \frac{1}{\omega_k} \right)^{\zeta_{ik}} \left( \frac{\omega_k - 1}{\omega_k} \right)^{y_{ik}}, \tag{2.2.2} \]

and the log-likelihood

\[ L = \sum_{i=1}^{N} \sum_{k=1}^{K} \left( LG(y_{ik} + \zeta_{ik}) - LG(\zeta_{ik}) - LG(y_{ik} + 1) - \zeta_{ik} \log(\omega_k) ight. \\
\left. + y_{ik} \left[ \log(\omega_k - 1) - \log(\omega_k) \right] \right), \]

where \( LG(\cdot) \) here denotes the log-gamma function \( \log(\Gamma(\cdot)) \) and \( \zeta_{ik} = a_i b_k / (\omega_k - 1) \) as stated before. Since we observe the stock holdings \( \{z_{ik}\} \) of investors in different groups and \( y_{ik} = h(z_{ik}) \) under the transformation \( h \), the (log-)likelihood function in terms of \( z = \{z_{ik}\} \) can then be expressed as

\[ L = \sum_{i=1}^{N} \sum_{k=1}^{K} \left( LG(h(z_{ik}) + \zeta_{ik}) - LG(\zeta_{ik}) - LG(h(z_{ik}) + 1) ight. \\
\left. - \zeta_{ik} \log(\omega_k) + h(z_{ik}) \left[ \log(\omega_k - 1) - \log(\omega_k) \right] \right). \tag{2.2.3} \]

The parameters of interest in our model are \( \theta = (\{\omega_k\}_{k=1}^{K}, \{a_i\}_{i=1}^{N}, \{b_k\}_{k=1}^{K})' \), a \((N + 2K) \times 1\) vector. We also have \( N \times K \) observations of \( z_{ik} \). In addition, in our baseline scenario, \( y_{ik} = z_{ik} / c \), so that the transformation parameter \( c \) will be estimated jointly with \( \theta \).

We shall estimate our model parameters using the method of maximum likelihood (MLE) based on (2.2.3), and we normalize \( \sum_{k=1}^{K} b_k \) to one to separately identify \( \{a_i\} \) and \( \{b_k\} \).

\[ ^4 \text{This normalization is needed because } \{a_i\} \text{ and } \{b_k\} \text{ enter the log-likelihood function together only as a joint entity } a_i b_k. \]

\[ ^5 \text{The intuition behind our MLE is similar to the idea behind the profile maximum likelihood technique (see, e.g. Murphy, Rossini, and van der Vaart (1997), Murphy and van der Vaart (2000)) used in transformation models where the underlying interest } y \text{ is a increasing transform of the observable variable } z. \text{ It can be understood as follows. For each possible value of } c, \text{ we first compute the maximum likelihood estimate of } \theta \text{ and the} \]

65
appendix, we will discuss the estimation mechanism and the associated results for the case where $h$ is a general increasing transform. Additionally, we will also consider the estimation for another interesting scenario where $z$ is a censored version of $y$.

### 2.3 Data and Estimation

#### 2.3.1 Data

Our data on stock holdings of mutual funds are obtained from the CDA/Spectrum Mutual fund Common Stock Holdings database provided by Thompson Reuters for the period 1990–2011. The database sources from semi-annually mandatory filings to the SEC and quarterly voluntary disclosure by mutual funds. We then merge the CDA/Spectrum database with survivorship-bias free CRSP mutual fund database. The CRSP mutual fund database provides information on a variety of mutual fund characteristics such as fund locations, investment objectives, monthly fund returns and assets under management. Additionally, we augment our mutual fund data with the database used in [Hong and Kostovetsky] (2012), which contains managerial demographic information on age, gender, name of undergraduate college, median SAT score of the undergraduate college attended, having a graduate degree or not, and political affiliation.

In order to keep only actively managed, non-sector domestic equity funds in our sample, we apply the following detailed screening procedures. Firstly, to exclude international, bond and index funds, we require (1) funds’ investment objective code reported by CDA/Spectrum to be aggressive growth, growth or growth and income, (2) their investment objectives in CRSP to be equity (E) and domestic (D) at the first two levels, (3) their CRSP objectives not to be EDCL, which indicates S&P500 index corresponding maximal value of the log-likelihood, then we find the value of $c$ such that the log-likelihood (2.2.3) attains the maximum with the associated $\theta$ estimate.
Secondly, to exclude sector funds, we require funds’ CRSP investment objectives at the third level to be either (C) or (Y). Thirdly, to exclude the possible presence of hedge funds, we require funds’ CRSP investment objectives not to be (H) or (S) at the last level. This screening leaves us with a sample of 1680 unique actively managed, non-sector domestic equity funds, or 111144 fund-quarter observations on stock holdings.\footnote{On average, approximately 1263 funds reported their portfolio holdings information in a single quarter. The frequency of reporting peaked at 2005Q2 when around 1550 funds filed their holdings information.}

Besides the institutional investor holdings data, we will also employ the retail investor holdings data from \textit{Barber and Odean (2001)} in the Appendix. Their dataset contains the monthly investments of 78,000 households between January 1991 and December 1996 from a large discount brokerage firm. It includes all investment accounts opened by each household at this discount brokerage firm, thus we aggregate the account information if a household had multiple accounts. Moreover, we focus on the the common stock investments of these households and exclude investments in mutual funds (both open- and closed-end), American depository receipts, warrants, and options. In addition, we only consider those households that had 10 or more stocks in their monthly portfolios on average. This is because the subsequent analyses performed on retail investors are only meaningful if their numbers of stock picks are not too small. Finally, we will use the demographic information contained in \textit{Barber and Odean (2001)}’s dataset on age, gender and household income for these retail investors.\footnote{Please refer to \textit{Barber and Odean (2001)} and also \textit{Barber and Odean (2000)} for a detailed description on the dataset.} Overall, we have 1609 unique retail investors (households) with demographic information and monthly holdings of 10 or more stocks on average, or about 93600 household-month observations.\footnote{At the monthly level, there are about 1300 households with a monthly average holdings of 10 or more stocks who had their holdings reported in the dataset.}
Next, we categorize the stocks held by mutual funds or retail investors into city groups and industry groups. We use the information on companies’ headquartered cities and their SIC industry codes that is available from the CRSP stock database. To obtain city groups for stocks, we match the city information of companies with the location information from COMPUSTAT, which maps cities into metropolitan statistical areas (MSAs).\footnote{We would like to thank Hyun-Soo Choi from the Singapore Management University for providing us with the MSA information.} On the other hand, to obtain industry groups, we utilize the industry definitions of Fama-French 30 industry portfolios and classify each stock into a particular industry.\footnote{The detailed information on Fama-French 30 industry portfolios is available at Kenneth R. French’s website\url{http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_30_ind_port.html}.}

We shall only consider the largest 20 cites (MSAs) or largest 20 industries in terms of the number of located companies. The reason is because the 20 largest groups, either cities or industries, already cover approximately 80% of all the stocks held by mutual funds or retail investors in our sample. There is no significant value added by allowing for more groups in our study. Hence in what follows, the number of groups $K$ is fixed at 20.

### 2.3.2 Rolling Estimation

We shall conduct a rolling maximum likelihood estimation on the model’s parameters $\theta = (\{\omega_k\}, \{a_i\}, \{b_k\})'$ and the transformation parameter $c$ using both the mutual fund and the retail investor holdings data. To be more precise, at each point in time (quarter for mutual funds and month for retail investors), we will use the past 12 quarters or months of holdings data as a rolling subsample to estimate $\theta$ and $c$ based on the log-likelihood (2.2.3). The observations $z_{ik}$ are then the number of unique stock picks from a group $k$ made by an investor $i$ during the past 12 quarters or
months. Therefore, our rolling estimates start at 1993Q1 (resp. Jan 1992) and end at 2011Q4 (resp. Dec 1996) for mutual funds (resp. retail investors).

After obtaining the rolling estimates, we will follow Fama and MacBeth (1973) in taking the time series means of the rolling estimates to form our overall estimates of $\theta$ and $c$. We denote these Fama-MacBeth estimates as our estimated parameter values.

### 2.4 Main Results

In this section, we report our main estimation results based on the mutual fund data. We shall concentrate on the results having cities as groups, and relegate the results with industry groups to the Appendix. We also perform the same set of analyses on retail investor data and the associated results will be shown in the Appendix as well.

#### 2.4.1 Transformation Parameters

Table 2.4.1 presents the estimates (Fama-MacBeth means of the quarterly rolling estimates) and related summary statistics of the transformation parameter for city groups. It shows that the mean of the transformation parameter $c$ is 1.37 with a standard deviation of .1 over time. There is not much variation over time in this parameter. This parameter estimates suggest that the number of contacts in a group is the number of holdings in that group divided by 1.37.

<table>
<thead>
<tr>
<th>City groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>City groups</td>
<td>1.37</td>
<td>0.10</td>
<td>1.38</td>
<td>1.13</td>
<td>1.87</td>
</tr>
</tbody>
</table>
2.4.2 Gregariousness Parameters

Next, Table 2.4.2 shows the summary statistics of the estimated values of the gregarious parameters \( a_i \) and Figure 2.4.1 illustrates the histogram of their Fama-MacBeth averages. We observe that the mean of \( a_i \) is 105 using city groups. This estimate can be interpreted literally as the typical manager having around 100 friends in the mutual fund industry overall and just in our sample. But there is a fairly sizeable standard deviation of around 120 or so friends. The estimate does not seem out of bounds relative to results in the sociology literature on the number of friends people have more generally.

Nevertheless, we view the estimates of gregariousness parameters as more akin to investor fixed effects for some investors having more stocks than others. They are separate from and do not affect our inference on whether investors belong to a network. In other words, having a lot of friends is not the same as being part of a network since it could also be affected by other factors such as investment style.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>City groups</td>
<td>105.0</td>
<td>116.0</td>
<td>76.8</td>
<td>0.6</td>
<td>1497.0</td>
</tr>
</tbody>
</table>

Note: summary statistics are based on individual time-series averages of \( a_i \).

2.4.3 Group Size Parameters

We then report the parameter estimates for \( b_k \) that gauge the relative sizes of cities. Table 2.4.3 and Figure 2.4.2 demonstrate the values of \( b_k \) for the 20 cities.

Two aspects of the estimates are noticeable. First, there are a few groups that have a much larger number of potential social connections attached to them comparing to the rest, for example, New York and Chicago. However, a group having a larger \( b_k \)
Figure 2.4.1: Histogram of $a_i$ estimates

![Histogram of $a_i$ estimates]

Table 2.4.3: Estimates of relative city size parameter $b_k$

<table>
<thead>
<tr>
<th>City</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>0.168</td>
<td>0.009</td>
<td>0.167</td>
<td>0.154</td>
<td>0.186</td>
</tr>
<tr>
<td>LA</td>
<td>0.066</td>
<td>0.003</td>
<td>0.066</td>
<td>0.060</td>
<td>0.074</td>
</tr>
<tr>
<td>Bos</td>
<td>0.069</td>
<td>0.003</td>
<td>0.069</td>
<td>0.064</td>
<td>0.075</td>
</tr>
<tr>
<td>SF</td>
<td>0.065</td>
<td>0.007</td>
<td>0.067</td>
<td>0.048</td>
<td>0.075</td>
</tr>
<tr>
<td>Chi</td>
<td>0.080</td>
<td>0.012</td>
<td>0.072</td>
<td>0.065</td>
<td>0.098</td>
</tr>
<tr>
<td>SJ</td>
<td>0.088</td>
<td>0.012</td>
<td>0.085</td>
<td>0.066</td>
<td>0.109</td>
</tr>
<tr>
<td>Dal</td>
<td>0.063</td>
<td>0.004</td>
<td>0.062</td>
<td>0.055</td>
<td>0.069</td>
</tr>
<tr>
<td>Hou</td>
<td>0.067</td>
<td>0.006</td>
<td>0.066</td>
<td>0.057</td>
<td>0.078</td>
</tr>
<tr>
<td>Phi</td>
<td>0.044</td>
<td>0.005</td>
<td>0.044</td>
<td>0.035</td>
<td>0.054</td>
</tr>
<tr>
<td>Was</td>
<td>0.047</td>
<td>0.004</td>
<td>0.047</td>
<td>0.041</td>
<td>0.054</td>
</tr>
<tr>
<td>Mia</td>
<td>0.020</td>
<td>0.001</td>
<td>0.020</td>
<td>0.017</td>
<td>0.023</td>
</tr>
<tr>
<td>Atl</td>
<td>0.037</td>
<td>0.002</td>
<td>0.038</td>
<td>0.033</td>
<td>0.040</td>
</tr>
<tr>
<td>Min</td>
<td>0.038</td>
<td>0.003</td>
<td>0.039</td>
<td>0.031</td>
<td>0.043</td>
</tr>
<tr>
<td>Den</td>
<td>0.020</td>
<td>0.003</td>
<td>0.020</td>
<td>0.014</td>
<td>0.027</td>
</tr>
<tr>
<td>SD</td>
<td>0.017</td>
<td>0.003</td>
<td>0.017</td>
<td>0.012</td>
<td>0.024</td>
</tr>
<tr>
<td>Stfd</td>
<td>0.029</td>
<td>0.007</td>
<td>0.025</td>
<td>0.022</td>
<td>0.039</td>
</tr>
<tr>
<td>Sea</td>
<td>0.024</td>
<td>0.002</td>
<td>0.025</td>
<td>0.021</td>
<td>0.027</td>
</tr>
<tr>
<td>Phx</td>
<td>0.018</td>
<td>0.003</td>
<td>0.019</td>
<td>0.011</td>
<td>0.021</td>
</tr>
<tr>
<td>SL</td>
<td>0.022</td>
<td>0.001</td>
<td>0.022</td>
<td>0.020</td>
<td>0.024</td>
</tr>
<tr>
<td>Det</td>
<td>0.015</td>
<td>0.003</td>
<td>0.015</td>
<td>0.012</td>
<td>0.021</td>
</tr>
</tbody>
</table>

does not imply that the degree of overdispersion in the group would necessarily be higher. To put it another way, just because a city has a substantial (relative) size does not mean that investors are more likely to form structured social networks with individuals from that group. Second, most of the standard deviations of the Fama-MacBeth $b_k$ estimates are small, implying that the sizes of various groups are stable across time.

Figure 2.4.2: Boxplot of $b_k$ estimates

Note: the green marker is the mean, the red line is the median, the box is the interquartile range, and the tails extend to the min and the max. For an explanation to the abbreviated city names, please refer to the note under Table 2.4.3.

2.4.4 Overdispersion Parameters

Now we turn to the estimates of our main parameter of interest – the degree of overdispersion $\omega_k$ among different groups. Recall that we introduced the overdispersions in our model in an attempt to estimate the variability in investors’ relative propensities to form ties to members of different groups. For groups where $\omega_k$ is closer to 1, it is
quite possible that there is no much variation in these relative propensities. However, larger values of $\omega_k$ would imply dissimilarities in individuals’ relative propensities to make connections.

Table 2.4.4 and Figure 2.4.3 display the estimated overdispersions $\omega_k$ for city groups. There are three evident features. Firstly, New York, Los Angeles, San Jose and San Diego stand out as the most overdispersed cities compared to the rest. This suggests that investors are more likely to form and be part of structured networks with acquaintances from these cities. Secondly, cities being larger (in terms of $b_k$) does not necessarily imply cities being more overdispersed. The correlation between the Fama-MacBeth estimates of $\omega_k$ and those of $b_k$ is about 0.37, and the rank correlation between them is merely about 0.23. Thirdly, although the majority of the cities do not exhibit a substantial degree of overdispersion, the $t$-statistics of testing the null Poisson distribution of $\omega = 1$ are all significant at the 5% level. Hence it implies that some investors do belong to certain integrated social networks even in the smaller cities (in terms of $b_k$) such as Miami or Minnesota.

The overdispersion estimates therefore signify that a number of investors live among some intricate social networks. They do not have to be the most gregarious investors, nor are they necessarily tied to the largest cities or industries.

2.4.5 Robustness Checks: Controlling for Local Bias and Fund Styles

Lastly, we report the results of two robustness checks on the overdispersion estimates using mutual fund data with city groups. The first check is a local-bias check, where we exclude managers’ local stock holdings from the estimation to ensure that our overdispersion results are not due to local biases. The other one is a verification where we dropped all growth funds from the estimation to ensure our results are not driven by fund styles.
Table 2.4.4: Estimates of overdispersion parameter $\omega_k$

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>1.502</td>
<td>0.302</td>
<td>1.396</td>
<td>1.013</td>
<td>2.386</td>
<td>5.621</td>
</tr>
<tr>
<td>LA</td>
<td>1.411</td>
<td>0.313</td>
<td>1.331</td>
<td>1.012</td>
<td>3.167</td>
<td>4.820</td>
</tr>
<tr>
<td>Bos</td>
<td>1.182</td>
<td>0.191</td>
<td>1.117</td>
<td>1.002</td>
<td>1.660</td>
<td>4.488</td>
</tr>
<tr>
<td>SF</td>
<td>1.161</td>
<td>0.187</td>
<td>1.092</td>
<td>1.002</td>
<td>1.812</td>
<td>3.985</td>
</tr>
<tr>
<td>Chi</td>
<td>1.191</td>
<td>0.163</td>
<td>1.188</td>
<td>1.001</td>
<td>1.740</td>
<td>5.223</td>
</tr>
<tr>
<td>SJ</td>
<td>2.625</td>
<td>0.663</td>
<td>2.550</td>
<td>1.471</td>
<td>5.396</td>
<td>12.059</td>
</tr>
<tr>
<td>Dal</td>
<td>1.024</td>
<td>0.064</td>
<td>1.007</td>
<td>1.001</td>
<td>1.445</td>
<td>3.351</td>
</tr>
<tr>
<td>Hou</td>
<td>1.034</td>
<td>0.071</td>
<td>1.016</td>
<td>1.002</td>
<td>1.479</td>
<td>4.475</td>
</tr>
<tr>
<td>Phi</td>
<td>1.048</td>
<td>0.076</td>
<td>1.018</td>
<td>1.001</td>
<td>1.424</td>
<td>3.996</td>
</tr>
<tr>
<td>Was</td>
<td>1.025</td>
<td>0.067</td>
<td>1.010</td>
<td>1.001</td>
<td>1.437</td>
<td>3.494</td>
</tr>
<tr>
<td>Mia</td>
<td>1.283</td>
<td>0.233</td>
<td>1.251</td>
<td>1.003</td>
<td>2.182</td>
<td>5.269</td>
</tr>
<tr>
<td>Atl</td>
<td>1.059</td>
<td>0.082</td>
<td>1.016</td>
<td>1.001</td>
<td>1.419</td>
<td>4.175</td>
</tr>
<tr>
<td>Min</td>
<td>1.206</td>
<td>0.203</td>
<td>1.081</td>
<td>1.002</td>
<td>1.662</td>
<td>4.304</td>
</tr>
<tr>
<td>Den</td>
<td>1.158</td>
<td>0.156</td>
<td>1.111</td>
<td>1.001</td>
<td>1.696</td>
<td>4.927</td>
</tr>
<tr>
<td>SD</td>
<td>1.624</td>
<td>0.355</td>
<td>1.613</td>
<td>1.018</td>
<td>3.496</td>
<td>9.343</td>
</tr>
<tr>
<td>Stfd</td>
<td>1.018</td>
<td>0.059</td>
<td>1.005</td>
<td>1.001</td>
<td>1.407</td>
<td>2.828</td>
</tr>
<tr>
<td>Sea</td>
<td>1.084</td>
<td>0.078</td>
<td>1.076</td>
<td>1.002</td>
<td>1.476</td>
<td>7.071</td>
</tr>
<tr>
<td>Phx</td>
<td>1.069</td>
<td>0.085</td>
<td>1.051</td>
<td>1.002</td>
<td>1.518</td>
<td>5.118</td>
</tr>
<tr>
<td>SL</td>
<td>1.050</td>
<td>0.077</td>
<td>1.018</td>
<td>1.004</td>
<td>1.422</td>
<td>3.989</td>
</tr>
<tr>
<td>Det</td>
<td>1.048</td>
<td>0.075</td>
<td>1.023</td>
<td>1.003</td>
<td>1.494</td>
<td>4.769</td>
</tr>
</tbody>
</table>

Note: for an explanation to the abbreviated city names, please refer to the note under Table 2.4.3. The t-statistics are adjusted for serial correlation using Newey and West (1987) lags of order twelve since we use past twelve quarters as our rolling estimation window size. They test the null hypothesis of $\omega_k = 1$ (Poisson) against the alternative of $\omega_k > 1$ (overdispersion).

As can be seen clearly from Table 2.4.5, the results from the two robustness checks echo our earlier findings in Table 2.4.4. Thus it implies that the overdispersions we find are not subject to the influence of either local biases or fund styles.

2.5 Networks, Demographics and Performances

2.5.1 What Explains Overdispersions?

Overdispersion is often treated as a nuisance rather than something fundamentally informative. But in the context of social networks, the Poisson null model has a very
Our major concern is that overdispersion might simply capture data error or some outliers but that are otherwise uninformative. To show that this is not the case, we relate our model’s output to demographic information about the investors as well as the performance of their investments.

We first conjecture that managers who went to colleges in a particular city are more likely to form social connections within that city via their college-friend networks.\cite{Cohen2008} To give an idea about the local university representation of managers, Table 2.5.1 tabulates the summary statistics of proportions of managers who attended un-

\footnote{This is related to the idea of fund manager-corporate director college links in Cohen, Frazzini, and Malloy (2008).}
Table 2.4.5: Robustness checks on overdispersion $\omega_k$ estimates

<table>
<thead>
<tr>
<th></th>
<th>No Local Response</th>
<th>No Growth Fund</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>t-stat</td>
</tr>
<tr>
<td>NY</td>
<td>1.548</td>
<td>5.768</td>
</tr>
<tr>
<td>LA</td>
<td>1.473</td>
<td>5.287</td>
</tr>
<tr>
<td>Bos</td>
<td>1.160</td>
<td>5.351</td>
</tr>
<tr>
<td>SF</td>
<td>1.141</td>
<td>3.082</td>
</tr>
<tr>
<td>Chi</td>
<td>1.170</td>
<td>5.731</td>
</tr>
<tr>
<td>SJ</td>
<td>2.560</td>
<td>11.413</td>
</tr>
<tr>
<td>Dal</td>
<td>1.027</td>
<td>4.012</td>
</tr>
<tr>
<td>Hou</td>
<td>1.064</td>
<td>4.819</td>
</tr>
<tr>
<td>Phi</td>
<td>1.038</td>
<td>4.128</td>
</tr>
<tr>
<td>Was</td>
<td>1.027</td>
<td>3.293</td>
</tr>
<tr>
<td>Mia</td>
<td>1.237</td>
<td>5.302</td>
</tr>
<tr>
<td>Atl</td>
<td>1.041</td>
<td>3.776</td>
</tr>
<tr>
<td>Min</td>
<td>1.165</td>
<td>3.978</td>
</tr>
<tr>
<td>Den</td>
<td>1.171</td>
<td>4.661</td>
</tr>
<tr>
<td>SD</td>
<td>1.600</td>
<td>9.337</td>
</tr>
<tr>
<td>Stfd</td>
<td>1.016</td>
<td>3.473</td>
</tr>
<tr>
<td>Sea</td>
<td>1.072</td>
<td>6.459</td>
</tr>
<tr>
<td>Phx</td>
<td>1.063</td>
<td>4.467</td>
</tr>
<tr>
<td>SL</td>
<td>1.034</td>
<td>4.476</td>
</tr>
<tr>
<td>Det</td>
<td>1.060</td>
<td>3.473</td>
</tr>
</tbody>
</table>

Note: this table demonstrates the results of two robustness checks on the overdispersion estimates, using mutual fund data with city groups. “No Local Response” denotes the case where managers’ local holdings have been dropped from the estimation, and “No Growth Fund” indicates that all growth funds have been dropped from the estimation.

dergraduate schools in different cities and Figure [2.5.1] gives the scatter plot of $\omega_k$’s against these proportions.

To examine our conjecture, we first analyze how the overdispersion parameter $\omega_k$ of each city depends on the number of managers who attended colleges in that city. We employ the following regression specification:

$$
\log(\omega_{k,t}) = \alpha + \beta prop_{k,t} + \gamma Ctrl_{k,t} + \eta DotCom_t + \varepsilon_{k,t},
$$

(2.5.1)
Table 2.5.1: Proportion of managers who attended undergraduate schools in a city

<table>
<thead>
<tr>
<th>City</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>11.78%</td>
<td>2.36%</td>
<td>11.67%</td>
<td>8.25%</td>
<td>15.08%</td>
</tr>
<tr>
<td>LA</td>
<td>3.75%</td>
<td>0.63%</td>
<td>3.62%</td>
<td>2.81%</td>
<td>4.69%</td>
</tr>
<tr>
<td>Bos</td>
<td>9.82%</td>
<td>1.96%</td>
<td>9.73%</td>
<td>6.87%</td>
<td>12.57%</td>
</tr>
<tr>
<td>SF</td>
<td>1.67%</td>
<td>0.24%</td>
<td>1.54%</td>
<td>1.32%</td>
<td>2.02%</td>
</tr>
<tr>
<td>Chi</td>
<td>2.95%</td>
<td>0.49%</td>
<td>2.82%</td>
<td>2.21%</td>
<td>3.69%</td>
</tr>
<tr>
<td>SJ</td>
<td>4.02%</td>
<td>0.67%</td>
<td>3.97%</td>
<td>3.02%</td>
<td>5.03%</td>
</tr>
<tr>
<td>Dal</td>
<td>1.32%</td>
<td>0.17%</td>
<td>1.17%</td>
<td>1.07%</td>
<td>1.57%</td>
</tr>
<tr>
<td>Hou</td>
<td>0.63%</td>
<td>0.07%</td>
<td>0.59%</td>
<td>0.53%</td>
<td>0.73%</td>
</tr>
<tr>
<td>Phi</td>
<td>4.23%</td>
<td>0.85%</td>
<td>4.11%</td>
<td>2.96%</td>
<td>5.50%</td>
</tr>
<tr>
<td>Was</td>
<td>1.46%</td>
<td>0.18%</td>
<td>1.40%</td>
<td>1.19%</td>
<td>1.73%</td>
</tr>
<tr>
<td>Mia</td>
<td>1.25%</td>
<td>0.16%</td>
<td>1.19%</td>
<td>1.02%</td>
<td>1.48%</td>
</tr>
<tr>
<td>Atl</td>
<td>0.56%</td>
<td>0.06%</td>
<td>0.53%</td>
<td>0.48%</td>
<td>0.64%</td>
</tr>
<tr>
<td>Min</td>
<td>1.88%</td>
<td>0.37%</td>
<td>1.84%</td>
<td>1.33%</td>
<td>2.43%</td>
</tr>
<tr>
<td>Den</td>
<td>0.98%</td>
<td>0.11%</td>
<td>0.87%</td>
<td>0.83%</td>
<td>1.13%</td>
</tr>
<tr>
<td>SD</td>
<td>0.63%</td>
<td>0.07%</td>
<td>0.63%</td>
<td>0.53%</td>
<td>0.73%</td>
</tr>
<tr>
<td>Stfd</td>
<td>0.52%</td>
<td>0.05%</td>
<td>0.50%</td>
<td>0.45%</td>
<td>0.59%</td>
</tr>
<tr>
<td>Sea</td>
<td>0.98%</td>
<td>0.11%</td>
<td>0.87%</td>
<td>0.83%</td>
<td>1.13%</td>
</tr>
<tr>
<td>Phx</td>
<td>0.28%</td>
<td>0.03%</td>
<td>0.27%</td>
<td>0.24%</td>
<td>0.32%</td>
</tr>
<tr>
<td>SL</td>
<td>0.56%</td>
<td>0.06%</td>
<td>0.56%</td>
<td>0.48%</td>
<td>0.64%</td>
</tr>
<tr>
<td>Det</td>
<td>0.21%</td>
<td>0.02%</td>
<td>0.19%</td>
<td>0.18%</td>
<td>0.24%</td>
</tr>
</tbody>
</table>

Figure 2.5.1: $\omega_k$ against local university representation, observations over all periods
where \( prop_{k,t} \) is the proportion of managers who attended undergraduate schools in city \( k \) at quarter \( t \), \( Ctrl_{k,t} \) is a vector of relevant control variables, and \( DotCom_t \) is a time dummy variable that equals one if quarter \( t \) belong to the dotcom bubble period of 1997Q1 to 2001Q4. We restrict the coefficients in front of \( DotCom_t \)'s to be all equal to keep it parsimonious. This is then performed as a panel regression with city fixed effects, and the estimation results are displayed in Table 2.5.2.

Table 2.5.2: Relationship between \( \omega_k \) and no. of managers graduating from city \( k \)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( prop )</td>
<td>2.061**</td>
<td>1.818**</td>
<td>1.198**</td>
<td>1.799**</td>
<td>1.310**</td>
<td>1.714**</td>
</tr>
<tr>
<td></td>
<td>(2.03)</td>
<td>(2.45)</td>
<td>(1.98)</td>
<td>(2.25)</td>
<td>(1.96)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>( StkNums )</td>
<td>0.044</td>
<td>0.030</td>
<td>0.047</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.51)</td>
<td>(1.09)</td>
<td>(0.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( RGDP )</td>
<td>0.219</td>
<td>0.201*</td>
<td>0.358**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.46)</td>
<td>(1.87)</td>
<td>(2.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( RepCity )</td>
<td>0.025</td>
<td>0.043</td>
<td>0.040</td>
<td>0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.84)</td>
<td>(1.28)</td>
<td>(1.28)</td>
<td>(0.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( DotCom )</td>
<td>0.029**</td>
<td>0.066**</td>
<td>0.077**</td>
<td>0.094**</td>
<td>0.053**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.27)</td>
<td>(2.44)</td>
<td>(2.51)</td>
<td>(2.35)</td>
<td>(2.30)</td>
<td></td>
</tr>
<tr>
<td>( MktCap )</td>
<td>-0.364</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.604)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: this table reports the estimation results of various forms of the regression \( \log(\omega_{k,t}) = \alpha + \beta prop_{k,t} + \gamma' Ctrl_{k,t} + \varepsilon_{k,t} \), with \( t \)-statistics shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively. \( StkNums \) denotes the number of stocks headquartered in a city. \( RGDP \) is the real GDP per capita of a city. \( RepCity \) is a dummy variable that equals one if the political affiliation of a city is Republican (based on the most recent election results). \( MktCap \) denotes the market-cap Herfindahl index of a city, which is a Herfindahl index computed based on the market capitalization of each individual stock located in that given city. The vector of controls \( Ctrl_{k,t} \) includes either some or all of the variables among \( StkNums \), \( RGDP \), \( RepCity \) and \( MktCap \), depending on the specifications. City-level (or MSA-level to be more precise) real GDP data are obtainable from Bureau of Economic Analysis. Political affiliations of different counties where cities are located are obtained from the website of David Leip’s atlas of U.S. presidential elections at [http://uselectionatlas.org/](http://uselectionatlas.org/).

It is clear from Table 2.5.2 that the number of managers who went to college in city \( k \) has a positive and statistically significant effect on the overdispersion \( \omega_k \) of that city. When only \( prop_{k,t} \) is included (column (1)), the coefficient estimate of \( \beta \)
equals 2.06 with a t-statistic equal to 2.03 (hence significant at the 5% level). It indicates that if the proportion of managers who attended colleges in city \( k \) increases by 1 percentage point, \( \omega_k \) would rise by 2.06%. The effect does not diminish if we include other city-level control variables such as the number of stocks headquartered in a city, the political affiliation of a city, and/or the real GDP per capita of a city (column (2) to (4)). The estimate of \( \beta \) is in the range of 1.2 to 2.0 and is always significant at the usual significance levels. This result thus supports our conjecture since if more managers ended up going to schools in city \( k \), more of them were likely to skew their social networks into that city, resulting in a higher overdispersion of acquaintance connections over there.

Furthermore, in column (6) of Table 2.5.2, we incorporate the market-cap Herfindahl index of a city (\( MktCap \)) as an extra control. The market-cap Herfindahl index is a Herfindahl index calculated based on the market capitalization of each individual stock located in a given city. The purpose to include this particular variable is to make sure that the overdispersion of a city is not \textit{mechanically} driven by the presence of a few dominant, large-sized companies with many other small companies, since it might be the case that most managers hold the stock of one large, well-known company while there are some managers holding stocks in a lot of the small companies. The displayed result shows that the presence of some large companies in a city does not have any positive effect on the city’s overdispersion, thus overdispersions could not be mechanically caused by such presence.

In addition, Table 2.5.2 also tells us that our overdispersion parameters \( \{\omega_k\} \) indeed capture social network effects, since their subtlety are not simply explained by other standard economic variables such as a city’s real GDP level or the number of stocks it has. As discussed before, overdispersion does not have to arise just because cities are large (in terms of number of stocks), and we can see here that cities being richer does not necessarily lead to more profound overdispersions either.
2.5.2 Managerial RPC Measure and Demographics

We next use our model to generate for each manager his propensity to be connected in a non-i.i.d. way to groups in these cities and relate these managerial RPC scores to managerial demographics. Recall that in our model, investors’ expected relative propensities to know a member in group $k$, $g_{ik} = \lambda_{ik}/(a_ib_k)$, cannot be identified or estimated individually. The RPC measures that we construct, $RPC_{ik} = y_{ik}/(a_ib_k)$, can then be considered as a proxy for $g_{ik}$. In other words, the RPC measures can be thought of as investors’ realized relative propensities to know a member from a specific group. The RPC measure for any investor in a particular group $k$ is computed as $g_{ik} = y_{ik}/(a_ib_k)$. Our model predicts that an investor should have an expected number of $a_ib_k$ connections in a given group, and that $y_{ik}$ should be very close to $a_ib_k$ if connections are formed in an Erdős and Rényi (1959) i.i.d. manner. On the other hand, an investor who holds a (much) higher number of stocks and hence knows a (much) larger number of acquaintances than expected in a group is more likely to be part of and has $g_{ik} > 1$ in that group, i.e. being part of that network.

Then we sum up investors’ RPC measures across all the groups, i.e. $g_{sum_i} = \sum_{k=1}^{K}[y_{ik}/(a_ib_k)]$. We shall label this the RPC score for each investor and will use $g_{sum_i}$ interchangeably with RPC score. Furthermore, if social connections are formed in an i.i.d. fashion so that $y_{ik}/(a_ib_k)$ are around 1 for each $(i,k)$ pair, we would expect all the RPC scores $\{g_{sum_i}\}$ to be close to 20 as we have $K = 20$ groups. However, if there are structured social networks among various groups, we would anticipate $g_{sum_i} > 20$ for an investor $i$ who is part of networks. This is because his underlying true $\sum_k g_{ik} = \lambda_{ik}/(a_ib_k)$ is likely to be greater than 20 as a result of social influences.

\footnote{Strictly speaking, this should be denoted as $\hat{g}_{ik} = y_{ik}/(\hat{a}_i\hat{b}_k)$ (where $\hat{a}_i$ and $\hat{b}_k$ are our estimates) since it is not the real $g_{ik}$ that equals $\lambda_{ik}/(a_ib_k)$. However, as stated before, we do not estimate individual $g_{ik}$ value in our model. Hence this notation is unlikely to cause any major confusion in what follows and we will denote $g_{ik}$ to mean $y_{ik}/(\hat{a}_i\hat{b}_k)$. In addition, we will use $g_{ik}$ and RPC measure interchangeably.}
Table 2.5.3 illustrates the correlations between our RPC measures \( g_{ik} \) and our gregariousness parameter estimates \( a_i \), using their respective Fama-MacBeth averages. It is clear from the Table that the correlations between \( g_{ik} \) and \( a_i \) are rather mild for city groups. Such weak correlations further confirm that being gregarious and being part of a network are not one and the same.

Table 2.5.3: Correlations between \( g_{ik} \) and \( a_i \), mutual funds, city groups

<table>
<thead>
<tr>
<th>NY</th>
<th>LA</th>
<th>Bos</th>
<th>SF</th>
<th>Chi</th>
<th>SJ</th>
<th>Dal</th>
<th>Hou</th>
<th>Phi</th>
<th>Was</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.043</td>
<td>0.040</td>
<td>0.027</td>
<td>0.147</td>
<td>-0.022</td>
<td>0.058</td>
<td>0.013</td>
<td>0.004</td>
<td>0.041</td>
<td>0.006</td>
</tr>
<tr>
<td>Mia</td>
<td>Atl</td>
<td>Min</td>
<td>Den</td>
<td>SD</td>
<td>Std</td>
<td>Sea</td>
<td>Phx</td>
<td>SL</td>
<td>Det</td>
</tr>
<tr>
<td>0.185</td>
<td>-0.017</td>
<td>-0.010</td>
<td>0.012</td>
<td>0.220</td>
<td>-0.030</td>
<td>0.027</td>
<td>0.202</td>
<td>0.033</td>
<td>0.110</td>
</tr>
</tbody>
</table>

Note: correlations are based on the Fama-MacBeth time-series means of \( g_{ik} \) (for each \( k \)) and \( a_i \). For explanations on abbreviated city group names, please refer to the notes under Table 2.4.3.

The summary statistics for our RPC scores \( g_{sum_i} \) are demonstrated in Table 2.5.4 as well as in Figure 2.5.2. We notice that the mean of RPC scores are close to 20 with city groupings, yet the standard deviation (around 5) is sizeable. Once more, this is another piece of evidence showing that certain investors have non-i.i.d. propensities to form ties with members from different cities. Furthermore, we find in Table 2.5.5 that for investors who have RPC scores greater than 20 (i.e. they are part of certain networks), the number of cities in which they have RPC measures larger than 1 is approximately nine. It indicates that for investors who are part of networks, they have higher propensities of making connections to certain cities only but not to all of the cities.

Table 2.5.4: Summary statistics of \( g_{sum_i} \), mutual funds

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>City groups</td>
<td>19.20</td>
<td>5.18</td>
<td>18.91</td>
<td>2.07</td>
<td>359.9</td>
</tr>
</tbody>
</table>

Note: summary statistics are based on individual time-series averages of \( g_{sum_i} \).
Having calculated our RPC measure and score for each manager, we are now in a position to study how they are related to demographic information about the mutual fund managers. To assess what type of managers are more likely to have higher RPC scores, we regress manager RPC score $gsum_i$ on a vector of manager demographic characteristics with the following regression equation:

$$
\log(gsum)_i = \alpha + \beta_1 MedSAT_i + \beta_2 Adv_i + \beta_3 Female_i + \beta_4 Old_i + \beta_5 Rep_i + \varepsilon_i, \quad (2.5.2)
$$

where MedSAT is the median SAT score of the undergraduate school that a manager went to, Adv is a dummy variable that equals 1 if a manager attended graduate school, Female is a dummy for being a female, Old is a dummy for being over 45 years old, and Rep is a dummy for being a Republican. We estimate this for each quarter and
then take the Fama-MacBeth time-series means and Newey-West standard errors of the resulting quarterly estimates. The estimation result is shown in Table 2.5.6.

Table 2.5.6: RPC Score and Manager Demographic Characteristics

<table>
<thead>
<tr>
<th>Dependent variable: log(gsum)</th>
<th>const</th>
<th>MedSAT</th>
<th>Adv</th>
<th>Female</th>
<th>Old</th>
<th>Rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.917***</td>
<td>-0.004</td>
<td>0.032***</td>
<td>0.022*</td>
<td>-0.029***</td>
<td>0.016***</td>
<td></td>
</tr>
<tr>
<td>(45.03)</td>
<td>(-0.48)</td>
<td>(10.12)</td>
<td>(3.38)</td>
<td>(-9.59)</td>
<td>(4.13)</td>
<td></td>
</tr>
</tbody>
</table>

Note: the table reports the Fama-MacBeth estimation results of the regression log(gsum) = α + β1 MedSAT + β2 Adv + β3 Female + β4 Old + β5 Rep + ε, with t-statistics based on Newey-West HAC standard errors shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

It can be seen from Table 2.5.6 that managers who went to graduate school have gsum’s that are 3.2% higher on average, and the effect is statistically significant. This suggests that managers with a graduate degree tend to be more socially connected, which is essentially consistent with our earlier conjecture of the college-friend networks. Furthermore, we find that female as well as Republican managers are likely to have more social ties than others, while older managers seem to have fewer. Having an advanced degree and being young are by far the strongest variables, which is consistent with the role of the influence of a city’s local universities in the mutual fund universe in explaining ω’s.

In the same vein as addressing what explains {ωk}, we next analyze whether the RPC measures {gi,k} of managers are higher in those cities where they attended college. For this purpose, we use the following regression

\[ g_{i,k} = \alpha + \beta D_{i,k} + \gamma SameStyle_{i,k} + \eta Local_{i,k} + \varepsilon_{i,k}, \]  

(2.5.3)

where \( D_{i,k} \) is a dummy variable that equals one if manager \( i \) went to college in city \( k \), \( SameStyle_{i,k} \) is the average \( g_{i,k} \) in city \( k \) of all funds that have the same CRSP fund style categorization as fund \( i \), and \( Local_{i,k} \) is a dummy variable that equals one
if fund \( i \) is also headquartered in city \( k \). If fund managers tilt their social connections towards their college cities, then we should find that \( \beta \) is positive and statistically significant.

We estimate the above regression quarter by quarter and then take the Fama-MacBeth time-series means and Newey-West standard errors of the quarterly estimates. The estimated value of \( \beta \) equals 0.17 and is significant at the 1% level (with a t-statistic equal to 6.93). Hence this tells us that the RPC measure of a typical fund is 0.17 point higher in the city where the manager attended college than in the rest of the cities on average. Therefore, it confirms our conjecture that managers tend to have more network ties from cities where they received college education and as a result, prefer to have disproportionally more stock picks in these cities as well.

### 2.5.3 Overdispersion in Counts vs. Portfolio Weights

One natural question that arises is whether fund managers also tilt their portfolio weights towards cities where they attended undergraduate schools, since the result we just obtained focuses on their RPC measures \( \{g_{i,k}\} \) (and hence stock picks). To this end, we estimate a similar regression of the form

\[
w_{i,k} = \alpha + \beta D_{i,k} + \gamma \text{SameStyle}_{i,k} + \eta \text{Local}_{i,k} + \varepsilon_{i,k},
\]

but with the dependent variable changed to portfolio weight \( w_{i,k} \) (measured in percentages) and the regressor \( \text{SameStyle}_{i,k} \) to the average \( w_{i,k} \) in city \( k \) of all funds that have the same CRSP fund style categorization as fund \( i \). What we find this time is that the estimated \( \beta \) equals 0.96, yet the t-statistic is only 0.982. This indicates that although managers overweight stocks to some extent in cities where they went to college, the overweighting effect is noisy and insignificant.
So why the social network effect shows up in our RPC measures \( \{g_{i,k}\} \) and hence stock picks, but not in portfolio weights? We consider the possible reason to be as follows. Social connections are defined in terms of counts, and our RPC measures (and also stock picks) are criteria that aim at capturing the numbers of such network ties. They are likely to be a cleaner and a more direct gauge of the links between people in networks within any particular city than portfolio weights. This is because portfolio weights might as well be affected by other factors such as the need for benchmarking that are not immediately related to network connections. To put it in another way, a manager may have quite a few stock picks in a city based on his social networks, yet he might not allocate a large portfolio weight to each of these stocks because his preference for benchmarking.

### 2.5.4 Managerial RPC and Fund Performance

Now we turn our attention to a more important question, which is how social networks, i.e. our RPC scores \( g_{sum} \), are related to mutual fund performances. There is a range of existing literature suggesting that social networks could exert positive values on investment performances, e.g. Hong, Kubik, and Stein (2005), Cohen, Frazzini, and Malloy (2008) and Feng and Seasholes (2008). Networks, such as knowing someone who is the CEO of a company, are not easy to obtain and may contain valuable investment information not accessible by the common public. Based on these ideas, the presence of structured networks in our model would imply that investors with RPC scores (much) larger than 20 should earn higher returns on their investment portfolios. Consequently, active equity funds with larger RPC scores should enjoy higher performances than their counterpart with smaller scores.

To test such implications, we utilize the following regression specification from Chen, Hong, Huang, and Kubik (2004) to examine the effect of social networks on mutual fund performance:
\[ pfm_{i,t} = \alpha + \beta RPC_{dummy_{i,t-1}} + x'_{i,t-1}\gamma + \varepsilon_{i,t}, \]  

(2.5.5)

where the dependent variable \( pfm_{i,t} \) is fund \( i \)'s net return in quarter \( t \). The regressor \( RPC_{dummy_{i,t-1}} \) is a dummy variable that equals one if fund \( i \)'s RPC score \( gsum_i \) is greater than 20 in quarter \( t \). Furthermore, \( x'_{i,t-1} \) is a vector of standard fund characteristic controls at \( t - 1 \). They include: (1) fund \( i \)'s lagged \( pfm \) at \( t - 1 \), (2) log of total net asset of fund \( i \), (3) log of one plus the total net asset of other funds in fund \( i \)'s family, (4) expense ratio of fund \( i \), (5) turnover ratio of fund \( i \), and (6) fund \( i \)'s age. Additionally, we also control for the gregariousness of a manager via his \( \log(a_i) \) and for whether a fund is located in a financial center (which is found by Christoffersen and Sarkissian (2009) to be associated with superior performance). They are contained in the regressor \( x \) as well. Finally, \( \alpha \) is a constant term and \( \varepsilon_{i,t} \) is a generic error term uncorrelated with all other explanatory variables in 2.5.5. We will carry out the regression 2.5.5 quarter by quarter and then take the Fama-MacBeth time-series means and Newey-West standard errors of the quarterly estimates.

Table 2.5.7 depicts our fund performance regression results with cities as groups. Most of the coefficients come in with the expected signs given the results in Chen, Hong, Huang, and Kubik (2004). For instance, fund size (log TNA) is associated with poor returns. There is persistence in performance and expense ratio is associated with poor returns. Moreover, we find consistent with Christoffersen and Sarkissian (2009) that a fund located in the financial center has superior performance.

Most relevant for us, it is evident that fund managers with higher RPC scores (i.e. with \( gsum > 20 \)) outperform substantially, by about 2.5% a year using city groups. However, we notice that being gregariousness does not necessarily lead to outperformance, as the coefficient on \( \log(a_i) \) is close to zero and is insignificant. Thus this difference in generating superior performance supports our prediction that being gregarious is not the same as being part of networks.
Table 2.5.7: RPC scores and mutual fund performances

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>const</strong></td>
<td>0.012*</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(1.06)</td>
</tr>
<tr>
<td><strong>FundReturn(_{t-1})</strong></td>
<td>0.063**</td>
<td>0.064*</td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td>(1.95)</td>
</tr>
<tr>
<td><strong>logTNA(_{t-1})</strong></td>
<td>-0.0010**</td>
<td>-0.0008*</td>
</tr>
<tr>
<td></td>
<td>(-2.23)</td>
<td>(-1.94)</td>
</tr>
<tr>
<td><strong>logFamSize(_{t-1})</strong></td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(0.65)</td>
</tr>
<tr>
<td><strong>ExpRatio(_{t-1})</strong></td>
<td>-0.003***</td>
<td>-0.002**</td>
</tr>
<tr>
<td></td>
<td>(-4.28)</td>
<td>(-2.15)</td>
</tr>
<tr>
<td><strong>Turnover(_{t-1})</strong></td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(-0.59)</td>
<td>(-0.64)</td>
</tr>
<tr>
<td><strong>FundAge(_{t-1})</strong></td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(-1.24)</td>
<td>(-1.38)</td>
</tr>
<tr>
<td><strong>gsum &gt; 20</strong></td>
<td>0.0057**</td>
<td>0.0050***</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(2.51)</td>
</tr>
<tr>
<td><strong>FinCenter</strong></td>
<td>0.0007***</td>
<td>0.0012***</td>
</tr>
<tr>
<td></td>
<td>(2.74)</td>
<td>(3.31)</td>
</tr>
<tr>
<td><strong>log((a_i))</strong></td>
<td>-0.0006</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>(-0.77)</td>
<td>(-0.86)</td>
</tr>
<tr>
<td><strong>ICI</strong></td>
<td>0.0078</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.93)</td>
<td></td>
</tr>
</tbody>
</table>

Note: this table reports the Fama-MacBeth estimates of the regression coefficients in specification 2.5.5 with \(t\)-statistics based on Newey-West HAC standard errors (of lag order 12) shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively. The dependent variable is fund’s net return at quarter \(t\). \(gsum > 20\) is a dummy variable that equals 1 if a fund’s RPC score \(gsum\) is larger than 20. FinCenter is a dummy variable that equals 1 if a fund is located in a financial center. The following six cities are defined to be financial centers: Boston, Chicago, Los Angeles, New York, Philadelphia, and San Francisco, in the spirit of Christoffersen and Sarkissian (2009). ICI denotes the Industry Concentration Index (ICI), which is constructed in a similar manner as in Kacperczyk, Sialm, and Zheng (2005). But for simplicity, we use an equally weighted index instead.
The findings on the influence of RPC scores on mutual fund performances here are reminiscent of the Industry Concentration Index (ICI) of Kacperczyk, Sialm, and Zheng (2005). They find that managers who hold concentrated positions outperform those that do not. Their interpretation on ICI is along the lines of closet indexing as those with concentrated portfolio holdings are less likely to be index-fund mimickers. However, our RPC scores and ICI are not very correlated and including ICI in the performance regression does not change the coefficient in front of our RPC scores. This is shown in column (2) of Table 2.5.7 where ICI is included as an extra explanatory variable in the regression specification of (2.5.5). In addition, our result that social networks are valuable to the tune of 2.5% a year for mutual fund returns is evocative of earlier studies documenting the value of investor and CEO networks such as Cohen, Frazzini, and Malloy (2008) and Engelberg, Gao, and Parsons (2012).

2.6 Extensions

In this section, we detail the estimation process for the general model in which (1) the transform $h : h(z) = y$ is of a general increasing form and (2) $z$ is a censored version of $y$.

2.6.1 A General Increasing Transform

When $h$ is a general increasing transform, we follow the approach of Murphy (1994) and Murphy (1995). We denote the unique realizations of $z_{ik}$ by $z^u_s$, where $s = 1, 2, \ldots$ are respectively the first, second, \ldots unique counts of $z$. A normalization is applied so that $h(0) = 0$, i.e. if $z_{ik}$ is zero, so is $y_{ik}$. Then we let the step sizes for the transformation $h$ be $\Delta_s := h(z^u_{s+1}) - h(z^u_s)$, $s = 1, 2, \ldots$. Next, we define a matching function
The matching function transforms the value of $z_{ik}$ by summing up all the $\Delta_s$'s up to the index $m$ where $z_{ik}$ equals the unique count number $z^u_{m+1}$. Thus we could use the matching function to map each $z_{ik}$ back to its corresponding value of $y_{ik}$. In other words, if we know all the $\{\Delta_s\}$, then we know the increasing transform $h$ (hence $y_{ik}$) for each unique $z_{ik}$ that we observe. Therefore, the step sizes $\{\Delta_s\}$ are the additional parameters that we would like to estimate.

In this situation, let us denote the dimensionality of $\Delta = \{\Delta_s\}$ by $J$, where $J$ equals the number of unique realizations of $z_{ik}$ minus one. Our main parameters of interest are $\theta = (\{\omega_k\}, \{a_i\}, \{b_k\})'$, an $N + 2K$ vector. Hence we will use an $N \times 2K$ number of observations $\{z_{ik}\}$ to estimate a $N + 2K + J$ number of parameters $\theta$ and $\Delta$. The log-likelihood function in terms of $z = \{z_{ik}\}$ then becomes

$$L = \sum_{i=1}^{N} \sum_{k=1}^{K} \left( LG(H(z_{ik}, \Delta) + \zeta_{ik}) - LG(\zeta_{ik}) - LG(H(z_{ik}, \Delta) + 1) - \zeta_{ik}\log(\omega_k) + H(z_{ik}, \Delta)[\log(\omega_k - 1) - \log(\omega_k)] \right),$$

(2.6.2)

where $H(z_{ik}, \Delta)$ is defined as above. To estimate the parameters under a general transform $h$, we will adopt a technique similar to the profile maximum likelihood (see, e.g. Murphy, Rossini, and van der Vaart (1997), Murphy and van der Vaart (2000)). This method has been used in transformation models where the underlying interest $y$ is a increasing transform of the observable variable $z$. The intuition has been discussed in the main body of this article and can be understood as follows. For each possible values of $\{\Delta_s\}$, we first compute the maximum likelihood estimate of $\theta$ and the corresponding maximal value of the log-likelihood, then we find the
values of \( \{\Delta_s\} \) such that the log-likelihood attains the maximum with the associated \( \theta \) estimate.

We estimated our model under a general transform \( h \) using the mutual fund data with city groups, and we illustrate the results here. Since most of the results are similar to the ones from our baseline case, we will keep the illustration parsimonious and concise.

To start with, Table 2.6.1 show the estimates for the gregariousness parameters. We can notice that compared to the baseline result in Table 2.4.2, the estimates are somewhat larger, but the differences are quite small\footnote{The estimates of \( b_k \) are very close to the baseline estimates and are hence not shown.}

\[
\text{Table 2.6.1: Estimates of } a_i \text{ under a general } h \\
\begin{array}{cccccc}
\text{mean} & \text{s.d.} & \text{med} & \text{min} & \text{max} \\
131.0 & 142.5 & 104.8 & 0.6 & 1571.6 \\
\end{array}
\]

Next, the overdispersion parameter estimates are reported in Table 2.6.2. We observe that the overdispersions have increased in magnitude. However, New York, Los Angeles, San Jose and San Diego remain as the cities that appear to have substantial overdispersions, which is in line with our baseline results.

Lastly, we can see from Table 2.6.3 and Table 2.6.4 that the RPC scores based on the current estimates under a general transform \( h \) are similarly related to managers’ demographic characteristics as in the baseline case, and these RPC scores have a positive impact on mutual fund returns too. In particular, the outperformance number from the RPC score is 2.54\% a year, consistent with what we found earlier in our baseline model.
Table 2.6.2: Estimates of $\omega_k$ under a general $h$

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>1.81</td>
<td>0.36</td>
<td>1.79</td>
<td>1.03</td>
<td>2.51</td>
<td>6.06</td>
</tr>
<tr>
<td>LA</td>
<td>1.77</td>
<td>0.38</td>
<td>1.52</td>
<td>1.03</td>
<td>3.51</td>
<td>4.93</td>
</tr>
<tr>
<td>Bos</td>
<td>1.62</td>
<td>0.23</td>
<td>1.40</td>
<td>1.01</td>
<td>1.94</td>
<td>4.68</td>
</tr>
<tr>
<td>SF</td>
<td>1.55</td>
<td>0.26</td>
<td>1.37</td>
<td>1.01</td>
<td>2.35</td>
<td>4.38</td>
</tr>
<tr>
<td>Chi</td>
<td>1.54</td>
<td>0.20</td>
<td>1.57</td>
<td>1.01</td>
<td>2.19</td>
<td>5.39</td>
</tr>
<tr>
<td>SJ</td>
<td>2.93</td>
<td>0.71</td>
<td>2.68</td>
<td>1.48</td>
<td>5.58</td>
<td>12.22</td>
</tr>
<tr>
<td>Dal</td>
<td>1.18</td>
<td>0.09</td>
<td>1.15</td>
<td>1.01</td>
<td>1.69</td>
<td>3.44</td>
</tr>
<tr>
<td>Hou</td>
<td>1.19</td>
<td>0.12</td>
<td>1.16</td>
<td>1.02</td>
<td>1.88</td>
<td>4.76</td>
</tr>
<tr>
<td>Phi</td>
<td>1.22</td>
<td>0.14</td>
<td>1.26</td>
<td>1.02</td>
<td>1.70</td>
<td>4.45</td>
</tr>
<tr>
<td>Was</td>
<td>1.17</td>
<td>0.12</td>
<td>1.12</td>
<td>1.01</td>
<td>1.93</td>
<td>3.87</td>
</tr>
<tr>
<td>Mia</td>
<td>1.48</td>
<td>0.28</td>
<td>1.58</td>
<td>1.01</td>
<td>2.35</td>
<td>5.56</td>
</tr>
<tr>
<td>Atl</td>
<td>1.28</td>
<td>0.14</td>
<td>1.24</td>
<td>1.01</td>
<td>1.60</td>
<td>4.60</td>
</tr>
<tr>
<td>Min</td>
<td>1.56</td>
<td>0.25</td>
<td>1.45</td>
<td>1.01</td>
<td>1.83</td>
<td>4.53</td>
</tr>
<tr>
<td>Den</td>
<td>1.33</td>
<td>0.20</td>
<td>1.27</td>
<td>1.01</td>
<td>2.04</td>
<td>5.28</td>
</tr>
<tr>
<td>SD</td>
<td>1.97</td>
<td>0.40</td>
<td>1.68</td>
<td>1.03</td>
<td>3.86</td>
<td>9.64</td>
</tr>
<tr>
<td>Stfd</td>
<td>1.18</td>
<td>0.10</td>
<td>1.14</td>
<td>1.03</td>
<td>1.63</td>
<td>3.14</td>
</tr>
<tr>
<td>Sea</td>
<td>1.31</td>
<td>0.14</td>
<td>1.28</td>
<td>1.02</td>
<td>1.82</td>
<td>7.19</td>
</tr>
<tr>
<td>Phx</td>
<td>1.28</td>
<td>0.15</td>
<td>1.19</td>
<td>1.02</td>
<td>1.74</td>
<td>5.54</td>
</tr>
<tr>
<td>SL</td>
<td>1.27</td>
<td>0.13</td>
<td>1.11</td>
<td>1.02</td>
<td>1.80</td>
<td>4.48</td>
</tr>
<tr>
<td>Det</td>
<td>1.26</td>
<td>0.14</td>
<td>1.16</td>
<td>1.01</td>
<td>1.89</td>
<td>5.18</td>
</tr>
</tbody>
</table>

Table 2.6.3: RPC scores and demographics of mutual fund managers, general $h$

<table>
<thead>
<tr>
<th>Dependent variable: $\log(gsum)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
</tr>
<tr>
<td>2.904***</td>
</tr>
<tr>
<td>(51.74)</td>
</tr>
</tbody>
</table>
Table 2.6.4: RPC scores and mutual fund performances, general $h$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$const$</td>
<td>0.011</td>
<td>(1.49)</td>
</tr>
<tr>
<td>$FundReturn_{t-1}$</td>
<td>0.064**</td>
<td>(2.38)</td>
</tr>
<tr>
<td>$logTNA_{t-1}$</td>
<td>-0.0011**</td>
<td>(-2.28)</td>
</tr>
<tr>
<td>$logFamSize_{t-1}$</td>
<td>0.0000</td>
<td>(0.69)</td>
</tr>
<tr>
<td>$ExpRatio_{t-1}$</td>
<td>-0.003***</td>
<td>(-4.20)</td>
</tr>
<tr>
<td>$Turnover_{t-1}$</td>
<td>0.000</td>
<td>(-0.49)</td>
</tr>
<tr>
<td>$FundAge_{t-1}$</td>
<td>0.000</td>
<td>(-1.32)</td>
</tr>
<tr>
<td>$gsum &gt; 20$</td>
<td>0.0063**</td>
<td>(2.29)</td>
</tr>
<tr>
<td>FinCenter</td>
<td>0.0008***</td>
<td>(2.82)</td>
</tr>
<tr>
<td>$log(a_i)$</td>
<td>-0.0003</td>
<td>(-0.67)</td>
</tr>
</tbody>
</table>

2.6.2 A Censored Model

In the second part of this section, we discuss the scenario where our observable $\{z_{ik}\}$ is a censored version of the underlying $\{y_{ik}\}$. To be more specific, we consider a censoring threshold $U$, such that

$$
\begin{align*}
  y_{ik} &= z_{ik} & \text{if we observe } z_{ik} \text{ is strictly less that } U \\
  y_{ik} &\geq U & \text{if we observe } z_{ik} \text{ is greater than or equal to } U.
\end{align*}
$$

(2.6.3)
Therefore, with such a censoring, the log-likelihood becomes

$$\mathcal{L} = \sum_{i=1}^{N} \sum_{k=1}^{K} \left[ d_{ik} \left( LG(z_{ik} + \zeta_{ik}) - LG(\zeta_{ik}) - LG(z_{ik} + 1) - \zeta_{ik} \log(\omega_k) \right) \\
+ z_{ik} \left[ \log(\omega_k - 1) - \log(\omega_k) \right] \right] + (1 - d_{ik}) \log \left( 1 - P(U|\zeta_{ik}, \omega_k) \right)$$

(2.6.4)

where $d_{ik}$ is an indicator variable such that $d_{ik} = 1$ if $z_{ik} < U$, and $P(U|\zeta_{ik}, \omega_k)$ is the negative binomial cumulative distribution function

$$P(U|\zeta_{ik}, \omega_k) = \sum_{x=0}^{U} \frac{\Gamma(x + \zeta_{ik})}{\Gamma(\zeta_{ik}) \Gamma(x + 1)} \left( \frac{1}{\omega_k} \right)^{\zeta_{ik}} \left( \frac{\omega_k - 1}{\omega_k} \right)^{x}.$$

(2.6.5)

This model can then be estimated using the usual maximum likelihood method. Additionally, one could also adapt it to the more complex case where $y$ is an increasing transform of $z$ if $z < U$.

We estimated the censored version of our model as illustrated above, with a range of censoring levels $U = 50, 75, 100, 125$. In general, we find that the estimation results under censoring are all qualitatively similar to the results from our baseline model. The main noticeable difference is that the estimates for the overdispersion parameters $\{\omega_k\}$ become larger due to the effect of censoring.

### 2.7 Conclusion

In this paper, we extend the overdispersed Poisson regression models used in statistics and sociology to study social networks in finance. Even though detailed network data is not typically available in finance settings, we show that we can model the count of an investor’s social connections in different groups, such as cities or industries, as proportional to the number of stocks an investor holds that are headquartered in these cities or part of these industries. When connections are formed randomly,
the count of these connections in any group follows a Poisson distribution. When connections are formed in a non-i.i.d. manner, the count of these connections in any group follows an overdispersed Poisson. Using data from institutional and retail investors’ holdings, we estimate the degree of overdispersion for different groups. We find substantial overdispersion for some city groups such as San Diego, Los Angeles and San Jose, and for some industry groups such as Finance and Utilities. Our model also allows us to predict the relative propensity of any investor to be connected to a group. We show that these propensities are tied to investor demographics and are highly correlated with superior investor performance, suggesting that such networks are valuable.

These models can be used to study any financial network where investment data are available. Our set-up can be easily applied to many other contexts in finance such as banking networks where one can count trades between a bank with other banks in different countries or lending volume between banks and companies in different industries. In other words, while we do not have answers to survey questions about how many people investors know in different groups, we can proxy for answers to these questions by counting their investments across different categories. In short, the value of our set-up is that it connects the study of social networks in finance, which is hampered by limited data on social connections, to the study of social networks in statistics and sociology, which is hampered by data on performance. We hope this application of count models of social networks in financial markets might be useful for many different endeavors.
2.8 Appendix

In the appendix, we first display the results for industry groups using mutual fund holdings data. Then we depict our findings using retail investor data with both city and industry groups.

2.8.1 Mutual Fund Results for Industry Groups

Table 2.8.1 presents the estimates of the transformation parameter for mutual funds with industry groupings, while Table 2.8.2 and Figure 2.8.1 show the estimates of the gregariousness parameters $a_i$. Compared to the city-group results in the main paper, we can see that the transformation parameter and the gregariousness parameters are larger when industry groups are considered for stocks. However, the differences between the estimates in the two group classifications are not substantial.

Table 2.8.1: Summary statistics of transformation parameter estimates, mutual funds

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry groups</td>
<td>1.53</td>
<td>0.16</td>
<td>1.55</td>
<td>1.18</td>
<td>1.79</td>
</tr>
</tbody>
</table>

Table 2.8.2: Summary statistics, estimates of $a_i$, mutual funds

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry groups</td>
<td>124.7</td>
<td>134.6</td>
<td>87.7</td>
<td>0.2</td>
<td>1220.9</td>
</tr>
</tbody>
</table>

Note: summary statistics are based on individual time-series averages of $a_i$.

Table 2.8.3 and Figure 2.8.2 show the estimated values of $b_k$ for the 20 industries. Similar to the city results, there are a few groups that have a much larger number of potential social connections attached to them comparing to the rest, and the sizes of various groups are stable across time.

Table 2.8.4 and Figure 2.8.3 present the estimated overdispersions $\omega_k$ for industry groups. Compared to the $\omega_k$ estimates from city groups, we notice a couple of key
Figure 2.8.1: Histogram of $a_i$ estimates, industry groups, mutual funds

![Histogram of $a_i$ estimates](image1)

Figure 2.8.2: Boxplot of $b_k$ estimates, industry groups, mutual funds

![Boxplot of $b_k$ estimates](image2)

Note: the green marker is the mean, the red line is the median, the box is the interquartile range, and the tails extend to the min and the max. For an explanation to the abbreviated industry names, please refer to the note under Table 2.8.3
Table 2.8.3: Estimates of $b_k$, industry groups, mutual funds

<table>
<thead>
<tr>
<th>Industry</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>0.172</td>
<td>0.009</td>
<td>0.173</td>
<td>0.152</td>
<td>0.185</td>
</tr>
<tr>
<td>Serv</td>
<td>0.131</td>
<td>0.024</td>
<td>0.140</td>
<td>0.083</td>
<td>0.164</td>
</tr>
<tr>
<td>Hlth</td>
<td>0.103</td>
<td>0.010</td>
<td>0.105</td>
<td>0.086</td>
<td>0.120</td>
</tr>
<tr>
<td>BsEq</td>
<td>0.126</td>
<td>0.012</td>
<td>0.124</td>
<td>0.105</td>
<td>0.151</td>
</tr>
<tr>
<td>Rtail</td>
<td>0.073</td>
<td>0.006</td>
<td>0.072</td>
<td>0.066</td>
<td>0.085</td>
</tr>
<tr>
<td>Whsl</td>
<td>0.035</td>
<td>0.002</td>
<td>0.035</td>
<td>0.031</td>
<td>0.040</td>
</tr>
<tr>
<td>Telcm</td>
<td>0.038</td>
<td>0.007</td>
<td>0.037</td>
<td>0.028</td>
<td>0.054</td>
</tr>
<tr>
<td>Oil</td>
<td>0.046</td>
<td>0.006</td>
<td>0.044</td>
<td>0.038</td>
<td>0.058</td>
</tr>
<tr>
<td>EEq</td>
<td>0.017</td>
<td>0.004</td>
<td>0.017</td>
<td>0.013</td>
<td>0.024</td>
</tr>
<tr>
<td>FabP</td>
<td>0.041</td>
<td>0.005</td>
<td>0.042</td>
<td>0.035</td>
<td>0.050</td>
</tr>
<tr>
<td>Cnstr</td>
<td>0.025</td>
<td>0.003</td>
<td>0.025</td>
<td>0.022</td>
<td>0.030</td>
</tr>
<tr>
<td>Trans</td>
<td>0.026</td>
<td>0.002</td>
<td>0.027</td>
<td>0.021</td>
<td>0.028</td>
</tr>
<tr>
<td>Game</td>
<td>0.018</td>
<td>0.002</td>
<td>0.018</td>
<td>0.016</td>
<td>0.022</td>
</tr>
<tr>
<td>Meal</td>
<td>0.021</td>
<td>0.002</td>
<td>0.022</td>
<td>0.017</td>
<td>0.025</td>
</tr>
<tr>
<td>Util</td>
<td>0.029</td>
<td>0.003</td>
<td>0.029</td>
<td>0.024</td>
<td>0.034</td>
</tr>
<tr>
<td>Food</td>
<td>0.021</td>
<td>0.004</td>
<td>0.020</td>
<td>0.017</td>
<td>0.031</td>
</tr>
<tr>
<td>Hshld</td>
<td>0.020</td>
<td>0.004</td>
<td>0.018</td>
<td>0.016</td>
<td>0.028</td>
</tr>
<tr>
<td>Chems</td>
<td>0.023</td>
<td>0.005</td>
<td>0.022</td>
<td>0.017</td>
<td>0.033</td>
</tr>
<tr>
<td>Book</td>
<td>0.015</td>
<td>0.004</td>
<td>0.013</td>
<td>0.009</td>
<td>0.025</td>
</tr>
<tr>
<td>Paper</td>
<td>0.018</td>
<td>0.005</td>
<td>0.015</td>
<td>0.013</td>
<td>0.028</td>
</tr>
</tbody>
</table>


Similarities. One is that there are some industries that show up as being much more overdispersed, such as Finance, Service and Utility. The other is that the $t$-statistics of testing the null Poisson distribution of $\omega = 1$ are also all significant at the 5% level, indicating the existence of delicate networks within each of the industry groups. On the other hand, it is observable that there is more overdispersion along industry classifications than city categories, which suggests that network connections are more prevalent along industry lines than city lines. Nonetheless, overdispersions are present independent of the type of group categorizations under consideration.
Table 2.8.4: Estimates of $\omega_k$, industry groups, mutual funds

<table>
<thead>
<tr>
<th>Industry</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>4.398</td>
<td>0.558</td>
<td>4.500</td>
<td>3.163</td>
<td>5.512</td>
<td>25.176</td>
</tr>
<tr>
<td>Serv</td>
<td>4.785</td>
<td>1.265</td>
<td>4.516</td>
<td>3.065</td>
<td>7.975</td>
<td>11.760</td>
</tr>
<tr>
<td>Hlth</td>
<td>2.225</td>
<td>0.427</td>
<td>2.209</td>
<td>1.641</td>
<td>3.927</td>
<td>14.404</td>
</tr>
<tr>
<td>BsEq</td>
<td>2.067</td>
<td>0.690</td>
<td>1.921</td>
<td>1.259</td>
<td>4.846</td>
<td>6.890</td>
</tr>
<tr>
<td>Rtail</td>
<td>2.043</td>
<td>0.093</td>
<td>1.009</td>
<td>1.001</td>
<td>1.524</td>
<td>3.342</td>
</tr>
<tr>
<td>Whsl</td>
<td>1.452</td>
<td>0.329</td>
<td>1.438</td>
<td>1.003</td>
<td>2.590</td>
<td>6.395</td>
</tr>
<tr>
<td>Telcm</td>
<td>2.069</td>
<td>0.718</td>
<td>1.777</td>
<td>1.484</td>
<td>4.939</td>
<td>6.222</td>
</tr>
<tr>
<td>Oil</td>
<td>1.215</td>
<td>0.181</td>
<td>1.185</td>
<td>1.008</td>
<td>2.135</td>
<td>6.073</td>
</tr>
<tr>
<td>EEq</td>
<td>1.445</td>
<td>0.353</td>
<td>1.399</td>
<td>1.005</td>
<td>2.669</td>
<td>5.700</td>
</tr>
<tr>
<td>FabP</td>
<td>1.024</td>
<td>0.087</td>
<td>1.004</td>
<td>1.001</td>
<td>1.587</td>
<td>1.873</td>
</tr>
<tr>
<td>Cnstr</td>
<td>1.455</td>
<td>0.326</td>
<td>1.429</td>
<td>1.006</td>
<td>2.570</td>
<td>6.133</td>
</tr>
<tr>
<td>Trans</td>
<td>1.031</td>
<td>0.126</td>
<td>1.007</td>
<td>1.002</td>
<td>2.023</td>
<td>1.809</td>
</tr>
<tr>
<td>Game</td>
<td>1.031</td>
<td>0.073</td>
<td>1.009</td>
<td>1.002</td>
<td>1.400</td>
<td>3.015</td>
</tr>
<tr>
<td>Meal</td>
<td>1.109</td>
<td>0.135</td>
<td>1.035</td>
<td>1.002</td>
<td>1.784</td>
<td>4.197</td>
</tr>
<tr>
<td>Util</td>
<td>4.330</td>
<td>0.748</td>
<td>4.428</td>
<td>2.331</td>
<td>5.796</td>
<td>20.664</td>
</tr>
<tr>
<td>Food</td>
<td>1.215</td>
<td>0.145</td>
<td>1.189</td>
<td>1.004</td>
<td>1.879</td>
<td>7.221</td>
</tr>
<tr>
<td>Hshld</td>
<td>1.020</td>
<td>0.055</td>
<td>1.006</td>
<td>1.002</td>
<td>1.316</td>
<td>2.418</td>
</tr>
<tr>
<td>Chems</td>
<td>1.219</td>
<td>0.262</td>
<td>1.149</td>
<td>1.003</td>
<td>2.221</td>
<td>3.670</td>
</tr>
<tr>
<td>Book</td>
<td>1.216</td>
<td>0.097</td>
<td>1.234</td>
<td>1.002</td>
<td>1.470</td>
<td>13.072</td>
</tr>
<tr>
<td>Paper</td>
<td>1.203</td>
<td>0.135</td>
<td>1.219</td>
<td>1.006</td>
<td>1.794</td>
<td>7.923</td>
</tr>
</tbody>
</table>

Note: for an explanation to the abbreviated industry names, please refer to the note under Table 2.8.3. The $t$-statistics are adjusted for serial correlation using Newey and West (1987) lags of order twelve since we use past twelve quarters as our rolling estimation window size. They test the null hypothesis of $\omega_k = 1$ (Poisson) against the alternative of $\omega_k > 1$ (overdispersion).

Table 2.8.5 illustrates the correlations between the RPC measures $g_{ik}$ and the gregariousness parameter estimates $a_i$ of mutual funds for industry groups, while the summary statistics of the RPC scores $gsum_i$ are demonstrated in Table 2.8.6 as well as in Figure 2.8.4 and Table 2.8.7. They resemble the results in the main paper with cities as groups. In addition, we compute the correlation between fund managers’ RPC scores using city groups and those using industry groups. Interestingly, managers’ scores from city groups are not very correlated with their scores from industry groups and the correlation coefficient is approximately 0.2. Thus it suggests that city and industry networks can be dissimilar for investors.
Figure 2.8.3: Boxplot of $\omega_k$ estimates, industry groups, mutual funds

Note: the green marker is the mean, the red line is the median, the box is the interquartile range, and the tails extend to the min and the max. For an explanation to the abbreviated industry names, please refer to the note under Table 2.8.3.

Table 2.8.5: Correlations between $g_{ik}$ and $a_i$, mutual funds, industry groups

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>Fin</th>
<th>Serv</th>
<th>Hlth</th>
<th>BsEq</th>
<th>Retail</th>
<th>Whsl</th>
<th>Telcm</th>
<th>Oil</th>
<th>EEq</th>
<th>FabP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>0.020</td>
<td>0.035</td>
<td>-0.006</td>
<td>0.016</td>
<td>0.090</td>
<td>0.328</td>
<td>-0.040</td>
<td>0.003</td>
<td>0.165</td>
<td>0.003</td>
</tr>
<tr>
<td>Cntr</td>
<td>0.291</td>
<td>0.200</td>
<td>0.148</td>
<td>0.222</td>
<td>0.003</td>
<td>-0.039</td>
<td>-0.024</td>
<td>0.016</td>
<td>0.057</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Note: correlations are based on the Fama-MacBeth time-series means of $g_{ik}$ (for each k) and $a_i$. For explanations on abbreviated industry group names, please refer to the notes under Table 2.8.3.

Table 2.8.6: Summary statistics of $g_{sum_i}$, mutual funds

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>19.61</td>
<td>4.79</td>
<td>19.28</td>
<td>0.30</td>
<td>167.5</td>
</tr>
</tbody>
</table>

Note: summary statistics are based on individual time-series averages of $g_{sum_i}$. 

99
Table 2.8.7: Number of groups with $g_{ik} > 1$ for mutual funds having $gsum > 20$

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry groups</td>
<td>9.02</td>
<td>0.50</td>
<td>9.14</td>
<td>7.08</td>
<td>9.94</td>
</tr>
</tbody>
</table>

Note: the table contains the summary statistics on the average of the number of groups where $g_{ik} > 1$, for mutual funds that have $gsum > 20$.

Table 2.8.8 depicts our fund performance regression results using industry groups. Fund managers with higher RPC scores outperforms by about 2% per annum. As in the case of city groups, this result does not diminish even if ICI is included in the performance regression. Moreover, when we include both city RPC scores and industry RPC scores in the regression, both scores entail significant outperformance numbers. Nevertheless, the effect from city RPC scores is somewhat larger and more significant than the effect from industry RPC scores.

### 2.8.2 Retail Investor Results for City Groups

To ensure that our results on institutional investor networks are not driven by unique mutual fund industry considerations, we perform the same set of analyses on Barber and Odean (2001)’s retail investor stock holdings data. The relevant results with cities as groups are reported first, and the industry-group results are shown next.
Table 2.8.8: RPC scores and mutual fund performances

<table>
<thead>
<tr>
<th></th>
<th>Industry</th>
<th>Both gsum's</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{const})</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(1.16)</td>
<td>(1.10)</td>
</tr>
<tr>
<td>(\text{FundReturn}_{t-1})</td>
<td>0.061**</td>
<td>0.061**</td>
</tr>
<tr>
<td></td>
<td>(2.36)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>(\log TNA_{t-1})</td>
<td>-0.0011**</td>
<td>-0.0011**</td>
</tr>
<tr>
<td></td>
<td>(-2.05)</td>
<td>(-2.16)</td>
</tr>
<tr>
<td>(\log \text{FamSize}_{t-1})</td>
<td>0.0000</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>(\text{ExpRatio}_{t-1})</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(-4.43)</td>
<td>(-4.33)</td>
</tr>
<tr>
<td>(\text{Turnover}_{t-1})</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(-0.32)</td>
<td>(-0.46)</td>
</tr>
<tr>
<td>(\text{FundAge}_{t-1})</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(-1.25)</td>
<td>(-1.06)</td>
</tr>
<tr>
<td>(\text{gsum} &gt; 20)</td>
<td>0.0044**</td>
<td>0.0040**</td>
</tr>
<tr>
<td></td>
<td>(1.98)</td>
<td>(2.09)</td>
</tr>
<tr>
<td>(\text{FinCenter})</td>
<td>0.0009***</td>
<td>0.0036*</td>
</tr>
<tr>
<td></td>
<td>(3.64)</td>
<td>(1.92)</td>
</tr>
<tr>
<td>(\log(a_i))</td>
<td>0.0017</td>
<td>0.0010***</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(2.58)</td>
</tr>
<tr>
<td>(\text{ICI})</td>
<td>0.0085</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.01)</td>
<td></td>
</tr>
</tbody>
</table>

Note: this table reports the Fama-MacBeth estimates of the regression coefficients in specification 2.5.5 with \(t\)-statistics based on Newey-West HAC standard errors (of lag order 12) shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively. The dependent variable is fund’s net return at quarter \(t\). \(\text{gsum} > 20\) is a dummy variable that equals 1 if a fund’s RPC score \(\text{gsum}\) is larger than 20. \(\text{FinCenter}\) is a dummy variable that equals 1 if a fund is located in a financial center. The following six cities are defined to be financial centers: Boston, Chicago, Los Angeles, New York, Philadelphia, and San Francisco, in the spirit of Christoffersen and Sarkissian (2009). \(\text{ICI}\) denotes the Industry Concentration Index (ICI), which is constructed in a similar manner as in Kacperczyk, Sialm, and Zheng (2005). But for simplicity, we use an equally weighted index instead.
Because all the results on retail investor data are qualitatively similar to those on mutual fund data, we shall keep our discussions brief.

To start with, Table 2.8.9 and 2.8.10 illustrate respectively the transformation and the gregariousness parameter estimates with city groups for retail investors. In general, both of the two sets of estimates are smaller compared to those from mutual fund data. This is because retail investors hold a much smaller number of stocks in contrast to mutual fund managers.

Again, we are interpreting these coefficients as investor fixed effects. Though one could also reasonably conclude that retail investors are likely to have much smaller investor networks than mutual fund managers.

Table 2.8.9: Summary statistics, transformation parameter estimates, retail investors

<table>
<thead>
<tr>
<th>City groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.06</td>
<td>0.03</td>
<td>1.06</td>
<td>1.02</td>
<td>1.12</td>
</tr>
</tbody>
</table>

Table 2.8.10: Summary statistics, estimates of $a_i$, retail investors

<table>
<thead>
<tr>
<th>City groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14.1</td>
<td>8.0</td>
<td>12.5</td>
<td>3.5</td>
<td>194.9</td>
</tr>
</tbody>
</table>

Next, Table 2.8.11 shows the estimated values of relative city sizes $b_k$ for retail investors. They are very close to the estimates from using mutual fund data. In particular, the correlation between the “mutual-fund” city group sizes and the “retail-investor” city group sizes is 0.96.

Table 2.8.12 then documents the overdispersion estimates with city group classifications for retail investors. In general, we find that the overdispersion parameters become smaller when retail investor data are used, suggesting that social networks effects are weaker for retail investors. However, qualitatively, the estimates are still similar to the ones from mutual fund data in that every city is overdispersed to some extent and that there are major overdispersions in certain cities (e.g. San Jose).
Table 2.8.11: Estimates of $b_k$, city groups, retail investors

<table>
<thead>
<tr>
<th>City</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>0.201</td>
<td>0.013</td>
<td>0.208</td>
<td>0.173</td>
<td>0.214</td>
</tr>
<tr>
<td>LA</td>
<td>0.066</td>
<td>0.007</td>
<td>0.064</td>
<td>0.055</td>
<td>0.079</td>
</tr>
<tr>
<td>Bos</td>
<td>0.061</td>
<td>0.002</td>
<td>0.061</td>
<td>0.057</td>
<td>0.065</td>
</tr>
<tr>
<td>SF</td>
<td>0.071</td>
<td>0.005</td>
<td>0.070</td>
<td>0.064</td>
<td>0.079</td>
</tr>
<tr>
<td>Chi</td>
<td>0.113</td>
<td>0.003</td>
<td>0.114</td>
<td>0.107</td>
<td>0.117</td>
</tr>
<tr>
<td>SJ</td>
<td>0.086</td>
<td>0.018</td>
<td>0.079</td>
<td>0.069</td>
<td>0.126</td>
</tr>
<tr>
<td>Dal</td>
<td>0.061</td>
<td>0.002</td>
<td>0.061</td>
<td>0.055</td>
<td>0.065</td>
</tr>
<tr>
<td>Hou</td>
<td>0.045</td>
<td>0.006</td>
<td>0.044</td>
<td>0.036</td>
<td>0.056</td>
</tr>
<tr>
<td>Phi</td>
<td>0.034</td>
<td>0.001</td>
<td>0.034</td>
<td>0.031</td>
<td>0.037</td>
</tr>
<tr>
<td>Was</td>
<td>0.032</td>
<td>0.002</td>
<td>0.031</td>
<td>0.029</td>
<td>0.036</td>
</tr>
<tr>
<td>Mia</td>
<td>0.019</td>
<td>0.001</td>
<td>0.019</td>
<td>0.017</td>
<td>0.021</td>
</tr>
<tr>
<td>Atl</td>
<td>0.045</td>
<td>0.003</td>
<td>0.045</td>
<td>0.039</td>
<td>0.049</td>
</tr>
<tr>
<td>Min</td>
<td>0.028</td>
<td>0.001</td>
<td>0.028</td>
<td>0.026</td>
<td>0.029</td>
</tr>
<tr>
<td>Den</td>
<td>0.013</td>
<td>0.002</td>
<td>0.013</td>
<td>0.010</td>
<td>0.015</td>
</tr>
<tr>
<td>SD</td>
<td>0.018</td>
<td>0.002</td>
<td>0.018</td>
<td>0.012</td>
<td>0.022</td>
</tr>
<tr>
<td>Stfd</td>
<td>0.038</td>
<td>0.002</td>
<td>0.038</td>
<td>0.035</td>
<td>0.042</td>
</tr>
<tr>
<td>Sea</td>
<td>0.027</td>
<td>0.001</td>
<td>0.028</td>
<td>0.025</td>
<td>0.029</td>
</tr>
<tr>
<td>Phx</td>
<td>0.013</td>
<td>0.002</td>
<td>0.012</td>
<td>0.010</td>
<td>0.015</td>
</tr>
<tr>
<td>SL</td>
<td>0.014</td>
<td>0.001</td>
<td>0.014</td>
<td>0.013</td>
<td>0.015</td>
</tr>
<tr>
<td>Det</td>
<td>0.015</td>
<td>0.001</td>
<td>0.015</td>
<td>0.014</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Lastly, we depict for retail investors their RPC scores and how these RPC scores are tied to their investment portfolio returns. Importantly, Table 2.8.14 suggests that retail investors' RPC scores also lead to outperformance in their common-stock portfolios, similar to our institutional investor results. For retail investors, the outperformance is about 1.33% per year using city groups.\footnote{14}

Through comparing the institutional and retail investor results, we can see that social networks are more prevalent among mutual fund managers than among ordinary households. However, the qualitative resemblance between the two sets of results implies that the impact of social networks seems to be universal and is not confined to the particular system of the mutual fund industry.

\footnote{14}Remember the holdings data of retail investors are at the monthly level. Hence the outperformance number — the coefficient in front of $gsum > 5$ in Table 2.8.14 is at the monthly level too.
Table 2.8.12: Estimates of $\omega_k$, city groups, retail investors

<table>
<thead>
<tr>
<th>City</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>1.032</td>
<td>0.020</td>
<td>1.016</td>
<td>1.010</td>
<td>1.111</td>
<td>1.99</td>
</tr>
<tr>
<td>LA</td>
<td>1.110</td>
<td>0.067</td>
<td>1.096</td>
<td>1.009</td>
<td>1.275</td>
<td>6.91</td>
</tr>
<tr>
<td>Bos</td>
<td>1.175</td>
<td>0.116</td>
<td>1.099</td>
<td>1.053</td>
<td>1.390</td>
<td>4.98</td>
</tr>
<tr>
<td>SF</td>
<td>1.156</td>
<td>0.107</td>
<td>1.205</td>
<td>1.027</td>
<td>1.317</td>
<td>4.94</td>
</tr>
<tr>
<td>Chi</td>
<td>1.154</td>
<td>0.065</td>
<td>1.177</td>
<td>1.014</td>
<td>1.255</td>
<td>9.61</td>
</tr>
<tr>
<td>SJ</td>
<td>2.006</td>
<td>0.309</td>
<td>1.884</td>
<td>1.616</td>
<td>2.567</td>
<td>7.79</td>
</tr>
<tr>
<td>Dal</td>
<td>1.014</td>
<td>0.011</td>
<td>1.005</td>
<td>1.002</td>
<td>1.077</td>
<td>3.24</td>
</tr>
<tr>
<td>Hou</td>
<td>1.043</td>
<td>0.030</td>
<td>1.049</td>
<td>1.002</td>
<td>1.107</td>
<td>12.71</td>
</tr>
<tr>
<td>Phi</td>
<td>1.028</td>
<td>0.021</td>
<td>1.026</td>
<td>1.001</td>
<td>1.106</td>
<td>4.26</td>
</tr>
<tr>
<td>Was</td>
<td>1.013</td>
<td>0.008</td>
<td>1.012</td>
<td>1.001</td>
<td>1.063</td>
<td>3.86</td>
</tr>
<tr>
<td>Mia</td>
<td>1.244</td>
<td>0.041</td>
<td>1.252</td>
<td>1.169</td>
<td>1.343</td>
<td>21.86</td>
</tr>
<tr>
<td>Atl</td>
<td>1.006</td>
<td>0.005</td>
<td>1.004</td>
<td>1.001</td>
<td>1.017</td>
<td>1.87</td>
</tr>
<tr>
<td>Min</td>
<td>1.436</td>
<td>0.094</td>
<td>1.465</td>
<td>1.209</td>
<td>1.578</td>
<td>13.28</td>
</tr>
<tr>
<td>Den</td>
<td>1.087</td>
<td>0.040</td>
<td>1.082</td>
<td>1.004</td>
<td>1.153</td>
<td>11.26</td>
</tr>
<tr>
<td>SD</td>
<td>1.127</td>
<td>0.098</td>
<td>1.147</td>
<td>1.010</td>
<td>1.291</td>
<td>4.77</td>
</tr>
<tr>
<td>Stfd</td>
<td>1.004</td>
<td>0.003</td>
<td>1.004</td>
<td>1.001</td>
<td>1.010</td>
<td>1.91</td>
</tr>
<tr>
<td>Sea</td>
<td>1.341</td>
<td>0.076</td>
<td>1.358</td>
<td>1.181</td>
<td>1.454</td>
<td>13.24</td>
</tr>
<tr>
<td>Phx</td>
<td>1.048</td>
<td>0.034</td>
<td>1.032</td>
<td>1.010</td>
<td>1.158</td>
<td>5.92</td>
</tr>
<tr>
<td>SL</td>
<td>1.125</td>
<td>0.040</td>
<td>1.130</td>
<td>1.050</td>
<td>1.267</td>
<td>15.24</td>
</tr>
<tr>
<td>Det</td>
<td>1.045</td>
<td>0.016</td>
<td>1.043</td>
<td>1.012</td>
<td>1.074</td>
<td>7.02</td>
</tr>
</tbody>
</table>

Table 2.8.13: Summary statistics of $gsum_i$, retail investors

<table>
<thead>
<tr>
<th>City groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18.57</td>
<td>7.00</td>
<td>17.62</td>
<td>4.82</td>
<td>161.9</td>
</tr>
</tbody>
</table>

Table 2.8.14: RPC scores and retail investor portfolio performances, city groups

<table>
<thead>
<tr>
<th>Dep Var: $return_t$</th>
<th>$const$</th>
<th>$return_{t-1}$</th>
<th>$portv$</th>
<th>comms</th>
<th>turnover</th>
<th>$hheq$</th>
<th>$gsum &gt; 5$</th>
<th>$\log(a_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.008**</td>
<td>0.031</td>
<td>0.001</td>
<td>-0.0011**</td>
<td>0.001</td>
<td>0.0002**</td>
<td>0.0011**</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>(2.05)</td>
<td>(1.47)</td>
<td>(0.29)</td>
<td>(-1.97)</td>
<td>(0.81)</td>
<td>(2.36)</td>
<td>(2.13)</td>
<td>(-1.10)</td>
</tr>
</tbody>
</table>

Note: dependent variable is $return_t$, the monthly gross return on a investor’s common-stock portfolio. Each column stands for one specific regressor. $t$-statistics based on Newey-West HAC standard errors (of lag order 12) are shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively. $portv$ is the market value of an investor’s stock portfolio measured in logs, comms is the monthly commissions paid (from trades) as a percentage of the $portv$, and $hheq$ is the total household equity value of an investor measured in logs. For retail investors, the RPC score dummy, $gsum > 5$, is over the largest 5 groups only where investors actually have a meaningful number of stock picks.
2.8.3 Retail Investor Results for Industry Groups

Table 2.8.15 and 2.8.16 illustrate the transformation and gregariousness parameter estimates respectively for retail investors, using industry as groups. These estimates are similar to the city-group estimates shown in the main paper.

Table 2.8.15: Summary statistics, transformation parameter estimates, retail investors

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry groups</td>
<td>1.10</td>
<td>0.05</td>
<td>1.09</td>
<td>1.03</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Table 2.8.16: Summary statistics, estimates of $a_i$, retail investors

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry groups</td>
<td>16.7</td>
<td>10.0</td>
<td>14.7</td>
<td>1.2</td>
<td>270.1</td>
</tr>
</tbody>
</table>

Table 2.8.17 shows the estimated values of $b_k$ of industry groups for retail investors. They are very close to the estimates from using mutual fund data. In particular, the correlation between the “mutual-fund” industry group sizes and the “retail-investor” industry group sizes is 0.87.

Table 2.8.18 then documents the overdispersion estimates with industry group classifications for retail investors. Similar to the case of city groups, we find that the overdispersion parameters become smaller when retail investor data are used instead of mutual fund data, but qualitatively the estimates are still similar to the ones from mutual fund data.

Lastly, we depict for retail investors their RPC scores based on industry groups and how these RPC scores are related to their investment portfolio returns. Table 2.8.20 demonstrates that the industry RPC scores lead to an outperformance of about 1.45% per annum in retail investors’ common-stock portfolios, which is very close to the number found in city RPC scores.
Table 2.8.17: Estimates of $b_k$, industry groups, retail investors

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>0.119</td>
<td>0.006</td>
<td>0.119</td>
<td>0.104</td>
<td>0.128</td>
</tr>
<tr>
<td>Serv</td>
<td>0.080</td>
<td>0.011</td>
<td>0.078</td>
<td>0.067</td>
<td>0.101</td>
</tr>
<tr>
<td>Hlth</td>
<td>0.146</td>
<td>0.010</td>
<td>0.149</td>
<td>0.129</td>
<td>0.160</td>
</tr>
<tr>
<td>BsEq</td>
<td>0.145</td>
<td>0.012</td>
<td>0.139</td>
<td>0.131</td>
<td>0.173</td>
</tr>
<tr>
<td>Retail</td>
<td>0.073</td>
<td>0.004</td>
<td>0.074</td>
<td>0.065</td>
<td>0.076</td>
</tr>
<tr>
<td>Whsl</td>
<td>0.020</td>
<td>0.002</td>
<td>0.021</td>
<td>0.017</td>
<td>0.023</td>
</tr>
<tr>
<td>Telcm</td>
<td>0.051</td>
<td>0.002</td>
<td>0.051</td>
<td>0.048</td>
<td>0.057</td>
</tr>
<tr>
<td>Oil</td>
<td>0.046</td>
<td>0.006</td>
<td>0.045</td>
<td>0.035</td>
<td>0.054</td>
</tr>
<tr>
<td>EEq</td>
<td>0.021</td>
<td>0.005</td>
<td>0.020</td>
<td>0.014</td>
<td>0.033</td>
</tr>
<tr>
<td>FabP</td>
<td>0.027</td>
<td>0.001</td>
<td>0.027</td>
<td>0.026</td>
<td>0.030</td>
</tr>
<tr>
<td>Cnstr</td>
<td>0.025</td>
<td>0.001</td>
<td>0.025</td>
<td>0.023</td>
<td>0.028</td>
</tr>
<tr>
<td>Trans</td>
<td>0.018</td>
<td>0.001</td>
<td>0.018</td>
<td>0.017</td>
<td>0.021</td>
</tr>
<tr>
<td>Game</td>
<td>0.016</td>
<td>0.004</td>
<td>0.016</td>
<td>0.011</td>
<td>0.024</td>
</tr>
<tr>
<td>Meal</td>
<td>0.020</td>
<td>0.001</td>
<td>0.020</td>
<td>0.019</td>
<td>0.022</td>
</tr>
<tr>
<td>Util</td>
<td>0.039</td>
<td>0.002</td>
<td>0.039</td>
<td>0.035</td>
<td>0.044</td>
</tr>
<tr>
<td>Food</td>
<td>0.050</td>
<td>0.004</td>
<td>0.050</td>
<td>0.043</td>
<td>0.056</td>
</tr>
<tr>
<td>Hshld</td>
<td>0.044</td>
<td>0.003</td>
<td>0.044</td>
<td>0.038</td>
<td>0.050</td>
</tr>
<tr>
<td>Chems</td>
<td>0.029</td>
<td>0.002</td>
<td>0.029</td>
<td>0.023</td>
<td>0.031</td>
</tr>
<tr>
<td>Book</td>
<td>0.016</td>
<td>0.001</td>
<td>0.016</td>
<td>0.014</td>
<td>0.017</td>
</tr>
<tr>
<td>Paper</td>
<td>0.015</td>
<td>0.001</td>
<td>0.015</td>
<td>0.013</td>
<td>0.018</td>
</tr>
</tbody>
</table>
Table 2.8.18: Estimates of $\omega_k$, industry groups, retail investors

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fin</td>
<td>1.575</td>
<td>0.087</td>
<td>1.546</td>
<td>1.454</td>
<td>1.723</td>
<td>17.11</td>
</tr>
<tr>
<td>Serv</td>
<td>1.364</td>
<td>0.204</td>
<td>1.281</td>
<td>1.110</td>
<td>1.995</td>
<td>5.44</td>
</tr>
<tr>
<td>Hlth</td>
<td>1.865</td>
<td>0.129</td>
<td>1.855</td>
<td>1.646</td>
<td>2.231</td>
<td>19.15</td>
</tr>
<tr>
<td>BsEq</td>
<td>1.939</td>
<td>0.469</td>
<td>1.704</td>
<td>1.427</td>
<td>2.832</td>
<td>4.71</td>
</tr>
<tr>
<td>Rtai</td>
<td>1.248</td>
<td>0.045</td>
<td>1.260</td>
<td>1.147</td>
<td>1.332</td>
<td>19.73</td>
</tr>
<tr>
<td>Whsl</td>
<td>1.137</td>
<td>0.127</td>
<td>1.084</td>
<td>1.005</td>
<td>1.363</td>
<td>2.98</td>
</tr>
<tr>
<td>Telcm</td>
<td>1.928</td>
<td>0.090</td>
<td>1.908</td>
<td>1.793</td>
<td>2.065</td>
<td>24.84</td>
</tr>
<tr>
<td>Oil</td>
<td>1.399</td>
<td>0.053</td>
<td>1.408</td>
<td>1.300</td>
<td>1.498</td>
<td>21.38</td>
</tr>
<tr>
<td>EEq</td>
<td>1.200</td>
<td>0.165</td>
<td>1.186</td>
<td>1.003</td>
<td>1.482</td>
<td>3.82</td>
</tr>
<tr>
<td>FabP</td>
<td>1.160</td>
<td>0.068</td>
<td>1.143</td>
<td>1.035</td>
<td>1.304</td>
<td>8.64</td>
</tr>
<tr>
<td>Cnstr</td>
<td>1.068</td>
<td>0.055</td>
<td>1.048</td>
<td>1.007</td>
<td>1.240</td>
<td>4.95</td>
</tr>
<tr>
<td>Trans</td>
<td>1.115</td>
<td>0.075</td>
<td>1.102</td>
<td>1.018</td>
<td>1.302</td>
<td>4.63</td>
</tr>
<tr>
<td>Game</td>
<td>1.068</td>
<td>0.064</td>
<td>1.079</td>
<td>1.002</td>
<td>1.184</td>
<td>6.37</td>
</tr>
<tr>
<td>Meal</td>
<td>1.059</td>
<td>0.030</td>
<td>1.060</td>
<td>1.011</td>
<td>1.125</td>
<td>6.42</td>
</tr>
<tr>
<td>Util</td>
<td>2.547</td>
<td>0.166</td>
<td>2.531</td>
<td>2.324</td>
<td>2.932</td>
<td>23.24</td>
</tr>
<tr>
<td>Food</td>
<td>1.313</td>
<td>0.028</td>
<td>1.319</td>
<td>1.237</td>
<td>1.374</td>
<td>43.39</td>
</tr>
<tr>
<td>Hshld</td>
<td>1.026</td>
<td>0.019</td>
<td>1.023</td>
<td>1.004</td>
<td>1.071</td>
<td>5.17</td>
</tr>
<tr>
<td>Chems</td>
<td>1.143</td>
<td>0.077</td>
<td>1.162</td>
<td>1.017</td>
<td>1.252</td>
<td>9.39</td>
</tr>
<tr>
<td>Book</td>
<td>1.109</td>
<td>0.032</td>
<td>1.106</td>
<td>1.037</td>
<td>1.178</td>
<td>19.27</td>
</tr>
<tr>
<td>Paper</td>
<td>1.043</td>
<td>0.033</td>
<td>1.022</td>
<td>1.003</td>
<td>1.243</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Table 2.8.19: Summary statistics of $gsum_i$, retail investors

<table>
<thead>
<tr>
<th>Industry groups</th>
<th>mean</th>
<th>s.d.</th>
<th>med</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry groups</td>
<td>18.60</td>
<td>6.72</td>
<td>17.87</td>
<td>8.70</td>
<td>196.2</td>
</tr>
</tbody>
</table>
Table 2.8.20: RPC scores and retail investor portfolio performances

<table>
<thead>
<tr>
<th></th>
<th>Industry</th>
<th>Both $gsum$’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>$const$</td>
<td>0.008**</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td>(1.88)</td>
</tr>
<tr>
<td>$return_{t-1}$</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(1.41)</td>
<td>(1.47)</td>
</tr>
<tr>
<td>$PortVal$</td>
<td>0.0002</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$commission$</td>
<td>-0.0011**</td>
<td>-0.0011**</td>
</tr>
<tr>
<td></td>
<td>(-2.05)</td>
<td>(-2.26)</td>
</tr>
<tr>
<td>$turnover$</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.93)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>$HHequity$</td>
<td>0.0002**</td>
<td>0.0003**</td>
</tr>
<tr>
<td></td>
<td>(2.52)</td>
<td>(2.18)</td>
</tr>
<tr>
<td>$gsum &gt; 5$</td>
<td>0.0012**</td>
<td>0.0006*</td>
</tr>
<tr>
<td></td>
<td>(2.12)</td>
<td>(1.81)</td>
</tr>
<tr>
<td>$log(a_i)$</td>
<td>-0.0006</td>
<td>0.0010**</td>
</tr>
<tr>
<td></td>
<td>(-1.36)</td>
<td>(2.06)</td>
</tr>
</tbody>
</table>

Note: dependent variable is $return_{t}$, the monthly gross return on a investor’s common-stock portfolio. $t$-statistics based on Newey-West HAC standard errors (of lag order 12) are shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively. $PortVal$ is the market value of an investor’s stock portfolio measured in logs, $commission$ is the monthly commissions paid (from trades) as a percentage of the $PortVal$, and $HHequity$ is the total household equity value of an investor measured in logs. For retail investors, the RPC score dummy, $gsum > 5$, is over the largest 5 groups only where investors actually have a meaningful number of stock picks.
3.1 Introduction

This chapter studies the relationship between security trading and price movement at the high frequency level through developing and estimating a continuous-time Hawkes jump-diffusion model. As seen in Chapter 1, high frequency trading has been growing rapidly in recent years. At the same time, the field of high frequency financial econometrics has also received a lot of attention, aided by the increasing availability of asset transaction data that are measured in seconds or even fractions of seconds. This strand of literature is mainly concerned with studying the statistical properties of price processes in continuous time, developing estimators for the volatility components and deriving test statistics for the jump components of asset prices.\footnote{See, for instance, \textit{A"{i}t-Sahalia and Jacod (2012)} and the reference therein.} Despite the large number of newly developed high frequency econometric methods, less effort has been devoted to rationalizing why asset price processes are actually exhibiting these statistical regularities. Since security prices are formed predominantly as a result
of trading in financial markets, essentially there is a gap in explaining how trading activities affect price processes at the high frequency level.

On the other hand, there exists an older strand of literature – market microstructure theory, which aims to explain how asset prices are influenced by the structure of the underlying financial markets. In particular, the literature focuses on analyzing the role information plays in the formation of prices. O’Hara (1998) provides a thorough survey of the existing work in this area. However, this literature has not been integrating the recent discoveries on price processes from high frequency financial econometrics into their theoretical studies, nor has it been utilizing the tools developed by the econometrics literature in their empirical studies.

Therefore, the current chapter attempts to provide the missing link between market microstructure theories and high frequency financial econometrics, as well as to understand more generally how trading activities in financial markets shape the stochastic behaviors of asset prices at the high frequency level. Specifically, I will examine how asset trading volume affects asset price. There are two reasons why I choose to focus on volume. Firstly, besides price, trading volume is the indicator that attracts the most attention from financial market participants. It is an old Wall Street adage that “It takes volume to make prices move”. Indeed, trading volume is generally thought to be an important factor in the process of asset price adjustment, if not the most important one. Secondly, in microstructure theories, there are several models that have implications on the role of volume in the movement of price. Sequential trade models based on asymmetric information between informed and uninformed traders with information event uncertainty (for example, Easley and O’Hara (1992)) suggest that, the higher the volume, the less frequent no-trade outcomes are, and the more likely that new information exists. Thus volume has two distinctive effects on price adjustment: unexpected volume moves price due to its information content, and high volume tends to be associated with high price volatility. Blume.
Easley, and O’Hara (1994) develops a model in which traders learn by watching both price and volume. Volume itself is informative because it provides data on the quality or the precision of information in past price changes.

To investigate their relationship at the high frequency level, I will develop a continuous-time model for the joint stochastic processes of price and volume. In order to model the effect of volume on price as implied by microstructure theories, I will use the Hawkes jump process. In a bivariate Hawkes process for volume and price, a jump in volume increases the probability of future jumps in both price and volume. Moreover, I will make the jump intensities, which increase with each jump, mean-revert to their steady-state levels until the next jump occurs. Therefore, we can interpret the Hawkes process in the model as follows. When new information arrives, asset volume jumps upwards due to the sudden eruption of trading, and this cross-excites the movement of asset price because of the information content the jump in volume carries. As more people learn from the new information, its dissemination raises the likelihood of some more bursts of trading, which causes the jump in volume to self-excite. In the absence of any further information arrival, jump intensities then revert back to their steady-state levels. Hence in my model, a rapid, sudden increase in trading volume has the potential capability of making price jump through the bivariate Hawkes process. In addition, I will incorporate a Brownian diffusion component for both the volume and the price process so as to capture their evolution in normal times without arrivals of news. Thus my model is a Hawkes jump-diffusion model.

I will implement a generalized method of moments (GMM) estimation procedure for my model to quantify the extent of the volume effect on price, based on the moment functions of the model, which I will derive in closed-form. Then I estimate the model using the high frequency data for one of the most heavily traded company stocks in the US. The estimation results suggest that, among other things, volume
cross-excites price powerfully and both volume and price vigorously self-excite. These findings are mostly in agreement with the predictions of microstructure theories.

The model that I analyze in this chapter builds on the work of Aït-Sahalia, Cacho-Diaz, and Laeven (2013), who first proposed the Hawkes jump-diffusion model as a way to study financial contagions. Though Hawkes processes themselves have seen applications in many different socioeconomic areas, they were always modelled as pure jump processes. Hawkes processes were originally put forward in the earlier papers of Hawkes (1971a,b); Hawkes and Oakes (1974); Oakes (1975), and used in the modelling of earthquake occurrences (see, for example, Ogata and Akaike (1982)). In finance, self-exciting and cross-exciting Hawkes processes have been employed to examine, for example, the joint defaults in portfolios of credit derivatives (Errais, Giesecke, and Goldberg (2010)) and how tradings at different maturities of the yield curve affect themselves and each other (Salmon and Tham (2007)). Pure jump Hawkes processes are appropriate devices for analyzing financial defaults or earthquake events since their sample paths are piecewise constant. However, this is certainly not the case for modelling asset price and trading volume. Hence I follow the strategy of Aït-Sahalia, Cacho-Diaz, and Laeven (2013) and add Brownian diffusion components into my model.

The rest of this chapter is organized as follows. In Section 3.2 I describe in detail my model of asset price and trading volume. I derive the explicit moment functions of the model in Section 3.3. In Section 3.4 I explain my estimation strategies. I then take the model to real world data in Section 3.5 and present my empirical analysis. Conclusions are offered in Section 3.6. Finally, the Appendix materials are collected in Section 3.7.
3.2 The Dynamics of Asset Price and Volume

In order to generate those features of the price-volume relationship as implied by market microstructure theories, I employ a bivariate Hawkes jump-diffusion process in my model. A Hawkes jump-diffusion process is a combination of a standard Brownian-driven drift-diffusion process with a Hawkes jump process. Hawkes jump processes are attractive here because not only do they help produce the volume-move-price effect, they can also capture the empirical properties of self-clustering and serial dependence that we observe at the high frequency level for price and volume (Russell and Engle (2010)).

3.2.1 Mutually Exciting Jump Processes

Hawkes processes are mutually exciting point processes introduced by Hawkes (1971a,b). The jump intensities of Hawkes processes are stochastic processes that depend on the paths of the point processes themselves. The jump process and its intensity together form a Markov process.

To be more precise for my model, I consider a bivariate Hawkes process, \( N_{i,t} \), \( i = 1, 2 \), where 1 denotes price and 2 denotes volume. Similar to a Poisson process, \( N_{i,t} \) is a point process characterized by its intensity process \( \lambda_{i,t} \), such that

\[
\begin{align*}
\mathbb{P}[N_{i,t+du} - N_{i,t} = 1|\mathcal{F}_t] &= \lambda_{i,t}du + o(du) \\
\mathbb{P}[N_{i,t+du} - N_{i,t} > 1|\mathcal{F}_t] &= o(du) \\
\mathbb{P}[N_{i,t+du} - N_{i,t} = 0|\mathcal{F}_t] &= 1 - \lambda_{i,t}du + o(du),
\end{align*}
\]

(3.2.1)

where \( \mathcal{F}_t = \sigma\{N_{i,s} : s \leq t\} \) and \( \{\mathcal{F}_t\}_{t \geq 0} \) is the filtration. The jump intensity \( \lambda_{i,t} \) has the dynamics.
\[ \lambda_{i,t} = \lambda_{i,\infty} + \int_{-\infty}^{t} g_{i,1}(t-s) dN_{1,s} + \int_{-\infty}^{t} g_{i,2}(t-s) dN_{2,s}, \quad i = 1, 2, \]  

(3.2.2)

where \( \lambda_{i,\infty} \) are the constant steady-state intensity levels and \( g_{i,j}(s), s \geq 0 \) are real-valued functions. Let \( \lambda_i := \mathbb{E}\lambda_{i,t} \). From (3.2.1) and by the law of iterated expectation, \( \mathbb{E}dN_{i,t} = \lambda_i dt \), so that the compensated process \( N_{i,t} - \int_{-\infty}^{t} \lambda_{i,s} ds \) is a \( \mathcal{F}_t \)-adapted local martingale. To have non-negative intensity processes with probability one, I will make the assumption that \( \lambda_{i,\infty} \geq 0 \) for all \( i \), and that the functions \( g_{i,j}(s) \geq 0 \) for all \( s \geq 0 \) and for all pairs of \( \{i, j\} \). Therefore, the jump intensities rise whenever there is a jump in price or volume.

If we apply the expectation operator to (3.2.2), it becomes

\[ \lambda_i = \lambda_{i,\infty} + \sum_{j=1}^{2} \lambda_j \int_{-\infty}^{t} g_{i,j}(t-s) ds. \]  

(3.2.3)

In vector form, we can write \( \lambda = \lambda_\infty + \Gamma \lambda \), where \( \lambda := [\lambda_1, \lambda_2]' \), \( \lambda_\infty := [\lambda_{1,\infty}, \lambda_{2,\infty}]' \) and \( \Gamma_{i,j} = \int_{-\infty}^{t} g_{i,j}(t-s) ds = \int_{0}^{\infty} g_{i,j}(u) du \). Thus \( \lambda = (I - \Gamma)^{-1} \lambda_\infty \), with \( I \) being the identity matrix. I will assume that the vector \( \lambda \) has positive and finite elements, which is sufficient to guarantee that our bivariate Hawkes process is stationary.\(^2\)

### 3.2.2 Joint Process of Price and Volume

I now specify the full joint process of asset price and volume. The asset log-return \( dX_t \) is assumed to follow the dynamics

\[ dX_t = \mu dt + \sigma dW_{1,t} + Z_{1,t}dN_{1,t}, \]  

(3.2.4)

where \( X_t \) is the asset log-price. For the asset log-volume \( V_t \), I assume its dynamics have the form

---

\(^2\)See [Hawkes and Oakes (1974)](#).
\[ dV_i = \phi(\omega - V_i)dt + \eta \sqrt{V_i} dW_{2,t} + Z_{2,t} dN_{2,t}. \] (3.2.5)

Here, \([W_{1,t}, W_{2,t}]\)' is a 2-dimensional vector of independent standard Brownian motions, \([Z_{1,t}, Z_{2,t}]\)' is a vector of jump sizes, mutually and serially independent with laws \(F_{Z_i}\) supported on \((−∞, ∞)\), and \([N_{1,t}, N_{2,t}]\)' is a vector of the bivariate Hawkes jump process described in the last subsection. Furthermore, \((\mu, \sigma, \phi, \omega, \eta)\) are all constant parameters. \(\mu\) is the drift of the log-return, \(\sigma\) and \(\eta\) are instantaneous volatility parameters, \(\phi\) measures the speed of mean reversion of the log-volume and \(\omega\) is the steady-state log-volume level. The vector of Brownian motions, the vector of jump sizes and the vector of Hawkes processes are assumed to be mutually independent.

### 3.2.3 Further Specifications on Jumps

In order to derive closed-form expressions for the moments of my model and perform estimations, I need some further parameterizations on the jump processes. Referring back to (3.2.2), I consider the kernel \(g_{i,j}(·)\) in the jump intensities to have the form

\[ g_{i,j}(u) = \alpha_{i,j} e^{-\beta_{i,j}u}, \quad u \geq 0, \; i, j = 1, 2, \] (3.2.6)

where \(\alpha_{i,j} \geq 0, \; \beta_{i,j} > 0, \; \forall i, j\). Therefore, the intensity processes exhibit mean reversion. The intensity \(\lambda_{i,t}\) jumps up by \(\alpha_{i,1}\) if there is a jump in price or by \(\alpha_{i,2}\) if there is a jump in volume, and then it decays exponentially back towards the level of \(\lambda_{i,∞}\) at the speed of \(\beta_{i,1}\) or \(\beta_{i,2}\). Furthermore, we have \(\int_0^\infty g_{i,j}(u)du = \alpha_{i,j}/\beta_{i,j}\), so that the \(\Gamma\) matrix in the model becomes

\[ \int_0^\infty g_{i,j}(u)du = \frac{\alpha_{i,j}}{\beta_{i,j}}, \]
In this case, a simple sufficient condition for the bivariate Hawkes process to be stationary will be given in the appendix. This condition will be imposed in the analysis that follows.

The reasons to choose this functional form for $g_{i,j}(\cdot)$ are as follows. First, it allows us to keep the model tractable. This tractability will prove to be particularly helpful when later I derive the moment conditions of my model and carry out estimations. Second, it has a clear analytical interpretation. Based on market microstructure theories, I treat the jumps as a result of the arrivals of new information, thus the cross-excitation parameter $\alpha_{1,2}$ measures the degree of ability of volume to move price. If this effect is strong, we will find $\alpha_{1,2}$ to be large. The self-excitation parameter $\alpha_{2,2}$ then gauges the effect from consecutive waves of eruption of trading when new information arrives, and $\alpha_{1,1}$ can be thought of as describing the “after-shock” impact on price movement due to new information/ramp in trading. These self-excitations are also used to capture the empirical regularities of jump-clustering and serial dependence at the high frequency level. In addition, the $\beta_{i,j}$’s would account for the speeds of information dissemination in the model.\footnote{Here, $\alpha_{2,1}$ is meant to keep the generality of the model. I will introduce a restriction on this parameter in the next section, as well as the economic reason behind this.}

As will be seen in the next section, the moment expressions of my model involve the moments of the jump sizes $\{Z_{i,t}\}$. I will thus parameterize them so as to have a full parametric model that can be used for estimation later. To this end, I will assume that $Z_{i,t}$ is a random variable with a probability density function

$$f_{Z_i}(x) = \begin{cases} 
  p_i \nu_{i,1} e^{-\nu_{i,1} x}, & -\infty < x \leq 0 \\
  (1 - p_i) \nu_{i,2} e^{-\nu_{i,2} x}, & 0 < x < \infty. 
\end{cases} \quad (3.2.8)$$
This means $Z_{i,t}$ has an asymmetric double exponential distribution on the negative and the positive real line, and we can easily verify that

$$
\mathbb{E}Z_{i,t}^k = (-1)^k \frac{k!p_i}{\nu_{i,1}^k} + \frac{k!(1-p_i)}{\nu_{i,2}^k}, \quad k = 1, 2, \ldots
$$

(3.2.9)

The double exponential distribution has two useful properties for my modelling purposes. First, it has the leptokurtic feature, which accounts well for the fat-tailness of asset returns. Second, it is memoryless. This corresponds well to the idea that jumps in my model are the result of responses to new information or new market conditions.

3.2.4 Comments on the Model

In my Hawkes jump-diffusion model, occurrences of jumps lead to increases in jump intensities. This, together with the mean-reverting properties of jump intensities, resembles what ARCH/GARCH does in the context of volatility (see Engle (1982) and Bollerslev (1986)). In ARCH/GARCH models, high returns induce high volatility that makes high returns more likely to happen. Similarly, in my setting, jumps themselves would cause jump intensities to rise, which then raises the likelihood of future jumps. In addition, jump intensities move back to their steady-state levels without further jumps, in the same way as how volatility would evolve in ARCH/GARCH models when larger returns do not realize.

3.3 Closed-form Moment Expressions

I am now in a position to derive the explicit (unconditional) moment functions of the model. For this purpose, I will utilize the mutual independence among $\{W, N, Z\}$ repeatedly and use the fact that without jumps, the steady state distribution function

---

6This jump size distribution was first used by Kou (2002) for pricing options.
of the log-volume process $V_t$ is a gamma distribution with shape parameter $2\phi \omega / \eta^2$ and scale parameter $\eta^2 / 2\phi$.\footnote{Since without jumps, $V_t$ is a Feller’s square root process. See Feller (1951).}

### 3.3.1 Moment Functions of the Model

The moments of $(\Delta X_t, \Delta V_t) = (X_{t+\Delta} - X_t, V_{t+\Delta} - V_t)$ are derived up to leading orders of $\Delta$. For the mean, notice that $\mathbb{E} \Delta W_{i,t} = 0$, $\mathbb{E} V_t = \omega$ and $\mathbb{E} \Delta N_{i,t} = \lambda_i \Delta$, hence

$$\mathbb{E} \Delta X_t = (\mu + \lambda_1 \mathbb{E} Z_1) \Delta + o(\Delta^2),$$
$$\mathbb{E} \Delta V_t = (\lambda_2 \mathbb{E} Z_2) \Delta + o(\Delta^2).$$

For the variances, notice that $\mathbb{E} (\Delta W_{i,t})^2 = \Delta$ and $\mathbb{E} (\Delta N_{i,t})^2 = \lambda_i \Delta$, so that

$$\mathbb{E} (\Delta X_t - \mathbb{E} \Delta X_t)^2 = (\sigma^2 + \lambda_1 \mathbb{E} Z_1^2) \Delta + o(\Delta),$$
$$\mathbb{E} (\Delta V_t - \mathbb{E} \Delta V_t)^2 = (\eta^2 \omega + \lambda_2 \mathbb{E} Z_2^2) \Delta + o(\Delta).$$

For the third and fourth central moments, notice that $\mathbb{E} (\Delta N_{i,t})^3 = \lambda_i \Delta$ and $\mathbb{E} (\Delta N_{i,t})^4 = \lambda_i \Delta$, thus

$$\mathbb{E} (\Delta X_t - \mathbb{E} \Delta X_t)^3 = (\lambda_1 \mathbb{E} Z_1^3) \Delta + o(\Delta),$$
$$\mathbb{E} (\Delta V_t - \mathbb{E} \Delta V_t)^3 = (\lambda_2 \mathbb{E} Z_2^3) \Delta + o(\Delta),$$
$$\mathbb{E} (\Delta X_t - \mathbb{E} \Delta X_t)^4 = (\lambda_1 \mathbb{E} Z_1^4) \Delta + o(\Delta),$$
$$\mathbb{E} (\Delta V_t - \mathbb{E} \Delta V_t)^4 = (\lambda_2 \mathbb{E} Z_2^4) \Delta + o(\Delta).$$

Let $\Xi(\tau)$ be the covariance density matrix of our bivariate Hawkes process, i.e.

$$\Xi_{i,j}(\tau) = (\mathbb{E}[dN_{i,t+\tau}dN_{j,t}] - \mathbb{E}[dN_{i,t+\tau}]\mathbb{E}[dN_{j,t}])/(dt)^2 = \mathbb{E}[dN_{i,t+\tau}dN_{j,t}] / dt^2 - \lambda_i \lambda_j,$$

(3.3.1)
then the autocovariances of \((\Delta X_t, \Delta V_t)\) are given by the following expressions, up to the order of \(\Delta^2\):

\[
\begin{pmatrix}
\mathbb{E}[\Delta X_{t+\tau} \Delta X_t] - (\mathbb{E}[\Delta X_t]^2) & \mathbb{E}[\Delta X_{t+\tau} \Delta V_t] - \mathbb{E}[\Delta X_t] \mathbb{E}[\Delta V_t] \\
\mathbb{E}[\Delta V_{t+\tau} \Delta X_t] - \mathbb{E}[\Delta V_t] \mathbb{E}[\Delta X_t] & \mathbb{E}[\Delta V_{t+\tau} \Delta V_t] - (\mathbb{E}[\Delta V_t]^2)
\end{pmatrix}
\]

\[
= \begin{pmatrix}
(E Z_1)^2 \Xi_{11}(\tau) \Delta^2 & E[Z_1]E[Z_2] \Xi_{12}(\tau) \Delta^2 \\
E[Z_2]E[Z_1] \Xi_{21}(\tau) \Delta^2 & \frac{\eta^2 \omega \phi}{2} e^{-\phi \tau} + (E Z_2)^2 \Xi_{22}(\tau) \Delta^2
\end{pmatrix}
\]

From these moment functions, we can see that the higher order moments (third and fourth) separate out the jump parameters at the leading order, while the variances put them on an equal footing with the diffusive parameters. This feature is important for identification purposes.

### 3.3.2 Covariance Density Matrix of the Hawkes Process

We need to determine the matrix \(\Xi(\tau)\) in order to have a closed-form expression for the autocovariances. In vector form, this matrix is given by, for \(\tau > 0\),

\[
\Xi(\tau) = \mathbb{E}[dN_{t+\tau} dN_t]/dt^2 - \lambda \lambda'
\]

\[
= \mathbb{E} \left[ \left( \lambda_{\infty} + \int_{-\infty}^{t+\tau} \Gamma(t + \tau - s) dN_s \right) \frac{dN_t}{dt} \right] - \lambda \lambda'
\]

\[
= \Gamma(\tau) \text{diag}(\lambda) + \int_{-\infty}^{\tau} \Gamma(\tau - s) \Xi(s) ds,
\]

where \(\text{diag}(\lambda)\) stands for the diagonal matrix formed by \((\lambda_1, \lambda_2)\). I follow Hawkes (1971b) here to derive the solution for \(\Xi(\tau)\). The key step is to introduce a supplementary matrix

\[
B(\tau) = \Gamma(\tau) \text{diag}(\lambda) + \int_{-\infty}^{\infty} \Gamma(\tau - s) \Xi(s) ds - \Xi(\tau), \quad -\infty < \tau < \infty,
\]
where $\Gamma(s) = 0$ for $s < 0$. Taking the fourier transform of \((3.3.3)\) and rearranging, we have
\[
\tilde{\Xi}(\omega) = \left( I - \tilde{\Gamma}(\omega) \right)^{-1} \left( \tilde{\Gamma}(\omega) \text{diag}(\lambda) - \tilde{B}(\omega) \right), \tag{3.3.4}
\]
where $\sim$ denotes the fourier transform. Since $\Xi(-\tau) = \Xi(\tau)'$, it follows that $\tilde{\Xi}(-\omega) = \tilde{\Xi}(\omega)'$, which implies
\[
\left( I - \tilde{\Gamma}(-\omega) \right)^{-1} \left( \tilde{\Gamma}(-\omega) \text{diag}(\lambda) - \tilde{B}(-\omega) \right) = \left( \tilde{\Gamma}(\omega) \text{diag}(\lambda) - \tilde{B}(\omega) \right)' \left( I - \tilde{\Gamma}(\omega)' \right)^{-1},
\]
or
\[
\left( I - \tilde{\Gamma}(-\omega) \right) \tilde{B}(\omega)' + \tilde{\Gamma}(-\omega) \text{diag}(\lambda) = \tilde{B}(-\omega) \left( I - \tilde{\Gamma}(\omega)' \right) + \text{diag}(\lambda) \tilde{\Gamma}(\omega)' \tag{3.3.5}
\]
after simplification. Based on \((3.3.5)\), it then follows from Hawkes (1971b) that
\[
\tilde{B}(\omega) = -\text{diag}(\lambda) \tilde{\Gamma}(\omega)' \left( I - \tilde{\Gamma}(-\omega) \right)^{-1}. \tag{3.3.6}
\]
Hence substituting this in \((3.3.4)\) yields
\[
\tilde{\Xi}(\omega) = \left( I - \tilde{\Gamma}(\omega) \right)^{-1} \left( \tilde{\Gamma}(\omega) \text{diag}(\lambda) + \text{diag}(\lambda) \tilde{\Gamma}(\omega)' - \tilde{\Gamma}(\omega) \text{diag}(\lambda) \tilde{\Gamma}(\omega)' \right) \cdot \\
\left( I - \tilde{\Gamma}(-\omega)' \right)^{-1} \tag{3.3.7}
\]
from which we obtain $\Xi(\tau)$ by the inverse fourier transform. The explicit expression of $\Xi(\tau)$ derived is given in the Appendix, due to its long and complicated form.

3.4 Estimation

The availability in closed-form of the moment functions of my model allows me to estimate the model parameters through a GMM-based procedure. I will use the
following moment conditions to carry out my GMM estimation:

\[
\begin{align*}
\mathbb{E}[\Delta X_t], \mathbb{E}[\Delta V_t] \\
\mathbb{E}[(\Delta X_t - \mathbb{E}[\Delta X_t])^r], \mathbb{E}[(\Delta V_t - \mathbb{E}[\Delta V_t])^r], \quad r = 2, 3, 4 \\
\mathbb{E}[\Delta X_{t+\tau} \Delta X_t - (\mathbb{E}[\Delta X_t])^2], \mathbb{E}[\Delta X_{t+\tau} \Delta V_t - \mathbb{E}[\Delta X_t] \mathbb{E}[\Delta V_t]], \quad \tau > 0 \\
\mathbb{E}[\Delta V_{t+\tau} \Delta X_t] - \mathbb{E}[\Delta V_t] \mathbb{E}[\Delta X_t], \mathbb{E}[\Delta V_{t+\tau} \Delta V_t] - (\mathbb{E}[\Delta V_t])^2, \quad \tau > 0.
\end{align*}
\]

As explained earlier in the previous section, the primary reason for using these moment conditions is that each of them plays a specific role in identifying parts of the model. This is important since the number of parameters to be estimated is large, thus it reduces the effort spent on minimizing the GMM criterion function.

### 3.4.1 The GMM Procedure

Here I describe my GMM estimation procedure. Let

\[
y_n = (\Delta_n X, \Delta_n V) = (X_{n\Delta} - X_{(n-1)\Delta}, V_{n\Delta} - V_{(n-1)\Delta}),
\]

\[n = 1, \ldots, N, \Delta > 0, \text{ with } N\Delta = T \text{ on the time interval } [0, T].\]

Let \(\theta \in \Theta\) denote the \(K\)-dimensional parameter vector we are interested in, with \(\theta_0 \in \Theta\) being its true value, where \(\Theta\) is a compact parameter space. I consider \(M\) moment conditions \(h(y_n, \theta)\), \(M \geq d\), continuously differentiable in \(\theta\), where \(h\) is the vector of differences between the “moments” of \(y_n\) and their closed-form expectations derived under my model.\(^8\) This ensures that the global identification condition \(\mathbb{E}[h(y_n, \theta)] = 0\) only when \(\theta = \theta_0\) is satisfied.

Let \(g_N(\theta)\) denote the sample analogue of \(h(y_n, \theta)\), that is, the differences between the sample moments of \(y_n\) and its closed-form population moment functions in the

\(^8\)Note that \(\Delta\) does not depend on \(N\) or \(T\).

\(^9\)For example, one of the elements of \(h(y_n, \theta)\) is \((\Delta_n X - \mathbb{E}\Delta_n X)^2 - \mathbb{E}(\Delta_n X - \mathbb{E}\Delta_n X)^2\).
model. The GMM estimator \( \hat{\theta}_N \) is the value of \( \theta \in \Theta \) that minimizes the GMM criterion function:

\[
g_N(\theta)' W_N g_N(\theta),
\]

where \( W_N \) is a positive definite weighting matrix converging in probability to a positive definite limit \( W \).

Assume that the process \( \{h(y_n, \theta_0)\}_{n=-\infty}^{\infty} \) is strictly stationary with the \( j \)-th autocovariance matrix \( \Psi_j = \mathbb{E}[h(y_n, \theta_0)h(y_{n-j}, \theta_0)'] \), and the autocovariances are absolutely summable, i.e. each element of the sequence of matrices \( \{\Psi_j\}_{n=0}^{\infty} \) forms an absolutely summable scalar sequence. Let \( S = \sum_{j=-\infty}^{\infty} \Psi_j \) be the long run variance of \( h(y_n, \theta_0) \). Under standard regularity conditions (see Hansen (1982) or Newey and McFadden (1994)), the GMM estimator \( \hat{\theta}_N \) is consistent and asymptotically normal, with

\[
\sqrt{N}(\hat{\theta}_N - \theta_0) \implies \mathcal{N}(0, \Omega),
\]

(3.4.2)

where \( \Omega = (D'WD)^{-1}(D'WSWD)(D'WD)^{-1} \) and \( D \) is the gradient of \( \mathbb{E}[h(y_n, \theta)] \) with respect to \( \theta' \) evaluated at \( \theta_0 \). The weighting matrix \( W_N \) can be chosen optimally to minimize the asymptotic variance \( \Omega \), by taking it to be a consistent estimator of \( S^{-1} \). When \( W_N \) is chosen optimally, \( W = S^{-1} \) and \( \Omega \) reduces to \( (D'S^{-1}D)^{-1} \).

To estimate the long run variance \( S \) (and hence to have an optimal weighting matrix \( W_N \)), I will use the HAC estimator of Newey and West (1987) (see also Newey and West (1994)). It has the following form:

\[
\hat{S} = \hat{\Psi}_{0,T} + \sum_{j=1}^{m} \left( 1 - \frac{j}{m+1} \right) \left( \hat{\Psi}_{j,T} + \hat{\Psi}_{j,T}' \right),
\]

(3.4.3)

where

\[
\hat{\Psi}_{j,T} = \frac{1}{N} \sum_{n=j+1}^{N} h(y_n, \hat{\theta})h(y_{n-v}, \hat{\theta})',
\]

(3.4.4)
and $\tilde{\theta}$ is any consistent estimator of $\theta_0$. This suggests that I use the 2-step GMM estimation procedure. In the first step, I put $W_N = I$ to obtain $\hat{\theta}(1)$, an initial consistent estimate of $\theta_0$. In the second step, I insert $\hat{\theta}(1)$ into the HAC estimator $\hat{S}$ and put $W_N = \hat{S}^{-1}$ to compute $\hat{\theta}_N$.

### 3.4.2 Reduction of Parameter Space

As it stands, the model has 21 parameters to be estimated: 10 parameters of the bivariate Hawkes process ($\alpha_{i,j}$, $\beta_{i,j}$ and $\lambda_{i,\infty}$ for $i, j = 1, 2$), 6 parameters concerning the jump sizes ($p_i$ and $\nu_{i,j}$ for $i, j = 1, 2$), and $\{\mu, \sigma, \omega, \eta, \phi\}$. I will introduce a few restrictions on the parameters before I perform the actual estimation, and the main reasons are as follows:

- The number of parameters to be estimated is large. I would like to keep the model parsimonious and not to use too many moment conditions, which may cause difficulties in numerical optimization.

- More importantly, I will be tracking down the finer structure of the Hawkes jump process, i.e. the effects of self-excitation and cross-excitation, not just whether jumps occur or not. While in theory, the 2-step GMM procedure seems straightforward to implement, when taking the model to real data, one should always be aware of the inherent difficulties associated with estimating parameters about jump structures (which are our principal interest here), since there is a non-zero probability that jumps do not occur in any given finite time interval, and if the number of jumps is small, the identification could be weak.

---

10 need to choose the number of lags $m$ in implementing the HAC estimator. Newey and West (1994) develop a nonparametric method to automatically select $m$ based on the sample. Because our primary interest here is not about choosing $m$, to maintain simplicity and reduce computation burden, I set $m = \lfloor 4(N/100)^{1/4} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the greatest lesser integer function.
Therefore, I need to maintain the identification power of the model and the degree of accuracy during estimation.

The parameter restrictions that I impose are as much guided by microstructure theories as possible.

- I set \( \alpha_{2,1} = 0 \). This is because in theory, it is the jumps in volume that could cause price to change, and also what I am after is how certain trades make prices shift in a particular fashion, not the other way around.

- Since it is the unexpected increase in volume or number of trades that contains new information, if volume is low, it is incorrect to say that volume “jumps down” due to the arrival of information. Thus I will set \( p_2 = 0 \), i.e. volume does not “jump down”.

- This is not a restriction, but I will estimate the parameter \( \eta^2 \omega \) instead of \( \{\omega, \eta\} \) separately, as it is \( \eta^2 \omega \) that enters the moment conditions. Similarly, I will estimate \( 1/\nu_{i,j} \) directly in place of \( \nu_{i,j} \).

- I restrict \( \nu_{1,2} \) to be equal to \( \nu_{1,1} \), mainly to avoid identification problems.

As a result, this leaves us with 15 parameters to estimate:

\[
\theta = (\alpha_{1,1}, \alpha_{1,2}, \alpha_{2,2}, \beta_{1,1}, \beta_{1,2}, \beta_{2,2}, \lambda_{1,\infty}, \lambda_{2,\infty}, p_1, 1/\nu_{1,1}, 1/\nu_{2,2}, \mu, \sigma, \eta^2 \omega, \phi)'.
\]

3.5 Empirical Analysis

In this section, I estimate the model parameters on high frequency financial data. I will first give an overview of the data that I use, then explain my empirical results and perform hypothesis tests.
3.5.1 Description of Data

I use the price and the volume data on Intel (ticker: INTC) during the full trading month (21 days) of October, 2010\(^{[11]}\). All data are from the TAQ database. The price and the volume are the actual traded price and number of shares from 9:30am to 16:00pm, sampled every 30 seconds.\(^{[13]}\) I average all the traded prices at every 30-second mark to obtain the single price data for that time stamp, and aggregate the number of traded shares from the previous 29 seconds with the number at the 30-second mark to obtain the volume data at each 30-second time stamp. I then take the differences of logs of price and volume to arrive at my data for $\Delta X_t$ and $\Delta V_t$. Overall, there are $N = 780 \times 21 = 16380$ data points in my sample. The summary statistics of my data are presented in the table below.

3.5.2 Empirical Results

I employ 16 moment conditions to identify the 15 parameters in the model: the means and the variances, the third and fourth central moments and 8 autocovariances (2 lags, \(^{[11]}\)I have picked the month of October for my analysis, since Intel announces its third quarter earnings in the second week of October, which might reduce the possibility of observing no jumps (of course, it cannot/will not alter the actual probability of no jumps in any given time interval) and help us avoid the problem of weak identification. However, I do not include other earning-announcement months, primarily because in the model, $\mu, \sigma, \eta$ are all set constant in order to keep simplicity, whereas they are normally treated as stochastic processes too in high frequency financial econometrics. Thus I want to minimize the impact due to this loss of generality and including other months may cause the parameters’ changes to be too large to be ignored.\(^{[12]}\)It might as well be argued that I could choose a sample period when there was major economic news, for example, September 2008, in order to (possibly) minimize the chances that jumps did not occur. I did not opt to do this type of event study however, because my goal is to illustrate the applicability of the model, even during normal time periods.\(^{[13]}\)I choose 30 seconds as the sampling frequency for two reasons. On the one hand, I need to avoid encountering the problem of observing zero trade in any particular sampling interval. Sampling at ultra-high frequency (e.g. 10- or 5-second mark) would inevitably lead us to deal with this additional nuisance issue. On the other hand, sampling at somewhat lower frequency (say, 5 or 10 minutes) could work, but it runs the bigger risk of reducing the ability of my model to identify jumps when they actually occurred, which I would like to prevent as well.

125
Table 3.5.1: Summary statistics of data

<table>
<thead>
<tr>
<th></th>
<th>$\Delta X_t$</th>
<th>$\Delta V_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$-8.0497 \times 10^{-7}$</td>
<td>0.011</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.3248 $\times 10^{-4}$</td>
<td>1.9812</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.5555</td>
<td>0.0653</td>
</tr>
<tr>
<td>Excess Kurtosis</td>
<td>19.7164</td>
<td>0.1711</td>
</tr>
</tbody>
</table>

30- and 60-second) as shown in the last section, approximated at the leading order in $\Delta$. The GMM parameter estimates are depicted in the following table.

Table 3.5.2: Parameter estimates of the model

<table>
<thead>
<tr>
<th>$\alpha_{1,1}$</th>
<th>$\alpha_{1,2}$</th>
<th>$\alpha_{2,2}$</th>
<th>$\beta_{1,1}$</th>
<th>$\beta_{1,2}$</th>
<th>$\beta_{2,2}$</th>
<th>$\lambda_{1,\infty}$</th>
<th>$\lambda_{2,\infty}$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$p_1$</th>
<th>$1/n_{1,1}$</th>
<th>$1/n_{2,2}$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\eta^{2}\omega$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.258***</td>
<td>0.022***</td>
<td>5.635***</td>
<td>0.0192</td>
<td>6.965***</td>
<td>10.776***</td>
<td>35.187***</td>
</tr>
<tr>
<td>(0.0456)</td>
<td>(0.00416)</td>
<td>(0.917)</td>
<td>(0.0271)</td>
<td>(0.177)</td>
<td>(0.938)</td>
<td>(7.854)</td>
</tr>
</tbody>
</table>

Note: this table reports the GMM parameter estimates of the 15 parameters in the model, where standard errors are in parentheses. *, ** and *** denote statistical significance at the 90%, 95% and 99% significance levels respectively.

The estimated value of $\alpha_{1,2}$ is found to be large and statistically significant. It shows that a jump in volume can cause a strong increase in the probability of a successive jump in price. Thus an unexpected rise in volume has the ability to move price, which broadly supports the prediction from microstructure theoretical models. It also demonstrates that Hawkes processes are capable of quantifying a specific transmission mechanism from changes in asset trading to movements in price. Because of the magnitude of this estimate, it is likely that the findings from traditional empirical microstructure research that high volume tends to be associated with high price volatility are in fact the result of those cross-excited jumps in my model.

Furthermore, I obtain large and significant estimates for $\alpha_{1,1}$ and $\alpha_{2,2}$, the self-excitation parameters. This implies that the effect of “after-shocks” due to new information (as explained in Section 3.2) is large, as well as the fact that both price
and volume exhibit jump-clustering, a pattern also regularly found in other studies. The sizes of these estimates also imply that what traditional market microstructure literature considered as volatility may partially be seized by the Hawkes jumps in the model. In addition, I find the estimates of $\beta_{i,j}$’s to be substantial, so that the rate of information dissemination is high. This should come at no surprise, since Intel is a very liquid and heavily traded stock.

By virtue of my GMM estimation results, we can test the following null hypotheses using Wald tests: $H^1_0: \alpha_{i,i} = 0, i = 1, 2$ (no self-excitation); $H^2_0: \alpha_{1,1} = \alpha_{1,2} = \alpha_{2,2} = 0$ (no excitation); $H^3_0: \lambda_{i,\infty} = 0, i = 1, 2$ (no jumps). The test statistic has an asymptotic $\chi^2_q$ distribution, where $q$ is the number of restrictions. The next table shows my test results.

<table>
<thead>
<tr>
<th>$H^1_0$: No Self-Excitations</th>
<th>$H^2_0$: No Excitations</th>
<th>$H^3_0$: No Jumps</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.156***</td>
<td>18.306***</td>
<td>29.539***</td>
</tr>
</tbody>
</table>

Note: this table reports the values of the Wald test statistics, where ** and *** denotes rejection of the null hypothesis at the 95% and 99% significance levels respectively.

As can be seen, the null hypotheses are firmly rejected in all three cases. The rejection of $H^2_0$ indicates that my Hawkes model is preferred over a Poisson jump-diffusion model, while the rejection of $H^3_0$ provides clear evidence against the simple drift-diffusion model without jumps.

### 3.6 Conclusion

By means of a Hawkes jump-diffusion model for asset trading volume and asset price, I demonstrate a particular technique that one could use to quantify how asset trading affects asset price movement at the high frequency level. In the model, jumps in volume due to arrivals of new information raise the probability of jumps in price, so
that an unexpected, large increase in volume has the ability to stir the motion of price. In addition, jumps in volume and price self-excite, which captures the “after-shock” effects of new information arrivals.

I derive the moment functions of the model in closed-form and explain how to estimate the parameters of the model with a GMM procedure, relying on the closed-form moment conditions. Then I estimate my model on high frequency financial data (sampled every 30 seconds) and the estimation results show that volume strongly cross-excites price, which confirms the implication from microstructure theories. In addition, I find that both of the self-excitation parameters for volume and price are large and statistically significant, and simpler models of Poisson jump-diffusion and drift-diffusion without jumps are rejected against my model based on Wald tests.

My model attempts to build a connection between two seemingly separate strands of literature, namely the older market microstructure theory and the more recent high frequency financial econometrics. We should view this connection as an aid to resolve the more general and intriguing problem for high frequency financial econometrics, that is, the correspondence between trading behaviors and specific types of price movement at the high frequency level. With this consideration in mind, let us discuss some possible future work.

I have not yet fully exploited the newly developed methods from high frequency financial econometrics. Further work needs to be done on estimation and inference at the high frequency level for models where price and volume (or other measures of trade) are general Itô semimartingales. This would help us uncover the relations between the jump structures of volume and these of price. While an interest in itself, those relations could also lead to a better understanding of microstructure noise that is frequently encountered in studies of high frequency financial econometrics. More comprehensive insights into the stochastic process of microstructure noise would in
turn guide us to develop more robust estimators and test statistics based on high frequency data.

It is also desirable to improve market microstructure theoretical models (in continuous time) that try to grasp the trade-price link, in light of the findings from high frequency econometrics. This may not only prove to be useful from the standpoint of the asset pricing literature, but it could also yield practical and more structural underpinnings for financial econometric models.

3.7 Appendix

3.7.1 Stationarity of the Bivariate Hawkes Process under Exponential Decay

Referring back to Section 3.3 when \( g_{i,j}(u) = \alpha_{i,j} e^{-\beta_{i,j} u}, \ u \geq 0 \), the matrix \( \Gamma \) has each of its elements equal to \( \frac{\alpha_{i,j}}{\beta_{i,j}} \), \( i, j = 1, 2 \). Since \( \Lambda = (I - \Gamma)^{-1} \Lambda_\infty \), this yields

\[
\begin{bmatrix}
\lambda_1 \\
\lambda_2
\end{bmatrix}
= \begin{pmatrix}
\frac{\beta_{1,1}\beta_{2,1}(\beta_{1,2}(\alpha_{2,2} - \beta_{2,2})\lambda_{1,\infty} - \alpha_{1,2}\beta_{2,2}\lambda_{2,\infty})}{\alpha_{1,1}\beta_{1,2}\beta_{2,1}(\alpha_{2,2} + \beta_{2,2}) + \beta_{1,1}(\beta_{1,2}\beta_{2,1}(\alpha_{2,2} - \beta_{2,2}) + \alpha_{1,2}\alpha_{2,1}\beta_{2,2})} \\
\frac{\beta_{1,2}\beta_{2,2}(\alpha_{2,1}\beta_{1,1}\lambda_{1,\infty} + (\alpha_{1,1} - \beta_{1,1})\beta_{2,1}\lambda_{2,\infty})}{\alpha_{1,1}\beta_{1,2}\beta_{2,1}(\alpha_{2,2} + \beta_{2,2}) + \beta_{1,1}(\beta_{1,2}\beta_{2,1}(\alpha_{2,2} - \beta_{2,2}) + \alpha_{1,2}\alpha_{2,1}\beta_{2,2})}
\end{pmatrix}.
\]

(3.7.1)

Because the positiveness and finiteness of both \( \lambda_1 \) and \( \lambda_2 \) will ensure the stationarity of the bivariate Hawkes process as all of the \( \alpha_{i,j} \)'s, \( \beta_{i,j} \)'s and \( \lambda_{i,\infty} \)'s are positive and finite, a simple sufficient condition for \( 0 \leq \lambda_i < \infty, \ i = 1, 2 \) would be

\[
\alpha_{1,1} < \beta_{1,1} \quad \text{and} \quad \frac{\alpha_{1,2}\alpha_{2,1}}{\beta_{1,2}\beta_{2,1}} < \left( 1 - \frac{\alpha_{1,1}}{\beta_{1,1}} \right) \left( 1 - \frac{\alpha_{2,2}}{\beta_{2,2}} \right),
\]

(3.7.2)

which is equivalent to the positive definiteness of the matrix \((I - \Gamma)\).
3.7.2 Explicit Expression for the Covariance Density Matrix

As stated in Section 3.2, here I give the explicit expression for $\Xi(t), \ t > 0$, the covariance density matrix of the bivariate Hawkes process. For clarity purpose and, as I did in my GMM estimation, I will do the case for $\alpha_{2,1} = 0$.

Remember that the fourier transform of $\Xi(t), \ t > 0, \ \widetilde{\Xi}(\omega)$, is given by (3.3.7) in Section 3.2. Since $\widetilde{\Gamma}(\omega)$ is given by

$$
\widetilde{\Gamma}(\omega) = \begin{pmatrix} \frac{\alpha_{1,1}}{\omega + \beta_{1,1}} & \frac{\alpha_{1,2}}{\omega + \beta_{1,2}} \\ \frac{\alpha_{2,1}}{\omega + \beta_{2,1}} & \frac{\alpha_{2,2}}{\omega + \beta_{2,2}} \end{pmatrix}
$$

(3.7.3)

(where $i = \sqrt{-1}$), (3.3.7) delivers

$$
\widetilde{\Xi}(\omega) = \begin{pmatrix} \widetilde{\Xi}(\omega)_{1,1} & \widetilde{\Xi}(\omega)_{1,2} \\ \widetilde{\Xi}(\omega)_{2,1} & \widetilde{\Xi}(\omega)_{2,2} \end{pmatrix},
$$

(3.7.4)

where

$$
\widetilde{\Xi}(\omega)_{1,1} = -\frac{\lambda_1 \alpha_{1,1} (\alpha_{1,1} - 2\beta_{1,1}) (\omega^2 + \beta_{1,1}^2) (\omega^2 + (\alpha_{2,2} - \beta_{2,2})^2) + \lambda_2 \alpha_{1,2}^2 (\omega^2 + \beta_{1,2}^2) (\omega^2 + \beta_{2,2}^2)}{(\omega^2 + (\alpha_{1,1} - \beta_{1,1})^2) (\omega^2 + \beta_{1,2}^2) (\omega^2 + (\alpha_{2,2} - \beta_{2,2})^2)}
$$

$$
\widetilde{\Xi}(\omega)_{1,2} = -\frac{i \lambda_2 \alpha_{1,2} (\omega - i\beta_{1,1}) (\omega^2 + \beta_{2,2}^2)}{(\omega + i\alpha_{1,1} - i\beta_{1,1}) (\omega - i\beta_{1,2}) (\omega^2 + (\alpha_{2,2} - \beta_{2,2})^2)}
$$

$$
\widetilde{\Xi}(\omega)_{2,1} = \frac{i \lambda_2 \alpha_{1,2} (\omega + i\beta_{1,1}) (\omega^2 + \beta_{2,2}^2)}{(\omega - i\alpha_{1,1} + i\beta_{1,1}) (\omega + i\beta_{1,2}) (\omega^2 + (\alpha_{2,2} - \beta_{2,2})^2)}
$$

$$
\widetilde{\Xi}(\omega)_{2,2} = -\frac{\lambda_2 \alpha_{2,2} (\alpha_{2,2} - 2\beta_{2,2})}{\omega^2 + (\alpha_{2,2} - \beta_{2,2})^2}.
$$
By taking the inverse fourier transform of $\Xi(\omega)$, we have

$$
\Xi_{1,1}(t) = \frac{1}{2} \left[ e^{t(\alpha_{11}-\beta_{11})} \lambda_1 \alpha_{11} (\alpha_{11} - 2\beta_{11}) \right. \\
\left. + \lambda_2 \alpha_{12} \times \\
\left( -e^{t(\alpha_{22}-\beta_{22})} \alpha_{22} (-\beta_{11}^2 + (\alpha_{22} - \beta_{22})^2) (\alpha_{22} - 2\beta_{22}) (\alpha_{22} - \beta_{22})^{-1} \\
\left( -\beta_{11}^2 + (\alpha_{22} - \beta_{22})^2 \right) (\alpha_{11} + \alpha_{22} - \beta_{11} - \beta_{22}) (-\alpha_{11} + \alpha_{22} + \beta_{11} - \beta_{22}) \\
- e^{t(\alpha_{11}-\beta_{11})} \alpha_{11} (\alpha_{11} - 2\beta_{11}) \left( (\alpha_{11} - \beta_{11})^2 - \beta_{11}^2 \right) (\alpha_{11} - \beta_{11})^{-1} \\
\left( (\alpha_{11} - \beta_{11})^2 - \beta_{11}^2 \right) (\alpha_{11} + \alpha_{22} - \beta_{11} - \beta_{22}) (\alpha_{11} - \alpha_{22} - \beta_{11} + \beta_{22}) \\
+ \frac{e^{-t\beta_{12}} (-\beta_{11}^2 + \beta_{12}^2) (\beta_{12} - \beta_{22}^2) \beta_{11}^{-1}}{(\alpha_{11} - \beta_{11} + \beta_{12}) (-\alpha_{11} + \beta_{11} + \beta_{12}) (\alpha_{22} + \beta_{12} - \beta_{22}) (-\alpha_{22} + \beta_{12} + \beta_{22})} \right],
$$

$$
\Xi_{1,2}(t) = \\
- \frac{1}{2} \lambda_2 \alpha_{12} \left[ - e^{t(\alpha_{22}-\beta_{22})} \alpha_{22}^2 \left( -\alpha_{11} + \alpha_{22} + \beta_{11} - \beta_{22} \right) (\alpha_{22} + \beta_{12} - \beta_{22}) \\
+ e^{t(\alpha_{22}-\beta_{22})} \alpha_{22} (-\alpha_{22} \beta_{11} + 2 (\alpha_{22} + \beta_{11}) \beta_{22} - 2\beta_{22}^2) \\
\left( \alpha_{22} - \beta_{22} \right) (-\alpha_{11} + \alpha_{22} + \beta_{11} - \beta_{22}) (\alpha_{22} + \beta_{12} - \beta_{22}) \\
+ 2 e^{-t\beta_{12}} \left( \frac{\exp(t(\alpha_{11}-\beta_{11}+\beta_{12})) \alpha_{11} ((\alpha_{11} - \beta_{11})^2 - \beta_{12}^2) \alpha_{11}}{(\alpha_{11} + \alpha_{22} - \beta_{11} - \beta_{22}) (\alpha_{11} - \alpha_{22} - \beta_{11} + \beta_{22})} + \frac{\beta_{11} - \beta_{12} (\beta_{12}^2 - \beta_{22}^2)}{\beta_{11}^2 - (\alpha_{22} - \beta_{22})^2} \right) \right],
$$

and

$$
\Xi_{2,1}(t) = - \frac{e^{t(\alpha_{22}-\beta_{22})} \lambda_2 \alpha_{21} \alpha_{22} (\alpha_{22} - 2\beta_{22}) (\alpha_{22} - \beta_{11} - \beta_{22})}{2 (\alpha_{22} - \beta_{22}) (\alpha_{11} + \alpha_{22} - \beta_{11} - \beta_{22}) (\alpha_{22} - \beta_{12} - \beta_{22})},
$$

$$
\Xi_{2,2}(t) = \frac{e^{t(\alpha_{22}-\beta_{22})} \lambda_2 \alpha_{22} (\alpha_{22} - 2\beta_{22})}{2 (\alpha_{22} - \beta_{22})},
$$
where

\[
\lambda_1 = \frac{\beta_{1,1} (\beta_{1,2} (\alpha_{2,2} - \beta_{2,2}) \lambda_{1,\infty} - \alpha_{1,2} \beta_{2,2} \lambda_{2,\infty})}{(-\alpha_{1,1} + \beta_{1,1}) \beta_{1,2} (\alpha_{2,2} - \beta_{2,2})}
\]

\[
\lambda_2 = \frac{\beta_{2,2} \lambda_{2,\infty}}{-\alpha_{2,2} + \beta_{2,2}}
\]

from the computation in the first part of this Section.


