ABSTRACT

Title of dissertation: ESSAYS ON
EMPIRICAL MARKET MICROSTRUCTURE

Selahattin Tugkan Tuzun, Doctor of Philosophy, 2011

Dissertation directed by: Professor Albert S. Kyle
Department of Finance
Robert H. Smith School of Business

The first essay examines the events of May 6, 2010: the "Flash Crash". The Flash Crash, a brief period of extreme market volatility on May 6, 2010 raised questions about the current structure of the U.S. financial markets. Audit-trail data from U.S. Commodity Futures Trading Commission (CFTC) is used to describe the structure of the E-mini S&P 500 stock index futures market on May 6. In this study, three questions are asked. How did High Frequency Traders (HFTs) trade on May 6? What may have triggered the Flash Crash? What role did HFTs play in the Flash Crash? There is evidence which supports that HFTs did not trigger the Flash Crash, but their responses to the unusually large selling pressure on that day exacerbated market volatility.

The second essay examines the relationship between mutual fund trading and liquidity consumption in financial markets. Using Thompson Mutual Funds holdings data and the Trade and Quotes (TAQ) data, we relate the mutual fund trading to liquidity consumption. Mutual fund trading is positively correlated with liquidity

consumption. Mutual fund sensitivity to liquidity consumption differs based on mutual fund investment style. Large trades reveal the trading activity of actively managed mutual funds whereas the trading activity of index funds can be explained by small trades. This is consistent with a plausible explanation that index funds need to use small trades to rebalance their portfolios and information motivates the large trades of active mutual funds.

The third essay tests the predictions of trading game invariance using the sample of trades from TAQ dataset from 1993 to 2008. The theory of trading game invariance predicts that the distribution of trade sizes as a fraction of trading volume should vary across stocks proportionally to their trading activity in -2/3 power and that the number of trades should vary across stocks proportionally to their trading activity in 2/3 power. The data supports predictions of the invariance theory. For the number of trades, the estimated power coefficient of 0.69 (with standard errors of 0.001) is especially close to the predicted one of 2/3 on the subsample before 2001. These estimates increases to 0.79 (with standard errors of 0.004) after 2001 following a structural break related to a reduction in tick size and a consequent spread of algorithmic trading. Furthermore, the entire distribution of trade size shifts with the trading activity in a manner predicted by invariance theory. When trade sizes are adjusted for differences in trading activity, then their distribution is stable across stocks and similar to the distribution of a log-normal variable, truncated at the 100-share threshold.

ESSAYS ON EMPIRICAL MARKET MICROSTRUCTURE

by

Selahattin Tugkan Tuzun

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Albert S. Kyle, Chair/Advisor
Professor Russ Wermers
Professor Mark Loewenstein
Professor Anna Obizhaeva
Professor Peter Cramton

To “my beautiful and lonely country.”

.

Acknowledgments

This dissertation significantly underrepresents what I have learned from the faculty at the University of Maryland. It is also hard to summarize the story of my last six years which eventually produced this dissertation. But I will try anyway. Here it comes:

The courses I took in my first semester from Gerard Hoberg and Gordon Phillips changed my perception about Financial Economics. The more I read and learned in these courses, the better I realized how much I did not know and the stronger motivation I built to study more. Although studying more broadened my knowledge, it deepened my fear for the unknown. Advanced courses from Lemma Senbet, Mark Loewenstein and Pete Kyle expanded the boundaries of my intellect even further. At that time, I quickly realized that no matter how hard you study and how much you learn, the unknown will always be there. It is the ability to analyze critical issues I learned from these people will help me find and tackle the unknown. I am grateful to all these people for helping my transition from being a student to being a researcher.

Let me talk more about my advisor, Albert S. “Pete” Kyle. I think I had the best advisor any PhD student could ask for. What makes him special among PhD students in the Finance department is not just his intelligence, but his approach to PhD students. He spends long hours with PhD students not because he has to but he likes to. I have had meetings scheduled for thirty minutes becoming easily over three hours and two hour classes becoming over four hours. During all these years, I have not heard anyone complaining about overtime meetings or classes other than the people who reserved the room after us. I am indebted to Pete Kyle for his advice, guidance and personality.

The first chapter is my working paper titled *"The Flash Crash: The Impact of High Frequency Trading on an Electronic Market"* joint with Andrei Kirilenko, Albert S. "Pete" Kyle and Mehrdad Samadi. This chapter also uses the CFTC transaction level audit-trail data. I would like to thank the people at the Office of Chief Economist at the CFTC, especially Andrei Kirilenko and Mehrdad Samadi for providing an excellent collegiate atmosphere. Andrei Kirilenko has always had faith in me for becoming a good researcher and genuinely supported me every time I needed. Thanks to him for becoming a fantastic mentor. The second chapter is my single authored working paper titled *"Mutual Funds and Liquidity Consumption"*. Russ Wermers should have a special place in my acknowledgments. He was very generous with his time and guidance for all these years. My conversations with him benefited especially the second chapter. The third essay is my working paper titled *"Trading Game Invariance in the TAQ Dataset"* joint with Albert S. "Pete" Kyle and Anna Obizhaeva. It is essentially a product of my long and fun meetings with Anna Obizhaeva and Pete Kyle. I am also grateful to Mark Loewenstein and Gerard Hoberg for being extremely approachable and candid. I would also like to thank Peter Cramton for letting me participate in his valuable auction experiments and being in my PhD committee. Moreover, I am grateful to my professors at the University of Virginia, Charles Holt and Thomas W. Epps who endorsed my decision to pursue a Finance PhD in the first place. My fellow PhD students, especially Nitish

Sinha and Aysun Alp should get special thanks for offering true friendship in the school.

One can only be as lucky as I am to have friends like mine. My friends, Christina Arenas, Özgün Atasoy, Volkan Bayram, Burcu Ciğerli, Ozan Çağlargil, Dilek Günneç, Tuğba Gürcanlar, Serhan Kürkçü and Didem Tüzemen were never fed up with listening to mini-stories from my PhD experience. They have always motivated me in the right direction. My brother, Serkan Tüzün, and sister-in-law, Ebru Tüzün, made me truly feel they were standing by me. Many thanks to them. Lastly, my parents, Meral and Nazmi Tüzün, deserve special thanks from me for their support in every imaginable way. They have long waited for me to finish my PhD.

I shall now conclude with the song, "My Way", sung by Frank Sinatra which describes the journeys and moments like this one.

"And now, the end is near,

And so I face the final curtain.

...

...

...

Yes, it was my way."

Tuğkan Tüzün

April 18, 2011

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. FLASH CRASH AND HIGH FREQUENCY TRADING [1]

1.1 Introduction

On May 6, 2010, in the course of about 30 minutes, U.S. stock market indices, stock-index futures, options, and exchange-traded funds experienced a sudden price drop of more than five percent, followed by a rapid rebound. This brief period of extreme intraday volatility, commonly referred to as the "Flash Crash", raises a number of questions about the structure and stability of U.S. financial markets.

A survey conducted by Market Strategies International between June 23-29, 2010 reports that over 80 percent of U.S. retail advisors believe that "overreliance on computer systems and high-frequency trading" were the primary contributors to the volatility observed on May 6. Secondary contributors identified by the retail advisors include the use of market and stop-loss orders, a decrease in market maker trading activity, and order routing issues among securities exchanges.

Testifying at a hearing convened on August 11, 2010 by the Commodity Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC), representatives of individual investors, asset management companies, and market intermediaries suggested that in the current electronic marketplace, such an

[1] THIS CHAPTER DOES NOT REPRESENT THE VIEWS OF THE COMMODITY FUTURES TRADING COMMISSION, ITS COMMISSIONERS OR STAFF.

event could easily happen again.

In this paper, we describe trading in the bellwether E-mini Standard & Poor's (S&P) 500 equity index futures market on the day of the Flash Crash. We use audit-trail, transaction-level data for all regular transactions in the June 2010 E-mini S&P 500 futures contract (E-mini) during May 3-6, 2010 between 8:30 a.m. CT and 3:15 p.m. CT. This contract is traded exclusively on the Chicago Mercantile Exchange (CME) Globex trading platform, a fully electronic limit order market. For each transaction, we use data fields that allow us to identify the price, quantity and time of execution, the account id of the buyer and seller, order id, order type (market or limit), as well as the initiating side of the transaction (resting limit order or executable limit/market order).

Based on patterns of intraday volume, intraday inventory levels, and direction of trade, we classify each of more than 15,000 trading accounts that participated in transactions on May 6 into one of six categories which we name: High Frequency Traders (high volume and low inventory), Intermediaries (low inventory), Fundamental Buyers (consistent intraday net buyers), Fundamental Sellers (consistent intraday net sellers), Small Traders (low volume), Opportunistic Traders (all other traders not classified).

We investigate three questions. How did High Frequency Traders and other categories trade on May 6? What may have triggered the Flash Crash? What role did the High Frequency Traders play in the Flash Crash?

We find that on May 6, the 16 trading accounts that we classify as HFTs traded over 1,455,000 contracts, accounting for almost a third of total trading volume on that day. Yet, net holdings of HFTs fluctuated around zero so rapidly that they rarely held more than 3,000 contracts long or short on that day.

We also find that HFTs did not change their trading behavior during the Flash Crash. On the three days prior to May 6, and on May 6 itself—including specifically the period where prices were rapidly going down, the HFTs seem to exhibit the same trading patterns. Specifically, HFTs aggressively take liquidity from the market when prices were about to change and actively keep inventories near a target inventory level.

During the Flash Crash, High Frequency Traders initially bought contracts from Fundamental Sellers. After several minutes, HFTs proceeded to sell contracts and compete for liquidity with Fundamental Sellers. In this sense, the trading of HFTs, appears to have exacerbated the downward move in prices. In addition, HFTs appeared to rapidly buy and sell contracts from one another many times, generating a "hot potato" effect before Fundamental Buyers were attracted by the rapidly falling prices to step in and take these contracts off the market.

Each transaction in the Globex system results from a match of a executable order with a resting order. The CME audit-trail dataset explicitly labels the executable side of the transaction as aggressive and the non-executable side as passive. We find that approximately 46% of the volume High Frequency Traders trade is aggressively executed. For each category of traders, we define the aggressiveness imbalance of each trader category as the difference between the number of contracts aggressively bought and the number of contracts aggressively sold. We find that prices are more sensitive to the aggressiveness imbalances of High Frequency Traders and Opportunistic Traders than to the aggressiveness imbalances of Fundamental Buyers and Fundamental Sellers that take liquidity from the market. This may be due to High Frequency Traders ability to anticipate and react to price changes. Fundamental Traders do not have a large perceived price impact given their aggressiveness imbalance, possibly due to their desire to minimize their price impact and reduce transaction costs.

We find evidence of a significant increase in the number of contracts sold by Fundamental Sellers during the Flash Crash. Specifically, between 1:32 p.m. and 1:45 p.m. CT—the 13-minute period when prices rapidly declined—Fundamental Sellers were net sellers of more than 80,000 contracts while Fundamental Buyers were net buyers of only about 50,000 contracts. This level of net selling by Fundamental Sellers is about 15 times larger than their net selling over the same 13-minute interval on the previous three days, while this level of net buying by the Fundamental Buyers is about 10 times larger than their buying over the same time period on the previous three days.

In contrast, between 1:45 p.m. and 2:08 p.m. CT—the 23-minute period of the rapid price rebound of the E-mini—Fundamental Sellers were net sellers of more than 110,000 contracts and Fundamental Buyers were net buyers of more than 110,000 contracts. This level of net selling by Fundamental Sellers is about 10 times larger than their selling during same 23-minute interval on the previous three days, while this level of buying by the Fundamental Buyers is more than 12 times larger than their buying during the same interval on the previous three days.

The imbalance between Fundamental Buyers and Fundamental Buyers observed during the Flash Crash was many times larger than the inventories of High Frequency Traders. Opportunistic traders picked up the majority of the imbalance between Fundamental Buyers and Fundamental Sellers.

The CFTC-SEC May 6 report finds that the Flash Crash was triggered by a 75,000 contract sell program executed by a Fundamental Seller. Because net holdings of the HFTs were so small (rarely greater than 3000 contracts) relative to the selling pressure from the Fundamental Sellers on May 6, HFTs could have neither caused nor prevented the fall in prices without dramatically altering their trading strategies.

Nearly 40 years before the Flash Crash, Black (1971) conjectured that irrespective of the method of execution or technological advances in market structure, executions of large orders would always exert an impact on price. Black also conjectured that liquid markets exhibit price continuity only if trading is characterized by large volume coming from small individual trades.

This chapter proceeds as follows. In Section 1.2, we review the relevant literature. In Section 1.3, we summarize the public account of events on May 6, 2010. In Sections 1.4 and 1.5, we describe the E-mini S&P 500 futures contract and provide a description of the audit-trail, high frequency data we utilize. In Section 1.6, we describe our trader classification methodology. In Section 1.7, we present our analysis of the trading strategies of High Frequency Traders and Intermediaries. In Section 1.8, we describe the behavior of Fundamental Buyers and Sellers. In Section 1.9, we examine the activity of Opportunistic traders. In Section 1.10, we present the aggressivness imbalance regressions. In Section 1.11, we present our interpretation of the Flash Crash. Section 1.12 concludes the paper.

1.2 Literature

Nearly 40 years ago, when exchanges first contemplated switching to fully automated trading platforms, Fischer Black surmised that regardless of market structure, liquid markets exhibit price continuity only if trading is characterized by a large volume of small individual trades. Black (1971) also stated that large order executions would always exert an impact on price, irrespective of the method of execution or technological advances in market structure.

At that time, stock market "specialists" were officially designated market makers, obligated to maintain the order book and provide liquidity.[2] In the trading pits of the futures markets, many floor traders were unofficial, but easily identifiable market makers. Both the stock market specialists and futures market floor traders enjoyed a proximity advantage compared to traders who participated away from the trading floor. This advantage allowed specialists and floor traders to react more quickly to incoming order flow compared to other traders. Trading environments in which market makers are distinct from other traders are examined in the theoretical models of Kyle (1985) and Glosten and Milgrom (1985).

As markets became electronic, a rigid distinction between market makers and other traders became obsolete. Securities exchanges increasingly adopted a limit order market design, in which traders submit orders directly into the exchange's electronic systems, bypassing both designated and unofficial market makers. In

[2] Large orders were executed "upstairs" by block trading firms.

today's electronic markets, High Frequency Traders enjoy a latency advantage which allows them to react to changes in order flow more quickly than other traders. This occurred because of advances in technology, as well as regulatory requirements. Theoretical models of limit order markets include, among others, Parlour (1998), Foucault (1999), Biais, Martimor and Rochet (2000), Goettler, Parlour, and Rajan (2005, 2009), and Rosu (2009).

As more data became available, empirical research has confirmed a number of empirical regularities related to such issues as multiple characterizations of prices, liquidity, and order flow. Madhavan (2000), Biais, Glosten and Spatt (2005), and Amihud, Mendelson and Pedersen (2005) provide surveys of empirical market microstructure studies.

Most recently, Cespa and Foucault (2008) and Moallemi and Saglam (2010) propose theoretical models of latency - an increasingly important dimension of electronic trading. As low-latency, electronic limit order markets allowed for the proliferation of algorithmic trading strategies, a number of research studies aimed to examine algorithmic trading. Hendershott, Jones, and Menkveld (2010) and Hendershott and Riordan (2009) examine the impact of algorithmic traders in stock markets and find their presence beneficial. Chaboud et al (2009) study algorithmic traders in foreign exchange markets and reach similar conclusions. Hasbrouck and Saar (2010) and Brogaard (2010) examine certain types of algorithmic traders and find that they have a positive effect on market quality.

Another strand of literature examines optimal execution of large orders — a particular form of algorithmic trading strategies designed to minimize price impact and transaction costs. Studies on this issue include Bertsimas and Lo (1998), Almgren and Chriss (1999,2000), Engle and Ferstenberg (2007), Almgren and Lorenz (2006), and Schied and Schonenborn (2007).

Separately, Obizhaeva and Wang (2006) and Alfonsi and Schied (2008) study optimal execution by modeling the underlying limit order book. Brunnermier and Pedersen (2005), Carlin et al (2007), and Moallemi et al (2009) integrate the presence of an arbitrageur who can "front-run" a trader's execution. The majority of these studies find that it is optimal to split large orders into multiple executions to minimize price impact and transaction costs.

The effects of large trades on a market have also been thoroughly examined empirically by a multitude of authors starting with Kraus and Stoll (1972) who utilized data from the New York Stock Exchange.[3] These studies generally find that the execution of large orders exerts both permanent and temporary price impact, while reducing market liquidity.

1.3 Market Events on May 6, 2010: The Flash Crash

On May 6, 2010, major stock indices and stock index products rapidly dropped by more than 5 percent and then quickly recovered. The extreme intraday volatility in stock index prices is presented in Figure 1.1.

Between 13:45 and 13:47 CT, the Dow Jones Industrial Average (DJIA), S&P 500, and NASDAQ 100 all reached their daily minima. During this same period, all

[3] See, among others, Holthausen et al (1987, 1990), Chan and Lakonishok (1993, 1995), Chiyachantana et al (2004), Keim and Madhavan (1996, 1997), and Berkman (1996).

30 DJIA components reached their intraday lows. The DJIA components dropped from -4% to -36% from their opening levels. The DJIA reached its trough at 9,872.57, the S&P 500 at 1,065.79, and the NASDAQ 100 at 1,752.31. The E-mini S&P 500 index futures contract bottomed at 1,056.00.[4]

During a 13 minute period, between 13:32:00 and 13:45:27 CT, the front-month June 2010 E-mini S&P 500 futures contract sold off from 1127.75 to 1,070.00 , (a decline of 57.75 points or 5.1%). At 13:45:27, sustained selling pressure sent the price of the E-mini down to 1062.00. Over the course of the next second, a cascade of executed orders caused the price of the E-mini to drop to 1056.00 or 1.3%. The next executed transaction would have triggered a drop in price of 6.5 index points (or 26 ticks). This triggered the CME Globex Stop Logic Functionality at 13:45:28. The Stop Logic Functionality pauses executions of all transactions for 5 seconds, if the next transaction were to execute outside the price range of 6 index points either up or down. During the 5-second pause, called the "Reserve State," the market remains open and orders can be submitted, modified or cancelled, however, execution of pending orders are delayed until trading resumes.

At 13:45:33, the E-mini exited the Reserve State and the market resumed trading at 1056.75. Prices fluctuated for the next few seconds. At 13:45:38, price of the E-mini began a rapid ascent, which, while occasionally interrupted, continued until 14:06:00 when the price reached 1123.75, equivalent to a 6.4% increase from that day's low of 1056.00. At this point, the market was practically at the same

[4] For an in-depth review of the events of May 6, 2010, see the CFTC-SEC Staff Report entitled Preliminary Findings Regarding the Market Events of May 6, 2010.

price level where it was at 13:32:00 when the rapid sell-off began.

Trading volume of the E-mini increased significantly during the period of extreme price volatility. Figure 1.2 presents trading volume and transaction prices on May 6, 2010 over 1 minute intervals.

During the period of extreme market volatility, a large sell program was executed in the June 2010 E-mini S&P 500 futures contract. " At 2:32 p.m., against this backdrop of unusually high volatility and thinning liquidity, a large fundamental trader (a mutual fund complex) initiated a sell program to sell a total of 75,000 E-Mini contracts (valued at approximately $4.1 billion) as a hedge to an existing equity position...This large fundamental trader chose to execute this sell program via an automated execution algorithm (Sell Algorithm) that was programmed to feed orders into the June 2010 E-Mini market to target an execution rate set to 9% of the trading volume calculated over the previous minute...The execution of this sell program resulted in the largest net change in daily position of any trader in the E-Mini since the beginning of the year (from January 1, 2010 through May 6, 2010). Only two single-day sell programs of equal or larger size one of which was by the same large fundamental trader were executed in the E-Mini in the 12 months prior to May 6. When executing the previous sell program, this large fundamental trader utilized a combination of manual trading entered over the course of a day and several automated execution algorithms which took into account price, time, and volume. On that occasion it took more than 5 hours for this large trader to execute the first 75,000 contracts of a large sell program." [5]

[5] see Findings Regarding the Market Events of May 6, 2010

1.4 CME's E-mini S&P 500 Equity Index Contract

The CME S&P 500 E-mini futures contract was introduced on September 9, 1997. The E-mini trades exclusively on the CME Globex trading platform in a fully electronic limit order market. Trading takes place 24 hours a day with the exception of short technical break periods. The notional value of one E-mini contract is $50 times the S&P 500 stock index. The tick size for the E-mini is 0.25 index points or $12.50.

The number of outstanding E-mini contracts is created directly by buying and selling interests. There is no limit on how many contracts can be outstanding at any given time. At any point in time, there are a number of outstanding E-mini contracts with different expiration dates. The E-mini expiration months are March, June, September, and December. On any given day, the contract with the nearest expiration date is called the front-month contract. The E-mini is cash-settled against the value of the underlying index and the last trading day is the third Friday of the contract expiration month. Initial margin for speculators and hedgers(members) are $5,625 and $4,500, respectively. Maintenance margins for both speculators and hedgers(members) are $4,500. Empirically, it has been documented that the E-mini futures contract contributes the most to price discovery of the S&P 500 Index.[6]

The CME Globex matching algorithm for the E-mini offers strict price and time priority. Specifically, limit orders that offer more favorable terms of trade (sells at lower prices and buys at higher prices) are executed prior to pre-existing orders.

[6] See, Hasbrouck (2003).

Orders that arrived earlier are executed before other orders at the same price. This market operates under complete price transparency and anonymity. When a trader has his order filled, the identity of his counterparty is not available.

1.5 Data

We utilize audit trail, transaction-level data for all outright transactions in the June 2010 E-mini S&P 500 futures contract. These data come from the Computerized Trade Reconstruction (CTR) dataset, which the CME provides to the CFTC. We examine transactions occurring from May 3, 2010 through May 6, 2010, when the markets of the underlying equities of the S&P 500 index are open and before the daily halt in trading, i.e. weekdays between 8:30 a.m. CT and 3:15 p.m. CT. Price discovery typically occurs in the front month contract; the June 2010 contract was the nearby, most actively traded futures contract on May 6.

For each transaction, we use the following data fields: date, time (transactions are recorded by the second), executing trading account, opposite account, buy or sell flag, price, quantity, order ID, order type (market or limit), and aggressiveness indicator (indicates which trader initiated a transaction). These fields allow us to identify two trading accounts for each transaction: a buyer and seller, identify which account initiated a transaction, and whether the parties used market or limit orders to execute the transaction. We can also group multiple executions into an order. Table 1.1 provides summary of statistics for the June 2010 E-Mini S&P 500 futures contract during May 3-6, 2010.

According to Table 1.1, limit orders are the most popular tool for execution in this market. In addition, according to Table 1.1, trading volume on May 6 was significantly higher compared to the average daily trading volume during the previous three days.

1.6 Trader Categories

Financial markets are composed of traders that have different holding horizons and trading strategies. Some traders accumulate a position and hold it overnight. Other traders will accumulate a position and offset it within minutes. Yet another group of traders establish and offset a position within a matter of seconds.

Motivated by this and the absence of any designations in the E-mini market, we designate individual trading accounts into six categories based on their trading activity. Our classification method, which is described in detail below, produces the following categories of traders: High Frequency Traders (16 accounts), Intermediaries (179 accounts), Fundamental Buyers (1263), Fundamental Sellers (1276), Opportunistic Traders (5808) and Small Traders (6880).

We define Intermediaries as short horizon investors who follow a strategy of buying and selling a large number of contracts to stay around a relatively low target level of inventory. Specifically, we designate a trading account as an Intermediary if its trading activity satisfies the following two criteria. First, the account's net holdings fluctuate within 1.5% of its end of day level. Second, the account's end of day net position is no more than 5% of its daily trading volume. Together, these two criteria select accounts whose trading strategy is to participate in a large number of transactions, but to rarely accumulate a significant net position.

We define High Frequency Traders as a subset of Intermediaries, who individually participate in a very large number of transactions. Specifically, we order Intermediaries by the number of transactions they participated in during a day (daily trading frequency), and then designate accounts that rank in the top 7% as High Frequency Traders. This cutoff captures the significant difference in magnitude of trading activity between High Frequency Traders and Intermediaries. Once we designate a trading account as a HFT, we remove this account from the Intermediary category to prevent double counting. [7]

We define as Fundamental Traders trading accounts which mostly bought or sold in the same direction during May 6. Specifically, to qualify as a Fundamental Trader, a trading account's end of day net position on May 6 must be no smaller than 15% of its trading volume on that day. This criterion selects accounts that accumulate a significant net position by the end of May 6. Fundamental traders are further separated into Fundamental Buyers and Sellers, depending on whether their end of day net position is positive or negative, respectively. These traders appear to hold their positions for longer periods of time.

We define Small Traders as trading accounts which traded no greater than 9 contracts on May 6.

We classify the remaining trading accounts as Opportunistic Traders. Oppor-

[7] To account for a possible change in trader behavior on May 6, we classify HFTs and Intermediaries using trading data for May 3-5, 2010. We use data for May 6, 2010 to designate traders into other trading categories.

tunistic Traders may behave like Intermediaries (both buying and selling around a target net position) and at other times may behave like Fundamental traders (accumulating a directional long or short position).

Figure 1.3 illustrates the grouping of all trading accounts that transacted on May 6 into six categories of traders. The panels of Figure 1.3 presents individual trading accounts trading volume (vertical axis) and net position scaled by market trading volume (horizontal axis) for May 3-6.

Figure 1.3 shows that different categories of traders occupy quite distinct, albeit overlapping, positions in the "ecosystem" of a liquid, fully electronic market. HFTs, while very small in number, account for significant portion of trading volume. However, HFTs do not accumulate a large net position. Intermediaries also do not accumulate a large net position but trade much less volume than HFTs. Fundamental Traders accumulate directional positions. Some Fundamental Traders acquire large positions by executing many small-size orders, while others execute fewer large-size orders. Fundamental Traders which accumulate net positions by executing smaller orders may be disguising their trading activity in order to avoid being taken advantage of by the market. Opportunistic Traders at times act like Intermediaries (buying a selling around a given inventory target) and at other times act like Fundamental Traders (accumulating a directional position).

More formally, Table 1.2 presents descriptive statistics for these categories of traders and the overall market during May 3-5, 2010 and on May 6, 2010.

In order to characterize market participation of different categories of traders, we compute their shares of total trading volume. Table 1.2 shows that HFTs account for approximately 34% of total trading volume during May 3-5 and 29% of trading volume on May 6. Intermediaries account for approximately 10.5 % of trading volume during May 3-5 and 9% of trading volume on May 6. Trading volume of Fundamental Buyers and Sellers accounts for about 12% of the total trading volume during May 3-5. On May 6, Fundamental Buyers account for about 12% of total volume, while Fundamental Sellers account for 10% of total volume. We interpret the composition of this market as approximately 20% fundamental demand and 80% intermediation.

In order to further characterize whether categories of traders were primarily takers of liquidity, we compute the ratio of transactions in which they removed liquidity from the market as a share of their transactions.[8] According to Table 1.2, HFTs and Intermediaries have aggressiveness ratios of 45.68% and 41.62%, respectively. In contrast, Fundamental Buyers and Sellers have aggressiveness ratios of 64.09% and 61.13%, respectively.

This is consistent with a view that HFTs and Intermediaries generally provide liquidity while Fundamental Traders generally take liquidity. The aggressiveness ratio of High Frequency Traders, however, is higher than what a conventional definition of passive liquidity provision would predict.[9]

[8] When any two orders in this market are matched, the CME Globex platform automatically classifies an order as 'Aggressive' when it is executed against a 'Passive' order that was resting in the limit order book. From a liquidity standpoint, a passive order (either to buy or to sell) has provided visible liquidity to the market and an aggressive order has taken liquidity from the market. Aggressiveness ratio is the ratio of aggressive trade executions to total trade executions. In order to adjust for the trading activity of different categories of traders, the aggressiveness ratio is weighted either by the number of transactions or trading volume.

[9] This finding is consistent with that of Menkveld et al (2009). One possible explanation for the order aggressiveness ratios of HFTs is that some of them may actively engage in "sniping" orders resting in the limit order book. Cvitanic and Kirilenko (2010) model this trading behavior and conclude that under some conditions this trading strategy may have impact on prices. Similarly,

In order to better characterize the liquidity provision/removal across trader categories, we compute the proportion of each order that was executed aggressively.[10] Table 1.3 presents the distribution of ratios of order aggressiveness.

According to Table 1.3, the majority of High Frequency Traders' executed orders are entirely passive. Prior to May 6, about 79% of High Frequency Trader and Intermediary orders are resting orders. Executable limit orders are approximately 18% of total HFT orders and 20% of orders for Intermediaries.

As expected, Fundamental Traders utilize orders that consume more liquidity than the orders of HFTs and Intermediaries. During May 3-5, executable orders comprise 46% of the Fundamental Buyers' orders and 47% of the Fundamental Sellers' orders. On May 6, Fundamental Sellers use resting orders more often (59%) and executable orders less often (40%), whereas Fundamental Buyers use executable orders more often (63%) and resting orders less often (45%).

Moreover, during May 3-5, the average order size for both Fundamental Buyers and Sellers is approximately the same - about 15 contracts, while on May 6, the average executable order size of Fundamental Sellers (about 25 contracts) is more than 2.5 times larger than the average executable order size of Fundamental Buyers (about 9 contracts).

For all trader categories, order size exhibits an inverse U-shaped aggressiveness

Hasbrouck and Saar (2009) provide empirical support for a possibility that some traders may have altered their strategies by actively searching for liquidity rather than passively posting it.

[10] The following example illustrates how we compute the proportion of each order that was executed aggressively. Suppose that a trader submits an executable limit order to buy 10 contracts and this order is immediately executed against a resting sell order of 8 contracts, while the remainder of the buy order rests in the order book until it is executed against a new sell order of 2 contracts. This sequence of executions yields an aggressiveness ratio of 80% for the buy order, 0% for the sell order of 8 contracts, and 100% for the sell order of 2 contracts.

pattern: smaller orders tend to be either entirely aggressive or entirely passive. In contrast, larger orders result in both passive and aggressive executions. The number of trades per order also follows a similar pattern with larger orders being filled by a greater number of trade executions.

1.7 High Frequency Traders and Intermediaries

Together HFTs and Intermediaries account for over 40% of the total trading volume. Given that they account for such a significant share of total trading, we find it essential to analyze their trading behavior.

1.7.1 HFTs and Intermediaries: Net Holdings

Figure 1.4 presents the net position holdings of High Frequency Traders during May 3-6, 2010.

According to Figure 1.4, HFTs do not accumulate a significant net position and their position tends to quickly revert to a mean of about zero. The net position of the HFTs fluctuates between approximately ± 3000 contracts.

Figure 1.5 presents the net position of the Intermediaries during May 3-6, 2010.

According to Figure 1.5, Intermediaries exhibit trading behavior similar to that of HFTs. They also do not accumulate a significant net position. Compared to the HFTs, the net position of the Intermediaries fluctuates within a more narrow band of ± 2000 contracts, and reverts to a lower target level of net holdings at a slower rate.

On May 6, during the initial price decline, HFTs accumulated a net long position, but quickly offset their long inventory (by selling) before the price decline accelerated. Intermediaries appear to accumulate a net long position during the initial decrease in price, but unlike HFTs, Intermediaries did not offset their position as quickly. The decline in the net position of the Intermediaries occurred when the prices begin to rebound.

1.7.2 HFTs and Intermediaries: Profits and Losses

In addition, we calculate the profits and losses of High Frequency Traders and Intermediaries on a transaction by transaction basis by employing the following formula.

$$PL_y = \sum_{t=0}^{i} [y_{t-1} \times \Delta p_t] \tag{1.1}$$

Where y_{t-1} represents the net position of a trader at the time of market transaction t and Δp_t represents the change in price since the last transaction in the market. This measure is calculated from the first transaction of our sample where $t = 0$ through the last transaction, i. Our measure of profitability makes the assumption that trading accounts begin the day with no position. In addition, this measure is comprised of both realized gains and unrealized gains.

Figure 1.6 shows the profits and losses of High Frequency Traders on May 3-6.

High Frequency Traders are consistently profitable although they never accumulate a large net position. This does not change on May 6 as they appear to have been even more successful despite the market volatility observed on that day.

Figure 1.7 shows the profits and losses of Intermediaries on May 3-6.

Intermediaries appear to be relatively less profitable than HFTs. During the Flash Crash, Intermediaries also appeared to have incurred significant losses. This consistent with the notion that the relatively slower Intermediaries were run over by the decrease in price.

Overall, HFTs do not accumulate a significant net position and their position tends to quickly revert to a mean of about zero. Combined with their large share of total trading volume (34%), HFTs seem to employ trading strategies to quickly trade through a large number of contracts, without ever accumulating a significant net position. These strategies may be operating at such a high speed, that they do not seem to be affected by the price level or price volatility.

In contrast to HFTs, Intermediaries tend to revert to their target inventory levels more slowly. Because of this, on May 6, Intermediaries may have gotten caught on the wrong side of the market as they bought when prices rapidly fell.

1.7.3 HFTs and Intermediaries: Net Holdings and Prices

We formally examine the second-by-second trading behavior of HFTs and Intermediaries by examining empirical regularities between their net holdings and prices. Equation 1.2 presents this in a regression framework.

$$\Delta y_t = \alpha + \phi \Delta y_{t-1} + \delta y_{t-1} + \sum_{i=0}^{20} [\beta_{t-i} \times \Delta p_{t-i}/0.25] + \epsilon_t \tag{1.2}$$

where y_t denotes portfolio holdings of HFTs or Intermediaries during second t, where $t = 0$ corresponds to 8:30:00 CT. We utilize the price midpoint of an interval to calculate Price changes, Δp_{t-i}, $i = 0, ..., 20$ are in ticks (0.25 index points) and the change in inventories, Δy_t, is in the number contracts. We interpret δ and ϕ as long-term and short-term mean reversion coefficients.[11]

Table 1.4 presents estimated coefficients of the regression above. Panels A and B report the results for May 3-5 and May 6, respectively. The t statistics are calculated using the Newey-West (1987) estimator.

The first column of Panel A presents regression results for HFTs during May 3-5. The coefficient estimate for the long-term mean reversion parameter is -0.005, and is statistically significant. This suggests that HFTs reduce 0.5% of their position in one second. This long-term mean reversion coefficient corresponds to an estimated half-life of the inventory holding period of 137 seconds. In other words, holding prices constant, HFTs reduce half of their net holdings in 137 seconds. This is significantly smaller than the specialist inventory half-life measures of Hendershott and Menkveld (2010) who employ NYSE dataset from 1994-2005. This may be due to a dramatic increase in speed of intermediation over the last few years. Another explanation may be that this result is due to the fact that market makers are designated in

[11] Dickey-Fuller tests verify that HFT holdings level, Intermediary holdings level, as well as first differences are stationary. This is consistent with the intraday trading practices of HFTs and Intermediaries to target inventory levels close to zero. Results are available upon request.

equity markets and we classify our traders with a specific set of criteria. [12]

Changes in net holdings of HFTs are statistically significantly positively related to changes in prices for the contemporaneous price change and the first 4 lags. The estimated coefficients are positive, consistently decaying from the high of 32.089 for the contemporaneous price to the low of 3.909 for the price 4 seconds prior. This can be interpreted as follows: a one tick increase in current price corresponds to a increase of about 32 contracts in the net holdings of HFTs. Moreover, a one tick increase in the current price corresponds to an increase of up to 67 contracts during the next 4 seconds.

In contrast, estimated coefficients for lagged prices 10 to 20 seconds prior to the current holding period are negative and statistically significant. These estimated coefficients fall within a much more narrow range of -2.208 and -5.860. This, in turn, means that a one tick increase in price 10 to 20 seconds before corresponds to a maximum cumulative decrease in net holdings of about 39 contracts.

We interpret these results as follows. HFTs appear to trade in the same direction as the contemporaneous price and prices of the past four seconds. In other words, they buy, if the immediate prices are rising. However, after about ten seconds, they appear to reverse the direction of their trading - they sell, if the prices 10-20 seconds before were rising.

These regression results suggest that, possibly due to their speed advantage or superior ability to predict price changes, HFTs are able to buy right as the prices

[12] We calculate the estimated half-life of the inventory holding period as $\frac{ln(0.5)}{(\delta)}$.

are about to increase.[13] HFTs then turn around and begin selling 10 to 20 seconds after a price increase.

The second column of Panel A presents regression results for the Intermediaries on May 3-5. Similarly to HFTs, the long term mean reversion coefficient for the Intermediaries is -0.004 and is statistically significant. This suggests that the Intermediaries reduce their net holdings by 0.4% after one second. The half-life of their inventory is 173 seconds.

In marked contrast to HFTs, coefficient estimates for the contemporaneous price and the price one second before are negative (and significant) at -13.540 and -1.218, respectively. However, at prices 3 to 8 seconds prior, the estimated coefficients are positive and significant.

These coefficients could be interpreted as follows. The Intermediaries sell when the immediate prices are rising, and buy if the prices 3-8 seconds before were rising. These regression results suggest that, possibly due to their slower speed or inability to anticipate possible changes in prices, Intermediaries buy when the prices are already falling and sell when the prices are already rising.

Panel B presents the results of equation 1.2 on May 6. The first column of Panel B shows the results for HFTs. The coefficient for the lagged change in holdings parameter is positive but statistically insignificant at the 5% level. The coefficients for contemporaneous and 1st lagged price changes are positive at 10.808 and 4.625, respectively.

[13] We also introduce lead price changes up to 10 seconds in this regression framework. Prior to May 6, lead price change coefficients are positive and significant up to three seconds for HFTs while they are negative and significant for Intermediaries. Results are available upon request.

This result may suggest that that on May 6, HFTs repeatedly reversed the direction of their trading (e.g., become contrarian, switching from buying to selling, or otherwise) significantly sooner than during May 3-5.

The second column of Panel B reports the results for the change in holdings of Intermediaries on May 6th. The contemporaneous price change estimate is -8.164. The lagged price change coefficients become positive for the next 3 lagged price changes, decaying from 6.635 to 1.138.

We interpret the difference in results between these two samples to a change in Intermediary behavior during the Flash Crash. This may be due to a reduction in liquidity provision from this trader category during the Flash Crash.

1.7.4 HFTs and Intermediaries: Liquidity Provision/Removal

We consider Intermediaries and HFTs to be very short term investors. They do not hold positions over long periods of time and revert to their target inventory level quickly. Observed trading activity of HFTs can be separated into three parts. First, HFTs seem to anticipate price changes (in either direction) and trade aggressively to profit from it. Second, HFTs seem to submit resting orders in the direction of the anticipated the price move. Third, HFTs trade to keep their inventories within a target level. The inventory-management trading objective of HFTs may interact with their price-anticipation objective. In other words, at times, inventory-management considerations of HFTs may lead them to aggressively trade in the same direction as the prices are moving, thus, taking liquidity. At other times, in order to revert to their target inventory levels, HFTs may passively trade against price movements and, thus, provide liquidity.

In order to examine the liquidity providing and taking behavior of HFTs and Intermediaries, we separate their changes in holdings into aggressive changes (those incurred via aggressive acquisitions) and passive changes (those incurred via passive acquisitions). Specifically, when traders submit marketable orders into the order book, they are considered to be aggressive. Conversely, the traders' resting orders being executed by a marketable order result in passive execution.

Table 1.5 presents the regression results of the two components of change in holdings on lagged inventory, lagged change in holdings and lagged price changes over one second intervals. Panel A and Panel B report the results for May 3-5 and May 6th, respectively.

The dependent variable in the first column of Panel A is the aggressive change in holdings of HFTs on May 3-5. The short term and long term mean reversion coefficients are statistically significant, -0.042% and -.005%, respectively. In other words, HFTs aggressively reduce 0.5% of their holdings in one second. The coefficient estimates for price changes are positive for the contemporaneous and first 4 lagged prices, decaying from 57.778 to 3.290. This can be interpreted as follows: a one tick increase in current price corresponds to an aggressive increase of position of about 58 contracts by HFTs. Moreover, a one tick increase in the current price corresponds to an increase of up to 99 contracts during the next 4 seconds.

The second column of Panel A presents the regression results for the passive change in holdings of HFTs on May 3-5. The coefficient for lagged change in holdings is 0.036 and statistically significant. The long term mean reversion estimate is

-0.001, which is smaller than the coefficient from the aggressive holdings change regression. The coefficient estimates for the price changes are almost always negative. The contemporaneous and first lagged price changes are negative and statistically significant; ranging from -25.689 for the contemporaneous price change to -5.371 for the 1st lagged price change.

Given the difference in magnitude between the aggressive and passive long term mean reversion coefficients, we interpret these results as follows, HFTs may be reducing their positions and reacting to anticipated price changes by submitting marketable orders. In addition, passive holdings changes of HFTs reflect liquidity provision.

The dependent variable in the third column of Panel A is the aggressive holdings change of the Intermediaries on May 3-5. The coefficients for lagged change in holdings and lagged inventory level are 0.007 and -0.002, respectively. This result corresponds to Intermediaries reducing 0.2% of their holdings aggressively in one second. The coefficients for the current and lagged price changes are positive; decreasing from 6.377 for the current price change to 1.007 for the 10th lagged price change.

These estimates are smaller than the estimates for HFTs. Accordingly, we interpret these results as evidence suggesting that Intermediaries are slower than HFTs in responding to anticipated price changes.[14]

The fourth column of Panel A presents the results for the passive position

[14] We also introduce lead price changes up to 10 seconds into this regression framework. Price change coefficients are positive and significant for the aggressive trading of High Frequency Traders before May 6. Results are available upon request.

change component of Intermediaries' activity. The coefficient estimates for lagged change in holdings and lagged level of holding of Intermediaries are -0.013 and -0.002, respectively. These coefficients are similar to those we observe from the passive trading of Intermediaries. The coefficient estimates for price changes are statistically significant and negative through the 3rd lag. The coefficients range from -19.917 for the current price change to -1.117 for the 3rd lagged price change.

Our interpretation of these results suggests that given the similar passive and aggressive mean reversion coefficients, Intermediaries use primarily marketable orders to move to their target inventory level. The passive holdings change for Intermediaries is also contrarian to price fluctuations, suggesting that the passive holdings change can be a good proxy for the liquidity provision of Intermediaries.

In summary, the larger coefficient for the Aggressive long term mean reversion parameter, suggests that HFTs very quickly reduce their inventories by submitting marketable orders. They also aggressively trade when prices are about to change. Over slightly longer time horizons, however, HFTs sometimes act as providers of liquidity.

The first column of Panel B presents the results for aggressive holdings change of HFTs on May 6th. Only the coefficient on the current price change is positive and statistically significant; 23.703. The second column of Panel B shows the results for passive holdings change of HFTs. The contemporaneous price coefficient, -12.895, is statistically significant.

These results are qualitatively similar to those we observe on the 3 days prior to May 6. Therefore, we interpret these results as evidence that HFTs did not significantly alter their behavior during the Flash Crash. However, they may have executed their trading strategies faster as price volatility increased.

The third column of Panel B presents the results for the aggressive positions change of Intermediaries. The contemporaneous price change coefficient is 4.939 and statistically significant. The fourth column in Panel B displays the results for passive holdings change of Intermediaries. The contemporaneous price change coefficient is -13.103 and statistically significant.

The coefficients on price changes for the Intermediary passive holdings change regression are smaller than those we observe prior to May 6th. We interpret this as a possible decrease in liquidity provision by Intermediaries during the Flash Crash.

1.7.5 HFTs and Intermediaries: The Flash Crash

To examine these participants' activity at an even higher resolution during the Flash Crash. We employ equation 1.2 during the 36-minute period of the Flash Crash - starting at 13:32 p.m. and ending at 14:08 p.m. CT. We partition this sample into two sub samples, the price crash (DOWN, 13:32-13:45 p.m. CT) and recovery (UP, 13:45-14:08 CT), presented in Panels A and B, respectively of Table 1.6.

The first column of Panel A presents the results for aggressive holdings change of HFTs on May 6 during the rapid price decline. The long term mean reversion coefficient is -0.008 and statistically insignificant. The contemporaneous price change coefficient is positive and statistically significant at 24.226.

The second column of Panel A presents passive change in holding of HFTs during the price decline. The long term mean reversion coefficient is positive but statistically insignificant. The contemporaneous price coefficient is 8.533 and statistically significant.

We interpret these results as follows: As the price of the E-mini contract declined, High Frequency Traders were the counterparties to Opportunistic Traders' aggressive buying. However, the aggressive buying of Opportunistic Traders did not affect the direction of the price move. In addition, HFTs did not alter their behavior significantly when prices were rapidly going down. The shorter duration of statistical significance on price change coefficients may be a function of the price volatility observed during the Flash Crash.

The third column of Panel A presents the results for Intermediaries' aggressive position change on May 6th during as the price of the E-mini decreased rapidly. Price change coefficients are positive and statistically significant through the 2nd lag, ranging from 8.251 to 4.257.

The fourth column of Panel A presents the results for the passive position changes of Intermediaries during the decrease in price. The long term mean reversion coefficient is -0.012 and statistically significant. The coefficient for the contemporaneous price change is -9.603 and statistically significant.

These findings are not much different from those we obtain in previous regressions. Accordingly we interpret these results as evidence that Intermediaries did not seem to alter their trading strategies significantly as the price of the E-mini contract declined.

The dependent variable in the first column of Panel B is HFTs aggressive position change while the prices are rapidly going up. The long term mean reversion coefficient is -0.005 and statistically significant. The coefficient for the contemporaneous price change is -0.251 and statistically insignificant. These results are quantitatively different than those we observe in previous regressions.

We interpret this lack of statistical significance in the relationship between HFT aggressive net position changes and prices as being related to the increase in market volatility and the influx of Fundamental Buyers who bought as the price of the E-mini contract recovered after the trading pause.

The results in the second column of Panel B present the relation between prices and passive net position changes of HFTs when the prices were on their way up. The long term mean reversion coefficient is again insignificant. The statistically significant contemporaneous price change coefficient, -9.107, is similar to past regressions of passive holdings changes but differs from the result of 8.533 during the price decline.

We interpret these results as a continuation in liquidity provisions by HFTs as the price of the E-mini contract recovered to levels observed before the Flash Crash.

The third column of Panel B presents the regression results for the aggressive position change of Intermediaries. The long term mean reversion coefficient is -0.004 and is statistically significant. Coefficients are statistically significant and positive for the contemporaneous and first lagged price change at 2.912 and 2.150, respectively. This is smaller than the same coefficient during the regression of Intermediary aggressive holdings changes during the crash.

The fourth column of Panel B lists the regression results where the passive position changes of Intermediaries during the price recovery of the E-mini contract. Although the contemporaneous price coefficient is negative and statistically significant, the magnitude of this coefficient, -4.105, is considerably smaller the coefficient observed in the fourth column of Panel A.

We attribute this decrease in magnitude of contemporaneous price change to a decrease in liquidity provision by Intermediaries during this time period. However, the relatively smaller decrease in the aggressive holdings change coefficient compared to that of HFTs may be due to the increase in aggressiveness of Intermediaries who sought to offset their disadvantageous positions during the Flash Crash.

1.7.6 HFTs and Intermediaries: The Hot Potato Effect

A basic characteristic of futures markets is that they remain in zero net supply throughout the day. In other words, for each additional contract demanded, there is precisely one additional contract supplied. End of day open interest presents a single reading of the levels of supply and demand at the end of that day.

In intraday trading, changes in net demand/supply result from changes in net holdings of different traders within a specified period of time, e.g., one minute. These minute by minute changes in the net positions of individual trading accounts can be aggregated to get a minute by minute net change in holdings for our six trader categories. To change their net position by one contract, a trader may buy one contract or may buy 101 contracts and sell 100 contracts.

We examine the ratio of trading volume during one minute intervals to the change in net position over one second intervals to study the relationship between

High Frequency Trader trading volume and changes in net position. We calculate the same metric for Intermediaries and find that although High Frequency Traders are active before and during the Flash Crash, they do not significantly change their net positions.

Figure 1.8 presents the ratio of trading volume to net position change.

We find that compared to the three days prior to May 6, there was an unusually level of HFT "hot potato" trading volume — due to repeated buying and selling of contracts accompanied a relatively small change in net position. The hot potato effect was especially pronounced between 13:45:13 and 13:45:27 CT, when HFTs traded over 27,000 contracts, which accounted for approximately 49% of the total trading volume, while their net position changed by only about 200 contracts.

We interpret this finding as follows: the lack of Opportunistic and Fundamental Traders, as well as Intermediaries, with whom HFTs typically trade just before the E-mini price reached its trough, resulted in higher trading volume among HFTs, creating a hot potato effect. It is possible that during the period of high volatility, Opportunistic and Fundamental Traders were either unable or unwilling to efficiently submit orders. In the absence of their usual trading counterparties, HFTs were left to trade with other HFTs.

1.8 Fundamental Traders

Trading volume of the Fundamental Buyers and Sellers accounts for about 10-12% of the total trading volume both during May 3-5 and on May 6. However, Fundamental traders typically remove more liquidity from the market than they provide. As a result, a sizable program executed by the Fundamental traders is more likely to have a significant impact on the market.

In this section we examine the trading behavior of Fundamental traders. We ask the following question: Was the trading behavior of Fundamental Buyers and Sellers different on May 6, especially during the period of extreme price volatility?

Table 1.7 presents the average number of contracts bought and sold by different categories of traders during two time periods on May 3-5 and on May 6. For both May 3-5 and May 6, the period between 1:32 p.m. and 1:45 p.m. CT is defined as 'UP' and the period between 1:45 p.m. and 2:08 p.m. CT is defined as 'DOWN'.

According to Table 1.7, there a significant increase in the number of contracts sold by the Fundamental Sellers during the period of extreme price volatility on May 6 compared to the same period during the previous three days.

Specifically, between 1:32 p.m. and 1:45 p.m. CT, the 13-minute period when the prices rapidly declined, Fundamental Sellers sold more than 80,000 contracts net, while Fundamental Buyers bought approximately 50,000 contracts net. This level of net selling by the Fundamental Sellers is about 15 times larger compared to their net selling over the same 13-minute interval on the previous three days, while the level of net buying by the Fundamental Buyers is about 10 times larger compared to their net buying over the same time period on the previous three days.

In contrast, between 1:45 p.m. and 2:08 p.m. CT, the 23-minute period of the rapid price rebound, Fundamental Sellers sold more than 110,000 contracts net and Fundamental Buyers bought more than 110,000 contracts net. This level of selling by the Fundamental Sellers is about 10 times larger compared than their selling over the same 23-minute interval on the previous three days, while this level of buying by the Fundamental Buyers is more than 12 times larger compared to their buying over the same time period on the previous three days.

In order to visualize the activity of Fundamental and Opportunistic Traders, we calculate the change in net position of these traders during the time surrounding the Flash Crash.

As the price of the E-mini contract decreased, there was also an imbalance in trading activity between Fundamental Buyers and Sellers. Opportunistic Traders appear to have picked up the excess selling pressure. The price of the E-mini contract recovered as Fundamental Buyers entered the market.

1.9 Opportunistic Traders

Opportunistic Traders comprise approximately a third of trading accounts active on May 6. Accordingly, the trading behavior of Opportunistic Traders, especially during the Flash Crash, warrants discussion. These trading accounts' behavior differs from that of other trader categories.

1.9.1 Opportunistic Traders: Net Holdings

Opportunistic traders seem to exhibit mean reverting behavior similar to that of HFTs and Intermediaries, but also establish large net positions like Fundamental Traders. Figure 1.10 illustrates this point by presenting the net holdings of

Opportunistic traders on May 3-6.

Opportunistic traders increased their net position by approximately 70,000 contracts during the Flash Crash. This buying pressure came at an opportune time as prices had already fallen significantly.

1.9.2 Opportunistic Traders: Profits and Losses

Figure 1.11 shows the profits and losses of Opportunistic Traders on May 3-6.

The buying activity of Opportunistic Traders during the Flash Crash could have translated into substantial profits as a large portion of their buying was during the price rebound. However, it is important to note the assumptions of this calculation. We assume that traders begin the day with no preexisting position. Accordingly, the massive swings in profits and losses are a function of the large net position Opportunistic Traders established during the Flash Crash.

1.10 Aggressiveness Imbalance and Prices

We utilize the aggressiveness imbalance indicator to estimate the sensitivity of prices to the aggressiveness imbalances of various trader categories. Aggressiveness Imbalance is an indicator designed to capture the direction of the removal of liquidity from the market. Aggressiveness Imbalance is constructed as the difference between aggressive buy transactions minus aggressive sell transactions.

Figure 1.12 shows the relationship between price and cumulative Aggressiveness Imbalance (aggressive buys - aggressive sells).

In addition, we calculate aggressiveness imbalance for each category of traders over one minute intervals. For illustrative purposes, the Aggressiveness Imbalance indicator for HFTs and Intermediaries are presented in Figures 1.13 and 1.14, respectively.

According, to Figures 1.13 and 1.14, visually, HFTs behave very differently during the Flash Crash compared to the Intermediaries. HFTs aggressively sold on the way down and aggressively bought on the way up. In contrast, Intermediaries are about equally passive and aggressive both down and up.

More formally, we estimate sensitivity of prices to the aggressiveness imbalances of different categories of traders. The estimates are obtained by running the following minute-by-minute regressions:

$$\frac{\Delta P_t}{P_{t-1} \times \sigma_{t-1}} = \alpha + \sum_{i=1}^{5} [\lambda_i \times \frac{AGG_{i,t}}{Shr_{i,t-1} \times 100,000}] + \epsilon_t \tag{1.3}$$

The dependent variable in the regression is the price return scaled by the previous period's volatility.[15] The independent variables in the regression are the aggressiveness imbalance for each trader category scaled by the category's lagged share of market volume times 100,000. The Newey West (1987) estimator t is employed.Estimated coefficients are presented in Table 1.8.

Panel A of Table 1.8 presents regression results for the period May 3-5. The specification fits quite well with an R^2 of 36% and all estimated price sensitivity coefficients are statistically significant at 5% level.

[15] For the estimate of volatility, we use range - the natural logarithm of the maximum price over the minimum price.

HFTs and Opportunistic traders have the largest coefficients at 5.37 and 7.6, respectively. The coefficient of the Intermediaries is the lowest at 0.83. The coefficient for Fundamental Sellers (1.36) is about equal to that of the Fundamental Buyers (1.31).

Panel B of Table 1.8 presents regression results for May 6. The model seems to have a better fit with an R^2 of 59%. All slope coefficients are again statistically significant at 5% level. The coefficient for HFTs is smaller at 3.23. In contrast, the coefficients of the Intermediaries (5.99) is more than seven times larger on May 6 compared to the previous three days. The coefficient of Opportunistic traders on May 6 (7.49) is about the same as it is during May 3-5. However, the coefficient of Fundamental Sellers (0.53) is nearly double that of the Fundamental Buyers (0.53).

We interpret these results as follows. High Frequency Traders have a large, positive coefficient possibly due to their ability to anticipate price changes. In contrast, Fundamental Traders have much smaller coefficients, which is likely due to their explicit trading strategies that try to limit market impact, in order to minimize transaction costs.

To illustrate the fit of these regressions, we use the estimated coefficients from the aggressiveness imbalance regression during May 3-5 to fit minute-by-minute price changes on May 6 (Figure 1.15). According to Figure 1.15, the fitted price (marked line) is quite close to the actual price (solid line).

1.11 Discussion: The Flash Crash

We believe that the events on May 6 unfolded as follows. Financial markets, already tense over concerns about the European sovereign debt crisis, opened to news concerning the Greek government's ability to service its sovereign debt. As a result, premiums rose for buying protection against default on sovereign debt securities of Greece and a number of other European countries. In addition, the S&P 500 volatility index ("VIX") increased, and yields of ten-year Treasuries fell as investors engaged in a "flight to quality." By mid-afternoon, the Dow Jones Industrial Average was down about 2.5%.

Sometime after 2:30 p.m., Fundamental Sellers began executing a large sell program. Typically, such a large sell program would not be executed at once, but rather spread out over time, perhaps over hours. The magnitude of the Fundamental Sellers' trading program began to significantly outweigh the ability of Fundamental Buyers to absorb the selling pressure.

HFTs and Intermediaries were the likely buyers of the initial batch of sell orders from Fundamental Sellers, thus accumulating temporary long positions. Thus, during the early moments of this sell program's execution, HFTs and Intermediaries provided liquidity to this sell order.

However, just like market intermediaries in the days of floor trading, HFTs and Intermediaries had no desire to hold their positions over a long time horizon. A few minutes after they bought the first batch of contracts sold by Fundamental Sellers, HFTs aggressively sold contracts to reduce their inventories. As they sold contracts, HFTs were no longer providers of liquidity to the selling program. In fact, HFTs competed for liquidity with the selling program, further amplifying the price impact of this program.

Furthermore, total trading volume and trading volume of HFTs increased significantly minutes before and during the Flash Crash. Finally, as the price of the E-mini rapidly fell and many traders were unwilling or unable to submit orders, HFTs repeatedly bought and sold from one another, generating a "hot-potato" effect.

Yet, Fundamental Buyers, who may have realized significant profits from this large decrease in price, did not seem to be willing or able to provide ample buy-side liquidity. As a result, between 2:45:13 and 2:45:27, prices of the E-mini fell about 1.7%.

At 2:45:28, a 5 second trading pause was automatically activated in the E-mini. Opportunistic and Fundamental Buyers aggressively executed trades which led to a rapid recovery in prices. HFTs continued their strategy of rapidly buying and selling contracts, while about half of the Intermediaries closed their positions and got out of the market.

In light of these events, a few fundamental questions arise. Why did it take so long for Fundamental Buyers to enter the market and why did the price concessions had to be so large? It seems possible that some Fundamental Buyers could not distinguish between macroeconomic fundamentals and market-specific liquidity events. It also seems possible that the opportunistic buyers have already accumulated a significant positive inventory earlier in the day as prices were steadily declining.

Furthermore, it is possible that they could not quickly find opportunities to hedge additional positive inventory in other markets which also experienced significant volatility and higher latencies. An examination of these hypotheses requires data from all venues, products, and traders on the day of the Flash Crash.

1.12 Conclusion

In this paper, we analyze the behavior of High Frequency Traders and other categories of traders during the extremely volatile environment on May 6, 2010.

Based on our analysis, we believe that High Frequency Traders exhibit trading patterns inconsistent with the traditional definition of market making. Specifically, High Frequency Traders aggressively trade in the direction of price changes. This activity comprises a large percentage of total trading volume, but does not result in a significant accumulation of inventory. As a result, whether under normal market conditions or during periods of high volatility, High Frequency Traders are not willing to accumulate large positions or absorb large losses. Moreover, their contribution to higher trading volumes may be mistaken for liquidity by Fundamental Traders. Finally, when rebalancing their positions, High Frequency Traders may compete for liquidity and amplify price volatility.

Consequently, we believe, that irrespective of technology, markets can become fragile when imbalances arise as a result of large traders seeking to buy or sell quantities larger than intermediaries are willing to temporarily hold, and simultaneously long-term suppliers of liquidity are not forthcoming even if significant price concessions are offered.

We believe that technological innovation is critical for market development. However, as markets change, appropriate safeguards must be implemented to keep pace with trading practices enabled by advances in technology.

Tab. 1.1: Market Descriptive Statistics

	May 3-5	May 6th
Volume	2,397,639	5,094,703
# of Trades	446,340	1,030,204
# of Traders	11,875	15,422
Trade Size	5.41	4.99
Order Size	10.83	9.76
Limit Orders % Volume	95.45%	92.44%
Limit Orders % Trades	94.36%	91.75%
Volatility	1.54%	9.82%
Return	-0.02%	-3.05%

This table presents summary statistics for the June 2010 E-Mini S&P 500 futures contract. The first column presents averages calculated for May 3-5, 2010 between 8:30 and 15:15 CT. The second column presents statistics for May 6t, 2010 between 8:30 to 15:15 CT. Volume is the number of contracts traded. The number of traders is the number of trading accounts that traded at least once during a trading day. Order size and trade sizes are measured in the number of contracts. The use of limit orders is presented both in percent of the number of transactions and trading volume. Volatility is calculated as range, the natural logarithm of maximum price over minimum price within a trading day.

Tab. 1.2: Summary Statistics of Trader Categories

Panel A: May 3-5

Trader Type	% Volume	% of Trades	# Traders	Trade Size (Avg.)	Order Size (Avg.)	Limit Orders % Volume	Limit Orders % Trades	Agg Ratio Trade-Weighted	Agg Ratio Vol-Weighted
High Frequency Traders	34.22%	32.56%	15	5.69	14.75	100.000%	100.000%	49.91%	45.68%
Intermediaries	10.49%	11.63%	189	4.88	7.92	99.614%	98.939%	43.10%	41.62%
Fundamental Buyers	11.89%	10.15%	1,013	6.34	14.09	91.258%	91.273%	66.04%	64.09%
Fundamental Sellers	12.11%	10.10%	1,088	6.50	14.20	92.176%	91.360%	62.87%	61.13%
Opportunistic Traders	30.79%	33.34%	3,504	4.98	8.80	92.137%	90.549%	55.98%	54.71%
Small Traders	0.50%	2.22%	6,065	1.22	1.25	70.092%	71.205%	59.04%	59.06%
	Volume	# of Trades	# Traders	Trade Size (Avg.)	Order Size (Avg.)	Limit Orders % Volume	Limit Orders % Trades	Volatility	Return
All	2,397,639	446,340	11,875	5.41	10.83	95.45%	94.36%	1.54%	-0.02%

Panel B: May 6th

Trader Type	% Volume	% of Trades	# Traders	Trade Size (Avg.)	Order Size (Avg.)	Limit Orders % Volume	Limit Orders % Trades	Agg Ratio Trade-Weighted	Agg Ratio Vol-Weighted
High Frequency Traders	28.57%	29.35%	16	4.85	9.86	99.997%	99.997%	50.38%	45.53%
Intermediaries	9.00%	11.48%	179	3.89	5.88	99.639%	99.237%	45.18%	43.55%
Fundamental Buyers	12.01%	11.54%	1,263	5.15	10.43	88.841%	89.589%	64.39%	61.08%
Fundamental Sellers	10.04%	6.95%	1,276	7.19	21.29	89.985%	88.966%	68.42%	65.68%
Opportunistic Traders	40.13%	39.64%	5,808	5.05	10.06	87.385%	85.352%	61.92%	60.28%
Small Traders	0.25%	1.04%	6,880	1.20	1.24	63.609%	64.879%	63.49%	63.53%
	Volume	# of Trades	# Traders	Trade Size (Avg.)	Order Size (Avg.)	Limit Orders % Volume	Limit Orders % Trades	Volatility	Return
All	5,094,703	1,030,204	15,422	4.99	9.76	92.443%	91.750%	9.82%	-3.05%

This table presents summary statistics for trader categories and the overall market. The first column presents statistics prior to May 6 as the average over three trading days, May 3-5, 2010 from 8:30 to 15:15 CT. The second column presents statistics for May 6 from 8:30 to 15:15 CT.

Tab. 1.3: Order Properties

Panel A: May 3-5

	Order Aggressiveness Distribution						Avg Order Size						Avg # of Trades Per Order					
	HFT	INT	BUY	SELL	OPP.	SMALL	HFT	INT	BUY	SELL	OPP	SMALL	HFT	INT	BUY	SELL	OPP	SMALL
Agg=0	78.76%	78.61%	53.02%	52.22%	60.43%	43.11%	7.39	6.31	10.25	10.32	6.37	1.20	1.44	1.37	1.51	1.46	1.34	1.01
0<Agg<=0.1	0.19%	0.05%	0.04%	0.03%	0.04%	0.00%	88.22	52.64	218.07	106.97	239.06	0.00	11.61	8.29	47.71	10.58	40.98	0.00
0.1<Agg<=0.2	0.21%	0.06%	0.05%	0.03%	0.04%	0.00%	108.38	50.85	83.64	1089.06	319.81	0.00	14.10	9.67	9.06	195.18	49.90	0.00
0.2<Agg<=0.3	0.19%	0.08%	0.06%	0.04%	0.04%	0.00%	133.67	44.87	329.55	353.93	304.87	4.00	18.77	9.26	53.83	35.61	46.19	4.00
0.3<Agg<=0.4	0.27%	0.12%	0.07%	0.05%	0.05%	0.00%	103.16	24.16	150.01	474.38	275.38	0.00	16.29	6.31	19.48	61.70	45.78	0.00
0.4<Agg<=0.5	0.45%	0.25%	0.14%	0.10%	0.08%	0.00%	94.90	14.93	115.80	112.62	121.34	2.00	14.82	4.40	15.84	14.85	18.26	2.00
0.5<Agg<=0.6	0.19%	0.06%	0.05%	0.04%	0.05%	0.00%	219.60	50.06	273.76	344.49	238.27	0.00	35.78	11.07	45.81	54.19	37.54	0.00
0.6<Agg<=0.7	0.22%	0.10%	0.09%	0.06%	0.05%	0.00%	196.90	39.42	235.24	314.94	211.10	0.00	32.40	9.97	39.70	50.21	34.11	0.00
0.7<Agg<=0.8	0.21%	0.08%	0.06%	0.07%	0.05%	0.00%	252.42	58.44	259.47	242.90	214.29	0.00	43.75	14.29	38.68	37.52	35.43	0.00
0.8<Agg<=0.9	0.23%	0.07%	0.09%	0.07%	0.05%	0.00%	241.24	54.55	267.88	311.09	248.32	0.00	42.95	13.72	44.36	49.22	44.28	0.00
0.9<Agg<1	0.39%	0.09%	0.06%	0.06%	0.04%	0.00%	230.69	76.02	200.32	293.69	343.05	0.00	42.18	16.10	37.16	44.99	55.80	0.00
Agg=1	18.69%	20.43%	46.27%	47.23%	39.08%	56.89%	25.29	12.62	15.57	15.02	9.64	1.28	4.26	2.30	2.53	2.45	1.99	1.03

Panel B: May 6

	Order Aggressiveness Distribution						Avg Order Size						Avg # of Trades Per Order					
	HFT	INT	BUY	SELL	OPP	SMALL	HFT	INT	BUY	SELL	OPP	SMALL	HFT	INT	BUY	SELL	OPP	SMALL
Agg=0	78.92%	75.47%	44.67%	59.01%	53.95%	38.48%	5.49	5.07	7.72	12.74	7.35	1.17	1.35	1.34	1.38	1.58	1.44	1.01
0<Agg<=0.1	0.36%	0.12%	0.08%	0.12%	0.09%	0.00%	60.21	30.58	160.02	369.44	208.98	0.00	8.60	5.79	12.05	31.00	28.10	0.00
0.1<Agg<=0.2	0.33%	0.14%	0.12%	0.14%	0.11%	0.00%	49.13	18.71	238.71	290.55	191.90	0.00	8.37	4.61	32.89	31.08	35.67	0.00
0.2<Agg<=0.3	0.25%	0.13%	0.11%	0.16%	0.09%	0.00%	55.22	12.77	390.79	314.13	150.99	0.00	9.76	4.14	76.69	31.14	23.30	0.00
0.3<Agg<=0.4	0.30%	0.33%	0.15%	0.14%	0.10%	0.00%	56.70	12.19	110.54	218.36	105.87	0.00	10.20	3.93	14.36	32.52	16.09	0.00
0.4<Agg<=0.5	0.36%	0.63%	0.63%	0.25%	0.14%	0.00%	51.50	5.46	23.85	175.72	78.90	5.00	9.95	2.65	4.65	27.13	13.05	2.00
0.5<Agg<=0.6	0.20%	0.11%	0.13%	0.13%	0.09%	0.00%	106.72	18.19	115.93	224.16	155.77	0.00	19.15	5.75	19.48	31.79	25.37	0.00
0.6<Agg<=0.7	0.25%	0.17%	0.22%	0.11%	0.09%	0.00%	83.11	13.75	77.65	297.35	169.45	0.00	16.44	4.13	14.43	45.49	28.26	0.00
0.7<Agg<=0.8	0.22%	0.16%	0.23%	0.14%	0.10%	0.00%	101.36	16.57	94.00	252.70	139.10	0.00	19.01	5.48	16.28	43.67	27.56	0.00
0.8<Agg<=0.9	0.21%	0.09%	0.16%	0.14%	0.08%	0.00%	132.72	29.51	87.68	305.83	175.58	0.00	25.27	8.19	16.04	50.03	32.55	0.00
0.9<Agg<1	0.30%	0.05%	0.10%	0.15%	0.06%	0.00%	162.13	43.54	181.96	282.34	219.12	0.00	30.84	10.35	33.11	42.18	40.74	0.00
Agg=1	18.30%	22.60%	53.40%	39.51%	45.10%	61.52%	17.70	7.86	9.28	24.90	10.29	1.28	3.07	1.86	2.01	3.84	2.19	1.04

This table reports order statistics across trader categories. The order aggressiveness is defined as the ratio of size of the aggressive portion of the execution over order size. Order Aggressiveness, Average order size and average number of trades are distributed over various order aggressiveness ratios. Average order size is the average number of contracts traded per order. Average number of trades is the average number of trades an order generates.

Tab. 1.4: HFTs and Intermediaries: Net Holdings and Prices

	Panel A: May 3-5			Panel B: May 6	
	Δ NP HFT	**Δ NP INT**		**Δ NP HFT**	**Δ NP INT**
Intercept	-1.637	-0.529	**Intercept**	-3.222	0.038
	(-3.758)	**(-3.632)**		**(-3.429)**	(0.138)
$\Delta NPHFT_{t-1}$	-0.006		$\Delta NPHFT_{t-1}$	0.011	
	(-0.735)			(1.248)	
$NPHFT_{t-1}$	-0.005		$NPHFT_{t-1}$	-0.005	
	(-11.505)			**(-7.229)**	
$\Delta NPINT_{t-1}$		-0.006	$\Delta NPINT_{t-1}$		-0.035
		(-0.673)			**(-2.570)**
$NPINT_{t-1}$		-0.004	$NPINT_{t-1}$		-0.008
		(-10.043)			**(-8.426)**
ΔP_t	32.089	-13.540	ΔP_t	10.808	-8.164
	(18.380)	**(-21.992)**		**(5.142)**	**(-7.274)**
ΔP_{t-1}	17.178	-1.218	ΔP_{t-1}	4.625	6.635
	(12.983)	**(-2.708)**		**(3.639)**	**(9.784)**
ΔP_{t-2}	8.357	2.160	ΔP_{t-2}	-1.520	2.734
	(7.376)	**(5.107)**		(-1.384)	**(4.433)**
ΔP_{t-3}	5.086	2.525	ΔP_{t-3}	-1.360	1.138
	(4.998)	**(6.013)**		(-0.978)	**(3.031)**
ΔP_{t-4}	3.909	2.654	ΔP_{t-4}	-1.815	0.487
	(3.656)	**(6.583)**		(-1.680)	(1.270)
ΔP_{t-5}	1.807	2.499	ΔP_{t-5}	-0.228	-0.768
	(1.578)	**(5.898)**		(-1.680)	(-1.857)
ΔP_{t-6}	-0.078	2.163	ΔP_{t-6}	-0.312	-0.312
	(-0.072)	**(5.448)**		(-0.223)	(-0.826)
ΔP_{t-7}	-1.002	1.842	ΔP_{t-7}	-5.037	-0.617
	(-0.975)	**(4.969)**		**(-3.555)**	(-1.257)
ΔP_{t-8}	-1.756	1.466	ΔP_{t-8}	-1.775	-0.359
	(-1.535)	**(3.901)**		(-1.319)	(-1.044)
ΔP_{t-9}	-1.811	0.453	ΔP_{t-9}	-1.678	-1.105
	(-1.672)	(1.252)		(-1.432)	**(-2.736)**
ΔP_{t-10}	-3.899	0.525	ΔP_{t-10}	-1.654	-0.387
	(-3.795)	(1.366)		(-1.188)	(-0.936)
ΔP_{t-11}	-4.728	-0.026	ΔP_{t-11}	-1.076	-0.628
	(-4.752)	(-0.071)		(-0.903)	(-1.221)
ΔP_{t-12}	-3.456	0.152	ΔP_{t-12}	0.706	-1.171
	(-3.321)	(0.431)		(0.477)	**(-2.163)**
ΔP_{t-13}	-3.799	0.267	ΔP_{t-13}	2.261	-0.617
	(-3.772)	(0.738)		(1.354)	(-1.457)
ΔP_{t-14}	-4.769	0.317	ΔP_{t-14}	-2.664	-0.270
	(-4.708)	(0.822)		**(-2.346)**	(-0.735)
ΔP_{t-15}	-2.735	-0.195	ΔP_{t-15}	0.428	-0.833
	(-2.613)	(-0.544)		(0.330)	**(-2.442)**
ΔP_{t-16}	-2.208	-0.642	ΔP_{t-16}	-0.683	0.227
	(-2.123)	(-1.830)		(-0.385)	(0.638)
ΔP_{t-17}	-2.517	-0.100	ΔP_{t-17}	-0.657	0.293
	(-2.522)	(-0.261)		(-0.469)	(0.783)
ΔP_{t-18}	-4.358	0.044	ΔP_{t-18}	0.446	-0.769
	(-3.989)	(0.117)		(0.264)	**(-2.124)**
ΔP_{t-19}	-4.215	0.568	ΔP_{t-19}	-2.629	-0.296
	(-4.090)	(1.530)		**(-2.072)**	(-0.793)
ΔP_{t-20}	-5.860	-0.120	ΔP_{t-20}	-1.073	-0.706
	(-5.987)	(-0.343)		(-0.781)	(-1.576)
#obs	72837	72837	**#obs**	24275	24275
$Adj-R^2$	0.0194	0.0263	$Adj-R^2$	0.0101	0.0390

This table displays estimated coefficients of the following regression: $\Delta y_t = \alpha + \phi\Delta y_{t-1} + \delta y_{t-1} + \sum_{i=0}^{20}[\beta_{t-i} \times \Delta p_{t-i}/0.25] + \epsilon_t$. The dependent variable is changes in holdings of High Frequency Traders and Intermediaries, respectively. Both changes in holdings, Δy_t, and lagged holdings, $y_t - 1$, are in the number of contracts. Price changes, $\Delta p_t - i$, are in ticks. Estimates are computed for second-by-second observations. The t statistics are calculated using the Newey-West (1987) estimator. t values reported in parentheses are in bold if the coefficients are statistically significant at the 5% level.

Tab. 1.5: HFTs and Intermediaries: Liquidity Provision/Removal

	Panel A: May 3-5				Panel B: May 6			
	Δ A HFT	**Δ P HFT**	**Δ A INT**	**Δ P INT**	**Δ A HFT**	**Δ P HFT**	**Δ A INT**	**Δ P INT**
Intercept	-1.285	-0.352	-0.344	-0.185	-2.863	-0.359	-0.246	0.284
	(-2.855)	(-1.291)	**(-3.040)**	(-1.515)	**(-3.242)**	(-0.670)	(-1.277)	(1.212)
$\Delta NPHFT_{t-1}$	-0.042	0.036			-0.003	0.014		
	(-4.931)	**(6.805)**			(-0.286)	(1.770)		
$NPHFT_{t-1}$	-0.005	-0.001			-0.004	-0.001		
	(-9.619)	**(-3.204)**			**(-5.701)**	**(-2.924)**		
$\Delta NPINT_{t-1}$			0.007	-0.013			-0.003	-0.032
			(1.623)	(-1.683)			(-0.531)	**(-2.557)**
$NPINT_{t-1}$			-0.002	-0.002			-0.003	-0.004
			(-6.150)	**(-6.182)**			**(-4.540)**	**(-4.824)**
ΔP_t	57.778	-25.689	6.377	-19.917	23.703	-12.895	4.939	-13.103
	(29.925)	**(-28.850)**	**(17.751)**	**(-32.937)**	**(7.411)**	**(-5.281)**	**(7.807)**	**(-8.502)**
ΔP_{t-1}	22.549	-5.371	5.791	-7.009	-1.118	5.744	3.909	2.726
	(16.181)	**(-7.829)**	**(17.521)**	**(-18.574)**	(-0.946)	**(4.171)**	**(9.102)**	**(5.343)**
ΔP_{t-2}	9.614	-1.258	4.752	-2.592	-2.661	1.141	1.659	1.075
	(8.089)	(-1.826)	**(15.125)**	**(-7.739)**	**(-2.613)**	(1.101)	**(5.187)**	**(2.279)**
ΔP_{t-3}	5.442	-0.356	3.642	-1.117	-1.151	-0.209	0.536	0.602
	(5.142)	(-0.586)	**(12.586)**	**(-3.383)**	(-0.890)	(-0.175)	**(2.288)**	(1.675)
ΔP_{t-4}	3.290	0.619	3.114	-0.460	-2.814	0.999	0.229	0.258
	(2.937)	(0.949)	**(10.888)**	(-1.366)	**(-2.739)**	(0.994)	(1.004)	(0.690)
ΔP_{t-5}	1.926	-0.119	2.591	-0.092	-0.690	0.461	0.161	-0.929
	(1.664)	(-0.170)	**(8.656)**	(-0.266)	(-0.556)	(0.489)	(0.546)	(-1.822)
ΔP_{t-6}	-0.987	0.909	2.038	0.125	-1.824	1.512	0.053	-0.365
	(-0.872)	(1.374)	**(7.017)**	(0.373)	(-1.475)	(1.344)	(0.210)	(-1.058)
ΔP_{t-7}	-0.291	-0.711	2.101	-0.258	-2.688	-2.350	-0.516	-0.102
	(-0.257)	(-1.065)	**(8.333)**	(-0.812)	**(-2.295)**	(-1.754)	**(-2.345)**	(-0.244)
ΔP_{t-8}	-0.977	-0.779	1.740	-0.274	-2.216	0.441	-0.625	0.267
	(-0.797)	(-1.159)	**(6.540)**	(-0.850)	(-1.910)	(0.394)	**(-2.668)**	(0.815)
ΔP_{t-9}	-0.732	-1.078	1.158	-0.705	-0.801	-0.877	-0.099	-1.007
	(-0.643)	(-1.697)	**(4.541)**	**(-2.259)**	(-0.732)	(-0.896)	(-0.364)	**(-2.525)**
ΔP_{t-10}	-2.543	-1.356	1.007	-0.483	-2.958	1.304	-0.513	0.125
	(-2.370)	**(-2.246)**	**(3.858)**	(-1.538)	**(-2.519)**	(1.253)	(-1.949)	(0.291)
ΔP_{t-11}	-3.536	-1.193	0.425	-0.451	-1.099	0.023	-0.867	0.239
	(-3.356)	**(-1.963)**	(1.612)	(-1.463)	(-1.090)	(0.024)	**(-3.152)**	(0.509)
ΔP_{t-12}	-2.523	-0.934	0.207	-0.054	0.974	-0.268	-0.396	-0.775
	(-2.328)	(-1.436)	(0.781)	(-0.178)	(0.878)	(-0.203)	(-1.514)	(-1.532)
ΔP_{t-13}	-2.130	-1.669	0.502	-0.235	1.169	1.093	-0.293	-0.324
	(-2.040)	**(-2.712)**	(1.868)	(-0.786)	(0.904)	(0.716)	(-1.181)	(-0.838)
ΔP_{t-14}	-4.387	-0.382	0.107	0.210	-1.249	-1.415	-0.450	0.180
	(-4.154)	(-0.631)	(0.396)	(0.630)	(-1.223)	(-1.253)	(-1.892)	(0.522)
ΔP_{t-15}	-1.965	-0.770	0.099	-0.294	1.006	-0.579	-0.535	-0.298
	(-1.834)	(-1.231)	(0.368)	(-0.934)	(0.922)	(-0.638)	**(-2.153)**	(-0.857)
ΔP_{t-16}	-2.434	0.226	-0.182	-0.460	-1.300	0.617	0.215	0.012
	(-2.190)	(0.391)	(-0.673)	(-1.528)	(-1.028)	(0.560)	(0.859)	(0.037)
ΔP_{t-17}	-2.185	-0.332	0.238	-0.338	-1.707	1.051	-0.239	0.532
	(-2.019)	(-0.545)	(0.884)	(-1.066)	(-1.521)	(0.948)	(-0.957)	(1.595)
ΔP_{t-18}	-3.259	-1.099	0.311	-0.267	0.482	-0.036	0.051	-0.820
	(-2.862)	(-1.739)	(1.255)	(-0.824)	(0.440)	(-0.035)	(0.229)	**(-2.537)**
ΔP_{t-19}	-3.585	-0.631	0.544	0.024	-0.746	-1.883	-0.265	-0.0311
	(-3.297)	(-1.014)	**(2.085)**	(0.077)	(-0.761)	(-1.542)	(-1.070)	(-0.0782)
ΔP_{t-20}	-4.621	-1.240	0.211	-0.331	-0.535	-0.538	-0.501	-0.205
	(-4.493)	**(-2.144)**	(0.863)	(-1.114)	(-0.521)	(-0.570)	**(-2.276)**	(-0.484)
#obs	72837	72837	72837	72837	24275	24275	24275	24275
$Adj-R^2$	0.0427	0.0260	0.0202	0.0631	0.0252	0.0270	0.0457	0.0698

This table presents estimated coefficients of the following regression: $\Delta y_t = \alpha + \phi \Delta y_{t-1} + \delta y_{t-1} + \sum_{i=0}^{20} [\beta_{t-i} \times \Delta p_{t-i}/0.25] + \epsilon_t$. Dependent variables are changes in Aggressive and Passive holdings of High Frequency Traders and Intermediaries. Changes in holdings, Δy_t, and lagged holdings, $y_t - 1$, are in the number of contracts. Price changes, $\Delta p_t - i$, are in ticks. Estimates are computed for second-by-second observations. The t statistics are calculated using the Newey-West (1987) estimator. t values reported in parentheses are in bold if the coefficients are statistically significant at the 5% level.

Tab. 1.6: Aggressive and Passive Holdings: Flash Crash

	Panel A: Down				**Panel B: Up**			
	Δ A HFT	**Δ P HFT**	**Δ A INT**	**Δ P INT**	**Δ A HFT**	**Δ P HFT**	**Δ A INT**	**Δ P INT**
Intercept	-0.614	7.792	-1.320	9.992	2.111	-1.880	1.484	-1.477
	(-0.080)	**(2.306)**	(-0.440)	**(3.291)**	(0.676)	(-0.647)	(1.319)	(-1.837)
$\Delta NPHFT_{t-1}$	-0.023	-0.014			0.025	-0.026		
	(-0.748)	(-0.744)			(0.996)	(-1.130)		
$NPHFT_{t-1}$	-0.008	0.0010			-0.005	-0.001		
	(-1.947)	(0.370)			**(-2.258)**	(-0.336)		
$\Delta NPINT_{t-1}$			-0.043	-0.005			0.053	0.008
			(-1.585)	(-0.133)			**(2.563)**	(0.426)
$NPINT_{t-1}$			-0.0003	-0.012			-0.004	-0.0009
			(-0.079)	**(-2.812)**			**(-2.366)**	(-0.654)
ΔP_t	24.226	8.533	8.251	-9.603	-0.251	-9.107	2.912	-4.105
	(2.833)	(1.275)	**(3.864)**	**(-2.618)**	(-0.142)	**(-4.378)**	**(4.257)**	**(-6.296)**
ΔP_{t-1}	2.397	9.540	8.821	2.075	-0.993	6.350	2.150	2.934
	(0.557)	(1.710)	**(6.132)**	(0.977)	(-0.621)	**(2.773)**	**(4.446)**	**(5.790)**
ΔP_{t-2}	-4.273	3.669	4.257	0.298	-3.043	-0.445	0.402	0.457
	(-0.915)	(0.839)	**(2.307)**	(0.214)	**(-1.937)**	(-0.222)	(1.039)	(0.893)
ΔP_{t-3}	-2.891	1.747	0.759	-0.138	0.814	-1.763	-0.099	0.283
	(-0.681)	(0.569)	(0.865)	(-0.130)	(0.392)	(-0.686)	(-0.330)	(0.610)
ΔP_{t-4}	-2.040	-5.780	-2.175	0.009	-2.391	3.192	0.109	0.128
	(-0.510)	**(-2.053)**	**(-2.012)**	(0.007)	(-1.769)	**(2.022)**	(0.386)	(0.316)
ΔP_{t-5}	-4.990	-5.326	0.070	-1.314	0.586	1.898	0.007	-0.657
	(-1.046)	(-0.911)	(0.060)	(-1.302)	(0.403)	(1.088)	(0.019)	(-1.350)
ΔP_{t-6}	-7.924	6.621	-1.187	0.266	-0.426	2.800	0.282	-0.749
	(-1.847)	**(1.994)**	(-1.206)	(0.228)	(-0.345)	(1.515)	(0.873)	(-1.676)
ΔP_{t-7}	6.843	-11.357	0.597	-1.384	-4.091	-3.299	-0.708	-0.753
	(1.651)	**(-2.454)**	(0.640)	(-1.266)	**(-2.690)**	(-1.401)	**(-2.157)**	(-1.605)
ΔP_{t-8}	-6.903	6.837	-2.720	1.184	-0.049	-0.676	-0.401	0.183
	(-1.542)	(1.562)	**(-2.498)**	(0.892)	(-0.032)	(-0.365)	(-1.205)	(0.529)
ΔP_{t-9}	0.624	-7.531	-1.732	-0.761	0.219	-0.115	-0.444	-0.709
	(0.128)	(-1.623)	(-1.385)	(-0.646)	(0.189)	(-0.082)	(-1.244)	(-1.899)
ΔP_{t-10}	2.024	-3.278	-2.189	-0.300	-1.380	0.609	-0.299	-0.302
	(0.324)	(-0.583)	(-1.611)	(-0.194)	(-0.920)	(0.291)	(-0.962)	(-0.778)
ΔP_{t-11}	0.412	4.367	-5.216	-1.190	-0.157	1.102	-0.607	0.200
	(0.068)	(1.076)	**(-4.948)**	(-0.739)	(-0.135)	(0.607)	(-1.593)	(0.449)
ΔP_{t-12}	1.442	2.883	-2.684	1.850	0.700	-0.379	0.092	-0.986
	(0.220)	(0.577)	**(-1.984)**	(1.479)	(0.527)	(-0.163)	(0.288)	**(-2.480)**
ΔP_{t-13}	17.340	-9.284	-0.385	-4.370	2.551	3.614	-0.212	0.429
	(3.049)	(-1.613)	(-0.221)	**(-2.344)**	(1.351)	(1.418)	(-0.643)	(1.027)
ΔP_{t-14}	-11.389	-1.530	-1.904	2.974	0.378	-3.094	0.036	-0.349
	(-2.531)	(-0.226)	(-1.627)	(1.775)	(0.304)	(-1.571)	(0.108)	(-1.080)
ΔP_{t-15}	8.706	-2.304	-4.375	-1.206	1.317	-1.904	-0.297	0.043
	(1.281)	(-0.332)	**(-4.377)**	(-0.783)	(0.862)	(-1.287)	(-0.791)	(0.100)
ΔP_{t-16}	-3.908	-1.352	2.906	0.625	-1.480	0.541	0.372	0.234
	(-0.642)	(-0.229)	(2.064)	(0.369)	(-0.903)	(0.261)	(1.036)	(0.682)
ΔP_{t-17}	6.351	-2.788	-0.147	-1.420	0.765	1.750	-0.241	0.725
	(1.055)	(-0.652)	(-0.096)	(-0.915)	(0.505)	(0.921)	(-0.589)	(1.792)
ΔP_{t-18}	-8.521	-3.988	0.475	0.578	0.675	2.813	0.084	-0.584
	(-1.642)	(-0.647)	(0.375)	(0.356)	(0.452)	(1.533)	(0.252)	(-1.695)
ΔP_{t-19}	6.899	-11.448	1.279	-3.649	-1.076	-3.171	-0.098	-0.086
	(0.990)	**(-2.068)**	(0.936)	(-1.830)	(-0.835)	(-1.773)	(-0.300)	(-0.195)
ΔP_{t-20}	-14.611	6.997	-1.574	4.375	0.945	-1.366	-0.488	0.102
	(-3.011)	(1.226)	(-1.404)	**(2.650)**	(0.678)	(-0.922)	(-1.486)	(0.194)
#obs	808	808	808	808	1347	1347	1347	1347
$Adj-R^2$	0.0423	0.0593	0.1779	0.0739	0.0084	0.0583	0.0655	0.0816

This table displays the results of the regression of $\Delta y_t = \alpha + \phi\Delta y_{t-1} + \delta y_{t-1} + \sum_{i=0}^{20}[\beta_{t-i} \times \Delta p_{t-i}/0.25] + \epsilon_t$ over one second intervals. The dependent variables are aggressive and passive holdings changes of High Frequency Traders and Intermediaries. Changes in holdings (Δy_t) and lagged holdings($y_t - 1$) are defined in contracts. The price changes ($\Delta p_t - i$) are defined in ticks. DOWN period is defined as the interval between 13:32:00 (CT) and 13:45:28 (CT). UP period is defined as the interval between 13:45:33 (CT) and 14:08:00 (CT). The t statistics are calculated using the Newey-West (1987) estimator. t values reported in parentheses are in bold if the coefficients are statistically significant at 5% level.

Tab. 1.7: Trading Volume During the Flash Crash

	Panel A: May 3-5			
	DOWN		UP	
	Sell	Buy	Sell	Buy
High Frequency Traders	23,746	23,791	40,524	40,021
Intermediaries	6,484	6,328	11,469	11,468
Fundamental Buyers	3,064	7,958	6,127	14,910
Fundamental Sellers	8,428	3,118	15,855	5,282
Opportunistic Traders	20,049	20,552	37,317	39,535
Small Traders	232	256	428	504
	Panel B: May 6th			
	DOWN		UP	
	Sell	Buy	Sell	Buy
High Frequency Traders	152,436	153,804	191,490	189,013
Intermediaries	32,489	33,694	47,348	45,782
Fundamental Buyers	28,694	78,359	55,243	165,612
Fundamental Sellers	94,101	10,502	145,396	35,219
Opportunistic Traders	189,790	221,236	302,417	306,326
Small Traders	1,032	947	1,531	1,473

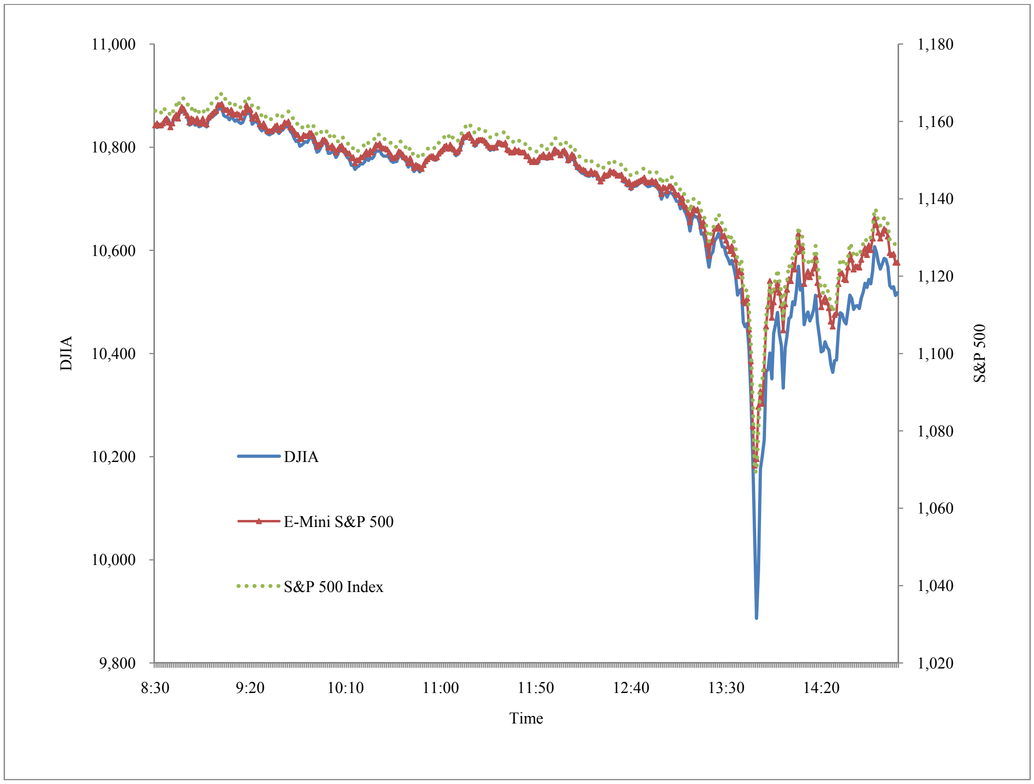
This table presents the number of contracts sold and bought by trader categories during DOWN and UP periods. DOWN period is defined as the interval between 13:32:00 and 13:45:28 CT. UP period is defined as the interval between 13:45:33 and 14:08:00 CT. Panel A reports the average number of contracts bought and sold between May 3 and May 5, 2010 during the DOWN and UP periods in the day. Panel B reports the number of contracts bought and sold on May 6, 2010 during the DOWN and UP periods.

Tab. 1.8: Aggressiveness Imbalance and Prices

	May 3-5	May 6
Intercept	-0.01	0.01
	(-0.19)	(0.31)
High Frequency Traders	5.37	3.23
	(6.43)	**(3.37)**
Intermediaries	0.83	5.99
	(1.08)	**(5.08)**
Fundamental Buyers	1.31	0.53
	(4.32)	**(2.20)**
Fundamental Sellers	1.36	0.92
	(5.81)	**(6.40)**
Opportunistic Traders	7.60	7.49
	(9.74)	**(10.61)**
# of Obs	1210	404
Adj-R2	0.36	0.59

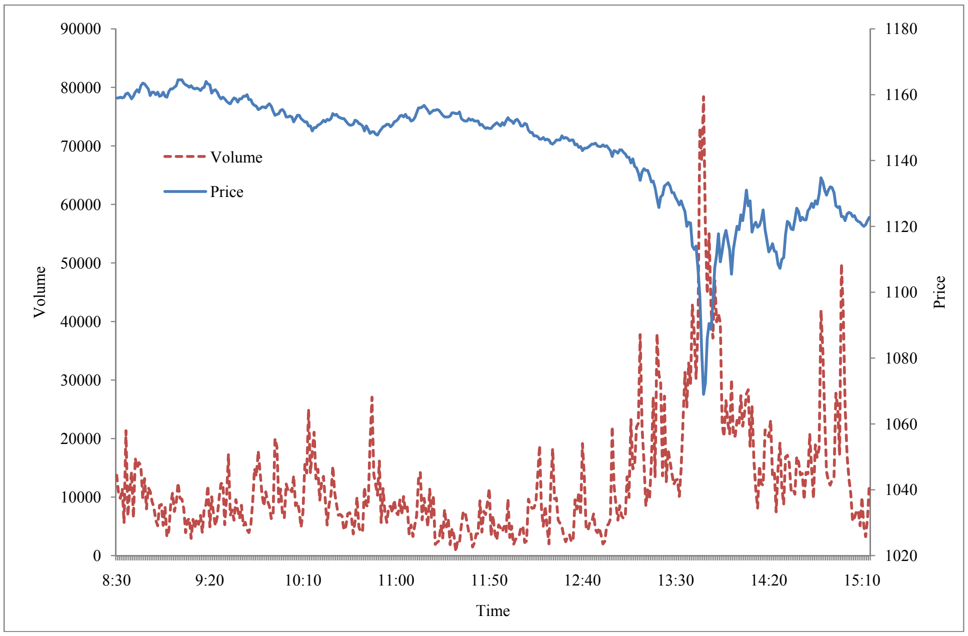
This table presents estimated coefficients of the following regression: $\frac{\Delta P_t}{P_{t-1} \times \sigma_{t-1}} = \alpha + \sum_{i=1}^{5} [\lambda_i \times \frac{AGG_{i,t}}{Shr_{i,t-1} \times 100{,}000}] + \epsilon_t$. The dependent variable is the return scaled by volatility over one minute interval. Independent variables are the aggressiveness imbalances of trader categories scaled by their market share times 100,000. t-values are corrected for serial correlation, up to three lags, using the Newey-West (1987) estimator. t-values, reported in parentheses, are in bold if the coefficients are statistically significant at the 5% level.

Fig. 1.1: U. S. Equity Indices on May 6, 2010



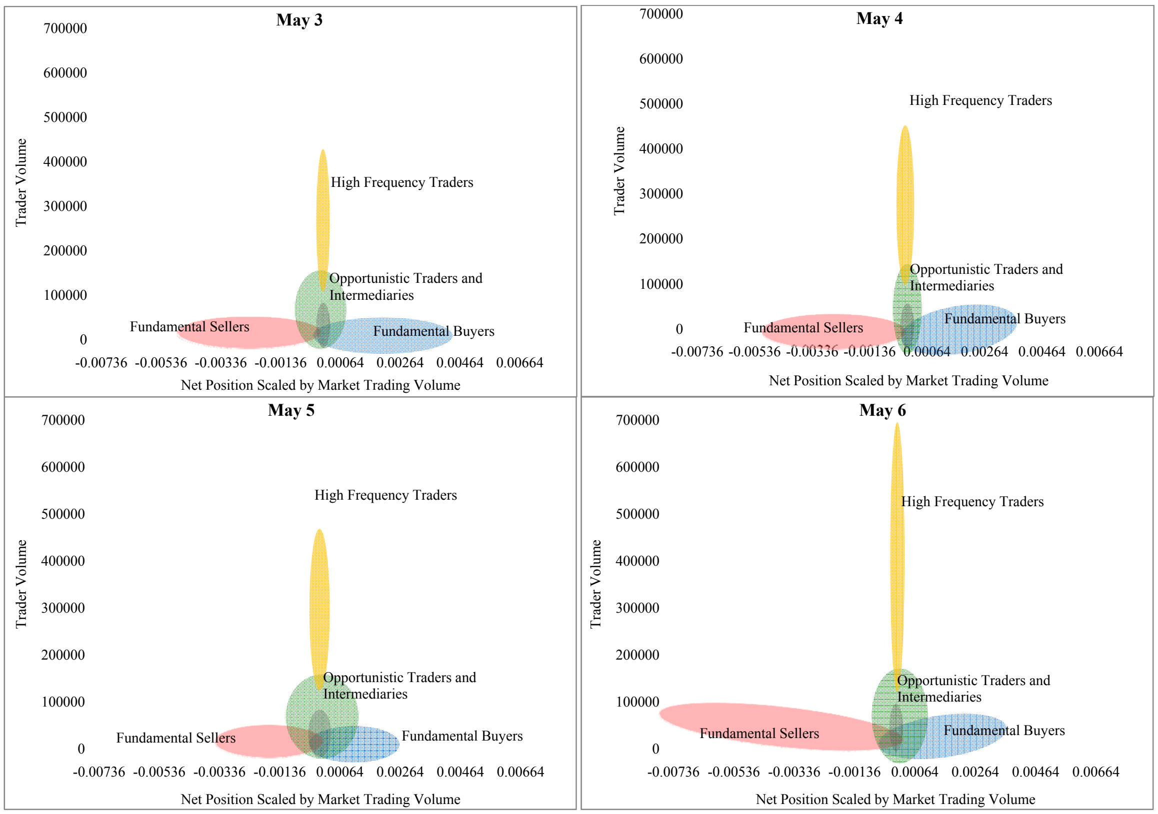
This figure presents end-of-minute transaction prices of the Dow Jones Industrial Average (DJIA), S&P 500 Index, and the June 2010 E-Mini S&P 500 futures contract on May 6, 2010 between 8:30 and 15:15 CT.

Fig. 1.2: Prices and Trading Volume of the E-Mini S&P 500 Stock Index Futures Contract



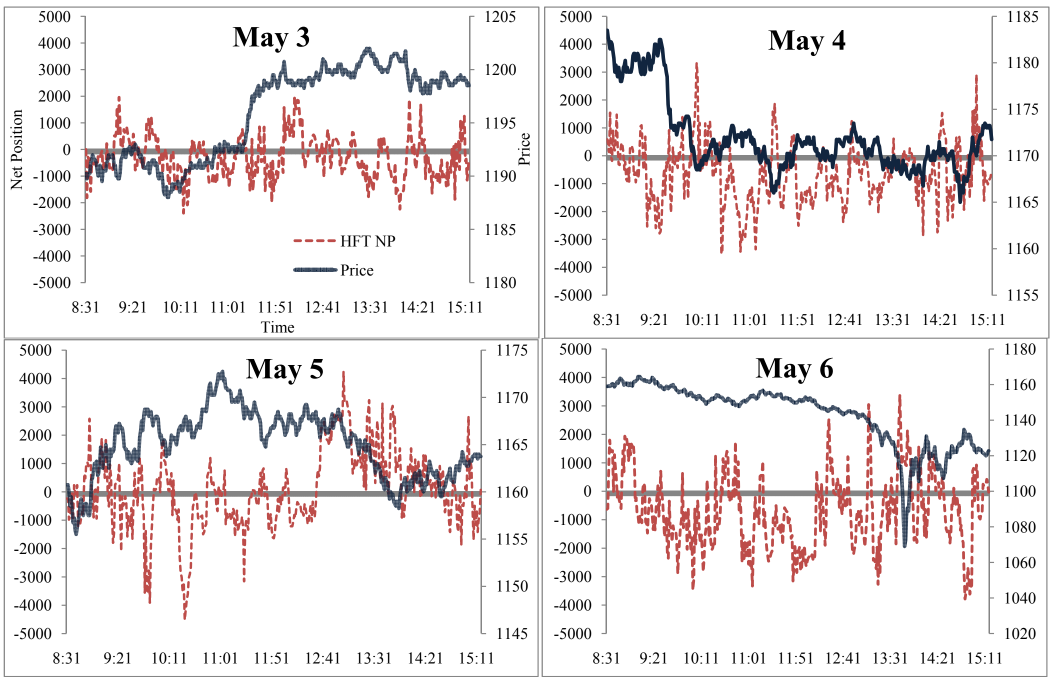
This figure presents minute-by-minute transaction prices and trading volume of the June 2010 E-Mini S&P futures contract on May 6, 2010 between 8:30 and 15:15 CT. Trading volume is calculated as the number of contracts traded during each minute. Transaction price is the last transaction price of each minute.

Fig. 1.3: Trading Accounts Trading Volume and Net Position Scaled by Market Trading Volume



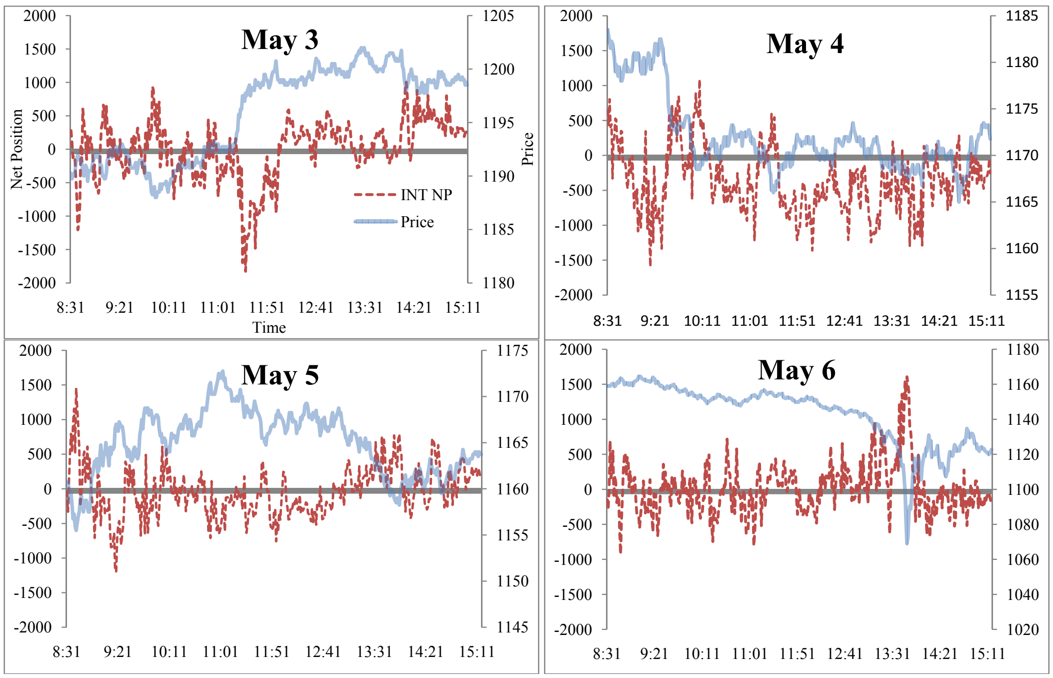
This figure presents trader categories superimposed (as shaded areas) over all individual trading accounts ranked by their trading volume and net position scaled by market trading volume. The figures reflect trading activity in the June 2010 E-Mini S&P 500 futures contract for May 3-6, 2010.

Fig. 1.4: Net Position of High Frequency Traders



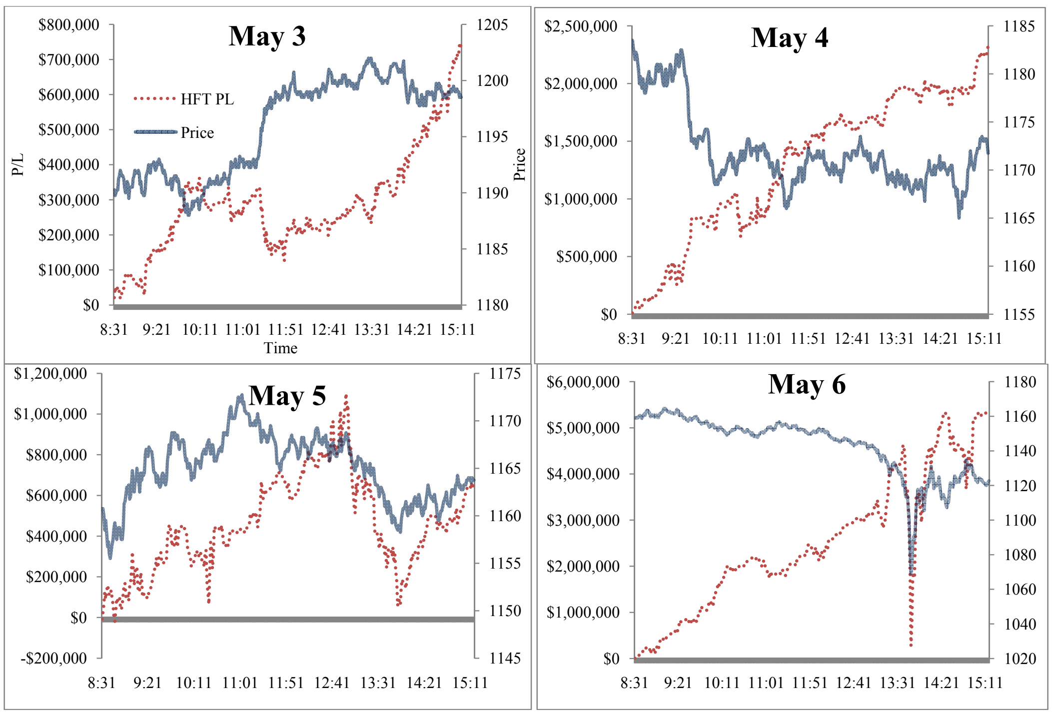
This figure presents the net position of High Frequency Traders (left vertical axis) and transaction prices (right vertical axis) in the June 2010 E-Mini S&P 500 futures contract over one minute intervals during May 3, 4 , 5, and 6 between 8:30 to 15:15 CT. Net position is calculated as the difference between total open long and total open short positions of High Frequency Traders at the end of each minute. Transaction price is the last transaction price of each minute.

Fig. 1.5: Net Position of Intermediaries



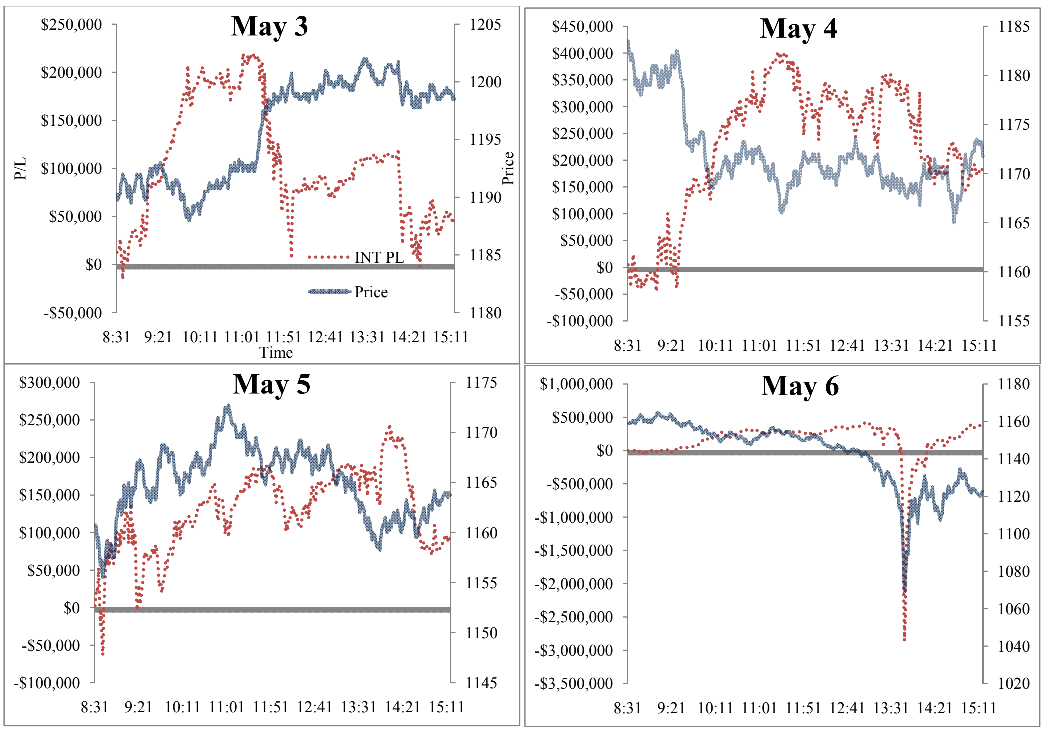
This figure presents the net position of Intermediaries (left vertical axis) and transaction prices (right vertical axis) in the June 2010 E-Mini S&P 500 futures contract over one minute intervals during May 3, 4, 5, and 6 between 8:30 to 15:15 CT. Net position is calculated as the difference between total open long and total open short positions of Intermediaries at the end of each minute. Transaction price is the last transaction price of each minute.

Fig. 1.6: Profits and Losses of High Frequency Traders



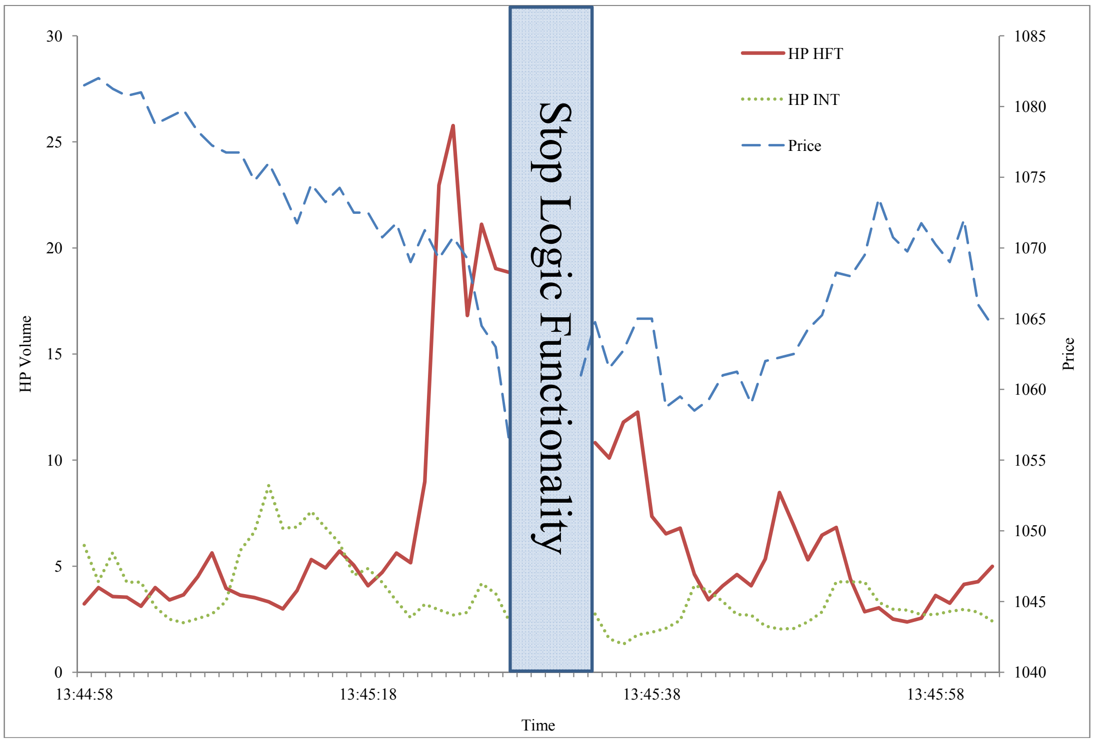
This figure presents the profits and losses of High Frequency Traders (left vertical axis)in the June 2010 E-Mini S&P 500 futures contract reported over one minute intervals during May 3, 4, 5, and 6 between 8:30 to 15:15 CT. Profits and losses are calculated by multiplying lagged net position by the change in price.

Fig. 1.7: Profits and Losses of Intermediaries



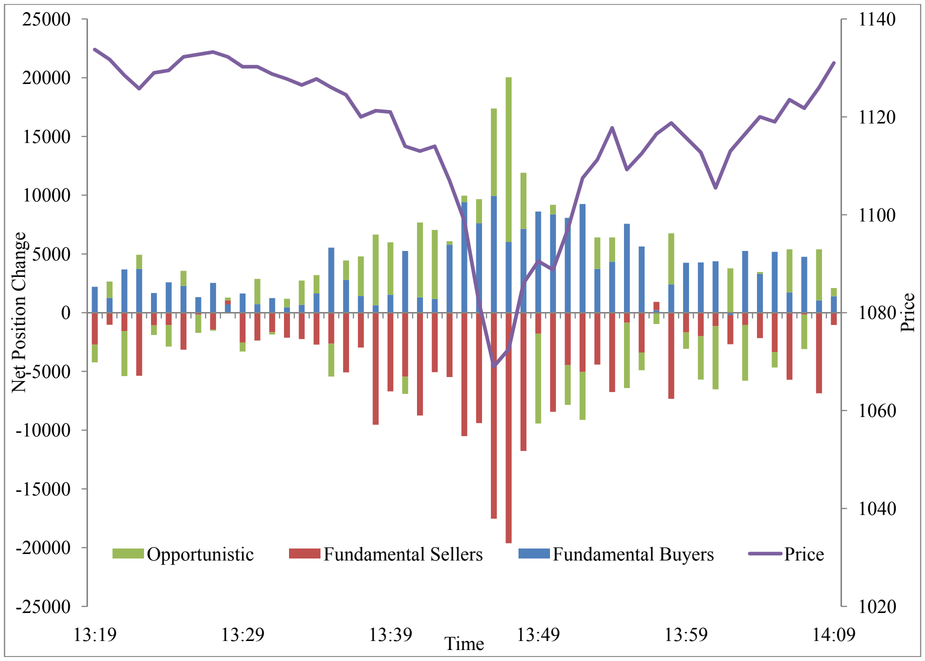
This figure presents the profits and losses of Intermediaries (left vertical axis)in the June 2010 E-Mini S&P 500 futures contract reported over one minute intervals during May 3, 4, 5, and 6 between 8:30 to 15:15 CT. Profits and losses are calculated by multiplying lagged net position by the change in price.

Fig. 1.8: Hot Potato Volume



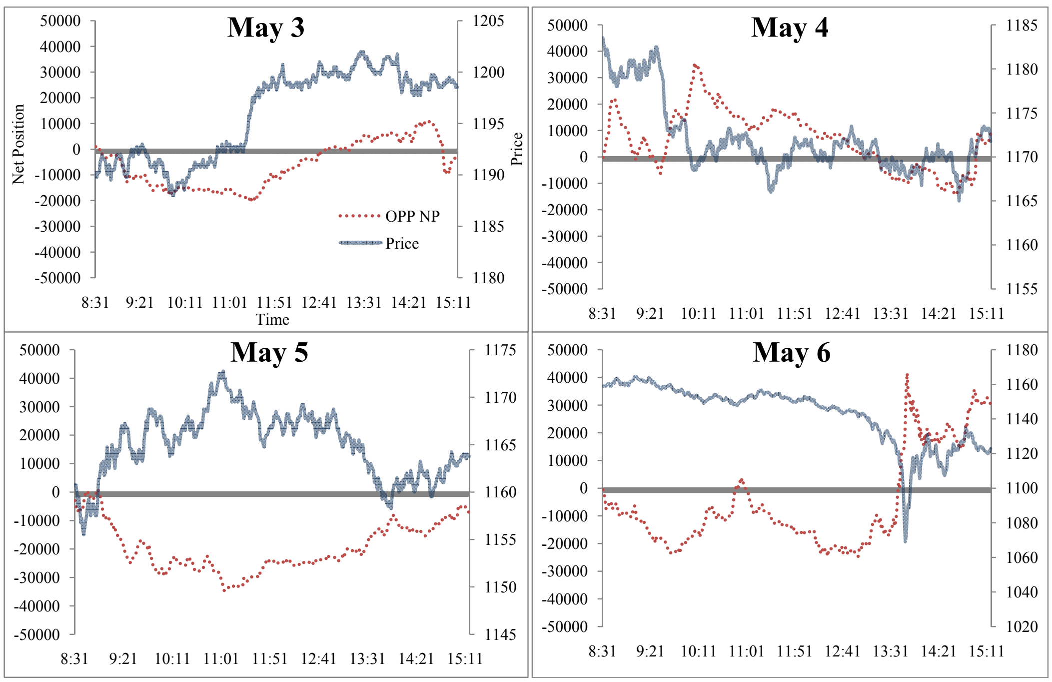
This figure shows the price and the scaled trading volume by HFTs and Intermediaries over one second intervals. Scaled trading volume is calculated as the 5 second moving average of contracts traded over absolute value net holdings. Price reflects the last transaction price during an interval. Prices and scaled trading volumes are reported from 13:44 to 13:46 CT.

Fig. 1.9: Change in Net Position of Fundamental and Opportunistic Traders



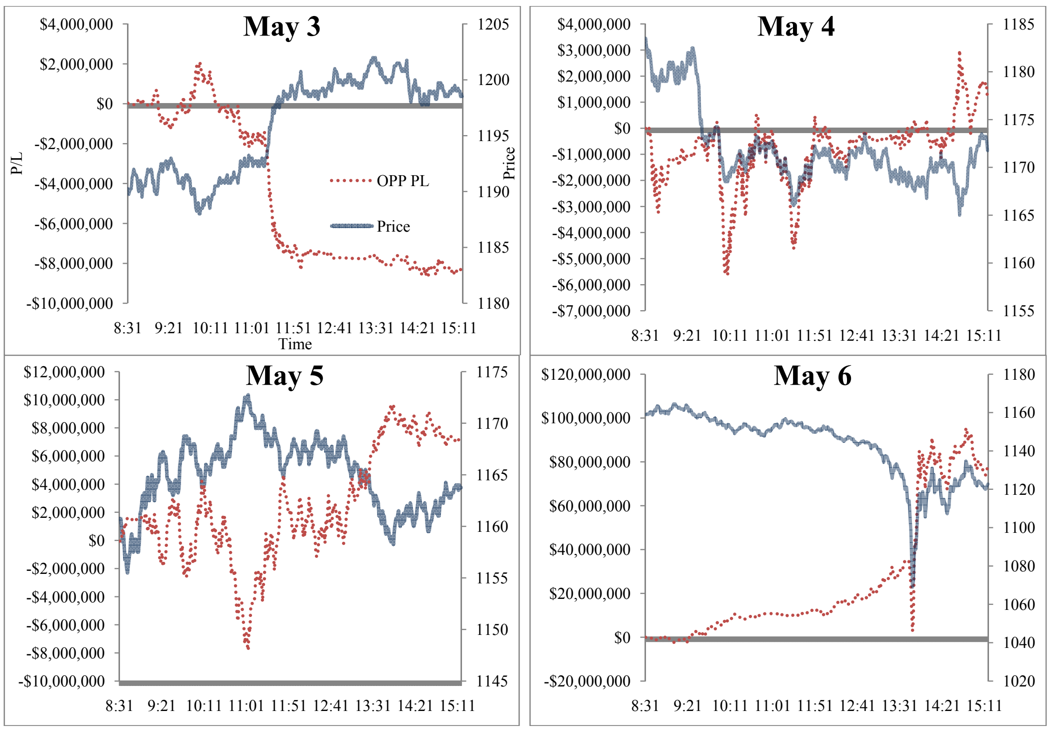
This figure presents the change in net position of Fundamental and Opportunistic Traders (left vertical axis) and transaction prices (right vertical axis) in the June 2010 E-Mini S&P 500 futures contract over one minute intervals on 6 between 13:19 to 14:09 CT. Net position is calculated as the difference between total open long and total open short positions of Opportunistic Traders at the end of each minute. Transaction price is the last transaction price of each minute.

Fig. 1.10: Net Position of Opportunistic Traders



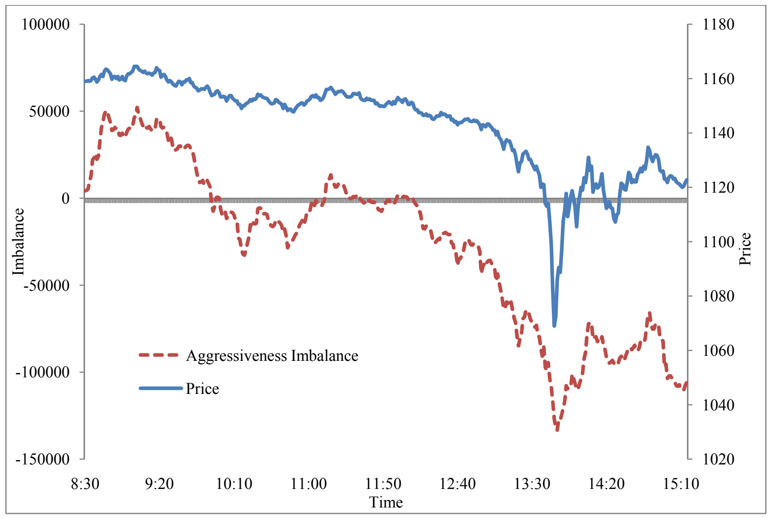
This figure presents the net position of Opportunistic Traders (left vertical axis) and transaction prices (right vertical axis) in the June 2010 E-Mini S&P 500 futures contract over one minute intervals during May 3, 4, 5, and 6 between 8:30 to 15:15 CT. Net position is calculated as the difference between total open long and total open short positions of Opportunistic Traders at the end of each minute. Transaction price is the last transaction price of each minute.

Fig. 1.11: Profits and Losses of Opportunistic Traders



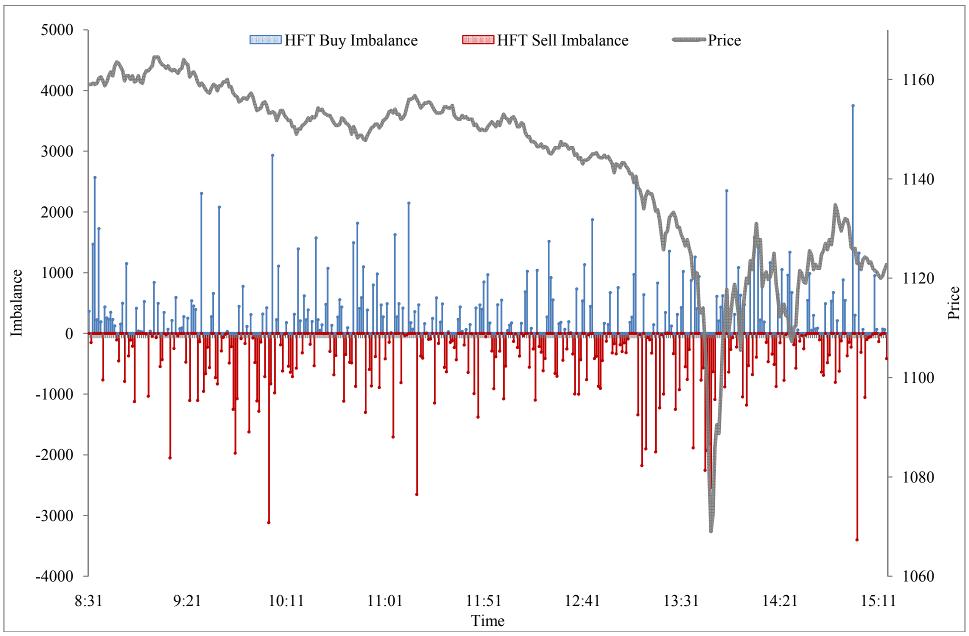
This figure presents the profits and losses of Opportunistic Traders (left vertical axis)in the June 2010 E-Mini S&P 500 futures contract reported over one minute intervals during May 3, 4, 5, and 6 between 8:30 to 15:15 CT. Profits and Losses are calculated by multiplying the lagged net position by the change in price.

Fig. 1.12: Total Aggressiveness Imbalance



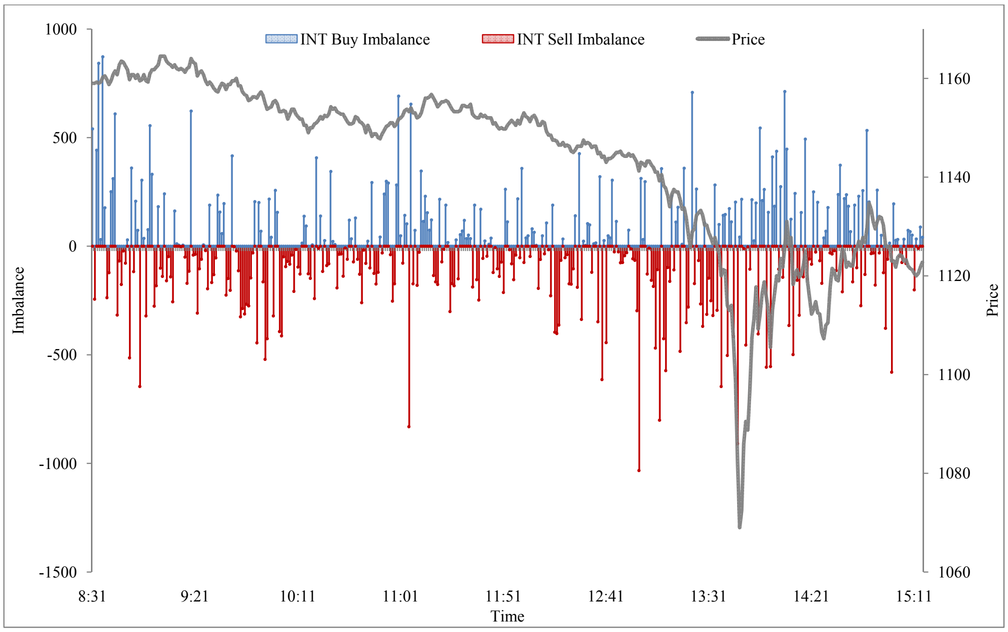
This figure presents the total aggressiveness imbalance and prices in the June 2010 E-Mini S&P 500 futures contract over one minute intervals between 8:30 to 15:15 CT on May 6, 2010. Aggressiveness Imbalance is calculated as cumulative total aggressive Buy transactions minus cumulative total aggressive Sell transactions at the end of each minute. Price is the last transaction price for each minute.

Fig. 1.13: Aggressiveness Imbalance of High Frequency Traders



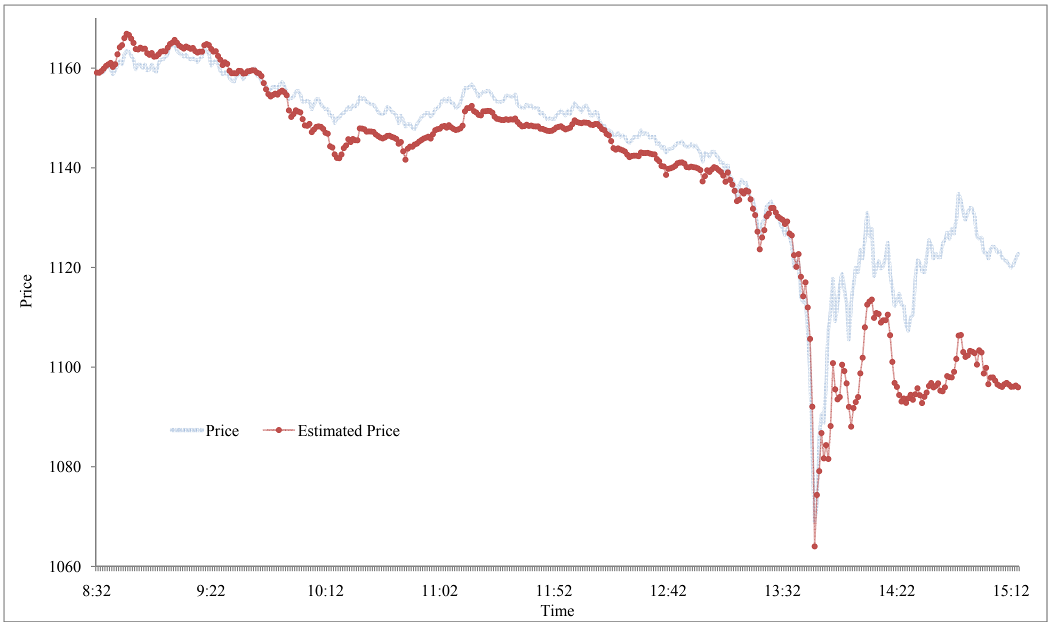
This figure presents the Aggressiveness Imbalance of High Frequency Traders (HFTs) and prices in the June 2010 E-Mini S&P 500 futures contract over one minute intervals between 8:30 to 15:15 CT on May 6, 2010. Aggressiveness Imbalance of HFTs is calculated as cumulative HFT aggressive Buy transactions minus cumulative HFT aggressive Sell transactions at the end of each minute. Price is the last transaction price for each minute.

Fig. 1.14: Aggressiveness Imbalance of Intermediaries



This figure presents the Aggressiveness Imbalance of Intermediaries and prices in the June 2010 E-Mini S&P 500 futures contract over one minute intervals between 8:30 to 15:15 CT on May 6, 2010. Aggressiveness Imbalance of Intermediaries is calculated as cumulative aggressive Buy transactions of Intermediaries minus cumulative aggressive Sell transactions of Intermediaries at the end of each minute. Price is the last transaction price for each minute.

Fig. 1.15: Fitted Price Based on Aggressiveness Imbalance



This figure presents actual and fitted prices in the June 2010 E-Mini S&P 500 futures contract over one minute intervals between 8:30 to 15:15 CT on May 6, 2010. The Solid line is the last actual transaction price for each minute. The Marked line is the fitted price calculated by applying estimated coefficients from the aggressiveness imbalance regressions (Equation (1.3)) using data for May 3-5, 2010 to realized Aggressive Imbalances of different trader categories on May 6, 2010.

2. MUTUAL FUNDS AND LIQUIDITY CONSUMPTION

2.1 Introduction

The way institutions trade in the financial markets has always been intriguing for researchers. Institutions are regarded as sophisticated and well-informed participants who significantly affect the prices and volatility. According to the Investment Company Institute, as of February 2011 Mutual Fund Industry size is $12 trillion of which equity funds hold $ 5.6 trillion. Institutions of this size have significant effect on market liquidity. This paper investigates the relationship between mutual fund trading and liquidity consumption in financial markets.

Price fluctuations are affected by not only the direction of trading but also the availability and the consumption of liquidity in financial markets. Liquidity consumption by institutions in an illiquid market is likely to cause higher volatility than liquidity provision in a liquid market.

Characterizing the liquidity sensitivity of institutions can help us understand the interactions of market participants and their effect on price volatility. In this study, we characterize the mutual fund sensitivity to liquidity consumption in the US Equity Markets. Mutual Funds are required to report their quarterly stock holdings with Securities Exchange Commission (SEC) through 13-F filings. However, the changes in stock holdings show the direction of mutual fund net trading, not their liquidity sensitivity. To calculate the mutual fund sensitivity to liquidity consumption, we use the information in the NYSE Trades and Quotes (TAQ) Dataset which reports the tick by tick data for all exchange listed stocks.

There are three main findings. First, mutual fund trades are generally correlated with order imbalance, suggesting that they trade in the same direction with liquidity consumption. Second, mutual fund liquidity sensitivity differs based on their investment styles. The liquidity sensitivity of growth-oriented funds is higher than the value oriented and passively managed funds. This is consistent with a hypothesis that information motivated trader consumes more liquidity. Third, trades of actively managed funds can be explained by larger trades consuming liquidity whereas small trades reveal the trades of index funds. Although explanations such as following other traders in the market can not be ruled out, this evidence supports a plausible explanation that information-motivated orders submitted by actively managed funds tend to be large and index funds need only small adjustments to track the index.

We use the method developed by Campbell, Ramadorai and Schwartz (2010), henceforth CRS, to estimate the liquidity sensitivity of mutual funds over different trade sizes. CRS develop a statistical method to explore the relationship between institutional trading and net order imbalance across different trade sizes. Their results show that trades under \$2,000 or over \$30,000 indicate institutional trading. However, we document that actively managed fund trades are correlated with larger trade sizes whereas small trades indicate index fund trades, suggesting that institutions with different investment styles can be accounted for the relationship between institutional trading and different trade sizes.

In August 2007, hedge funds received margin calls due to reduced stocks prices. Their liquidity needs required correlated sell pressure which further reduced the prices and led to a “liquidity spiral” (Brunnermeier 2009). The liquidity demand of an institution can be affected not only by its own liquidity shock but also the liquidity shocks of other institutions in the market. Another and the most recent example of a liquidity crisis is the “May 6: The Flash Crash”. According to the SEC-CFTC Joint Staff Report, buy-side of the S&P 500 E-Mini Futures limit order book was depleted throughout the day and market liquidity completely disappeared in just a matter of minutes during the crash. Orders removing liquidity from the book amplified the crisis and increased the volatility. These events alone have showed that characterizing the liquidity sensitivity of institutions has important implications for market stability.

This chapter is organized in the following way: Section II introduces the data and variables. Section III characterizes the liquidity sensitivity of institutions. Section IV concludes.

2.2 Literature

There is an extensive literature on institutional trading and the behavior of mutual funds. One line of this literature explores the effect of institutional trading on prices and the performance of institutions. Wermers (1999) shows that mutual funds earn abnormal returns from herding and their trades do not destabilize the prices. Nevertheless, these herds might be consuming extraordinary liquidity in the market while they build their positions. By executing large trades versus small trades, it is possible that they make the market less liquid for smaller traders. Grinblatt, Titman, and Wermers (1997), Chen, Jagadesh and Wermers (2000) and Bennett, Sias and Starks (2003) find that institutions add value by picking the high performing stocks.Da, Gao and Jagannathan (2009) argue that growth-oriented funds add value by informational trades but GNI funds add value by providing liquidity.

Another line of literature explores the trading patterns and transactions costs of institutions. Keim and Madhavan (1997) study the transaction costs of institutions. They find that transactions costs are different for different investment styles. CRS argue that aggregate trading patterns of institutions contrarian in the long run stabilizes the prices. They implicitly assume that the trading patterns of institutions are independent of their own liquidity shocks. The "liquidity spiral" phenomena showed that it is crucial to model the liquidity shocks when characterizing the trading patterns. In a theory paper, Wang (93) argues that liquidity shocks are one reason why uniformed trader trades with informed trader. If an uninformed trader cannot distinguish information-driven trade from liquidity-driven trade, then she has the incentive to trade with informed traders. Chordia et all (2002) and Grossman and Miller (1988) explore the determinants of market liquidity and conclude that order flows reduce the market liquidity.

Finance literature is generally silent on the trade initiation decision by institutions, more specifically the choice between liquidity providing orders versus liquidity demanding orders executed by institutions. The liquidity need of an institution is the immediacy to trade. Every financial transaction has one buyer and one seller. However, two parties are not symmetric in terms of immediacy to trade. The party that initiates the transaction demands liquidity from the other side. The simplest example is a market order hitting a limit order resting in the limit order book. The market order is an order which specifies the number of shares to be traded at the best available price. On the other hand, limit order specifies the number of shares to be traded at a particular price. Traders who have immediacy to trade tend to choose market orders over limit orders. On the other hand, by strategically submitting limit orders, traders may reduce their price impact and transactions costs.

2.3 Data

The data used in this study come from 4 different sources: the Center for Research on Security Prices (CRSP), Trade and Quotes (TAQ) data, Thomson Mutual Fund Holdings Data and CRSP Mutual Funds Dataset.

2.3.1 CRSP Dataset

Shares outstanding, returns, shares code and exchange code information come from the Center for Research in Security Prices (CRSP) dataset. Only the common stocks listed on NYSE, AMEX and NASDAQ are included in this study. The sample spans from March 1993 to December 2006. The daily volatility of a certain stock is calculated as the standard deviation of the past 20 days of return. The average quarterly volatility is calculated as the average daily volatility over the quarter. The PERMNO identification from CRSP files is used to match the datasets.

2.3.2 NYSE Trades and Quotes Dataset

NYSE Trade and Quotes (TAQ) Dataset reports the time and date of each trade along with the number of shares traded and the execution price for all the exchange listed stocks. TAQ data does not classify the trades as buy-initiated or sell-initiated trades. To achieve this classification, we use the method developed by Lee and Ready (1993). This method looks at the midpoint of bid-ask spread and the price of transaction. A trade is classified as buy (sell) if the price is higher (lower) than the midpoint of the bid-ask spread. If the transaction price is equal to the midpoint of bid-ask spread, then this method looks at the previous transaction price. Price increasing (decreasing) trades are classified as buy (sell). Furthermore, Lee and Ready (1993) suggest using the last recorded movement in prices to classify trades if there is no price change. After assigning each trade as buy or sell, we follow closely the method developed by CRS (2009). Specifically, buy and sell trades are placed into 19 different trade size bins. (The lower cut-offs are \$0, \$2000, \$3000, \$5000, \$7000, \$9000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000 and \$1 M.). After scaling shares traded with the daily number of shares outstanding reported in CRSP, within each bin, the quarterly buy (sell) volume is calculated as the aggregation of all buy (sell) trades. Using separate trade sizes is useful to characterize the trade size choices of institutions. The data is aggregated to the quarterly level to match the mutual funds data set which is in quarterly frequency. The difference between quarterly buy and sell volume within a certain trade size bin is the net order imbalance in that bin and denoted by $Net^{j}_{i,t}$ for stock i, quarter t and bin j. The total quarterly order imbalance is simply the sum of these order imbalances and denoted by $Net_{i,t}$ for stock i and quarter t.

2.3.3 Mutual Fund Holdings Dataset:

All Mutual Funds in the US are required to report their end of quarter portfolios (number of shares they own for each stock) through 13-F filing. This data is available on WRDS (Wharton Research Data Services) along with other characteristics of the funds such as investment style and total net assets. Mutual fund ownership for a particular stock is calculated as the number of common stocks owned by the mutual funds as a fraction of shares outstanding. Mutual fund ownership and change in mutual fund ownership variables are calculated separately for Growth, Aggressive Growth, Growth and Income (GNI) and Index Fund categories. The change in mutual fund ownership in a given quarter represents the mutual fund net trading amount and direction. This data is matched with TAQ data to infer the liquidity sensitivity of the mutual funds trades.

2.3.4 CRSP Mutual Fund Data:

CRSP Mutual Fund data contains the monthly return on asset (r_t), total net assets (TNA_t) of Mutual Funds. The dollar amount mutual fund investor flow at the monthly frequency is calculated in the following way:

$$\$Flow_t = TNA_t - (1 + r_t) \times TNA_{t-1}$$

The quarterly mutual fund investor flow in a given quarter is the sum of monthly flows in the quarter. CRSP Mutual Holdings data is matched with the holdings dataset by fund names and fund tickers. Later, the individual mutual fund flows are aggregated and scaled by the aggregate lagged TNA to calculate the aggregate quarterly mutual fund flow. The Flow variable is computed separately for Index, GNI, Growth and Aggressive Growth Fund categories.

2.3.5 Descriptive Statistics and Variable Construction:

Mutual fund quarterly net trades (Δy) are calculated as the quarterly change in aggregate ownership for each stocks and fund category. The past return (Ret_{t-1}) is defined as lagged quarterly return. To capture the possible asymmetry between high performing and poor performing stocks, we define Ret^+_{t-1} and Ret^-_{t-1} below.

$$Ret^+_{t-1} = Max\{Ret_{t-1}, 0\} \text{ and } Ret^-_{t-1} = Max\{-Ret_{t-1}, 0\}$$

Similarly, the mutual fund investor flows are formulated as positive and negative flows to identify possible asymmetries in the way institutions react to the flows.

$$Flow^+_f = Max\{Flow_f, 0\} \text{ and } Flow^-_f = Max\{-Flow_f, 0\}$$

$$f = \{Growth, Agg.Growth, GNI, Index\}$$

The quant crises in August 2007 provided evidence that institutional trading is correlated with liquidity shocks to the other institutions. Hence, the natural approach is to formulate the mutual fund flows that other institutions receive.

$$EFlow_f^{+} = \sum_{i \neq f} Max\{Flow_i, 0\} \text{ and } EFlow_f^{-} = \sum_{i \neq f} Max\{-Flow_i, 0\}$$
$$f = \{Growth, Agg.Growth, GNI, Index\}$$

A quick look at the flow variables reveals high degree of correlation among mutual fund industry flows. To address the multicollinearity problem, the flow variables are orthogonalized by regression $EFlow_f^{+}$ on $Flow_f^{+}$ and $EFlow_f^{-}$ on $Flow_f^{-}$ and residuals are used for further analysis.

Trade initiation decision of institution may also depend on the price volatility. Kyle (1985) shows that price impact is increasing with volatility. When price volatility increases, the traders may strategically place resting limit orders to reduce their price impact. If the offsetting orders, on the other hand, are unlikely to come to the market place, the risk-averse informed trader may choose to submit market orders to ensure trade execution. The quarterly price volatility is the mean of daily price volatility which is the standard deviation of return over the past 20 trading days. The stocks are divided into 5 size categories by the market capitalizations based on the NYSE stocks cutoffs. The trades are signed as buy and sell by the Lee and Ready (1993) algorithm with the prevailing quote assumed to be posted 2 seconds before the trade took place. The 2 seconds rule is consistent with 1-2 seconds rule developed by Piwowar and Wei (2006). The net order imbalance scaled by shares outstanding is calculated separately for separate trade size bins as the difference between buy and sell volumes.

Table 2.1 reports the summary statistics for the study. Panel A describes the stock level variables. The sample starts from March 1993 and ends in December 2006. There are 247,564 stock quarters in the sample. In the sample, there is a sell pressure in the small stocks and buy pressure in the large stocks. The net order imbalance ranges from -0.014 for smallest category to 0.0144 for the largest category. Index fund trading tends to correspond to the behavior in the net order imbalance. On average mutual funds are net sellers in the sample period. The mutual fund ownership is mostly concentrated in the larger stocks, consistent with the idea that institution hold large stocks for liquidity and corporate control purposes (Lo and Wang (2000)). The Ret_t^+ and Ret_t^- variables decrease monotonically as the market capitalization increases, matching nicely the behavior of price volatility which is 0.04 for smallest category and 0.0241 for the larger category.

Panel B describes the mutual fund investment style level variables. During the sample period, mutual funds are net receivers of investor flows. Index fund receives 2.98% inflow of TNA on average per quarter. Aggressive growth fund has 0.43% inflow of TNA which is the lowest among all mutual funds.

2.4 Mutual Fund Trading and Order Imbalance

2.4.1 Basic Framework

Institutional trades that remove liquidity from the book add the order imbalance calculations. If mutual funds generally consume liquidity then quarterly mutual fund trades and order imbalance variables should be positively correlated. To account for spurious correlation between these two variables, other effects such as high frequency momentum strategies should also be controlled. Following CRS, regression approach is used to control for other variables that might be correlated with order imbalance.

A trader buying on the up tick and selling on the down tick may appear to consume liquidity at quarterly frequency even if she only uses resting orders. Including Ret_t in the regression framework should control for high frequency momentum strategies. Other variables such as $Log(B/M)_{t-1}$, $Volatility_{t-1}$ and Ret_{t-1} also included to account for trading strategies that may resemble liquidity consumption even if these strategies involve resting orders but somehow correlate with order imbalance.

The natural first step to analyze the liquidity sensitivity of a mutual fund category is to estimate the following regression:

$$\begin{aligned}\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t-1} + \\ \gamma_4 Log(B/M)_{i,t-1} + \gamma_5 Log(MarketCap)_{i,t-1} + \beta \times Net_{i,t} + \epsilon_{i,t}\end{aligned} \tag{2.1}$$

This equation reveals how much of the variation in mutual fund trades can be explained by various variables including order imbalance. The β coefficient characterizes the liquidity sensitivity of mutual funds, a positive coefficient implying that they trade in the same direction with liquidity consumption. Table 2.2 reports the results from the equation (2.1) for different mutual fund categories for all stocks in the sample. The specifications include time-fixed effects and t statistics are calculated from the standard errors clustered at the stock level. Peterson (2008) shows that standard errors clustered by stock are unbiased whether the stock effect is temporary or permanent. $Adj - R^2$ of the regressions range from 0.71 for Index fund to 0.21 for Growth Funds. The mutual fund holdings have statistically significant mean reversion in the short and the long-run. The investment style differences among mutual funds appear in the coefficients. The coefficients on Ret_t and Ret_{t-1} are positive and significant for growth oriented funds. Same coefficients for GNI funds are negative. Growth-oriented funds buy momentum stocks whereas value oriented funds follow contrarian investment strategies. Positive coefficient for Ret_{t-1} of Index funds indicates that they tend to buy recent winners and sell recent losers. Negative coefficient for Ret_t of Index funds is consistent with the idea that they are contrarian within the same quarter. The coefficients for $Volatility_{t-1}$ describes the mutual funds' preference for volatility. All the funds except GNI sell high volatility stocks and buy lower volatility stocks. The coefficients for $Log(B/M)_{t-1}$ are positive for GNI and Index Fund suggesting that these funds buy value stocks and sell growth stocks. On the other hand, growth-oriented funds buy growth funds and sell value stocks. The coefficient for $Log(MarketCap)_{t-1}$ is positive for all mutual funds, indicating they buy larger stocks. This finding is consistent with Gompers and Metric

(2001) who argue that institutions have special interest for large stocks.

All the coefficients for Net_t are positive except the one for index funds. The positive coefficient indicates that mutual funds are positively correlated with the trades buying at the ask and selling at the bid price. The liquidity sensitivity coefficient for the index fund is further analyzed in the next section. The coefficients are larger for growth-oriented firms than value-oriented firms, suggesting that growth oriented funds are more sensitive to liquidity consumption than value oriented funds.

2.4.2 Basic Framework Across Stock Size Categories

Since mutual funds have special interest for holding larger stocks, their trading strategy and liquidity sensitivity may differ for different size categories. To explore the liquidity sensitivity of mutual funds across different stock sizes, equation (2.1), excluding the $Log(Market\_Cap)_{t-1}$ variable, is run for 5 different stock size categories. The categories are defined by NYSE market capitalization cutoff points.

Table 2.3 reports the results. The adjusted-R^2 of the regression is fairly large and monotonically increases from small stocks to the large stocks. The degree of trade detectability is especially high for the index funds mostly because their significant mean reversion in the short horizon. The liquidity sensitivity coefficients for index funds is monotonically increasing from -0.001 to 0.006. Keim (1999) shows that “9-10 Fund,” a fund tracking the CRSP 9-10 Index, generates excess returns over the index by deviating from tracking the index to potentially reduce transactions costs. The results in table 2.3 are consistent with the idea that index funds may avoid or even take advantage of transactions costs in illiquid stocks at expense of closely following the index and they demand liquidity in larger stocks whereas supply liquidity in smaller stocks.

The liquidity sensitivity coefficients of actively managed funds are positive in all stock size categories and they tend to be higher for larger stocks. Transaction costs in smaller stocks are higher, making liquidity consumption expensive. Institutions trying to minimize the transactions costs may restrict their liquidity demand in smaller stocks. Furthermore, liquidity sensitivity coefficients of growth oriented funds are larger than their counterparts for GNI fund. This difference is larger for smaller stocks. GNI funds are more likely to hold mature firms that have high earnings and pay dividends. Growth-oriented funds on the other hand invest in stocks that do not usually pay dividends but have large investments with risky future cash flows. Growth-oriented funds tend to rely on fundamental and technical analysis to acquire information about the future performance of stocks. The future performance of small-cap stocks is more uncertain than large-cap stocks simply because stock price of small cap stocks is not as informative as large cap stocks. Growth-oriented funds may invest in discovering the true value of the small cap stocks and produce information. Acting on this information, growth-oriented funds are more likely to initiate trades and demand immediacy even in small and illiquid stocks.

2.4.3 Sensitivity of Mutual Fund Trading to Order Imbalance

The natural next step is to analyze the sensitivity of mutual fund trading to order imbalance across stock and fund level characteristics. The following regression is considered for this purpose:

$$
\begin{aligned}
\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{t-1,i} + \gamma_2 Ret_{t,i} + \gamma_3 Volatility_{t-1,i} \\
+\gamma_4 Log(B/M)_{t,i} + \gamma_5 Log(Market\_Cap)_{i,t-1} + Net_{t,i} \times [\beta_1 + \beta_2 Log(Market\_Cap)_{i,t-1} \\
+\beta_3 Log(B/M)_{i,t-1} + \beta_4 y_{i,t-1} + \beta_5 Flow_t^+ + \beta_6 EFlow_t^+ + \beta_7 Flow_t^- \\
+\beta_8 EFlow_t^- + \beta_9 RET_{i,t-1}^+ + \beta_{10} RET_{i,t-1}^- + \beta_{10} Volatility_{i,t-1}] + \epsilon_{i,t}
\end{aligned}
\tag{2.2}
$$

The equation relates the mutual fund trading activity to the stock characteristics such as recent return, book-to-market and volatility as well as the net order imbalance and several cross-terms to explain the shifts in the liquidity sensitivity.

Table 2.4 reports the results from the equation (2.5) for all the entire sample. The results for stock variables reveal important characteristics of mutual fund trading at the quarterly frequency. The positive coefficients on the Ret_{t-1} variable suggest that growth-oriented and index funds are momentum-traders, buying recent winners and selling recent losers. GNI funds, on the other hand, appear to be contrarians, buying the recent losers and selling the recent winners. The negative coefficient for volatility suggests that mutual funds dislike the high volatility stocks and sell them. Growth-oriented funds also sell the ones that become value stocks and buy the ones which become growth stocks. GNI funds do the opposite, selling growth stocks and buying value stocks.

The cross terms with the Net_t variable show the sensitivity of mutual fund to liquidity consumption across stock and fund characteristics. Aggressive growth fund trades in low book-to-market stocks are revealed by liquidity consumption in those stocks. On the other hand, the trades of GNI funds in high book-to-market stocks are explained by liquidity consumption in these stocks. Similarly, when experienced buy flows, liquidity sensitivity of actively managed mutual funds increases. Index funds, however, reduce liquidity sensitivity in response to the flows. This is consistent with the following intuition. When index funds receive inflows, they buy S&P 500 future contracts to get immediate exposure to the index. Later, they buy the stocks which constitute the index slowly through liquidity providing trades. Actively managed funds, on the other hand, act on the information they acquire and pick stocks in which they invest. When they receive inflows, they initiate trades to buy the stocks they have already picked before the price moves away. The positive coefficient for the $Ret^{+}_{t-1} \times Net_t$ for aggressive growth funds show that liquidity consumption indicate the trades of these funds. The negative coefficient for the $Ret^{+}_{t-1} \times Net_t$ for GNI funds show that their trades are negatively correlated with liquidity consumption in these stocks. Since aggressive growth funds follow a momentum strategy and GNI funds follow a contrarian strategy, aggressive growth funds buy high performing stocks and GNI funds sell them. These results support a plausible explanation that in a likely transaction between these two funds, aggressive growth funds demand liquidity whereas GNI funds supply liquidity.

Mutual Funds also differ in the way they respond to the flows which go to the other funds. Growth funds tend to reduce their liquidity sensitivity whereas aggressive growth funds raise their liquidity sensitivity. If the aggressive growth funds try to buy the stocks before the mutual funds which experience inflow, they demand more immediacy from the market. The results are consistent with front-running or herding explanation. This strategy by aggressive growth funds may make market less liquid for the other institutions by raising the transactions costs.

Next, I compare the sensitivity of mutual funds to liquidity consumption in different market capitalization categories. Table 2.5 reports the results for different size categories. Growth-oriented funds, when received inflows, generally increase liquidity sensitivity in the larger stocks. Buying large stocks in response to flows allows the funds to reduce transactions costs as the larger stocks are more liquid. Higher liquidity sensitivity of growth-oriented funds for high performing stocks is concentrated in small stocks. Wermers (1995) show growth-oriented mutual funds buy small high performing stocks in herds. My results support a hypothesis that these herds are associated with liquidity consumption in markets when buying high performing small stocks.

Aggressive growth funds' liquidity sensitivity increases when other mutual funds receive outflows. If this is associated with front running, aggressive growth funds make the market less liquid on the sell side for other traders. Other funds, however, tend to reduce their liquidity sensitivity when other mutual funds are experiencing outflows. This behavior may help the market liquidity especially at the time when others receive adverse liquidity shock. Index funds reduce liquidity sensitivity in response to the inflows especially in the largest stock category. This is consistent with a plausible explanation. Index funds following S&P 500 index can trade S&P 500 Futures when they receive inflow to get an exposure to the index. On the sell side, nevertheless, this is not possible since selling S&P Futures contract does not generate cash immediately. They may utilize stock lending facilities to generate cash that they need to satisfy redemption. Index Funds, interestingly, increase liquidity consumption as the volatility increases in small stocks. When the volatility of the stock increases, institutions demand more risk premium and sell the stock. Thus, it is intriguing to find that trades of index funds can be explained by volatility.

2.5 Information in Trade Sizes

The equation 2.1 does not consider the information in different trade sizes. The general version of the sensitivity of mutual fund trading over different trade sizes is:

$$\begin{aligned}\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t} \\ + \gamma_4 Log(B/M)_{i,t} + \sum_{j=1}^{19} [\beta_j \times Net^j_{i,t}] + \epsilon_{i,t}\end{aligned} \tag{2.3}$$

where j is the index for trade sizes.

Estimating the equation 2.3 with simple linear regression is troublesome because net order imbalance is small stocks are rarely traded in large trade sizes and the information in the trade sizes is highly correlated which leads to the multicollinearity in the linear model. CRS estimate a smooth function which captures the relationship between institutional trading and various trade sizes . In this study, we choose to estimate this regression with a simpler function to assign the estimated coefficients meaning. To check the robustness, the analysis is done with the Nelson and Siegel (1987) function that CRS use and the results are qualitatively very similar.

$$\beta_j = b_{01} + b_{02} \times Log(Z_j) \tag{2.4}$$

where Z_j is the midpoint of the trade size bin. After plugging 2.4 into 2.3, the regression specification becomes:

$$\begin{aligned}\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t} + \\ \gamma_4 Log(B/M)_{i,t} + b_{01} \times Net_(i,t) + b_{02} \times \sum_{j=1}^{19} Log(Z_j) \times Net^j_{i,t} + \epsilon_{i,t}\end{aligned} \tag{2.5}$$

Net_Slope variable, $\sum_{j=1}^{19} Log(Z_j) \times Net^j_{i,t}$, is simply a sum of order imbalances weighted by their trade sizes. The parameter b_{02}, therefore, measures the sensitivity of correlation between order imbalance and institutional trading with respect to different trade sizes. The equation 2.5 is run for 5 different stock size categories across different mutual fund styles. Table 2.7 reports the results of the regression which includes Net_Slope variable. t statistics clustered at the stock level are reported in parentheses. The coefficients for Net_Slope variable are negative for all size categories. Their trading strategy is correlated more with small trade sizes than larger trade sizes. There are two competing explanations for this finding. First, index funds need immediacy in small trades because they need small adjustments to track the index. Hence, the coefficient for Net_Slope variable captures the liquidity demand of Index fund over different trade sizes. Second, Index funds follow investors who submit small orders and the negative coefficients for Net_Slope variable reveal their trading strategy with respect to the trade size choices. Following small trades may lead to random trading strategies for index funds which in turn result in significant deviations from the underlying index. It is conceivable to argue that for index funds, tracking the index is more important than adopting a statistical trading strategy to create value. Hence, the results for the index fund can be interpreted as liquidity demand sensitivity. This suggests that index fund demand liquidity demand is higher for small trade sizes.

The Net_Slope coefficients for actively managed funds are positive in general. They tend to be higher for growth-oriented funds rather than value oriented funds. Compared to small trades, larger trade sizes tend to explain the trades of these mutual funds. The trades of actively managed funds are more likely to be correlated with larger trade sizes then small trade sizes. This relationship might be caused by mutual funds following liquidity consumption in larger trade sizes. If private information has a short life, then informed traders have urgency to trade and demand liquidity in larger trade sizes. Growth-oriented funds invest heavily in fundamental and technical analysis to produce information about financial securities to become informed traders. The results are consistent with the idea that growth-oriented funds submit larger trades for information-related reasons. Nevertheless, the hypothesis that actively managed mutual funds follow a trading strategy that follows large orders cannot be rejected.

Figure 2.1 displays the estimated order imbalance coefficients , β_j, from 2.5 over different trade sizes for index and GNI funds. The coefficients for index fund are monotonically decreasing over trade sizes. Their trading strategy is generally correlated with order imbalances in small trade categories. Employing a strategy which involves trading in the same direction with small orders may cause large index tracking errors. The more likely explanation for this result is that index funds need immediacy especially in small trade sizes because they need small adjustments in their portfolio to closely follow the index. The coefficients for GNI funds, on the other hand, are increasing over trade sizes. Order imbalance in larger trades reveal GNI trading activity more than that in small trade sizes.

Figure 2.2 displays the estimated order imbalance coefficients from the equation 2.5 over different trade sizes for Growth and Aggressive Growth funds. Similar to the results from GNI funds, trades of growth-oriented funds positively correlate with order imbalances in large trade sizes. Such relationship may be a result of certain trading strategy sensitive to large orders. Another plausible interpretation would be that these coefficients indeed reveal the liquidity demand of the institutions over different trade sizes. The distinction between these two interpretations can not be tested in the data because separating these two hypothesis require observations in higher frequency such daily trades of mutual funds. Although the large order following strategy explanations can not be ruled out, these results are consistent with the following intuition: The trades of growth-oriented funds are motivated by the changes in the expected stocks returns. The information they have might be short-lived, therefore, they have to quickly rebalance their portfolio and submit large trades before price moves away. Index Funds, however, submit smaller trades to reduce transactions costs and follow volume-weighted average price (VWAP) trading strategies.

The results of CRS show that trades under $2,000 or over $30,000 indicate institutional trading. However, we document that actively managed funds are correlated with larger trade sizes whereas index fund trading is revealed by small trade sizes, suggesting that institutions with different investment styles can be accounted for the correlation of institutional trading and various trade sizes.

2.6 Conclusion

The sensitivity of mutual fund trading to liquidity consumption is studied. Mutual Funds are not only different in their investment strategy but also in their sensitivity to liquidity consumption. The three main results stand out.

- Mutual fund trades correlate with order imbalance, suggesting they trade in the same direction with liquidity consumption.

- Mutual fund liquidity sensitivity differs based on their investment styles. The liquidity sensitivity of growth-oriented funds is higher than the value oriented and passively managed funds. This is consistent with the hypothesis that information motivated trader consumes more liquidity.

- Actively managed fund trading can be explained by larger trades consuming liquidity whereas small trades reveal the trades of index funds. This evidence supports the plausible explanation that information-motivated orders submitted by actively managed funds tend to be large and index funds need only small adjustments to track the index.

Tab. 2.1: Descriptive Statistics

Panel A: Stock Level Statistics

	Size					
	Small	2	3	4	Large	All
Net_t	-0.0140	-0.0011	0.0093	0.0160	0.0144	-0.0052
Index Δy_t	-0.0001	0.0002	0.0001	0.00003	0.0001	-0.00001
Agg. Growth Δy_t	-0.0002	-0.0006	-0.0011	-0.0011	-0.0007	-0.0005
Growth Δy_t	-0.0003	-0.0002	-0.0003	-0.0012	-0.0012	-0.0004
GNI Δy_t	-0.0001	-0.0002	-0.0002	-0.0001	0.0000	-0.0001
Index y_t	0.0090	0.0127	0.0110	0.0101	0.0143	0.0103
Growth y_t	0.0080	0.0212	0.0258	0.0228	0.0123	0.0133
GNI y_t	0.0255	0.0609	0.0704	0.0714	0.0555	0.0412
Agg Growth y_t	0.0020	0.0070	0.0153	0.0327	0.0512	0.0100
$Volatility_{t-1}$	0.0431	0.0297	0.0269	0.0241	0.0224	0.0365
$Log(B/M)_{t-1}$	-0.4954	-0.7791	-0.8705	-0.9046	-1.0791	-0.6504
Ret_{t-1}	0.0359	0.0412	0.0384	0.0399	0.0357	0.0373
Ret^{+}_{t-1}	0.1269	0.1142	0.1047	0.0957	0.0831	0.1171
Ret^{-}_{t-1}	-0.0911	-0.0730	-0.0663	-0.0558	-0.0474	-0.0799
# of Obs.	147,757	37,081	25,105	20,180	17,441	247,564

Panel B: Fund Level Statistics

	Index Fund	GNI	Growth	Agg. Growth	All
Flow	0.0298	0.0054	0.0098	0.0043	0.0123
$Flow^{+}$	0.0413	0.0212	0.0264	0.0366	0.0314
$Flow^{-}$	0.0115	0.0158	0.0166	0.0324	0.0191
TNA_t ($ Millions)	219,589	492,908	685,450	172,546	392,623

Table reports the descriptive statistics for the sample from 1993 to 2006. Panel A presents the stock level variables at the quarterly frequency. Order Imbalance (Net_t), stock ownership (y_t) and change in ownership (Δy_t) are scaled by shares outstanding. $Volatility_t$ is calculated as the standard deviation of daily returns over the last 20 trading days. Panel B presents the quarterly mutual fund level variables at the investment style category. Flow variables are scaled by investment style aggregate lagged TNA.

Tab. 2.2: Basic Regression I

	Index Fund	GNI Fund	Growth Fund	Agg. Growth
Intercept	-0.01	-0.01	-0.03	-0.01
	-66.18	-43.45	-42.39	-27.31
y_{t-1}	-0.11	-0.05	-0.06	-0.11
	-47.5	-29.29	-40.85	-53.21
Δy_{t-1}	-0.62	-0.31	-0.31	-0.26
	-179.25	-53.11	-85.48	-55.65
Ret_t	0.0003	-0.0003	0.0028	0.0029
	6.83	-6.02	17.94	28.13
Ret_{t-1}	-0.0008	-0.0007	0.0028	0.0022
	-17	-15.1	18.16	22.24
$Volatility_t$	-0.0049	0.0004	-0.0274	-0.0030
	-9.62	0.72	-15.5	-3.48
$Log(B/M)_{t-1}$	0.0004	0.0002	-0.0003	-0.0003
	23.48	9.8	-6.98	-11.86
$Log(Market\_Cap)_{t-1}$	0.0002	0.0005	0.0004	0.0001
	32.88	39.44	15.05	10.55
Net_t	-0.0011	0.0025	0.0274	0.0127
	-2.86	5.99	18.74	14.53
Time Fixed Effects	Yes	Yes	Yes	Yes
# of Stocks	10,570	10,570	10,570	10,570
# of Obs	247,564	247,564	247,564	247,564
$Adj - R^2$	0.72	0.21	0.24	0.21

Table reports the regression results from the equation :

$$\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t} + \gamma_4 Log(B/M)_{i,t} + \gamma_5 Log(Market\_Cap)_{i,t-1} + \beta \times Net_{i,t} + \epsilon_{i,t}$$

t statistics calculated from stock level clustered standard errors are reported below the coefficients. Quarterly time fixed effects are included in all specifications. Sample runs from 1993 to 2006.

Tab. 2.3: Basic Regression II

Panel A

	Index Fund					GNI Fund				
	Small	2	3	4	5	Small	2	3	4	5
Intercept	-0.005	-0.009	-0.007	-0.004	-0.005	-0.001	-0.003	-0.005	-0.007	-0.012
	-48.66	-31.3	-21.51	-16.99	-15.55	-9.7	-10.71	-9.77	-12.28	-17.96
y_{t-1}	-0.104	-0.122	-0.102	-0.069	-0.146	-0.104	-0.102	-0.078	-0.054	-0.037
	-37.33	-15.99	-10.93	-8.76	-9.84	-18.69	-16.27	-17.23	-14.76	-10.91
Δy_{t-1}	-0.594	-0.517	-0.581	-0.711	-0.625	-0.331	-0.268	-0.281	-0.291	-0.260
	-129.26	-51.41	-41.97	-56.08	-35.6	-29.6	-18.42	-19.76	-24.75	-21.24
Ret_t	0.0004	0.0001	-0.0004	-0.00001	-0.0001	0.000	0.000	-0.001	-0.003	-0.005
	7.12	0.47	-2.57	-0.03	-0.47	-0.54	-1.9	-5.14	-6.7	-8.23
Ret_{t-1}	-0.0004	-0.001	-0.001	-0.0002	0.0005	0.0001	-0.0002	-0.0005	-0.001	-0.005
	-6.76	-5.42	-8.29	-1.45	2.63	3.05	-1.63	-2.07	-2.99	-7.75
$Volatility_{t-1}$	-0.013	-0.005	0.001	-0.002	-0.004	-0.004	-0.016	-0.024	-0.031	-0.033
	-22.5	-2.59	0.51	-0.71	-1.82	-12.87	-7.00	-5.48	-5.19	-3.1
$Log(B/M)_{t-1}$	0.00019	0.00054	0.00025	0.00006	0.00001	0.00002	-0.00002	0.00019	0.00022	0.00023
	11.41	11.99	5.87	1.14	0.39	1.61	-0.35	2.28	1.93	1.72
Net_t	-0.001	0.001	0.002	0.002	0.006	0.0003	0.003	0.008	0.015	0.018
	-2.30	1.30	2.23	2.00	4.15	0.77	2.65	4.65	5.37	4.3
# of Stocks	8195	3130	2109	1325	722	8195	3130	2109	1325	722
# of Obs	147,757	37,081	25,105	20,180	17,441	147,757	37,081	25,105	20,180	17,441
$Adj-R^2$	0.68	0.80	0.83	0.82	0.94	0.26	0.23	0.22	0.26	0.42

Panel B

	Growth Fund					Agg. Growth Fund				
Intercept	-0.013	-0.031	-0.034	-0.031	-0.024	-0.004	-0.011	-0.012	-0.010	-0.004
	-30.22	-31.14	-29.14	-25.24	-21.89	-20.41	-21.2	-19.49	-16.59	-11.31
y_{t-1}	-0.056	-0.064	-0.063	-0.063	-0.075	-0.116	-0.106	-0.114	-0.115	-0.134
	-26.34	-20.77	-16.65	-16.48	-13.9	-33.03	-24.65	-25.56	-19.02	-16.5
Δy_{t-1}	-0.313	-0.288	-0.280	-0.302	-0.341	-0.261	-0.236	-0.250	-0.271	-0.287
	-62.34	-34.81	-28.89	-25.64	-20.58	-37.21	-23.68	-23.24	-20.07	-15.66
Ret_t	0.001	0.004	0.007	0.011	0.011	0.001	0.006	0.008	0.010	0.007
	8.99	7.58	9.05	10.36	9.8	14.71	16.7	15.15	14.83	10.98
Ret_{t-1}	0.002	0.005	0.007	0.010	0.009	0.001	0.006	0.009	0.010	0.006
	14.87	9.19	9.61	9.72	8.25	14.81	18.44	15.68	12.56	8.02
$Volatility_{t-1}$	-0.038	-0.064	-0.031	-0.031	-0.010	-0.010	0.006	0.004	0.022	0.003
	-23.59	-7.02	-1.81	-2.38	-1.08	-14.58	1.02	0.48	2.43	0.59
$Log(B/M)_{t-1}$	-0.0002	-0.001	-0.001	-0.001	-0.001	-0.0002	-0.001	-0.001	-0.001	-0.001
	-3.36	-5.7	-6.3	-5.33	-4.38	-5.66	-6.8	-8.84	-6.08	-5.12
Net_t	0.013	0.039	0.054	0.045	0.061	0.005	0.023	0.020	0.019	0.016
	7.59	9.47	10.67	7.66	8.77	5.55	9.57	6.67	5.48	3.67
# of Stocks	8195	3130	2109	1325	722	8195	3130	2109	1325	722
# of Obs	147,757	37,081	25,105	20,180	17,441	147,757	37,081	25,105	20,180	17,441
$Adj-R^2$	0.21	0.27	0.29	0.32	0.40	0.22	0.22	0.26	0.28	0.30

Table reports the regression results for 5 different stock size categories from the equation :

$\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t} + \gamma_4 Log(B/M)_{i,t} + \beta \times Net_{i,t} + \epsilon_{i,t}$

Stock size cut-off points are calculated from NYSE market capitalization cut-off points. t statistics calculated from stock level clustered standard errors are reported below the coefficients. Quarterly time fixed effects are included in all specifications. Sample runs from 1993 to 2006.

Tab. 2.4: Sensitivity of Mutual Funds to Liquidity Consumption

	Index Fund	GNI Fund	Growth Fund	Agg Growth Fund
Intercept	-0.01	-0.01	-0.03	-0.01
	-67.24	-43.51	-41.73	-27.48
y_{t-1}	-0.11	-0.05	-0.06	-0.11
	-48.08	-30.37	-40.24	-53.72
Δy_{t-1}	-0.62	-0.31	-0.31	-0.26
	-178.82	-53.19	-85.49	-55.90
Ret_t	0.000	-0.0002	0.003	0.003
	6.28	-5.50	18.56	28.43
Ret_{t-1}	-0.001	-0.001	0.003	0.002
	-16.81	-14.85	16.66	21.84
$Volatility_{t-1}$	-0.0035	-0.0005	-0.03	-0.004
	-7.11	-0.97	-15.56	-4.64
$Log(B/M)_{t-1}$	0.0004	0.0002	-0.0003	-0.0003
	23.65	9.94	-6.18	-11.43
$Log(Market\_Cap)_{t-1}$	0.0003	0.0005	0.0004	0.0002
	33.65	39.42	14.56	10.90
Net_t	0.01	-0.04	-0.08	-0.05
	1.51	-7.38	-4.50	-4.81
$Net_t \times Log(Market_Cap)_{t-1}$	-0.0002	0.0023	0.0043	0.0023
	-0.84	7.75	4.57	4.87
$Net_t \times Log(B/M)_t$	0.0007	0.0017	0.0006	-0.0019
	1.88	4.17	0.46	-2.49
$Net_t \times y_{t-1}$	-0.02	0.14	0.26	0.34
	-0.47	2.52	6.42	7.68
$Net_t \times Flow_t^+$	-0.07	0.05	0.34	0.09
	-8.09	3.07	7.44	5.66
$Net_t \times EFlow_t^+$	-0.0030	0.01	-0.12	0.04
	-0.65	2.69	-5.06	4.07
$Net_t \times Flow_t^-$	-0.20	-0.0024	0.05	0.03
	-10.98	-0.15	0.97	1.61
$Net_t \times EFlow_t^-$	-0.05	-0.05	-0.27	0.11
	-7.64	-8.89	-4.87	6.01
$Net_t \times Ret_{t-1}^+$	-0.0036	-0.0076	0.0033	0.0129
	-2.52	-5.36	0.64	6.32
$Net_t \times Ret_{t-1}^-$	-0.0021	0.01	0.03	0.01
	-0.76	2.42	2.62	1.70
$Net_t \times Volatility_{t-1}$	0.07	0.0003	-0.04	-0.02
	4.86	0.02	-0.68	-0.80
Time Fixed Effects	Yes	Yes	Yes	Yes
# of Stocks	10,570	10,570	10,570	10,570
# of Obs	247,564	247,564	247,564	247,564
$Adj - R^2$	0.72153	0.210473	0.237293	0.214624

Table reports the regression results for the entire sample from the Equation 2.2. t statistics calculated from stock level clustered standard errors are reported below the coefficients. Quarterly time fixed effects are included in all specifications. Sample runs from 1993 to 2006.

Tab. 2.5: Sensitivity of Mutual Funds to Liquidity Consumption: Size Categories

	Index Fund					GNI Fund				
	Small	2	3	4	Large	Small	2	3	4	Large
Intercept	-0.005	-0.009	-0.007	-0.004	-0.005	-0.001	-0.003	-0.005	-0.007	-0.012
	-48.8	-31.59	-21.51	-16.57	-15.66	-9.68	-10.67	-10	-12.17	-17.08
y_{t-1}	-0.103	-0.123	-0.105	-0.069	-0.139	-0.101	-0.102	-0.077	-0.052	-0.035
	-37.09	-15.42	-11.05	-8.31	-9.08	-17.29	-16.24	-16.29	-12.88	-8.89
Δy_{t-1}	-0.593	-0.516	-0.581	-0.711	-0.625	-0.330	-0.268	-0.281	-0.291	-0.261
	-129.33	-51.01	-42.01	-55.91	-35.74	-29.56	-18.45	-19.77	-24.77	-21.35
Ret_t	0.0003	0.0001	-0.0004	0.0000	-0.0001	-0.00001	-0.0003	-0.0013	-0.0030	-0.0050
	6.64	0.8	-2.48	0.14	-0.34	-0.39	-1.8	-5.05	-6.75	-7.8
Ret_{t-1}	-0.00030	-0.00066	-0.00106	-0.00016	0.00028	0.00005	-0.00026	-0.00062	-0.00183	-0.00594
	-5.26	-5.29	-7.45	-0.87	1.4	1.57	-1.94	-2.45	-3.47	-7.85
$Volatility_{t-1}$	-0.01117	-0.0061	0.001168	-0.00103	-0.00425	-0.00399	-0.01577	-0.02545	-0.03409	-0.03468
	-19.8	-2.96	0.54	-0.31	-1.73	-11.52	-6.73	-5.54	-5.63	-3.01
$Log(B/M)_{t-1}$	0.00019	0.00054	0.00024	0.00006	0.000004	0.00002	-0.00002	0.00015	0.00001	0.00019
	10.91	12.06	5.54	1.2	0.1	1.88	-0.29	1.73	0.11	1.17
Net_t	-0.00351	0.007614	0.003486	0.006063	0.011081	0.000856	0.000622	0.017059	0.021889	0.03028
	-2.23	2.3	1.15	1.72	1.94	1.04	0.24	4.04	3.4	2.22
$Net_t \times Log(B/M)_{t-1}$	-0.00007	0.00239	0.00106	-0.00044	0.00084	0.000306	0.000982	0.004096	0.011749	0.002748
	-0.15	2.33	0.98	-0.33	0.54	0.99	0.74	2.02	3.45	0.62
$Net_t \times y_{t-1}$	0.092	0.022	0.157	-0.036	-0.299	0.270	0.145	-0.147	-0.094	-0.089
	1.29	0.19	1.22	-0.25	-1.63	1.83	0.91	-1.19	-0.79	-0.56
$Net_t \times Flow_t^+$	-0.041	-0.046	-0.025	-0.035	-0.048	-0.025	0.021	-0.148	0.128	-0.317
	-3.09	-1.94	-0.99	-1.32	-1.2	-1.77	0.6	-1.99	1.14	-1.83
$Net_t \times EFlow_t^+$	0.005	-0.001	0.005	0.008	0.024	0.007	-0.019	-0.001	0.008	0.063
	0.91	-0.1	0.33	0.49	1.18	2.15	-1.59	-0.03	0.26	1.23
$Net_t \times Flow_t^-$	-0.037	-0.120	-0.094	-0.044	-0.006	-0.022	0.022	-0.004	0.104	-0.223
	-1.27	-2.44	-2.07	-0.93	-0.14	-2.07	0.66	-0.05	0.99	-1.58
$Net_t \times EFlow_t^-$	0.018	-0.044	-0.023	-0.005	-0.024	0.003	-0.019	-0.039	-0.053	0.070
	1.8	-2.14	-1.1	-0.23	-0.98	0.83	-1.57	-1.99	-1.57	1.54
$Net_t \times Ret_{t-1}^+$	0.001	-0.005	-0.006	-0.008	-0.001	-0.001	0.002	0.014	0.004	-0.019
	0.4	-1.23	-1.28	-1.33	-0.16	-1.23	0.42	1.56	0.28	-0.97
$Net_t \times Ret_{t-1}^-$	-0.009	-0.003	0.018	-0.008	-0.019	0.001	0.005	0.027	-0.012	-0.089
	-2.59	-0.48	2.61	-0.92	-1.66	0.72	0.7	1.93	-0.6	-3.00
$Net_t \times Volatility_{t-1}$	0.091	-0.008	-0.033	-0.013	0.111	-0.003	0.016	-0.103	0.145	0.441
	5.48	-0.16	-0.7	-0.18	0.87	-0.32	0.29	-1.08	1.1	1.79
# of Stocks	8195	3130	2109	1325	722	8195	3130	2109	1325	722
# of Obs	147,757	37,081	25,105	20,180	17,441	147,757	37,081	25,105	20,180	17,441
$Adj-R^2$	0.68	0.80	0.83	0.82	0.94	0.27	0.23	0.22	0.26	0.42

Table reports the regression results for Index and GNI funds from the equation

$$\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{t-1,i} + \gamma_2 Ret_{t,i} + \gamma_3 Volatility_{t-1,i} + \gamma_4 Log(B/M)_{t-1,i} + Net_{t,i} \times [\beta_1 + \beta_2 Log(B/M)_{i,t-1} + \beta_3 y_{i,t-1} + \beta_4 Flow^+ + \beta_5 EFlow^+ + \beta_6 Flow^- + \beta_7 EFlow^- + \beta_8 RET^+_{i,t-1} + \beta_9 RET^-_{i,t-1} + \beta_{10} VolatilityL_{i,t-1}] + \epsilon_{i,t} \ .$$

t statistics calculated from clustered standard errors are reported below the coefficients. Quarterly time fixed effects are included in all specifications. Sample runs from 1993 to 2006.

Tab. 2.6: Trade Sensitivity of Mutual Funds II: Size Categories

	Growth Fund					Agg. Growth Fund				
Intercept	-0.013	-0.031	-0.033	-0.031	-0.025	-0.004	-0.011	-0.011	-0.010	-0.004
	-30.72	-31.02	-28.27	-24.41	-21.78	-20.84	-20.6	-18.84	-16.03	-10.44
y_{t-1}	-0.051	-0.064	-0.063	-0.064	-0.068	-0.108	-0.105	-0.114	-0.117	-0.137
	-24.1	-20.71	-15.95	-14.9	-10.73	-30.09	-24.58	-25.48	-17.43	-12.64
Δy_{t-1}	-0.312	-0.287	-0.280	-0.302	-0.342	-0.261	-0.236	-0.250	-0.269	-0.286
	-62.35	-34.76	-29.01	-25.69	-20.78	-37.35	-23.72	-23.25	-19.96	-15.85
Ret_t	0.0015	0.0044	0.0066	0.0105	0.0103	0.00135	0.0061	0.0082	0.0097	0.0071
	9.45	7.9	8.65	9.8	9.59	15.03	16.53	14.88	14.88	10.76
Ret_{t-1}	0.00228	0.00520	0.00750	0.01165	0.01130	0.00124	0.00647	0.00923	0.01020	0.00681
	13.43	9.22	10.01	11.06	9.77	14.12	18.36	15.75	13.3	8.96
$Volatility_{t-1}$	-0.03783	-0.06522	-0.02927	-0.01937	-0.00829	-0.01008	0.00674	0.007439	0.029644	0.00411
	-21.76	-6.99	-1.72	-1.43	-0.88	-13.59	1.24	0.87	3.34	0.84
$Log(B/M)_{t-1}$	-0.00018	-0.00092	-0.00140	-0.00134	-0.000922	-0.00018	-0.00063	-0.00105	-0.00088	-0.00046
	-3.41	-5.61	-6.21	-5.15	-4.66	-6.32	-6.45	-8.08	-5.23	-4.07
Net_t	-0.00674	0.030131	0.034338	0.047451	0.09994	-0.00201	0.009372	0.012841	0.013113	0.009376
	-1.67	2.66	2.38	3.32	5.45	-1.1	2.05	2.09	1.93	1.22
$Net_t \times Log(B/M)_{t-1}$	-0.00078	0.00654	0.00458	0.00673	0.00514	-0.001	0.000024	-0.00691	-0.00232	-0.00244
	-0.54	1.46	0.77	0.9	0.59	-1.81	0.01	-2.16	-0.5	-0.46
$Net_t \times y_{t-1}$	0.444	0.219	0.039	-0.013	-0.387	0.483	0.467	0.010	0.035	0.073
	6.57	2.75	0.42	-0.1	-2.19	5.89	5.36	0.1	0.28	0.33
$Net_t \times Flow_t^+$	0.025	-0.060	0.716	0.814	0.203	0.0002	0.038	0.174	0.296	0.090
	0.46	-0.41	3.94	4.17	0.88	0.01	0.81	2.61	4.71	0.91
$Net_t \times EFlow_t^+$	-0.108	-0.090	0.184	0.020	0.268	0.033	-0.018	0.101	0.006	0.059
	-4.14	-1.28	1.86	0.16	1.69	2.48	-0.53	2.18	0.12	0.95
$Net_t \times Flow_t^-$	0.147	-0.085	0.241	0.387	-0.091	-0.011	-0.066	0.120	0.138	0.014
	2.18	-0.55	1.4	1.98	-0.41	-0.54	-1.31	1.98	2.05	0.2
$Net_t \times EFlow_t^-$	-0.135	0.064	-0.333	-0.719	-0.125	0.082	0.114	0.108	0.208	0.222
	-1.9	0.37	-1.76	-4.09	-0.6	3.75	2.11	1.47	2.49	2.32
$Net_t \times Ret_{t-1}^+$	0.009	-0.003	-0.046	-0.077	-0.130	0.007	-0.010	-0.021	-0.037	-0.031
	1.65	-0.15	-1.68	-2.61	-3.3	4.33	-0.85	-1.24	-1.46	-1.38
$Net_t \times Ret_{t-1}^-$	0.012	0.013	-0.008	0.097	0.031	0.016	-0.006	-0.014	-0.011	0.042
	1.28	0.41	-0.19	1.95	0.55	2.92	-0.33	-0.55	-0.4	1.13
$Net_t \times Volatility_{t-1}$	0.0004	0.005	0.214	-0.624	0.331	-0.032	0.060	-0.154	-0.223	-0.046
	0.01	0.02	0.67	-2.01	0.91	-1.34	0.45	-0.76	-1.12	-0.18
# of Stocks	8195	3130	2109	1325	722	8195	3130	2109	1325	722
# of Obs	147,757	37,081	25,105	20,180	17,441	147,757	37,081	25,105	20,180	17,441
$Adj-R^2$	0.22	0.27	0.29	0.32	0.40	0.22	0.23	0.27	0.28	0.30

Table reports the regression results for Growth and Aggressive Growth funds from the equation

$$\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t-1} + \gamma_4 Log(B/M)_{i,t-1} + Net_{t,i} \times [\beta_1 + \beta_2 Log(B/M)_{i,t-1} + \beta_3 y_{i,t-1} + \beta_4 Flow^+ + \beta_5 EFlow^+ + \beta_6 Flow^- + \beta_7 EFlow^- + \beta_8 RET^+_{i,t-1} + \beta_9 RET^-_{i,t-1} + \beta_{10} Volatility_{i,t-1}] + \epsilon_{i,t} \ .$$

t statistics calculated from clustered standard errors are reported below the coefficients. Quarterly time fixed effects are included in all specifications. The stocks are divided into 5 size categories by the market capitalizations based on the NYSE stocks cutoffs. Sample runs from 1993 to 2006.

Tab. 2.7: Trade Size Regression

Panel A

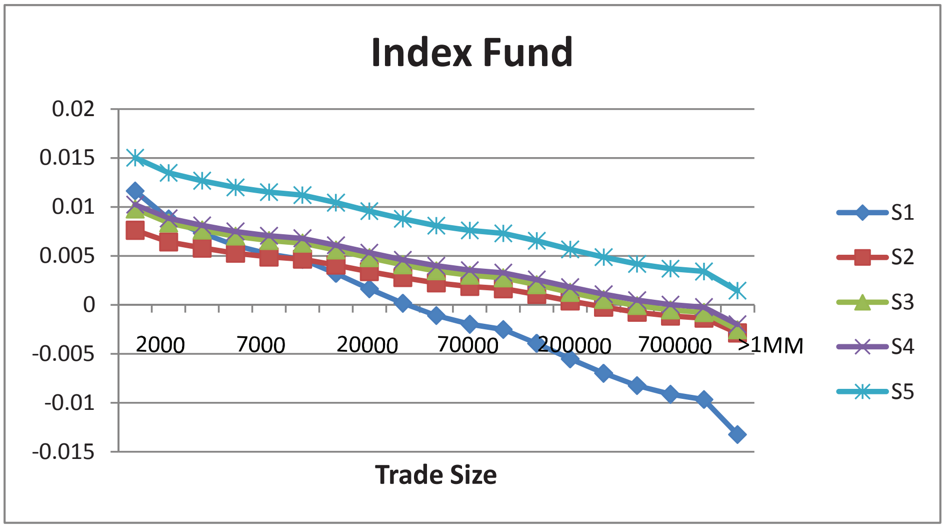
	Index Fund					GNI Fund				
	Small	2	3	4	Large	Small	2	3	4	Large
Intercept	-0.005	-0.009	-0.007	-0.004	-0.005	-0.001	-0.003	-0.005	-0.007	-0.012
	-48.48	-31.34	-21.54	-17.05	-15.58	-9.71	-10.71	-9.74	-12.22	-18.01
y_{t-1}	-0.105	-0.123	-0.103	-0.069	-0.145	-0.104	-0.102	-0.078	-0.054	-0.038
	-38.02	-16.11	-11.01	-8.76	-9.79	-18.68	-16.28	-17.24	-14.78	-11.18
Δy_{t-1}	-0.593	-0.517	-0.581	-0.711	-0.626	-0.331	-0.268	-0.281	-0.291	-0.260
	-129.14	-51.41	-41.94	-56.1	-35.66	-29.6	-18.41	-19.77	-24.75	-21.25
Ret_t	0.0002	0.00004	-0.0004	0.00002	-0.00003	-0.0002	-0.0003	-0.001	-0.003	-0.005
	3.93	0.29	-2.52	0.13	-0.15	-0.56	-1.87	-5.16	-6.79	-8.78
Ret_{t-1}	-0.0004	-0.0007	-0.0011	-0.0002	0.0005	0.0001	-0.0002	-0.001	-0.002	-0.005
	-7.3	-5.42	-8.21	-1.35	2.67	2.52	-1.9	-2.31	-3.42	-8.32
$Volatility_{t-1}$	-0.011	-0.006	-0.00006	-0.003	-0.005	-0.004	-0.016	-0.024	-0.029	-0.027
	-20.79	-2.82	-0.02	-1.13	-2.02	-12.20	-6.99	-5.21	-4.82	-2.91
$Log(B/M)_{t-1}$	0.0002	0.0005	0.0003	0.0001	0.00003	0.00002	-0.00002	0.0002	0.0002	0.0002
	12.38	12.07	6.13	1.41	0.70	1.61	-0.35	2.19	1.77	1.35
Net_t	0.033	0.017	0.020	0.021	0.027	0.000	0.001	-0.006	-0.013	-0.107
	12.74	3.29	3.15	2.2	2.53	0.28	0.13	-0.53	-0.64	-3.54
Net $Slope_t$	-0.003	-0.001	-0.002	-0.002	-0.002	-0.0002	0.0002	0.001	0.002	0.010
	-13.08	-3.14	-2.90	-1.99	-2.01	-0.11	0.33	1.21	1.37	4.13
# of Stocks	8195	3130	2109	1325	722	8195	3130	2109	1325	722
# of Obs	147,757	37,081	25,105	20,180	17,441	147,757	37,081	25,105	20,180	17,441
$Adj-R^2$	0.68	0.80	0.83	0.82	0.94	0.26	0.23	0.22	0.26	0.42

Panel B

	Growth Fund					Agg. Growth Fund				
Intercept	-0.013	-0.031	-0.034	-0.031	-0.024	-0.004	-0.011	-0.011	-0.010	-0.004
	-30.23	-31.15	-28.95	-25.02	-21.81	-20.39	-20.81	-19.12	-16.19	-11.11
y_{t-1}	-0.056	-0.064	-0.060	-0.060	-0.071	-0.116	-0.104	-0.110	-0.111	-0.129
	-26.12	-20.96	-16.05	-15.98	-13.59	-32.92	-24.07	-24.58	-18.61	-16.01
Δy_{t-1}	-0.313	-0.288	-0.281	-0.305	-0.346	-0.261	-0.237	-0.253	-0.273	-0.291
	-62.43	-34.82	-29.04	-26.14	-21.2	-37.28	-23.84	-23.5	-20.41	-16.19
Ret_t	0.001	0.004	0.007	0.010	0.010	0.001	0.006	0.008	0.009	0.007
	9.39	7.46	8.97	10.07	9.1	14.9	16.85	15.14	14.71	10.48
Ret_{t-1}	0.002	0.005	0.007	0.009	0.009	0.001	0.006	0.009	0.009	0.006
	14.95	9.19	9.5	9.52	8.23	14.87	18.44	15.52	12.53	7.99
$Volatility_{t-1}$	-0.038	-0.065	-0.022	-0.015	0.003	-0.010	0.007	0.009	0.029	0.009
	-23.55	-7.17	-1.29	-1.06	0.35	-14.6	1.34	1.08	3.37	1.71
$Log(B/M)_{t-1}$	-0.0002	-0.001	-0.001	-0.001	-0.001	-0.0002	-0.001	-0.001	-0.001	-0.001
	-3.52	-5.65	-6.64	-6.06	-5.08	-5.83	-6.8	-8.98	-6.53	-5.48
Net_t	-0.007	0.074	-0.097	-0.220	-0.236	-0.006	-0.041	-0.089	-0.115	-0.131
	-0.90	3.16	-3.01	-4.87	-3.93	-1.27	-2.66	-4.4	-4.32	-3.16
Net $Slope_t$	0.002	-0.003	0.013	0.022	0.024	0.001	0.005	0.009	0.011	0.012
	2.37	-1.46	4.75	5.91	5.00	2.3	4.21	5.38	5.03	3.62
# of Stocks	8195	3130	2109	1325	722	8195	3130	2109	1325	722
# of Obs	147,757	37,081	25,105	20,180	17,441	147,757	37,081	25,105	20,180	17,441
$Adj-R^2$	0.21	0.27	0.29	0.32	0.40	0.22	0.23	0.27	0.28	0.30

Table reports the regression results for 5 different stock size categories from the Equation 2.5. Net $Slope_t$ is calculated as $\sum_{j=1}^{19} Log(Z_j) \times Net^j_{i,t}$. Z_j is the mid point of trade size bin. Stock size cut-off points are calculated from NYSE market capitalization cut-off points. t statistics calculated from stock level clustered standard errors are reported below the coefficients. Quarterly time fixed effects are included in all specifications. Sample runs from 1993 to 2006.

Fig. 2.1: Trade Size Liquidity Sensitivity Coefficients I



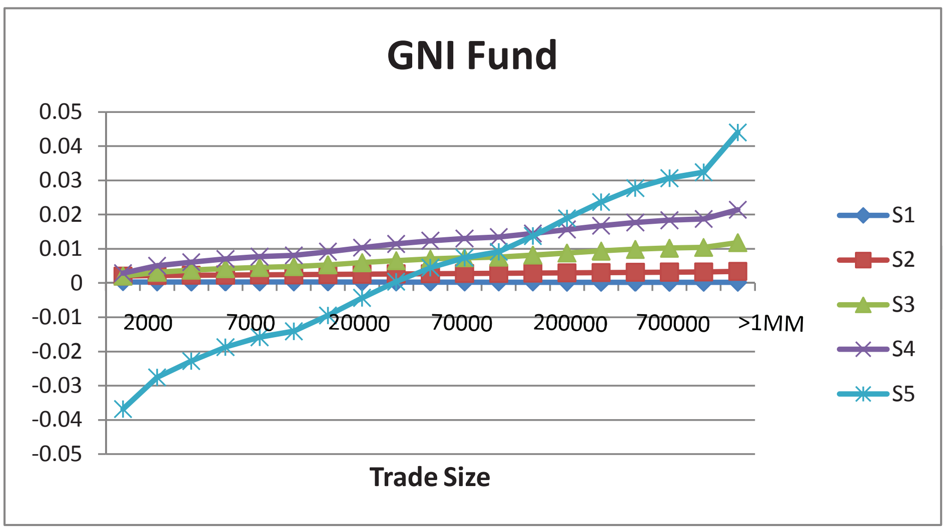
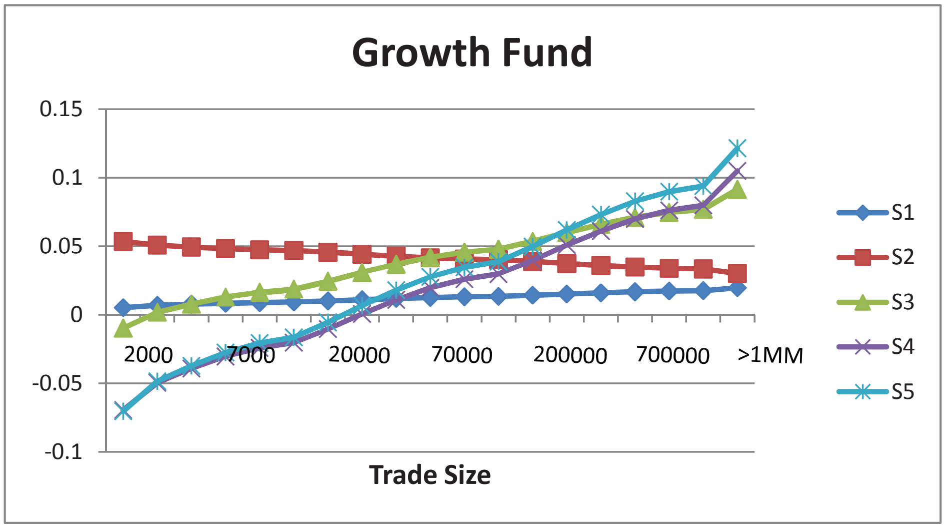


Figure plots the liquidity sensitivity coefficients (β_j) for Index and GNI funds from the equation

$$\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t-1}$$
$$+ \gamma_4 Log(B/M)_{i,t-1} + \sum_{j=1}^{19} [\beta_j \times Net_{i,t}^j] + \epsilon_{i,t}$$

where $\beta_j = b_{01} + b_{02} \times Log(Z_j)$ and Z_j is the midpoint of the trade size bin. Quarterly time fixed effects are included in all specifications. The stocks are divided into 5 size categories by the market capitalizations based on the NYSE stocks cutoffs. Sample runs from 1993 to 2006.

Fig. 2.2: Trade Size Liquidity Sensitivity Coefficients II



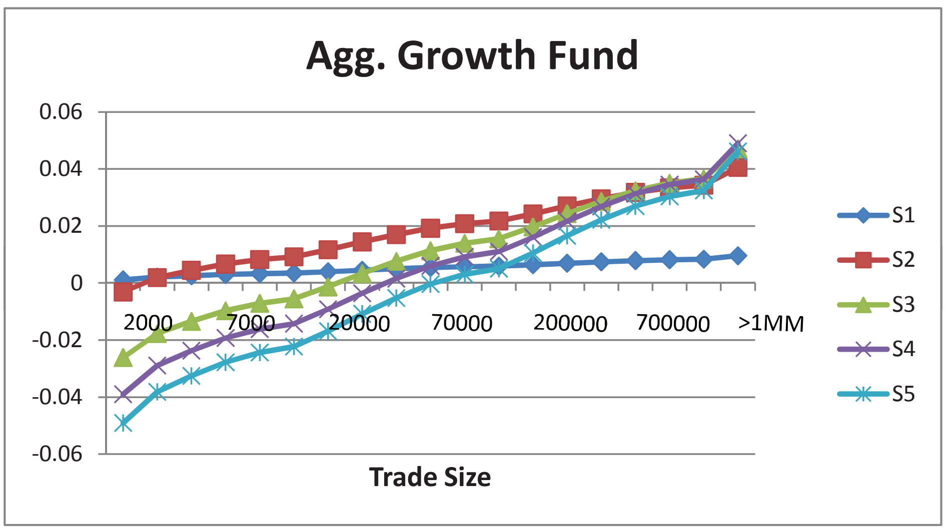


Figure plots the liquidity sensitivity coefficients (β_j) for Index and GNI funds from the equation

$$\Delta y_{i,t} = \alpha + \rho y_{i,t-1} + \phi \Delta y_{i,t-1} + \gamma_1 Ret_{i,t-1} + \gamma_2 Ret_{i,t} + \gamma_3 Volatility_{i,t-1}$$
$$+ \gamma_4 Log(B/M)_{i,t-1} + \sum_{j=1}^{19} [\beta_j \times Net^j_{i,t}] + \epsilon_{i,t}$$

where $\beta_j = b_{01} + b_{02} \times Log(Z_j)$ and Z_j is the midpoint of the trade size bin. Quarterly time fixed effects are included in all specifications. The stocks are divided into 5 size categories by the market capitalizations based on the NYSE stocks cutoffs. Sample runs from 1993 to 2006.

3. TRADING GAME INVARIANCE IN THE TAQ DATASET

3.1 Introduction

Even a quick look at financial markets reveals a significant variation in how securities are traded. During a month, hundred thousands of large orders are executed for some securities, but only a few small orders arrive to the marketplace for others. Kyle and Obizhaeva (2011) develop the hypothesis of market microstructure invariance which describes how the distribution of trade sizes and trade arrival rates should vary across stocks with different levels of trading activity. We examine the predictions of this hypothesis using the Trade and Quote (TAQ) dataset for the period from 1993 to 2008.

The main idea of the trading game invariance is that trading games played in different securities are fundamentally the same. The only difference between these games is the time horizon, called a "trading day", over which trading games are played. Trading days are short for actively traded stocks, perhaps corresponding to a few minutes; and trading days are long for inactively traded stocks, perhaps corresponding to a few months. This invariance hypothesis generates testable predictions concerning how the arrival rate of trades and the distribution of their sizes vary with trading activity across stocks and across time for the same stock: Market microstructure invariance hypothesizes that a one percent increase in trading activity can be decomposed into an increase in the arrival rate of trades by two-thirds of one percent and an increase in trade size of one-third of one percent. Thus, trade size as a fraction of average daily volume decreases by two-thirds of one percent. Market microstructure invariance conjectures that the shape of the distribution of trade size adjusted by trading activity is invariant across stocks and time. "Trading activity" is defined as the product of average dollar volume and daily standard deviation of returns ("volatility"). The product of dollar volume and volatility better capture the risk transferring taken place in the market than dollar volume.

The TAQ dataset contains tick-by-tick data recorded on the consolidated tape for all listed stocks from 1993. For each stock and each month, we calculate the number of trades per month and construct an empirical distributions of trade sizes (in shares). We describe the distribution of trade sizes by a list of attributes, including the means trade size, ten equally-spaced "trade-weighted" percentiles of trade size based on the distribution of trade sizes, and ten equally-spaced "volume-weighted" percentiles based on their contribution of trade size deciles to total trading volume. The "volume-weighted" deciles put more weight on larger and economically more significant trades than trade-weighted deciles, thus allowing us to examine the right tail of distributions of trade size in more detail. Our goal is to study whether the cross-sectional variation in the constructed variables is consistent with predictions of the invariance theory.

Our tests show that microstructure invariance explains the cross-sectional variation in the arrival rates of trades. We regress the logarithm of the number of trades on the logarithm of trading activity and find the coefficient equal to 0.74 for the entire sample from 1993 to 2008, 0.69 for the subsample from 1993 to 2001, and 0.79 for the subsample from 2001 to 2008 with standard errors smaller than 0.015. These estimate of 0.69 for the panel 1993-2001 is remarkably close to 2/3, the coefficient value predicted by the invariance theory. The higher coefficient of 0.79 after 2001 appear to be due to a structural break related to a reduction in tick size, which occurred in 2001 and unevenly affected small and large stocks. For example, the larger than predicted coefficient of 0.79 is consistent with the interpretation that order shredding has occurred more intensely for actively traded stocks than inactively traded ones. Our results also clearly reject alternative hypotheses of invariant bet size and invariant bet frequency. According to these models, the estimated coefficient should be close to either zero or one.

Our tests show that trading game invariance also provides a good explanation for the cross-sectional variation in the distribution of trade sizes. We regress the logarithm of the constructed attributes of the distribution of trade size (as a fraction of trading volume) on the logarithm of the trading activity. When we use the mean of distribution on the left-hand side, our estimates are equal to -0.74 before 2001 and -0.79 after 2001. When we use the trade-based percentiles, the coefficients range from -0.74 to -0.81. For the volume-weighted percentiles, the coefficients range from -0.51 to -0.80. These estimates are closer to the value of $-2/3$ predicted by invariance theory than to the coefficients of 0 and -1 predicted by alternative models.

To further investigate the share of the distribution of trade sizes, we examine empirical distributions of trade sizes graphically. According to the invariance theory, the distribution of trade sizes for any stock can be transformed, by a proper adjustment for differences in trading activity (i.e. multiplying by trading activity in two-third power), into a single distribution, common across stocks and across time. Can we observe this prediction in the data? For ten volume and four price volatility groups, we plot the distributions of the logarithms of the trades sizes, normalized as suggested by invariance theory. We find that these distributions are indeed quite stable across all subgroups. Furthermore, they closely resemble the bell-shaped density function of a normal random variable, suggesting that the normalized trade sizes are distributed as a log-normal random variable. Note that if we adjust trade sizes according to alternative models, then the means of the resulting distributions are much less stable across volume groups. This is further evidence in favor of the invariance theory.

Our finding that the distribution of trade sizes looks similar to log-normal has important economic implications.The variance of the log-normal is so large that a one standard deviation increase in trade size is approximating a factor of five. It suggests that the order flow is dominated by really large orders and that small orders are much smaller than large orders.

Our plots reveal, however, some systematic differences among empirical distributions for stocks with different levels of trading volume and volatility. Statistical tests reject the hypothesis that trade sizes are distributed as a log-normal random variable. We conjecture that various market frictions, such as a minimum tick size and a minimum lot size round of 100 shares, significantly distort trading patterns in a manner that leads to deviation from the invariance theory.

Market frictions are changing over time. Hendershott et al. (2010) and Chordia et al. (2009) discuss the recent transformation of trading activity after the decimalization of 2001 and a consequent spread of algorithmic trading. With the goal to examine these issues more closely, we plot distributions of normalized trade sizes for stocks sorted into ten volume and four price volatility in years 1993, 2001, and 2008.

Our distributions look similar to bell-shaped distributions being truncated from below at the 100-share threshold. The effect of the minimum lot size is especially pronounced after the reduction in tick size in 2001. For example, 100-share trades accounted for 16% of all trades executed before 2001 and 50% of trades executed after 2001. This number reaches 70% in 2008. Effectively, many orders, even the largest ones, are now being shredded into sequences of 100-share trades. This practice is facilitated by the fast electronic interfaces and algorithmic trading. came along with the introduction of electronic interface and algorithmic trading. Since the minimum lot size restriction and relative tick size certainly influence order shredding algorithms, they have to be taken into account when one tests the invariance theory. Invariance theory makes predictions about intended orders rather than actual TAQ "prints" generated by order shredding algorithms. Prior to 2001, some traders used electronic interfaces to submit orders, but this practice became much more pervasive after the tick size was cut in 2001. Furthermore, cutting the tick size provides added incentives to shred orders by placing scaled limit orders at different price points separated by only one cent. Since order shredding after 2001 is likely to affect active and inactive stock differently, with perhaps more order shredding for more active stocks, we are not surprised that empirical tests after 2001 show results that appear to deviate from invariance theory.

A minimum tick size of one penny and a minimum round-lot trade size of 100 shares have different effects on trade sizes observed in TAQ data. Higher returns volatility and higher share price make the penny tick size a less binding constraint. Therefore, we expect to see more evidence of order shredding in more volatile stocks and higher priced stocks. This should show up as a lower mean and median order size than predicted by invariance trades, with larger trades replaced by small trades equal to or greater in size than one hundred shares. Higher returns volatility and higher share price make the one hundred share minimum trade size more binding. We therefore expect to see more evidence of missing odd lots or trades rounded up to one hundred shares in the data. In fact, we see both of these effects.

A number of other market frictions can complicate testing the invariance theory using the TAQ dataset. Trade size tend to cluster at some "even" quantity levels (e.g., Alexander and Peterson (2007)). For example, there are more trades of 5000 shares than 4000 or 6000 shares, and far more than 4900 or 5100 shares. Some of this clustering is a result of various regulations. For example, from 1988 to 2001, Nasdaq market makers were required to fix the minimum quotation size at the level of 1,000-share for Nasdaq-listed stocks during the 1988-2001 period. This restriction is clearly reflected in our plots as a disproportionate number of 1,000-share trades for Nasdaq-listed during that period. These types of market frictions are certainly not captured by the invariance theory. An interesting topic for the further research is to design better econometric tests that will account for various market frictions.

Our paper adds to the results in Kyle and Obizhaeva (2011) that documented strong evidence for the invariance hypothesis using a sample of portfolio transition trades. Portfolio transition trades represent a special subset of market transactions with unique properties that make them especially valuable for testing invariance hypothesis, since the data makes it possible to observe actual orders intended for execution rather then sequences of executed trades where sizes may have been adjusted after the orders were placed. In contrast, our tests are based on a broader sample of trading data. Its broad coverage comes, however, at the expense of having to deal with data distorted by order shredding and ex post adjustments to order size.

There has been a literature on what determines trading frequency and trade sizes. Glosten and Harris (1988) found that average trade size (in shares) is negatively related to market depth. Brennan and Subrahmanyam (1998) documented that trade sizes (in dollars) are also related to other stock characteristics such as return volatility, the standard deviation of trading volume, market capitalization, the number of analysts following stocks, the number of institutional investors holding stocks, and the proportion of shares hold by them. Interestingly, the R-squared in their cross-sectional regressions is about 0.92, which is very similar to the R-squared in our regressions even when we restrict the power coefficient to be equal to 2/3, as implied by the invariance theory, leaving only a constant term to estimate. The similarity of R-squared indicates that the inclusion of additional explanatory variables does not help much in explaining the cross-sectional variations of the average trade sizes. The invariance theory explains a large percentage of the cross-sectional variation in trade size.

The remainder of this paper states the implications of trading game invariance, discusses the issues arising when these implications are tested using the Trades and Quotes dataset, and then describes results of empirical tests.

3.2 Testable Implications of Trading Game Invariance

In Kyle and Obizhaeva (2011), traders are thought as playing trading games. They arrive to the market and execute orders. Innovations in their order flow follow a compound Poisson process with arrival rate of γ innovations per day. A typical innovation in this order flow, called a "bet," is a random variable $\tilde{Q}$ with a zero mean. A positive value of $\tilde{Q}$ represents buying, and a negative value of $\tilde{Q}$ represents selling.

The invariance theory is formulated using the concept of "bet size" to measure risk transferred by a bet. "Bet size" is a random variable defined as the product of dollar bet value (dollar share price P times share quantity $\tilde{Q}$) and volatility σ (percentage standard deviation of returns per day),

$$\tilde{B} = \tilde{Q} \cdot P \cdot \sigma. \tag{3.1}$$

In a similar spirit, "trading activity" is defined as the product of daily volume V, share price P, and daily volatility σ,

$$W = V \cdot P \cdot \sigma = \frac{\zeta}{2} \cdot \gamma \cdot E|\tilde{B}|. \tag{3.2}$$

According to this measure of trading activity, active stocks are stocks with high volatility and high dollar trading volume per calendar day. Inactive stocks are stocks with low volatility and low dollar trading volume per calendar day. The last equality follows from the definition of expected trading volume V over a day as the product of expected arrival rate of bets and absolute value of bet size,

$$V = \frac{\zeta}{2} \cdot \gamma \cdot E|\tilde{Q}|, \tag{3.3}$$

The parameter ζ is a "volume multiplier," which shows how much of additional non-bet volume is generated endogenously as a response to bets. Non-bet volume includes trading by market makers, high frequency traders, and other arbitragers who intermediate among long-term bets.

Invariance theory describes how market microstructure characteristics like bet size and bet arrival rate vary with different levels of trading activity. Its main idea is that trading games are the same across stocks, up to some Modigliani-Miller transformation, except for the speed with which these games are being played. This speed is related to the level of trading activity: Trading games are played faster in active stocks and slower in inactive stocks. This claim is equivalent to the existing of a market microstructure invariant which can be represented as a random variable $\tilde{I}$, with the same distribution across stocks and across time. The random variable $\tilde{I}$ has a distribution defined by

$$\tilde{I} \approx \frac{\tilde{B}}{\gamma^{1/2}}, \tag{3.4}$$

which can be interpreted as the signed standard deviation of gains or losses on a bet in units of time rescaled so that one bet is expected to arrive per tick on the rescaled clock. Trading game invariance generates the following testable predictions concerning how bet arrival rate and bet size vary with trading activity:

$$\gamma = E\Big[\frac{\zeta}{2} \cdot |\tilde{I}|\Big]^{-2/3} \cdot W^{2/3}, \tag{3.5}$$

$$\frac{|\tilde{Q}|}{V} \approx E\Big[\frac{\zeta}{2} \cdot |\tilde{I}|\Big]^{-1/3} \cdot W^{-2/3} \cdot \tilde{I}. \tag{3.6}$$

Under the identifying assumption that ζ is constant across stocks and across time, equation (3.5) and equation (3.6) imply that when the number of bets γ and bet size $\tilde{Q}$ are normalized for differences in trading activity, the quantity $\gamma \cdot W^{-2/3}$ and the distribution of $W^{2/3}$, are invariant across stocks and across time.

This model says that changes in daily trading activity come from both changes in unsigned bet size and changes in bet frequency. In particular, the invariance of trading games requires that a one percent increase in trading activity W is associated with an increase of 2/3 of one percent in bet arrival frequency γ and an upward shift by 1/3 of one percent of the entire distribution of bet size $\tilde{B}$. The latter implication is equivalent to saying that distributions of (unsigned) trade sizes $|\tilde{Q}|$ as a fraction of trading volume V shifts downwards by 2/3 of one percent when trading activity increases by one percent.

Alternative Models. We consider two alternative models considered by Kyle and Obizhaeva (2011): the Model of Invariant Bet Frequency and the Model of Invariant bet Size. The model of Invariant Bet Frequency assumes that the variation in trading activity comes entirely from variation in bet sizes $\tilde{B}$, while the number of bets γ over a calendar day remai ns invariant across stocks. This model generates testable predictions concerning how (unsigned) bet size should varies with trading activity. Both trading frequency and the entire distribution of trade sizes,

$$\gamma \cdot W^0 \quad \text{and} \quad \frac{|\tilde{Q}|}{V} \cdot W^0,$$

are invariant across stocks and across time, even if no adjustments for differences in trading activity W are made (represented by the exponent of zero).

The model of Invariant Bet Size assumes that the variation in trading activity comes entirely from variation in the number of bets γ placed over a calendar day. The distribution of bet size $\tilde{B}$ over a calendar day remains the same across stocks. This model also generates testable predictions concerning how (unsigned) bet size should varies with trading activity. The trading frequency and the entire distribution of trade sizes, normalized for differences in trading activity,

$$\gamma \cdot W^{-1} \quad \text{and} \quad \frac{|\tilde{Q}|}{V} \cdot W^1,$$

are invariant across stocks and across time.

All three models imply specific relationships between the number of bets γ

and the distribution of bet sizes $\tilde{B}$ per calendar day on one side and the measure of trading activity W on the other side. The only difference between these predictions is the exponent of trading activity. We discuss next how to test which of three models describes best the transactions data in the TAQ database.

Testing Theories Using the TAQ Dataset. The TAQ database contains a time-stamped record of trades printed for NYSE and NASDAQ stocks. We can thus estimate trade frequency γ looking at the average number of prints per a calendar day and we can estimate the distribution of trade size $|\tilde{Q}|$ examining the empirical distribution of (unsigned) print sizes in the TAQ database. Equipped with this data, we can test the predictions of the invariance theory and alternatives. There are, however, several important problems that may present obstacles for our tests.

The assumption that the inventory of traders follows a compound Poisson process implies that their trades are independently distributed. In actual trading, one independent trading decision often generates multiple reports of order executions, since orders may be broken down into smaller pieces for execution and executed against several different counter-parties and at several different prices. The TAQ database gives a time-stamped records of trades as printed for NYSE and NASDAQ stocks. It allow us to observe neither independent intended orders nor the number of prints into which initial order has been split. At the same time, order shredding algorithms may vary across stocks in a complex manner, for example, as a function of stock price (based on tick size) or the 100 share minimum lot size constraint.

Dealing with various market frictions is another issue in designing good empirical tests, because market frictions most likely affect trading patterns. For example, there have been various restrictions on trade sizes during the period from 1993 to 2008. These restrictions certainly affect observed trade sizes and trading frequency. If trades cannot be smaller than 100 shares, then intended orders with less than 100 shares will be either not submitted at all or rounded up to satisfy a minimum lot size requirement. This restriction is expected to be especially binding for low-volume stocks with small orders traded at high price levels. Another binding restriction could be the fixing of market-maker minimum quotation sizes at the level of 1,000 shares at Nasdaq from 1988 to 2001. As we will see, this restriction is consistent with the clustering of Nasdaq trades at a level of 1,000 shares seen in the data, especially for Nasdaq stocks. Another important market friction is the tick size, or a minimum increment by which price can move. When tick size as a fraction of price volatility is large, traders tend to submit larger orders since fewer price levels are available. The minimum tick size was reduced from 1/8 of a dollar (12.5 cents) to 1/16 of a dollar (6.25 cents) in 1997 and to one cent in 2001.

In this study, we provide evidence on how these frictions influence our results. Exploring these frictions in more detail is an interesting issue for future research.

3.3 Data

3.3.1 Data Description

The NYSE TAQ database contains trade and quotes reported on the consolidated tape by each CTA participants for all stocks listed on exchanges starting from year 1993. Since we examine the distribution of unsigned trades, our analysis employs exclusively data on trades. We leave the interesting issues about signed trades, which may be reconstructed if we use both trades and quotes, for future research. For each trade, the TAQ data records the time, exchange, ticker symbol, number of shares traded, execution price, trade condition, and other parameters. This database is of a large size. Its subset from 1993 to 2008 contains over 19 billion records with over 5 million records per month in 1993 and over 500 million records per month in 2008.

We transform the raw TAQ data into another dataset, convenient for our subsequent analysis. We first remove bad records from the trades data using standard filters. The TAQ database provides information about the quality of recorded trades in their condition and correction codes. We eliminate trades with a condition codes of 8, 9, A, C, D, G, L, N, O, R, X, Z or with correction codes greater than 1. The correction code of 8 indicates, for example, that the trade was canceled.

The remaining trades are aggregated in a specific way. These trades are placed into 55 bins that we construct based on the number of shares traded. "Even" bins correspond to the orders of even sizes, i.e., trades with the following numbers of shares traded: 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 10000, 15000, 20000, 25000, 30000, 40000, 50000, 60000, 70000, 75000, 80000, 90000, 100000, 200000, 300000, 400000, 500000. "Odd" bins correspond to the trades with odd sizes, i.e., when the number of shares is in-between even bins. We chose these size bins so that their size grows approximately at a log-rate. The selection of these bins also reflects our intention to treat separately trades with even sizes, because these trades are especially frequent in the data. For each day and each symbol, we store the number of trades in each size bin. Once we assign a given trade to an appropriate size bin, we assume its size (in shares) is equal to a midpoint of that bin. If trade size is larger than 500,000 shares, it is assigned to the 55th bin and its size is assumed to be 1,000,000 shares. The aggregation of the data into size bins allows us to capture the main properties of trade size distribution and implement our analysis in a more efficient way. This convenience comes, however, at the expense of introducing additional noise into our analysis, which may affect our results.

For each day and each symbol, we also store other variables such as the open price, the close price, the number of trades per day, the dollar volume per day, the share volume per day, the close-to-close return, and the volatility defined as the standard deviation of returns over the past 20 trading days.

Since many stocks do not have enough transactions per day, to build a good empirical approximation for a theoretical distribution of trade sizes for individual stocks, we therefore aggregate our daily data by month. We sum up the number of trades within each bin and construct empirical distributions of trade sizes (in shares) for each stock and each month in the sample.

The theoretical distributions of trade sizes can be of quite a general form. We define several attributes of these distributions with the intention to capture their shapes. These variables are estimated from the empirically observed distributions. The attributes include the average number of trades per day. We also consider the average trade size and its various percentiles based on trade size distribution. We refer to these percentiles as trade-weighted percentiles. For example, the *xth* trade-based percentile corresponds to a trade size such that trades with sizes above this threshold constitute x% of all trades for a given stock in a given month. Note that trade-based percentiles effectively put the same weight onto trades of different sizes. This approach tends to emphasize small trades. For example, if there are 99 trades executed in 100-share lot and one large trade executed in 100,000-share lot, then the distribution of trade sizes will mostly concentrate at 100-share level. All trade-weighted percentiles below the $99th$ percentile will be equal to 100 shares. The total trading volume, however, is determined by one large trade.

Since large trades are economically more important than small ones, we need to investigate the right tail of trade size distributions in more detail as well. We therefore also consider percentiles based on trades' contributions to total volume. The contribution to the total volume by trades from a given trade size bin is calculated based on its midpoint. We refer to these percentiles as volume-weighted percentiles. The xth volume-based percentile corresponds to a trade size such that trades with sizes above this threshold constitute x% of total trading volume. In the previous example, the $99th$ volume-based percentile will correspond to the 100,000-share trade. It is worth mentioning that the volume-weighted distributions have an easy interpretation. Essentially, their plots show the actual distribution of trading volume across bins with trades of different sizes.

We report our results for distributions of both types because volume-weighted distributions allow us to focus on economically important trades. Of course, if we know a trade-weighted distribution of trade sizes, then we can easily calculate a volume-weighted distribution as well. For example, we will see that the distribution of trade sizes is close to be log-normal. It can be easily shown that if a random variable $\ln(\tilde{z})$ is normally distributed as $N(\mu, \sigma^2)$ and another random variable $\tilde{y}$ has the $\tilde{z}$-weighted density function of $\tilde{z}$, then the logarithm of this random log-normally distributed variable $\ln(\tilde{y})$ is distributed as $N(\mu + \sigma^2, \sigma^2)$. Applying this fact to our situation, we know that if trade size is distributed as a log-normal variable with the density of its logarithm being $N(\mu, \sigma^2)$, then the volume-weighted density is also log-normal with its logarithm being $N(\mu + \sigma^2, \sigma^2)$. Thus, the only difference between these distributions is the shift in the means.

Our monthly data is matched with the CRSP data set for the purpose of acquiring share and exchange codes for stocks in the sample. Only common stocks listed on the NYSE (New York Stock Exchange), AMEX (American Stock Exchange) or NASDAQ from year 1993 through year 2008 are included in our study. Stocks that had splits in a given month are eliminated from the sample in that month. Our data is also augmented with the data on the average daily volume (in dollars and in shares), the average price, and historical volatility for each stock and each month. Our final sample includes 1,107,990 stock-month observations. For each 192 months between 1993 and 2008, there are observations on about 5,800 stocks traded.

3.3.2 Descriptive Statistics

Table 1 provides a description of the data. Panel A reports statistics for the subsample from February 1993 to December 2000. Panel B reports statistics for the subsample from January 2001 to December 2008. We report these statistics separately, because the properties of the data have changed substantially following the decimalization in 2001. Statistics are calculated for all securities in aggregate as well as separately for ten groups of stocks sorted by average daily dollar volume. Instead of dividing the securities into ten deciles with the same number of securities, volume break points are set at the 30^{th}, 50^{th}, 60^{th}, 70^{th}, 75^{th}, 80^{th}, 85^{th}, 90^{th} and 95^{th} percentiles of trading volume for the universe of stocks listed in NYSE with CRSP share codes of 10 and 11. Group 1 contains stocks in the bottom 30^{th} percentile by dollar trading volume. Group 10 contains stocks in the top 5^{th} percentile. It approximately corresponds to the universe of S&P100 stocks. Smaller percentiles for the more active stocks make it possible to focus on the stocks which are economically the most important. For each month, the thresholds are recalculated and stocks are reshuffled across groups.

Panel A of Table 3.1 reports the statistical properties of trades and securities before 2001. For the entire sample of stocks, the average trading volume is \$6.28 million per day, ranging from \$0.14 million for the lowest volume deciles to \$181.98 million for the highest volume deciles. The average volatility for the entire sample is equal 4.1% per day. The volatility tends to be higher for smaller stocks. The volatility is 4.6% for the lowest volume decile and 3.3% for the highest volume group. Thus, the measure of trading activity, equal to the product of volume and volatility, increases from 0.14×0.046 to 181.98×0.033, a factor of 913.

The average trade size is equal to \$23,598 before 2001, ranging from \$11,428 for low-volume stocks to \$88,450 for for high-volume stocks, corresponding to a decrease from 8% to 0.05%, if considered as a fraction of daily volume. The median is much lower than the mean. This is consistent with the existence of several large orders that make the distribution of trade sizes positively skewed. The trade-weighted median ranges from \$5,706 for low-volume stocks to \$28,440 for high-volume stocks, corresponding to a decrease from 4% to 0.02% of daily volume from lowest to highest volume group. Note that the invariance theory predicts that the shape of the distribution of trade size as a fraction of volume, $|Q|/V$, should be similar across stocks. The only difference is that its mean is shifted downwards by two-thirds of the increase in a trading activity. Since from lowest to highest decile, trading activity increases by a factor of 913, a back-of-the-envelope calculation suggests that the distributions of trade sizes as a fraction of volume traded for low-volume stocks should be shifted downwards relative to high-volume stocks by a factor of $913^{-2/3} \approx 0.01$. This is roughly similar to the differences in the means and medians between low-volume and high-volume groups.

The average number of trades per day is 143 for the entire sample and monotonically increases from 16 to 2951 over the volume groups. The actual increase in number of trades is a factor of 2951/16=184.43. The invariance theory predicts that average number of trades should increase by two-thirds of the increase in a trading activity, i.e., $913^{2/3} \approx 100$. While this back-of-the envelope calculation suggests that number of trades increases more than our model predicts, potentially reflecting a more intensive order shredding in high-volume groups, further investigation is certainly warranted.

Trading is a subject to various restrictions. The minimum lot size, for example, is equal to 100 shares. Many orders are therefore executed in 100-share lots. These trades correspond to 16% of all trades executed or 2% of volume traded before 2001.

The 100-share trades represent 14% of trades for low-volume stocks and 25% for high-volume stocks. We see that the 100-share restriction is more binding for high-volume stocks. This happens because stocks with high volume usually have high prices, thus making 100-share threshold more significant in dollar terms. Another interesting observations is the large share of 1,000-share trades, especially for low-volume stocks, before 2001. This reflects the requirement for Nasdaq market makers to post quotes for at least 1,000 shares prior to 1997. The data also shows that even lots, corresponding to even-share bins with exact number of shares such as 100, 200, 300 shares etc. traded, account for more than 50% of volume traded and about 80% of trades executed. The prevalence of these trades validates our choice of trade size bins with the even-share trades being placed in separate bins.

Panel B of Table 3.1 reports the statistical properties of trades and securities for our sample after 2001. The difference between the data before and after 2001 is striking. After 2001, the average daily volume is over $19 million which is three times larger than before 2001. The average number of trades is 1761, having increased by a factor of 12, and the average trade size is only $7,945, a decrease by a factor of 3, compared to the earlier sample. A back-of-the-envelope calculation based on the invariance theory suggest that cross-sectional differences in trade frequencies and trade sizes after 2001 can be consistent with this theory as well.

The descriptive statistics show that order shredding and the minimum lot size became very important after 2001. The 100-share trades constitute, on average, 50% of all trades executed and accounts for 18% of volume traded in the latter sample. The migration of trades to smaller trade sizes continues throughout the period from

2001 to 2008. For example, we observe that 100-share trades represent about 70% of trades and account for 35 % of volume in 2008 (unreported). This indicates that order shredding can make it difficult for us to test the invariance theory using the data after 2001.

3.4 Results

All three invariance models make distinctively different predictions concerning the differences in the distributions of trade sizes and their frequencies across stocks. We use the TAQ dataset to determine which of the models is more reasonable in describing the data. We run our tests both based on trading frequencies and on trade sizes.

3.4.1 Tests Based on Trading Frequency

According to the theory of trading game invariance as well as the two alternative models, the number of trades will be constant across stocks if normalized appropriately for differences in trading activity W. Three models differ only in the suggested normalization. The theory of trading game invariance says that the number of trades per day γ has to be normalized with the trading activity W in a power of -2/3,

$$\ln(\gamma \times W^{-2/3}). \tag{3.7}$$

Alternative theories of invariant bet size and bet frequency propose other adjustments, namely

$$\ln(\gamma \times W^0) \quad \text{and} \quad \ln(\gamma \times W^{-1}). \tag{3.8}$$

Figure 3.1 plots the logarithm of average number γ of trades per day, normalized according to three models, against the logarithm of trading activity W, for each stock traded in April 1993, April 2001, and April 2008. Trading activity W is the product of average daily dollar volume and daily volatility for a given stock in a given month. We consider these three years for robustness because, as we have already mentioned, the trading process has changed significantly over the period under consideration. We also consider separately the NYSE-listed and Nasdaq-listed stocks for April 1993, as these two are also significantly different due to differences in market frictions.

If one of these theories correct, the plot of points should fall on a horizontal line. Figure 3.1 clearly shows that the theory of trading game invariance fits the data very well. Especially for the NYSE stocks traded in April 1993, all observations are lined up across a horizontal line. These pattern become slightly less pronounced after 2001.

When the number of trades is normalized according to the theory of invariant bet frequency, all observations are lined up across a line with a positive slope. This fact is easy to explain. This model assumes that differences in the trading activity come entirely from differences in trade sizes. In real data, however, changes in trading activity are partially explained by changes in trading frequency. Thus, this model tends to underestimate the number of trades for high-volume stocks and overestimate it for low-volume stocks.

When the number of trades is normalized according to the theory of invariant bet size, the results are the opposite. All observations are lined up along the line with a negative slope. Again, this fact is easy to explain. This model attributes differences in the trading activity entirely to differences in the trading frequencies. Some fraction of these differences comes, however, from differences in trade sizes. This model therefore tends to overestimate the number of trades for high-volume stocks and underestimate it for low-volume stocks.

The three theoretical models make distinctly different predictions concerning how arrival rates of trades vary with the level of activity. The predictions of the models can be nested into a simple linear regression,

$$\ln\Big[\gamma\Big] = \alpha + a_\gamma \times \ln\Big[\frac{W_i}{W_*}\Big] + \tilde{\epsilon}. \tag{3.9}$$

The equation relates the average number of trades γ per day to the level of trading activity W, defined as the product of average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for an arbitrary benchmark stock with price \$40 per share, trading volume of one million shares per day, and daily volatility of 2% per day. In the regression, the model of trading game invariance predicts $a_\gamma = 2/3$, the model of invariant bet frequency predicts $a_\gamma = 0$ and the model of invariant bet size predicts $a_\gamma = 1$. We run monthly regressions and report the Fama-MacBeth estimates with their Newey-West standard errors computed with 3 lags in Table 3.2. The first three columns of the table report the results for stocks and the entire sample period as well as for the two subsamples, before and after 2001. The next six columns report the results for the NYSE/AMEX-listed stocks and the NASDAQ-listed stocks.

The estimate of a_γ is equal to 0.74 for the entire sample with the standard errors of 0.011. As we see, it is far from to the value of 2/3 predicted by the invariance theory. A statistical test, however, rejects this theory with F-test of 44 and p-value of 0.001, due to very small standard errors. Note that alternative models are rejected with overwhelming margins. The magnitude of the F-tests indicates that if we do a Bayesian analysis, then almost regardless of our priors, we will conclude that the theory of trading game invariance provides a most likely explanation of the cross-sectional differences in trading frequencies than the other two theories.

When we break the sample into two subsamples before and after 2001, we observe that the results are closer to the predictions of the invariance theory in the first part of the sample. The point estimate of α_γ is equal to 0.69 before 2001 and 0.79 after 2001. The increase in coefficients can be most likely attributed to the fact that order shredding is more intensive for high-volume stocks.

Table 3.3 presents estimates of the monthly regressions

$$\ln\Big[\gamma\Big] = \alpha + \frac{2}{3}\ln\Big[\frac{W_i}{W_*}\Big] + b_1 \times \ln\Big[\frac{V_i}{(10^6)}\Big] + b_2 \times \ln\Big[\frac{P_i}{(40)}\Big] + b_3 \times \ln\Big[\frac{\sigma_{r,i}}{(0.02)}\Big] + \tilde{\epsilon}. \quad (3.10)$$

This regression imposes the restriction that the coefficient $a_0 = 2/3$ as predicted by the model of trading game invariance. It then allows the coefficient on the three components of W_i to vary freely. Thus, the model of trading game invariance predicts $b_1 = b_2 = b_3 = 0$. The model of invariant bet frequency predicts $b_1 = b_2 = b_3 = -2/3$, and the model of invariant bet size predicts $b_1 = b_2 = b_3 = 1/3$. Table 3.3 reports the Fama-MacBeth estimates and Newey-West standard errors based on monthly regressions.

The table reports that the estimate of $\hat{b}_1 = 0.18$ for the coefficient on volume V_i, the estimate $\hat{b}_2 = -0.3$ for the coefficient on price P_i, and the estimate of $\hat{b}_3 = -0.41$ for the coefficient on volatility $\sigma_{r,i}$. The corresponding standard errors imply that these estimates are significantly different from zero. Thus, the number of trades increases faster with trading volume and slower with dollar volatility than suggested by the invariance theory. Note that although we have here three explanatory variables, the increase in the R-square relative to the univariate regressions is small, but statistically insignificant.

3.4.2 Tests Based on Trade Size

Figure 3.2 and Figure 3.3 show the trade-weighted and volume-weighted distributions of normalized trade sizes for the NYSE-listed and the NASDAQ-listed stocks traded in April 1993. Our choice of the month April guarantees that trade size distribution figures are not influenced by seasonality in the market, because trade sizes tend to cluster less before the end of calendar quarter, as showed by Moulten (2005). We present the distributions from year 1993 only for illustrative purposes and closely examine data from 1993 to 2008 later.

Trade sizes are normalized as implied by the three models. These models predict that if trade sizes are adjusted appropriately for differences in the trading activity W, then their distributions will be similar across stocks. The models differ only in the adjustment they suggest. The theory of trading game invariance says that trade size $|Q|/V$ should be normalized with the trading activity W raised to the power of 2/3,

$$\ln(\frac{\tilde{|Q|}}{V} \times W^{2/3}). \tag{3.11}$$

Alternative theories of invariant bet size and bet frequency propose other adjustments,

$$\ln(\frac{\tilde{|Q|}}{V} \times W^{1}) \quad \text{and} \quad \ln(\frac{\tilde{|Q|}}{V} \times W^{0}). \tag{3.12}$$

Note that we calculate the trade size $|Q|$ based on the mid-point of its trade size bin. The empirical stock-level distributions of normalized trade sizes are pooled together for April 1993, averaged across stocks in each volume group, and plotted on the figures. The subplots of the trade-weighted distributions have the frequency of normalized trades on the vertical axis. The subplots with the volume-weighted distributions have the volume contribution of these trades on the vertical axis. We also superimpose the normal distribution with the same mean and the same variance equal to the mean and the variance of normalized trade sizes for the entire sample. We superimpose different distributions for trade-based and volume-based distributions. These superimposed distributions make it easy to interpret the results. Indeed, if the theories are correct, then distributions of normalized trade sizes should be identical across volume groups. They should also coincide with the superimposed normal distributions, if trade sizes are distributed as a log-normal random variable.

The three plots in the first column of figure 3.2 reveal that both the trade-weighted and volume-weighted distributions of trades, normalized according to the theory of trading game invariance, seem to be stable across volume groups. The support of these distributions is clearly similar for both low-volume and high-volume stocks. This suggests that the data fit the theory of trading game invariance quite well. These empirical distributions are also quite similar to the superimposed normal distributions, implying that the normalized trade sizes are indeed distributed similarly to a log-normal random variable.

Note that the distributions of trade sizes for the NASDAQ-listed stocks in Figure 3.3 are less smooth than those for the NYSE-listed stocks in Figure 3.2. We will see that this happens because of a particular market regulation existing at NASDAQ in the 90s. NASDAQ dealers were restricted to quote prices for at least 1,000 shares. This led to a disproportionably large fraction of 1,000-share trades at NASDAQ. These 1,000-share trades are responsible for the spikes on the graphs for the NASDAQ-listed stocks.

The second and third columns of figures 3.2 and 3.3 show that the distributions of trade sizes, normalized as suggested by alternative models, are not stable across volume groups. The theory of invariant bet frequency, for example, understates the magnitude of trade sizes for high-volume stocks. This happens because the high volume partially stems from high trading frequency, while this theory assumes that all variation in trading activity comes entirely from the variation in trade sizes.

The theory of invariant bet size, in contrast, overstates the magnitude of trade sizes for high-volume stocks. It assumes that all variations in trading activity come entirely from variations in trading frequency. We know, however, that it partially comes from variations in trade sizes. To summarize, alternative theories provide much worse explanations for observed variations in trade sizes comparing to the invariance theory.

It is also interesting to examine more closely the parameters of the superimposed normal distributions. For the NYSE-listed stocks, the distribution superimposed on the trade-weighted distributions has the mean of -1.01 and the variance of 1.78; the distribution superimposed on the volume-weighted distributions has the mean of 0.97 and the variance of 2.69. As mentioned in the previous section, when if trade sizes are distributed as log-normal random variables, then both the trade-weighted and volume-weighted distributions should be similar to that of normal random variables with the same variances but different means. These means should be such that the mean of volume-weighted distribution is equal to the mean of the trade-weighted distribution plus its variance. So, if normalized trade sizes are indeed distributed as log-normal random variables, then the volume-weighted distributions should be similar to the trade-weighted ones but shifted upwards. The magnitude of this upward shift should be equal to the variance. We can easily check whether these relations hold in our data. The volume-weighted mean of 0.97 is only slightly higher than the trade-weighted mean of -1.01 plus the variance of 1.78. But, the variance of 2.69 for the volume-weighted distributions is much higher than the variance of 1.78 for the trade-weighted distributions. We conclude that although the normalized trade sizes do seem to be distributed similarly to log-normal random variables, there are some deviations. These deviations might be consistent with the existence of some market frictions. For example, the log-normal distribution might be truncated at some level from below. When we examine distributions of normalized trade sizes in more detail, we will see that they are, indeed, truncated from below by a minimum lot size restricted to be 100 shares. Truncated trades are small in size and economically insignificant. This truncation thus will significantly affect the trade-weighted distributions but not the volume-weighted distributions, potentially resulting in a higher variance of the latter.

Next, we proceed to the tests how the data on trade sizes fit the model of trading game invariance using a regression analysis framework.

Table 3.4 shows the estimates based on the data between February 1993 and December 2000 from the following regression,

$$\ln\left[\frac{\tilde{Q}_i}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i, \tag{3.13}$$

where the left-hand side $\ln\left[\frac{\tilde{Q}_i}{V_i}\right]$ is the mean or the *pth* (20th, 50th and 80th) percentiles of the stock-level distributions of the logarithms of trade sizes as a fraction of daily volume. In particular, for each stock in a given month, we construct the empirical distribution of the logarithms of trade sizes over the logarithms of our trade-size bins. We then calculate various attributes of these empirical distributions. The means and the percentiles are calculated both for the trade-weighted distributions and the volume-weighted distributions. We run these regressions for each month and report the Fama-MacBeth estimates with their Newey-West standard errors. In the regression, the model of trading game invariance predicts that $a_Q = -2/3$, the model of invariant bet frequency predicts that $a_Q = 0$, and the model of invariant bet size predicts that $a_Q = -1$.

Table 3.4 shows that the estimates of α_Q range from -0.80 for the 20^{th} percentile to -0.74 for the 80^{th} percentile of trade-weighted distributions. Its estimate based on the mean is equal to -0.76. These coefficients are not too different from the value of -2/3 predicted by the theory of trading game invariance. Although the hypothesis a_Q=-2/3 is rejected, the F-tests suggest that the theory of trading game invariance has a better fit than other two theories for all the percentiles and the means of the trade size distributions. If we apply a Bayesian analysis, we will conclude that, almost regardless of our priors, the theory of trading game invariance is the most probable one.

The estimates of α_Q from the volume-weighted distributions are lower than those from the trade-weighted distributions, ranging between -0.69 to -0.51. We know that the volume-weighted distributions better describe the behavior of the large trades, while the trade-weighted distributions focus more on small trades. It seems that the size of smaller (larger) trades, as a fraction of volume, decreases with trading activity at a slower (faster) rate than the invariance theory predicts.

Although we use only one explanatory variable $\ln W$, the R-square of our regressions is fairly large, ranging between 0.90 and 0.93 for trade-weighted distributions. A similarly high R-squared has been documented in Brennan and Subrahmanyam (1998) in regressions with multiple explanatory variables such as the market depth, returns volatility, the standard deviation of trading volume, market capitalization, the number of analysts following stocks, the number of institutional investors holding stocks, and the proportion of shares hold by them and others. The similarity of R-squares suggests that these additional variables do not add too much additional explanatory power to the regression relative to our measure of trading activity W. Note also that the R-square is much lower for the volume-weighted distributions. One possible explanation for these lower R-squares is that our trade-size bins are much wider for larger trades, for instance, our thresholds are 100, 200, 300, 400, 500 shares for small trades and 100000, 200000, 300000, 400000, 500000 shares for large trades. In our analysis, each trade is assigned with a trade size equal to a mid-point of a corresponding bin. This procedure mechanically adds more noise into the regressions based on the volume-weighted distributions. Also, we expect that only a small number of large trades per month do not allow us to reconstruct a smooth distribution of trades in the right tail.

Table 3.5 reports the estimates from the same regressions as in Table 3.4 but for the sample from January 2001 to December 2008. As before, the estimates of α_Q are close to -2/3 as predicted by the model of trading game invariance. These estimates, however, are somewhat more negative than -2/3, indicating that the trade size, as a fraction of trading volume, decreases with increasing trading activity at a faster rate than the invariance theory suggests.

There is one noticeable difference between our results before and after 2001. After 2001, the estimates for the trade-weighted and volume-weighted percentiles are quite similar to each other. We believe that that the decimalization in 2001 and technological improvements in trading that allowed algorithmic traders to execute a large number of trades over a fairly short period of time, have resulted in a substantial amount of order shredding. Large trades became very infrequent because most of them are shredded now into sequences of 100-share trades. This made trade-weighted and volume-traded distributions more similar to each others after 2001 than before that. Since we expect order shredding to be more significant among high-volume stocks, the realized trade sizes will be lower for high-volume stocks than predicted by the invariance theory. This tilt might be reflected in lower estimates of α_Q after 2001.

Nevertheless, the data on the distributions of trade sizes support the model of trading game invariance and soundly reject assumptions made in alternative models. The reason is that the variation in trading activity in real markets is associated with both variations in trading frequency and trade sizes; neither remains constant as trading activity changes. Our analysis so far also suggests that a number of market frictions might have distorted our results and that the further investigation is warranted.

3.4.3 Market Frictions

Trade sizes and frequencies are certainly affected by various market frictions such as the minimum lot size, the clustering of trade sizes, the relative tick size, and the propensity for order shredding. These frictions might have distorted our results so far. In this section, we focus exclusively on the theory of trading game invariance. Our goal is to better understand how various market frictions affect observed trade sizes and trade frequencies. With this goal in mind, we examine the distributions of the logarithms of the trade sizes normalized according to the invariance theory as $\ln(\frac{Q}{V} \times W^{2/3})$ over various subsets of stocks.

As before, we consider ten volume groups. We also split our sample into four price volatility groups. For our purposes, "price volatility" $P\sigma_r$ is the standard deviation of daily change in prices expressed in dollars. In particular, each month we divide all observations into four equally-sized groups based on the variable $P\sigma_r/W^{1/3}$. This variable is equal to the price volatility $P\sigma_r$ normalized according to the theory of trading game invariance for the difference in trading activity W. Normalized price volatility is an important sorting variable because the effects of at least two market frictions are expected to depend on its levels.

First, the normalized price volatility $P\sigma_r/W^{1/3}$ is inversely related to a concept of the relative tick size. Usually, the relative tick size is defined as the tick size (in cents) divided by the price volatility $P\sigma_r$. According to the invariance theory, however, it is more reasonable to define the relative tick size as the tick size (in cents) divided by the price volatility over a trading day equal to H calendar days, i.e.,

$$\text{Relative Tick Size} := \frac{\text{Tick Size}}{P\sigma_r\sqrt{H}} \sim \left(\frac{P\sigma_r}{W^{1/3}}\right)^{-1}, \tag{3.14}$$

where H is related to a speed with which trading games are being played. According to the invariance theory, it is proportional to $W^{2/3}$. The relative tick size is therefore inversely proportional to normalized price volatility.

This clearly shows that the normalized relative tick size is inversely related to normalized price volatility. When the relative tick size is low, stocks can be traded at finer price levels. Their trade sizes are expected to be small and their trading frequencies are expected to be high. Indeed, as an order walks up the limit order book with limit orders placed at a finer grid, more prints of smaller sizes will be generated. Thus, trade size is expected to decrease and trading frequency is expected to increase with the normalized price volatility.

Second, the normalized price volatility $P\sigma_r/W^{1/3}$ is also related to a concept of minimum lot size of 100 shares. This is easy to see if we calculate the normalized 100-share trade as a fraction of trading volume. Effectively, we can normalize 100-share trade size as suggested by the invariance theory and get,

$$\text{Normalized 100-share Trade} := \frac{100}{V} \times W^{2/3} \sim \left(\frac{P\sigma_r}{W^{1/3}}\right). \tag{3.15}$$

The minimum lot size is proportionally related to the normalized volatility. When the normalized price volatility is high, the 100-share constraint is expected to be more binding. In other words, for high-volatility stocks, the realized trade sizes will be "too high" and realized trading frequency will be "too low", as many small trades will not be even submitted to the system. As we see, the effect of the 100-share threshold is opposite to the effect of the relative ticks size. We see that the 100-share effect is more pronounced in the trade-weighted data.

It is interesting to examine the distributions of normalized trade sizes over different volume and volatility groups as well as over different time periods.

Figure 3.4 shows the trade-weighted distributions of the logarithms of the normalized trade sizes for 10×4 groups sorted by volume and price volatility for the NYSE listed stocks in April 1993. Trade sizes are normalized as suggested by the theory of trading game invariance. To show the composition of trades, we highlight 100-share trades in light grey and 1,000-share trades in dark grey. We also superimpose a normal distribution with the same mean and the same variance, calculated based on the entire sample. If the invariance theory holds and trade sizes are distributed log-normally, then distributions should be identical across stocks and they should coincide with the superimposed normal distribution.

For most subgroups, the distributions are indeed close to the normal one. There is also a clear truncation from below at the 100-share threshold. A visual inspection suggests that holding the price volatility constant, the support of the distributions stays reasonably constant across volume groups. Holding the volume constant, however, the distributions change across volatility groups. When price volatility increases, the 100-share trades, shown in light grey, shift to the right and the number of average trades decreases. This indicated that the 100-share constraint is becoming more binding and small orders are not even being placed into the system. As for the tick size effect, when volatility increases, the relative tick size decreases making trade sizes smaller and trade frequency greater. We do not, however, observe these patterns. Obviously, the 100-share effect dominates the effect of the relative tick size. Note also that 100-share trades are placed close to each other on the charts, usually located in one or two adjacent columns. It means that there is not a lot of variation in the measure of trading activity W within the groups. The only exception is the first volume group, where the variation in W is quite significant and the 100-share trades are spread over more than four columns.

The trade sizes seem to deviate from the log-normal distribution when price volatility is low. In this case, there are too many 100-share trades and trade sizes are lower than suggested by the invariance theory. Otherwise, the distribution of the logarithms of the normalized trade sizes closely fit the truncated normal distribution.

Figure 3.5 is similar to Figure 3.4 but it shows the volume-weighted distributions of the logarithms of the normalized trade sizes for the NYSE-listed stocks in April 1993. The volume-weighted distributions put more weight onto large trades and allow us to examine in more detail the right tail of the distributions. These charts have an intuitive interpretation. They represent the distribution of trading volume across trade size bins.

We see that for most subgroups, the distributions of trade sizes are much closer to the imposed normal distribution for the volume-weighted distribution rather than for the trade-weighted distributions. There is a simple explanation. The distortions related to the minimum lot size have almost no effect on the volume-weighted distributions. Numerous 100-share trades contribute very little to the overall volume traded. The 100-share trades almost disappear from the charts and the 100-share constraint thus becomes invisible.

The volume-weighted distributions, by and large, provide a supportive evidence for the invariance theory, but there are a number of caveats. For low-volume stocks, large trades are somewhat smaller than they should be according to the invariance theory. There are also too few really large orders in the right tail of these distributions.

Figure 3.6 is similar to Figure 3.4 but it shows the trade-weighted distributions of the logarithms of the normalized trade sizes for the NASDAQ-listed stocks in April 1993.

The first thing to notice is a large fraction of 1,000-share trades, shown in dark grey. The clustering of these trades certainly distorts the distribution of trade sizes for NASDAQ-listed stocks, creating spikes and potentially contaminating the results of our analysis. A number of 1,000-share trades is especially significant in high-volume groups. Note that a disproportionably large number of 1,000-share trades is not observed after 2001 (unreported). This distortion most likely corresponds to the restriction on the minimum quotation sizes existing at NASDAQ since mid-1988. The Securities and Exchange Commission made it mandatory for market makers to have the quotation size of at least 1,000 shares. This rule affected mostly large stocks. For small stocks, the rule was slightly different. After 1996, this restriction has been gradually removed, first for a subset of securities and then for all securities traded at NASDAQ. We expect that this regulatory market friction is the reason why our results for the NASDAQ stocks are worse than those for the NYSE stocks before 2001.

Apart of the 1,000-share trades, the empirical distributions of trade sizes are reasonably close to the superimposed normal distribution. This is true especially for low-volume stocks with many stocks placed in those groups. For high-volume groups, the number of stocks decreases significantly and the empirical distributions are not very smooth.

Figure 3.7 shows the trade-weighted distributions of the logarithms of the normalized trade sizes for 10×4 groups sorted by volume and price volatility for stocks traded in April 2001. Figure 3.8 shows the same distributions for stocks traded in April 2008. One striking pattern clearly appears in the figures. Starting 2001, the distributions of trade sizes is becoming dominated by 100-share trades, shown in light grey on the plots. The block-order market seems to be disappearing.

The NYSE implemented a new pricing scheme on January 29, 2001, reducing the tick size from 1/16 to 1/100. The NASDAQ started to use decimal pricing on April 9, 2001. After decimalization and a consequent introduction of electronic interfaces, order shredding has become much more prevalent. Large trades are now broken into numerous small trades. It is not infrequent to see a million-share trade being shredded into a sequence of 100-share trades. In 2008, for example, the 100-share trades constitute about 70% of all trades executed and 35% of the volume traded. Both figures clearly show these changes.

In our sample, the trade size has decreased significantly over time. The distributions of normalized trades were centered around -1.01 and -0.18 for the NYSE and the NASDAQ-listed stocks in April 1993 (Figures 3.4 and 3.6). The log of average trade size as fraction of daily volume decreases from -1.31 in April 2001 to -2.66 in April 2008. The trading frequency, in contrast, has exploded. For high-volume low-volatility stocks, for example, the average number of trades has increased from 938 trades per month in April 1993 to 74,420 trades per month in April 2008.

The extent of order shredding makes it difficult to test the invariance theories using the TAQ dataset after the decimalization of 2001, unless one makes particular assumptions about the order shredding algorithms and how they depend on the trading activity W. Note that the invariance theory is formulated for intended orders or "ideas", independently arriving to the market. The TAQ dataset, however, contains prints. Thus, we effectively make an assumption that ideas generate the same number of prints, regardless of trading activity W. This assumption may not hold after 2001. Orders in actively-traded stocks may be shredded into more traders comparing to orders in actively-traded stocks.

Figure 3.10 shows how the trading process changed over time from a different angle. It shows the dynamics of the normalized number of trades per month and the distribution of normalized trade sizes over time between 1993 and 2008. As before, we normalize number of trades and their size according to the theory of trading game invariance. The figure shows the number of trades as well as the 20^{th}, 50^{th} and 80^{th} percentiles for trade-weighted and volume-weighted distributions. High-volume and low-volume stocks are examined separately.

We see that the normalized numbers of trades have increased and the normalized trade sizes have decreased significantly over time, even after adjustment for the invariance theory. For the high-volume stocks stocks, the normalized trade size distributions were stable until 2001 and then shifted downwards. For the low-volume stocks, the normalized trade sizes are decreasing gradually after 1997. These changes probably represent the impact of the reduction in tick sizes. If we recollect the history, both NYSE and NASDAQ went from 1/16 quotations to the decimal pricing in 2001. Earlier in 1997, NASDAQ had also announced the change from 1/8 to 1/16. Most of NASDAQ-listed stocks are placed into the low-volume group. These earlier changes at NASDAQ might therefore explain the decrease in trade sizes for low-volume group prior to 2001. Interestingly, with the exception of the volume-weighted distributions for large stocks, the effects of changes seem to start stabilizing in 2007 and 2008.

The effect of order shredding is certainly not uniform across stocks. Figure 3.10 shows that the number of trades has increased more significantly for high-volume stocks than for low-volume stocks. The distributions of trade sizes have changed in a different manner as well. For high-volume stocks, the distribution of trade sizes is somewhat tighter after 2001. For low-volume stocks, the distribution is stable and even widen for larger trades, as shown by the volume-weighted distributions. Comparing the trade size percentiles of low-volume and high-volume stocks, we observe that the latter have slightly larger trade sizes in trade-weighted distributions but smaller trade sizes in volume-weighted distributions. Order shredding might influence more significantly trading in high-volume stocks.

Figure 3.10 shows the dynamics of coefficients from monthly regressions for trading frequency (3.9) and for trade sizes (3.13). The coefficients predicted by the theory of trading game invariance are superimposed on the plots.

Panel A shows that the estimates of coefficients from (3.9) are not only close to the predicted $a_\gamma = 2/3$ on average but also for each month between 1993 and 2008. These coefficients are especially close to the predicted value before 2001 and they are slightly higher after 2001. This increase indicates that order shredding, which started to prevail after decimalization in 2001, is more intensive in high-volume stocks, probably because algorithmic trading is concentrated primarily in these stocks.

Panel B and C show the estimates of coefficients from (3.13) for the percentiles and means of the stock-level distributions of trade sizes. Again, we see that each month the estimates are quite close to $a_Q = -2/3$ predicted by the invariance theory. Before 2001, the coefficients tend to be closer to the predicted ones. After 2001, order shredding seems to contaminate the transaction data making the coefficients deviate from the values predicted by the invariance theory. In the recent period, the volume-weighted percentiles start behaving more similar to the trade-weighted percentiles, as the market for block traders has been disappearing.

3.5 Conclusions

We employ the TAQ dataset to test the theory of trading game invariance introduced by Kyle and Obizhaeva (2011). The theory puts forward the idea that financial markets have a particular structure. Securities are traded in such a way that trading games played by traders are the same across stocks. The only difference between these games is the speed with which these games are being played. In other words, the time clock may run at a different pace for different stocks. When trading games, equivalent in a trading time, are considered in a calender time, several specific cross-sectional patterns naturally appear. For example, if one stock has the trading activity that is one percent higher than another stock, then the invariance theory makes one prediction which can be expressed in two equivalent ways: Trade sizes, as a fraction of daily volume, should be two-third of one percent smaller and its number of trades per day should be two-third of one percent larger.

We test these predictions using the data on (unsigned) trades in the TAQ dataset from 1993 to 2008. Our tests based on trading frequencies show that the estimated coefficient 0.69 is remarkable close to the predicted value of 2/3, especially before 2001. After decimalization in 2001 and a consequent spread of algorithmic trading, the coefficient is slightly higher, possibly reflecting a more intensive order shredding in high-volume securities. Our tests based on the distribution of trade sizes also provide evidence in favor of the invariance theory. We find, for example, that the distributions of trade sizes, normalized according to this theory, are quite stable across stocks and that these distributions are similar to a log-normal distribution truncated from below at the level of 100 shares.

Why order sizes are log-normally distributed is an interesting question for the future research. There may be several explanations. Traders may be symmetric and each of them may be drawing quantities to trade from the same log-normal distribution. Alternatively, each trader may have a natural trade size, related to his own size, if for example large hedge funds submit large orders and small retail investors submit small orders. Although the distribution of trades may look normal if conditioning on the type of a trader, the variation in sizes of traders themselves may turn this distribution into a log-normal one.

There are several other issues that require further investigation. The invariance theory is formulated in terms of "bets" or "ideas" arriving to the market independently. We do not observe these ideas in the TAQ dataset; rather we observe realized prints. These prints are influenced by various market frictions such as minimum lot size or relative tick size as well as by order shredding, which became particularly prevailing after recent technological changes in the trading process. The interesting topic for further research, therefore, is how to design better econometric tests dealing with these issues.

So far, the predictions of the invariance theory about the cross-sectional pattern in trade frequencies and trade sizes have found support in the samples of trades from the portfolio transition dataset and the TAQ dataset. It would be interesting to see whether the predictions concerning quantities and frequencies also hold in the datasets containing changes in holdings of mutual funds or other reporting institutional traders as well as transactions from other markets.

Tab. 3.1: Descriptive Statistics.

Volume Groups:	All	1	2	3	4	5	6	7	8	9	10
						Panel A: Before 2001					
Avg. Trade Size, $\lvert\tilde{Q}\rvert$	23,598	11,428	27,348	36,370	43,387	49,167	53,390	60,404	67,649	77,789	88,450
Med. (TW) Trade Size	9,386	5,706	10,864	13,445	15,452	16,946	18,115	20,140	21,944	24,598	28,440
Med. (VW) Trade Size	105,819	48,176	153,785	178,670	193,535	207,665	230,830	242,800	273,980	308,606	337,984
Avg. # of Trades, γ	143	16	72	124	182	251	327	398	536	836	2951
Avg. Daily Volume($1000)	6,286	143	1,176	2,744	4,947	7,883	11,130	15,657	23,744	42,295	181,987
Avg. Volatility	0.041	0.046	0.034	0.033	0.033	0.033	0.033	0.032	0.032	0.033	0.033
Avg. Price	17.64	10.29	20.46	24.53	27.99	31.67	34.82	38.52	43.10	50.12	64.60
100-Shares: %Trades/%Vol	16/2	14/2	17/2	19/2	19/2	20/2	21/2	21/2	21/2	22/2	25/3
1000-Shares: %Trades/%Vol	18/14	18/15	18/13	17/12	16/12	15/11	15/11	14/10	14/10	13/10	13/11
Even Lots: %Trades/%Vol	80/61	79/63	80/60	80/59	80/58	80/57	79/57	79/56	79/55	79/56	82/58
# Obs	636,271	392,812	93,722	37,288	33,309	14,987	14,234	13,074	12,381	11,893	12,571
						Panel B: After 2001					
Avg. Trade Size, $\lvert\tilde{Q}\rvert$	7,945	4,337	8,323	10,961	13,235	15,666	17,794	19,966	23,218	27,049	34,743
Med. (TW) Trade Size	3,170	1,812	3,416	4,447	5,256	6,252	7,027	7,723	8,900	10,025	12,056
Med. (VW) Trade Size	34,358	29,124	27,884	30,489	36,134	41,724	48,484	56,585	66,787	82,316	118,826
Avg. Number of Trades, γ	1,761	267	1,370	2,143	2,951	3,720	4,611	5,663	7,309	10,318	24,090
Avg. Daily Volume($1000)	19,317	796	6,834	13,563	21,912	31,999	42,816	57,954	83,025	136,204	421,325
Avg. Volatility	0.034	0.038	0.031	0.029	0.028	0.028	0.027	0.026	0.026	0.026	0.027
Avg. Price	19.44	12.26	24.80	28.94	31.52	35.70	38.23	39.87	43.96	46.78	52.23
100-Shares: %Trades/%Vol	50/18	50/18	55/21	52/19	49/17	47/16	46/15	44/15	43/14	42/13	39/11
1000-Shares: %Trades/%Vol	5/7	5/8	3/6	3/5	3/6	3/6	4/6	4/5	4/5	4/5	5/6
Even Lots: %Trades/%Vol	86/61	86/61	90/64	89/63	88/61	87/60	86/59	85/58	85/57	84/56	82/54
# Obs	471,719	306,524	58,794	24,525	22,775	10,687	10,025	9,545	9,328	9,424	10,092

Table reports the properties of securities and trades in the two subsamples, before and after 2001. Panel A reports statistics for data from February 1993 to December 2000. Panel B reports statistics for data from January 2001 to December 2008. Both panels show the average of trade size, the trade-weighted median trade size, the volume-weighted median trade size, the average number of trades per day, the daily dollar volume (in thousands of $), the average volatility of daily returns, the average price, the percent of trades in the 100-share lot, the percent of volume in the 100-share lot, the percent of trades in the 1000-share lot, the percent of volume in the 1000-share lot, the percent of trades in the even lots, and the percent of volume in the even lots for all sample as well as for ten volume groups. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume.

Tab. 3.2: OLS Estimates of Number of Trades.

	All Stocks			NYSE/AMEX			NASDAQ		
	93/08	93/00	01/08	93/08	93/00	01/08	93/08	93/00	01/08
α	7.10	6.15	8.04	6.95	6.07	7.81	7.24	6.17	8.30
	(0.199)	(0.052)	(0.155)	(0.188)	(0.027)	(0.168)	(0.222)	(0.078)	(0.152)
a_γ	0.74	0.69	0.79	0.70	0.64	0.76	0.77	0.71	0.83
	(0.011)	(0.002)	(0.011)	(0.013)	(0.002)	(0.012)	(0.013)	(0.002)	(0.012)
Adj-R^2	0.92	0.91	0.94	0.94	0.93	0.95	0.92	0.90	0.94
# Obs	5,801	6,698	4,914	2,051	2,199	1,904	3,750	4,499	3,010
	Model of Trading Game Invariance: H_0: $a_\gamma = 2/3$								
F-Test	44.2	92.9	122.4	6.3	260.9	62.4	58.5	336.2	193.0
p-Value	<0.001	<0.001	<0.001	0.0129	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Frequency: H_0: $a_\gamma = 0$								
F-Test	4646.2	81391.5	5245.6	2994.2	141943.9	4317.6	3465.8	112044.3	5160.2
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $a_\gamma = 1$								
F-Test	581.5	16432.1	384.9	556.1	45201.6	441.2	322.6	19561.5	227.4
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α and a_γ from monthly regressions

$$\ln\left[\gamma_{1,i}\right] = \alpha + a_\gamma \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i.$$

Each observation corresponds to the stock i with $\gamma_{1,i}$ being the average number of trades per day and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns in a given month. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for the benchmark stock with price \$40 per share, trading volume of one million shares per day, and daily volatility of 0.02. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. # Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models. The estimates are reported for the entire sample from February 1993 to December 2008 as well as for two subsamples, before and after 2001.

Tab. 3.3: OLS Estimates of Number of Trades: Robustness Check.

	All Stocks			NYSE/AMEX			NASDAQ		
	93/08	93/00	01/08	93/08	93/00	01/08	93/08	93/00	01/08
α	2.07	2.51	1.64	3.30	3.94	2.67	1.25	1.98	0.52
	(0.157)	(0.057)	(0.262)	(0.196)	(0.057)	(0.302)	(0.188)	(0.116)	(0.225)
b_1	0.18	0.14	0.23	0.13	0.08	0.18	0.22	0.16	0.28
	(0.011)	(0.003)	(0.014)	(0.013)	(0.003)	(0.016)	(0.014)	(0.007)	(0.013)
b_2	-0.30	-0.35	-0.24	-0.26	-0.34	-0.17	-0.31	-0.34	-0.28
	(0.012)	(0.006)	(0.010)	(0.017)	(0.006)	(0.010)	(0.008)	(0.006)	(0.010)
b_3	-0.41	-0.46	-0.36	-0.45	-0.49	-0.41	-0.44	-0.41	-0.47
	(0.022)	(0.036)	(0.016)	(0.023)	(0.013)	(0.042)	(0.015)	(0.016)	(0.022)
Adj-R^2	0.95	0.94	0.96	0.96	0.95	0.97	0.96	0.94	0.97
# Obs	5,801	6,698	4,914	2,051	2,199	1,904	3,750	4,499	3,010
	Model of Trading Game Invariance: H_0: $b_1 = b_2 = b_3 = 0$								
F-Test	872.8	3072.4	324.1	1448.6	2692.0	504.8	1015.5	5496.5	299.7
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Frequency: H_0: $b_1 = b_2 = b_3 = -2/3$								
F-Test	1287.2	14521.5	1475.1	1245.5	14023.5	1108.9	3324.7	6546.2	3465.7
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $b_1 = b_2 = b_3 = 1/3$								
F-Test	2013.4	14062.9	1429.0	1226.0	13580.6	1074.2	3272.5	6339.5	3357.3
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α, b_1, b_2 and b_3 from monthly regression

$$\ln\left[\gamma_{1,i}\right] = \alpha + a_\gamma \times \ln\left[\frac{W_i}{W_*}\right] + b_1 \times \ln\left[\frac{V_i}{10^6}\right] + b_2 \times \ln\left[\frac{P_i}{40}\right] + b_3 \times \ln\left[\frac{\sigma_i}{0.02}\right] + \tilde{\epsilon}_i.$$

Each observation corresponds to the stock i with $\gamma_{1,i}$ being the average number of trades per day and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns in a given month. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the trading activity of the benchmark stock with price \$40 per share, trading volume of one million shares per day, and volatility of 0.02. Variables V_i, P_i and σ_i are the average trading volume (in shares), average price, and average daily volatility. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. # Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models. The estimates are reported for the entire sample from February 1993 to December 2008 as well as for two subsamples, before and after 2001.

Tab. 3.4: OLS Estimates of Trade Sizes, February 1993 - December 2000.

	Trade-Weighted Distribution				Volume-Weighted Distribution			
	Mean	$20th$	$50th$	$80th$	Mean	$20th$	$50th$	$80th$
α	-7.22	-8.47	-7.26	-6.24	-4.72	-6.39	-4.93	-3.38
	(0.033)	(0.039)	(0.045)	(0.036)	(0.073)	(0.053)	(0.084)	(0.087)
a_Q	-0.76	-0.80	-0.76	-0.74	-0.59	-0.69	-0.61	-0.51
	(0.006)	(0.008)	(0.006)	(0.005)	(0.003)	(0.002)	(0.004)	(0.006)
Adj-R^2	0.93	0.90	0.91	0.91	0.75	0.87	0.75	0.61
#Obs	6,698	6,698	6,698	6,698	6,698	6,698	6,698	6,698
	Model of Trading Game Invariance : H_0: $a_Q = -2/3$							
F-Test	254	269	319	226	503	93	180	644
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Frequency : H_0: $a_Q = 0$							
F-Test	17074	10030	19319	21827	30332	97083	21820	7261
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $a_Q = -1$							
F-Test	91674	50968	102846	120269	220002	584391	151664	63056
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α and a_Q from the monthly regressions of the mean trade size of its percentiles on the measure of trading activity W for the sample from February 1993 to December 2000. The coefficients α and a_Q are based on monthly regressions

$$\ln\left[\frac{|Q_i|}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of (unsigned) trade sizes $|Q_i|$, as a fraction of daily volume V_i in a given month. The means and percentiles are calculated both based on the distributions of trade size themselves (trade-weighted distribution) and based on the contribution to total trading volume (volume-weighted distribution). Each observation corresponds to the stock i with γ_i being the average daily number of trades and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the trading activity of the benchmark stock with price \$40 per share, trading volume of one million shares per day, and volatility of 0.02. Variables V_i, P_i and σ_i are the average trading volume (in shares), average price, and average daily volatility. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. #Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models.

Tab. 3.5: OLS Estimates of Trade Sizes, January 2001 - December 2008.

	Trade-Weighted Distribution				Volume-Weighted Distribution			
	Mean	$20th$	$50th$	$80th$	Mean	$20th$	$50th$	$80th$
α	-8.69	-9.42	-8.98	-8.07	-6.82	-8.55	-7.35	-5.56
	(0.103)	(0.044)	(0.111)	(0.162)	(0.079)	(0.188)	(0.242)	(0.237)
a_Q	-0.79	-0.79	-0.79	-0.81	-0.74	-0.80	-0.80	-0.72
	(0.005)	(0.007)	(0.004)	(0.008)	(0.008)	(0.012)	(0.025)	(0.030)
Adj-R^2	0.93	0.90	0.92	0.93	0.86	0.91	0.87	0.77
# bs	4,914	4,914	4,914	4,914	4,914	4,914	4,914	4,914
	Model of Trading Game Invariance : H_0: $a_Q = -2/3$							
F-Test	636	266	935	314	85	118	29	3
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.086
	Model of Invariant Bet Frequency : H_0: $a_Q = 0$							
F-Test	25116	11432	37000	10574	7994	4286	1056	565
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $a_Q = -1$							
F-Test	128423	58963	189234	53135	43978	21713	5347	3226
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α and a_Q from the monthly regressions of the mean trade size o its percentiles on the measure of trading activity W for the sample from January 2001 to December 2008. The coefficients α and a_Q are based on monthly regressions

$$\ln\left[\frac{|Q_i|}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of (unsigned) trade sizes $|Q_i|$, as a fraction of daily volume V_i in a given month. The means and percentiles are calculated both based on the distributions of trade size themselves (trade-weighted distribution) and based on the contribution to total trading volume (volume-weighted distribution). Each observation corresponds to the stock i with γ_i being the average daily number of trades and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the trading activity of the benchmark stock with price \$40 per share, trading volume of one million shares per day, and volatility of 0.02. Variables V_i, P_i and σ_i are the average trading volume (in shares), average price, and average daily volatility. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. #Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models.

Fig. 3.1: Comparison of Three Models based on Number of Trades

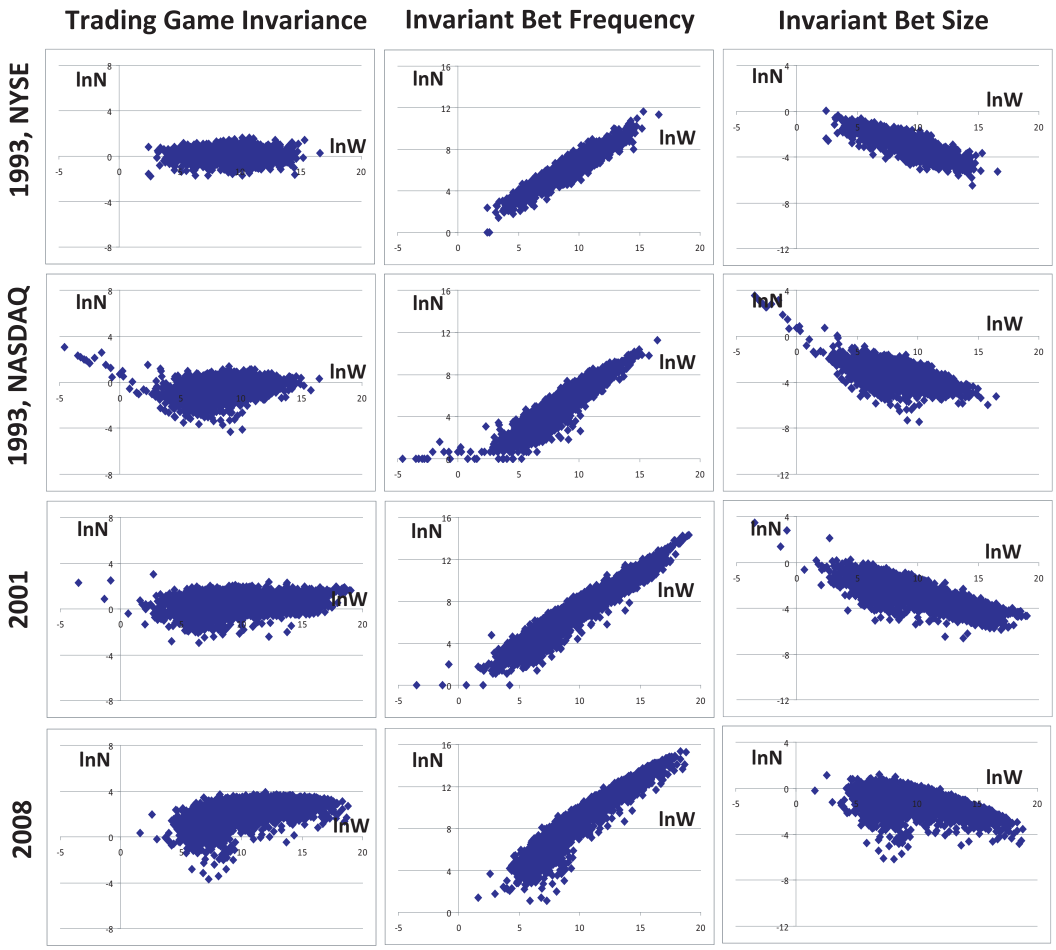


Figure shows the logarithm of normalized number N of trades across different levels of the logarithm of trading activity W. For the model of trading game invariance, the number γ of trades is normalized as $N = \gamma/W^{2/3}$. For the model of invariant bet frequency, the number of trades is shown as $N = \gamma$, without any adjustment. For the model of invariant bet size, the number of trades is normalized as $N = \gamma/W$. Four subsamples are considered: NYSE-listed stocks in April of 1993, Nasdaq-listed stocks in April of 1993, stocks in April of 2001 and in April of 2008. Trading activity W is calculated as the product of average daily dollar volume and daily returns standard deviation in that month.

Fig. 3.2: Comparison of Three Models based on Trade Size, NYSE-listed Stocks, April 1993

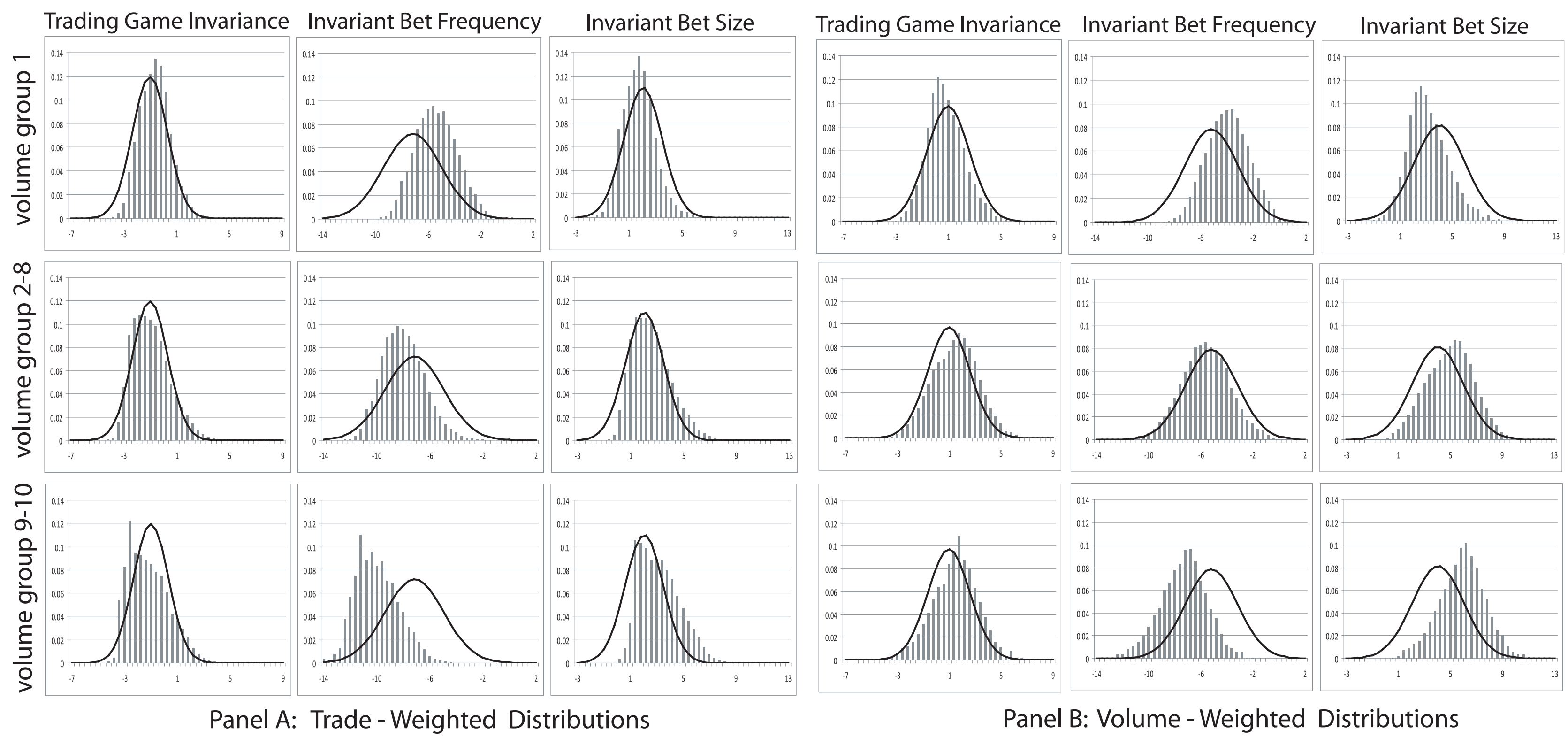
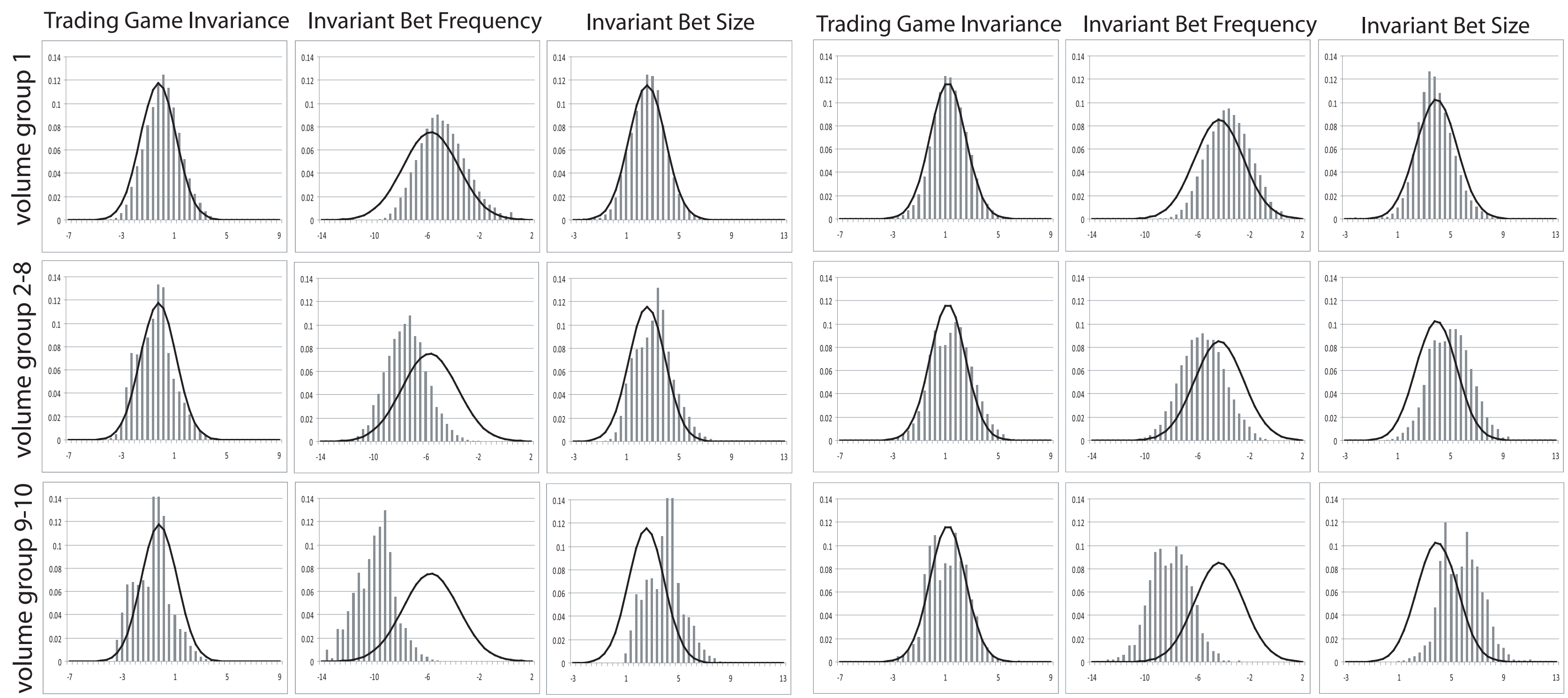


Figure shows the distribution of the logarithm of normalized trade sizes for three different models for the NYSE-listed stocks traded in April 1993. For the model of trading game invariance, the trade sizes are normalized as $|Q|/V \times W^{2/3}$. For the model of invariant bet frequency, the trade sizes are plotted as $|Q|/V$, without any adjustment. For the model of invariant bet size, the trade sizes are normalized as $|Q|/V \times W^{1}$. Trading activity W is calculated as the product of dollar volume and returns standard deviation. Panel A shows the trade-weighted distributions. Panel B shows the volume-weighted distributions. The stock-level distributions, averaged across stocks for volume groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume.

Fig. 3.3: Comparison of Three Models based on Trade Size, NASDAQ-listed Stocks, April 1993



Panel A: Trade - Weighted Distributions

Panel B: Volume - Weighted Distributions

Figure shows the distribution of the logarithm of normalized trade sizes for three different models for the NASDAQ-listed stocks traded in April 1993. For the model of trading game invariance, the trade sizes are normalized as $|Q|/V \times W^{2/3}$. For the model of invariant bet frequency, the trade sizes are plotted as $|Q|/V$, without any adjustment. For the model of invariant bet size, the trade sizes are normalized as $|Q|/V \times W^1$. Trading activity W is calculated as the product of dollar volume and returns standard deviation. Panel A shows the trade-weighted distributions. Panel B shows the volume-weighted distributions. The stock-level distributions, averaged across stocks for volume groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume.

Fig. 3.4: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 1993

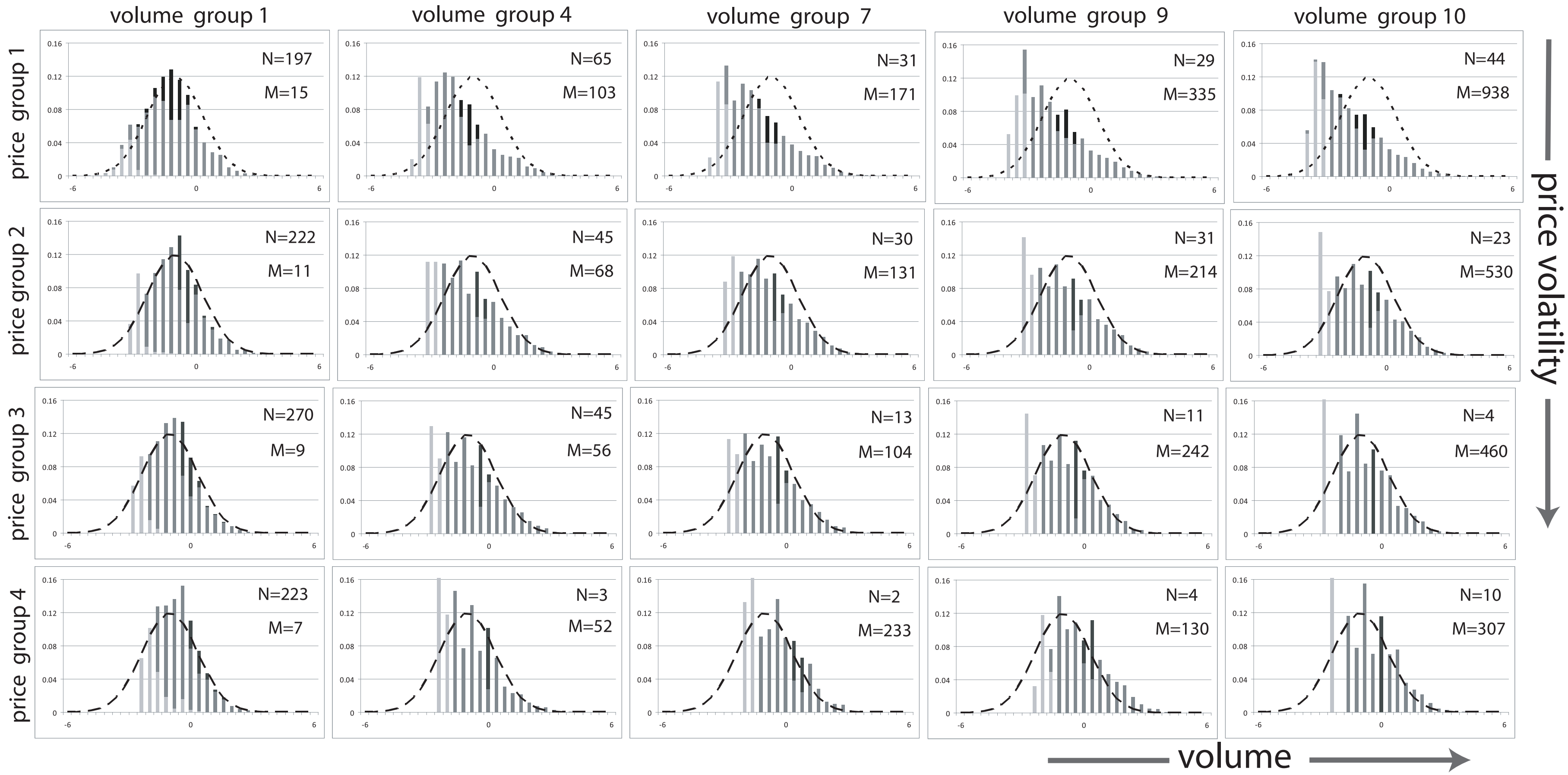


Figure shows distributions of the logarithms of normalized trade sizes for the NYSE stocks in April 1993. For each trade, the normalized trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -1.01 and averaged standard deviation of 1.33 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Fig. 3.5: Volume-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 1993

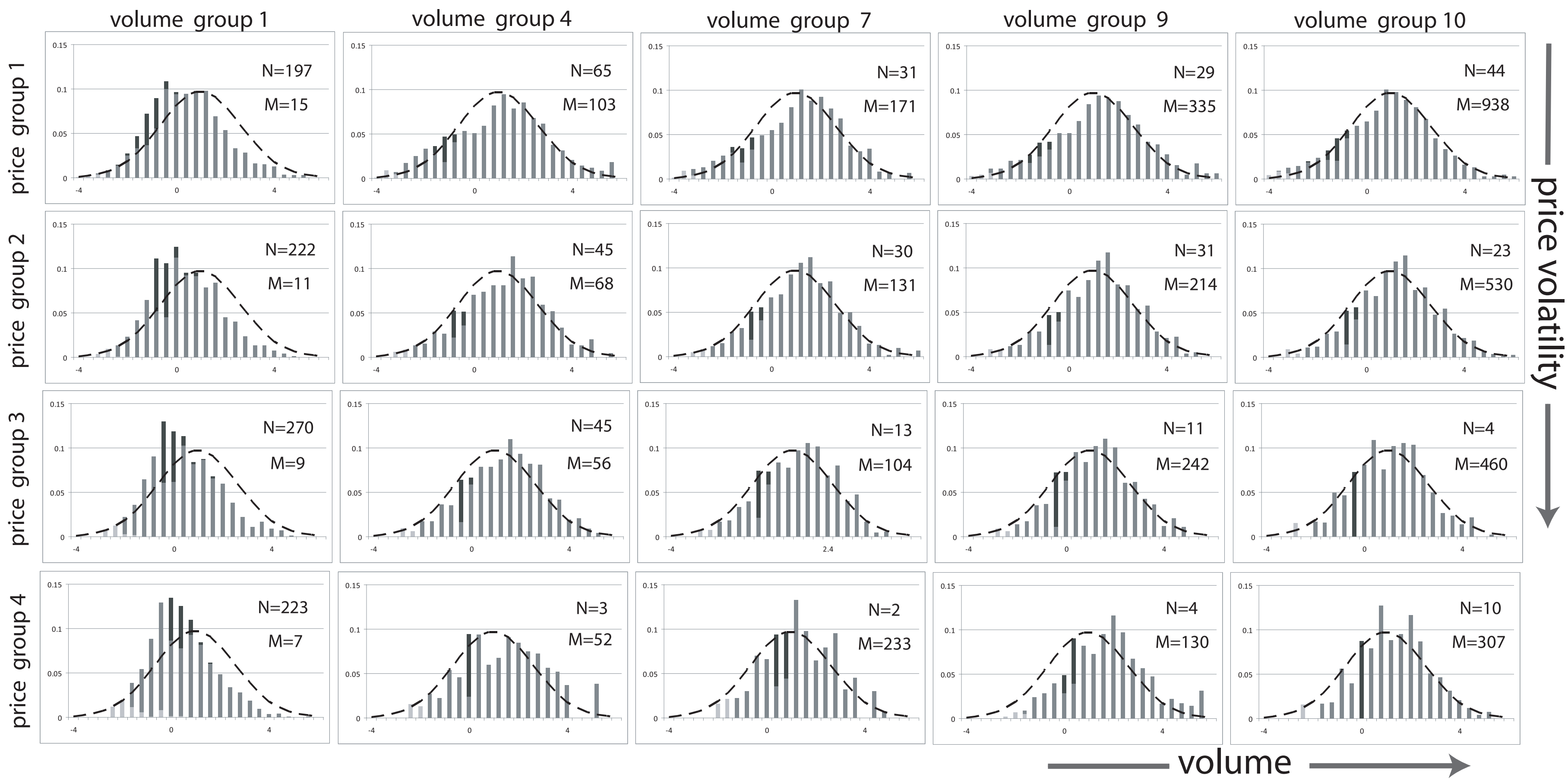


Figure shows distributions of total volume across different normalized trade size bins for the NYSE stocks in April 1993. For each stock, the volume distribution is calculated as the contribution to the total volume by trades from a given trade size bin. On x-axis, the logs of normalized trade sizes are depicted. The adjustment is done according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a trade size in shares (midpoint of a bin), V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of 0.97 and averaged standard deviation of 1.64 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Fig. 3.6: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NASDAQ-listed Stocks, April 1993

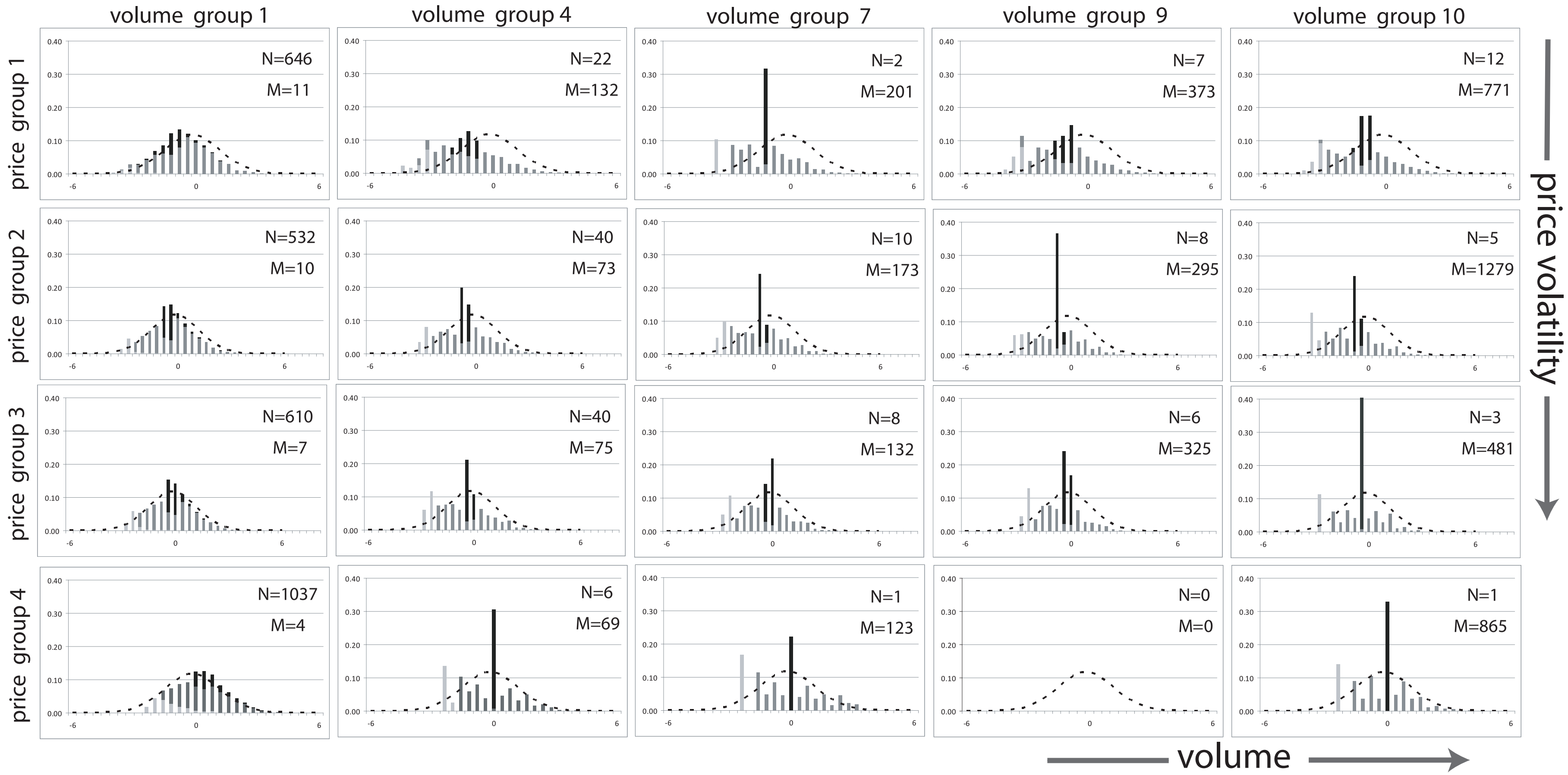


Figure shows distributions of the logarithms of normalized trade sizes for the Nasdaq stocks in April 1993. For each trade, the normalized trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -0.18 and averaged standard deviation of 1.36 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Fig. 3.7: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 2001

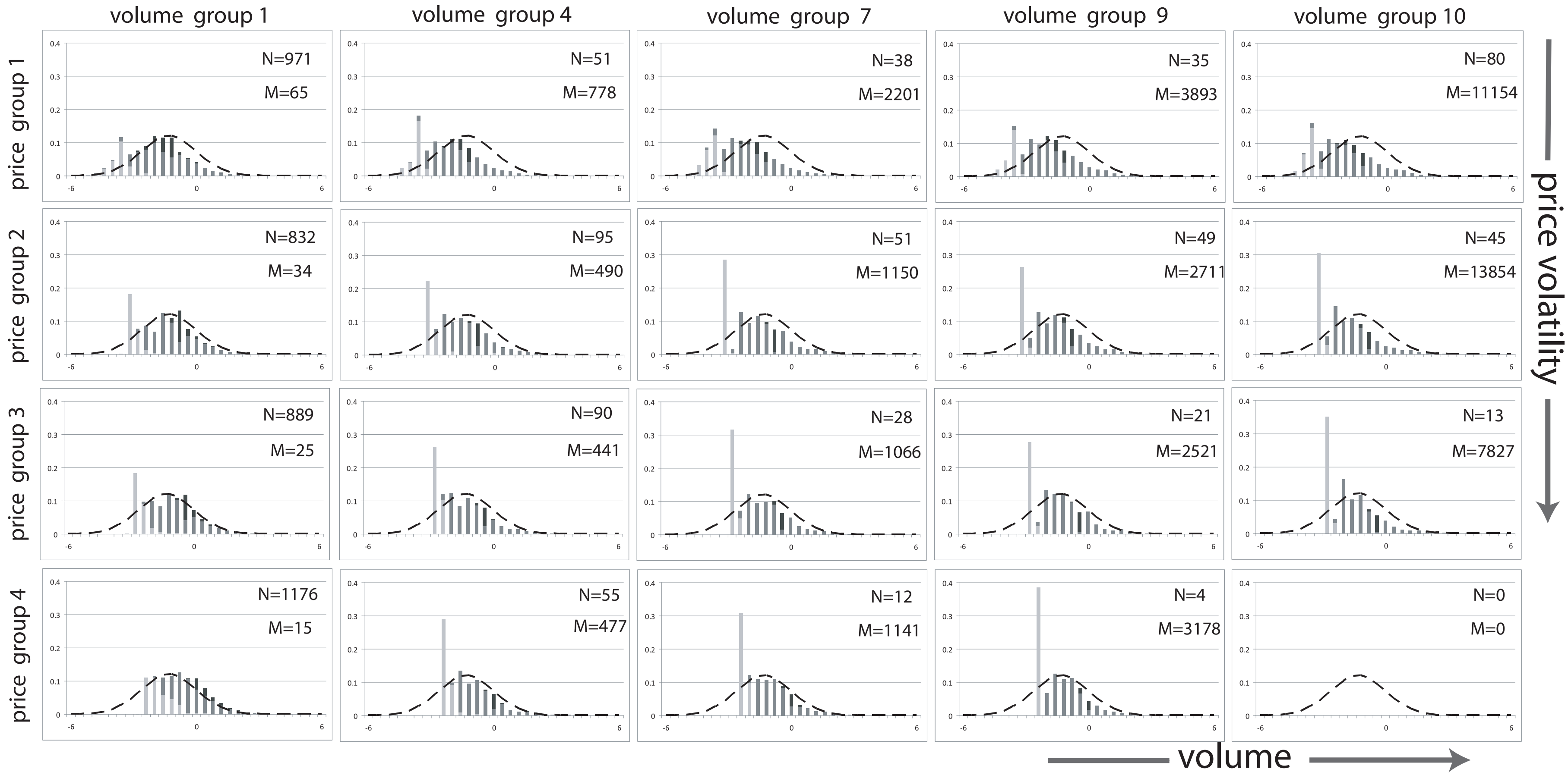


Figure shows distributions of the logarithms of adjusted trade sizes for stocks in April 2001. For each trade, the normalized trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -1.31 and averaged standard deviation of 1.32 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Fig. 3.8: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 2008

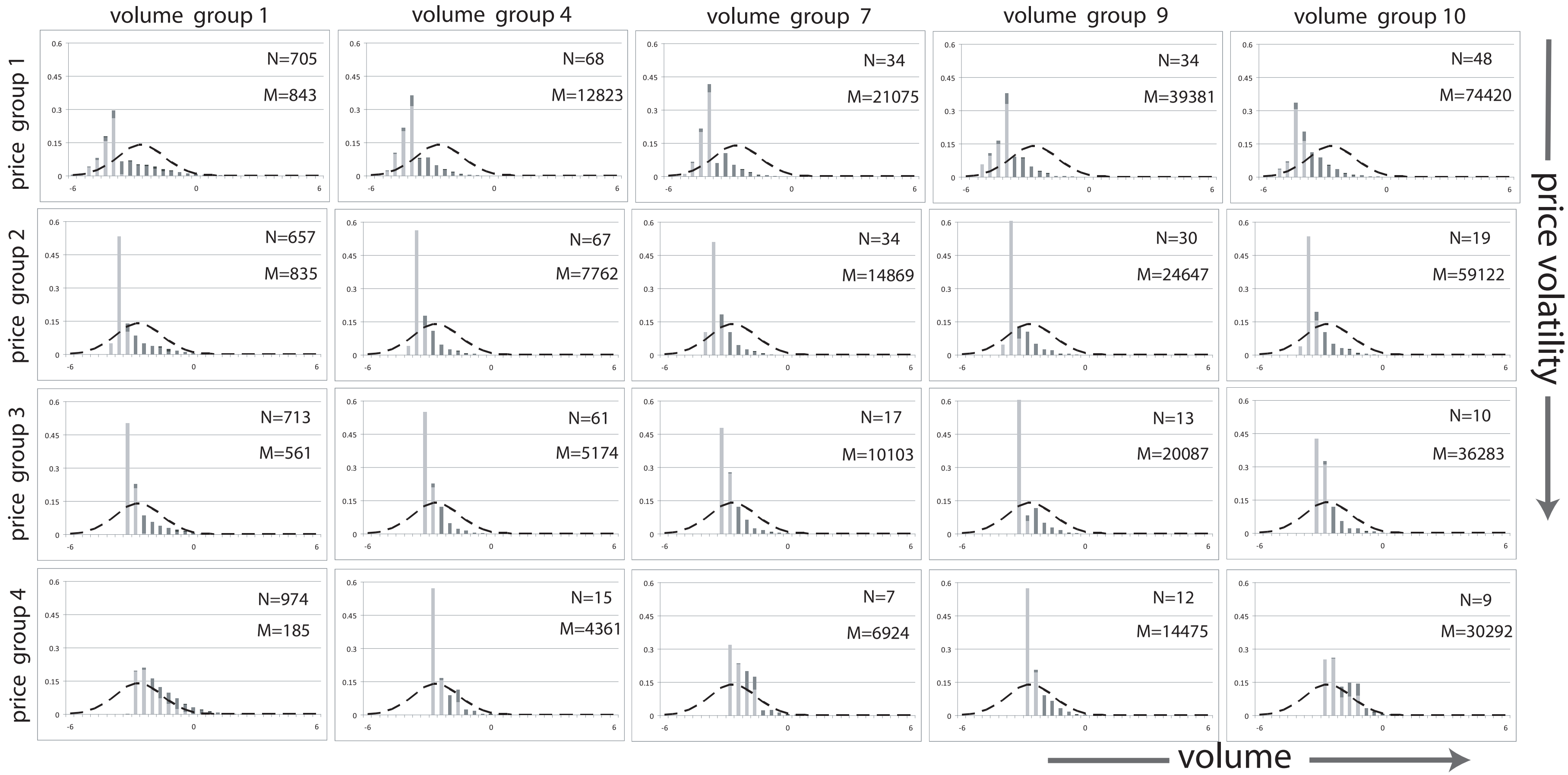


Figure shows distributions of the logarithms of normalized trade sizes for stocks in April 2001. For each trade, the adjusted trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -2.66 and averaged standard deviation of 1.15 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Fig. 3.9: Trading Patterns for Small and Large Stocks, February 1993 - December 2008.

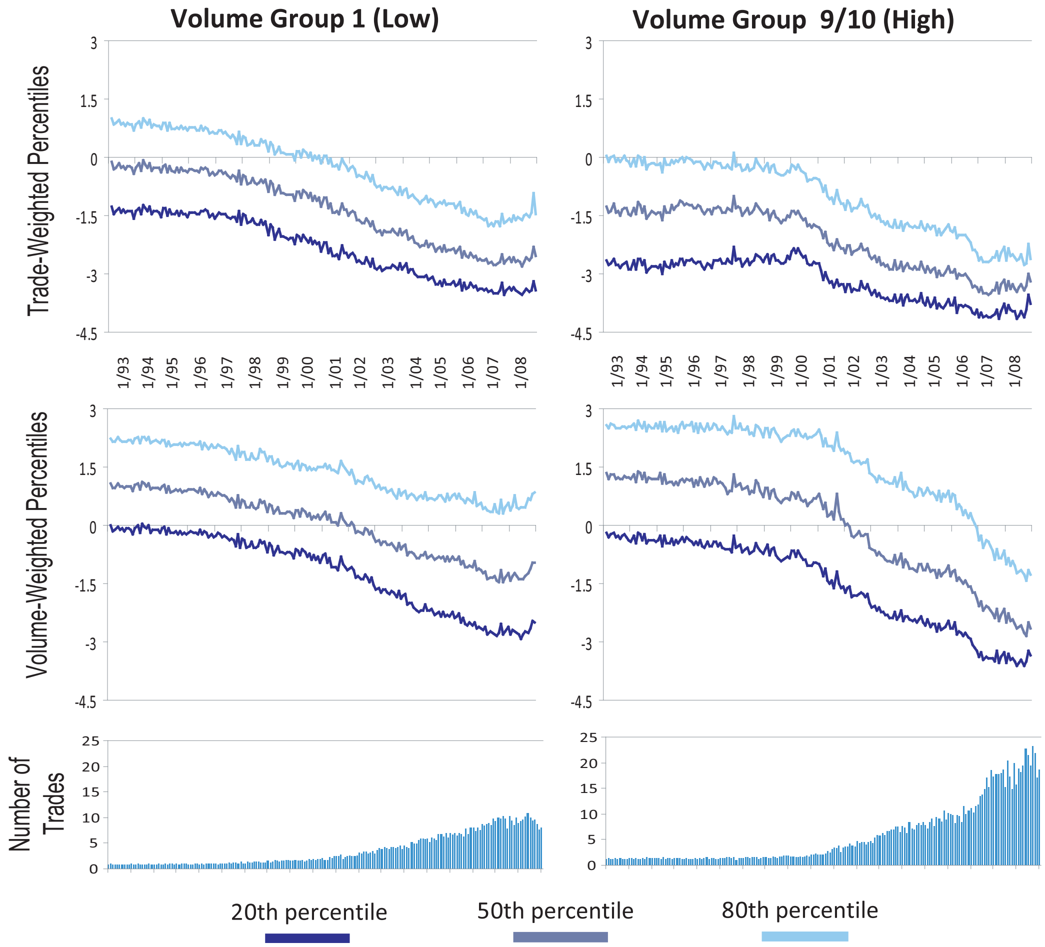


Figure shows the dynamics of average number of trades per month and the 20th, 50th and 80th percentiles for normalized trade size from 1993 to 2008. Trade-weighted percentiles and volume-weighted percentiles are shown for stocks in volume group 1 and volume groups 9 and 10. Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. For each trade, the normalized trade size is calculated based on the midpoint of a trade size bin, in which a trade locates, and normalized according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The stock-level distributions of normalized trade sizes are averaged across stocks for volume groups 1 and 9/10 in a given month. The trade-weighted and volume-weighted percentiles are plotted on this figure. W is calculated as the product of dollar volume and returns standard deviation. W_* is the measure of trading activity of the benchmark stock.

Fig. 3.10: Dynamics of OLS Estimates, February 1993 - December 2008.

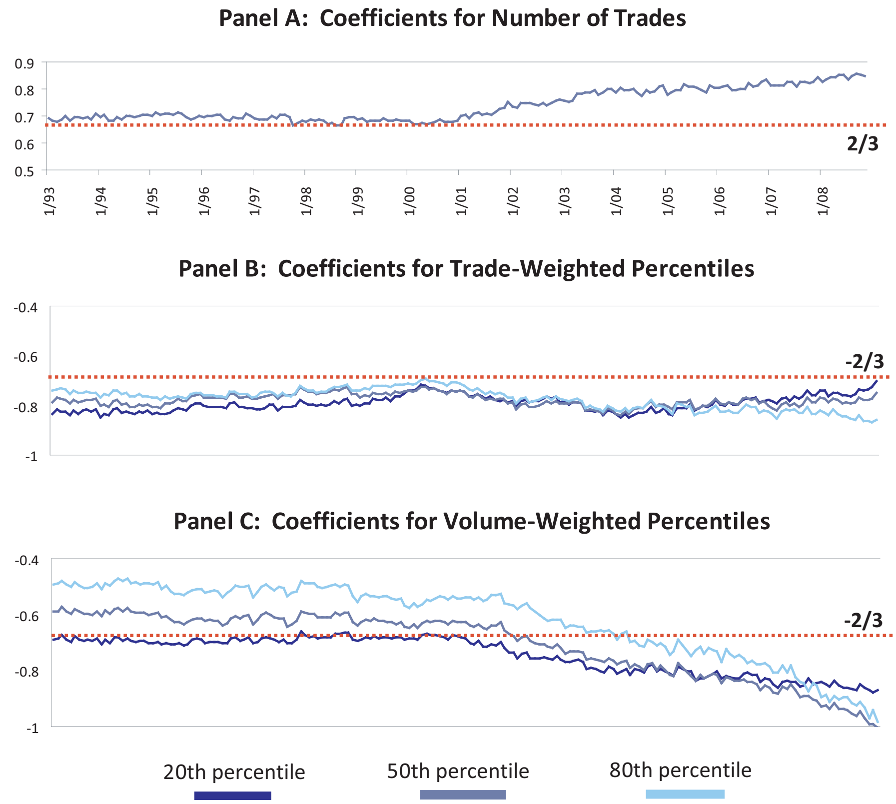


Figure shows the dynamics of coefficients from regressions of number of trades and various percentiles on the measure of trading activity W from 1993 to 2008. Panel A shows the coefficient a_γ from monthly regressions

$$\ln\Big[\gamma\Big] = \ln\left[\bar{q}\right] + a_\gamma \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where γ is the number of trades per month. The model of trading game invariance predicts $a_\gamma = 2/3$ and alternative models predict that $a_\gamma = 0$ or $a_\gamma = 1$. Panel B shows the coefficient a_Q from monthly regressions

$$\ln\left[\frac{\tilde{Q}_i}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where the left-hand side is the *pth* (20th, 50th and 80th) percentiles of the distribution of trade sizes $\tilde{Q}_i$. The model of trading game invariance predicts $a_Q = -2/3$ and alternative models predict that $a_Q = 0$ or $a_Q = -1$. Panel C shows the coefficient a_Q from similar monthly regressions but these regressions are based on percentiles Q_i^p, where percentiles are calculated based on the contribution to total trading volume. The model of trading game invariance predicts $a_Q = -2/3$ and alternative models predict that $a_Q = 0$ or $a_Q = -1$. W is calculated as the product of dollar volume and returns standard deviation. W_* is the measure of trading activity of the benchmark stock.

BIBLIOGRAPHY

[1] Alexander, Gordon, and Mark Peterson, 2007, *An analysis of trade-size clustering and its relation to stealth trading*, *Journal of Financial Economics*, 84, 435–471.

[2] Alfonsi, Aurelien, Antje Fruth, and Alexander Schied, 2008, *Constrained Portfolio Liquidation in a Limit Order Book Model*, Banach Center Publ. 83, 9–25.

[3] Almgren, Robert, and Neil Chriss, 1999, *Value Under Liquidation*, Risk 12, 61-63.

[4] Almgren, Robert, and Neil Chriss, 2000, *Optimal Execution of Portfolio Transactions*, Journal of Risk 12, 5-39.

[5] Almgren, Robert, and Julian Lorenz, 2006, *Bayesian Adaptive Trading With a Daily Cycle*, Journal of Trading 1, 38-46.

[6] Amihud, Yakov, Haim Mendelson, and Lasse Heje Pedersen, 2005,*Liquidity and Asset Pricing*, Foundations and Trends in Finance, 1, 269–364.

[7] Berkman, Henk, 1996, *Large Option Trades, Market Makers, and Limit Orders*, Review of Financial Studies, 9, 977–1002.

[8] Bertsimas, Dimitris, and Andrew, W. Lo, 1998, *Optimal Control of Execution Costs*, Journal of Financial Markets, 1, 1–50.

[9] Biais, Bruno, Larry Glosten, and Chester Spatt, 2005, *Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications*, Journal of Financial Markets, 8, 217–264.

[10] Brennan, Michael, and Avanidhar Subrahmanyam, 1998, *The determinants of average trade size*, Journal of Business, 71(1).

[11] Bruno Biais, David Martimort, and Jean-Charles Rochet, 2000, *Competing Mechanisms in a Common Value Environment*, Econometrica, 68, 799–838.

[12] Black, Fischer, 1971, *Towards a Fully Automated Exchange*, Part I, Financial Analysts Journal, 27, 29-34.

[13] Brogaard, Jonathan, 2010, *High Frequency Trading and its Impact on Market Quality*, Working Paper.

[14] Brunnermeier, Markus K., and Lasse Heje Pedersen, 2005, *Predatory Trading*, Journal of Finance, 60, 1825-1863.

[15] Campbell, J. Y., Grossman, S. J., Wang, J., 1993. *Trading volume and serial correlation in stock returns*. Quarterly Journal of Economics 108, 905.939.

[16] Campbell, John Y., Ramadorai, Tarun., Schwartz, Allie., 2009. *Caught on tape: Institutional Trading, stock returns and earnings announcements*. Journal of Financial Economics. 92, 66.91.

[17] Carlin, Bruce I., Miguel Sousa Lobo, and S. Viswanathan, 2007, *Episodic Liquidity Crises: Cooperative and Predatory Trading*, Journal of Finance, 62, 2235–2274.

[18] Cespa, Giovanni, and Thierry Foucault, 2008, *Insiders-Outsiders, Transparency, and the Value of the Ticker*, Working Paper.

[19] Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2009, *Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market*, Board of Governors of the Federal Reserve System International Finance Discussion Papers, 980.

[20] Chan, Louis K. C., and Josef Lakonishok, 1995, *The Behavior of Stock Prices Around Institutional Trades*, Journal of Financial Economics, 50, 1147–1174.

[21] Chan, Louis K. C., and Josef Lakonishok, 1993, *Institutional Trades and Intraday Stock Price Behavior*, Journal of Financial Economics, 33, 173–199.

[22] Chiyachantana, Chiraphol N., Pankaj K. Jain, Christine Jiang, and Robert A. Wood, 2004, *International Evidence on Institutional Trading Behavior and Price Impact*, Journal of Finance, 59, 869–898.

[23] Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2008, *Recent trends in trading activity*, Working Paper.

[24] Cvitanic, Jaksa, and Andrei A. Kirilenko, 2010, *High Frequency Traders and Asset Prices*, Working Paper.

[25] Da Z., Gao P., Jagannathan R., 2009, *Informed Trading, Liquidity Provision, and Stock Selection by Mutual Funds*, SSRN Working Paper

[26] Daniel, K. D., Grinblatt, M., Titman, S., Wermers, R., 1997. *Measuring mutual fund performance with characteristic-based benchmarks*. Journal of Finance 52, 1035.1058.

[27] Engle, Robert, and Robert Ferstenberg, 2007, *Execution Risk*, Journal of Portfolio Management, 33, 34–45.

[28] Foucault, Thierry, 1999, *Order Flow Composition and Trading Costs in a Dynamic Limit Order Market*, Journal of Financial Markets, 2, 99–134.

[29] Glosten, Lawrence R., and Paul Milgrom, 1985, *Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders*, Journal of Financial Economics, 14, 71–100.

[30] Glosten, Lawrence, and Lawrence Harris, 1988, *Estimating the components of the bid-ask spread*, Journal of Financial Economics, 21, 123-142.

[31] Goettler Ronald L., Christine A. Parlour, and Uday Rajan, 2005,*Equilibrium in a Dynamic Limit Order Market*, Journal of Finance, 60, 2149–2192.

[32] Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2009, *Informed Traders and Limit Order Markets*, Journal of Financial Economics, 93, 67–87.

[33] Gompers, P. A., Metrick, A., 2001. *Institutional investors and equity prices.* Quarterly Journal of Economics 116, 229.260.

[34] Grinblatt, M., Titman, S., Wermers, R., 1995. *Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior.* American Economic Review 85, 1088.1105.

[35] Hasbrouck, Joel, 2003, *Intraday Price Formation in U.S. Equity Index Markets*, *Journal of Finance* 58, 2375–2400.

[36] Hasbrouck, Joel, and Gideon Saar, 2009, *Technology and Liquidity Provision: The Blurring of Traditional Definitions*, Journal of Financial Markets, 12, 143–172.

[37] Hasbrouck, Joel, and Gideon Saar, 2010, *Low Latency Trading*, Johnson School Research Paper Series, 35.

[38] Hendershott, Terrence J., Charles M. Jones, and Albert J. Menkveld, 2010, *Does Algorithmic Trading Improve Liquidity?*, Journal of Finance.

[39] Hendershott, Terrence J., and Albert J. Menkveld, 2010, *Price Pressures*, WFA 2010 Paper.

[40] Hendershott, Terrence J., and Ryan Riordan, 2009, *Algorithmic Trading and Information*, Working Paper.

[41] Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1987, *The Effect of Large Block Transactions on Security Prices: A Cross-Sectional Analysis*, Journal of Financial Economics, 19, 237-267.

[42] Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1990, *Large-Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects*, Journal of Financial Economics, 26, 71–95.

[43] Keim, B. D., 1999 *An analysis of mutual fund design: the case of investing in small-cap stocks*. Journal of Financial Economics 51, 173-194.

[44] Keim, Donald B., and Ananth Madhavan, 1996, *The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects*, Review of Financial Studies, 9, 1–36.

[45] Keim, Donald B., and Ananth Madhavan, 1997, *Transactions Costs and Investment Style: an Inter-Exchange Analysis of Institutional Equity Trades*, Journal of Financial Economics, 46, 265–292.

[46] Kraus, Alan, and Hans R. Stoll, 1972, *Price Impacts of Block Trading on the New York Stock Exchange*, Journal of Finance, 27, 569–588.

[47] Kyle, Albert S., 1985, *Continuous Auctions and Insider Trading*, Econometrica, 53, 1315–1335.

[48] Kyle, Albert, and Anna Obizhaeva, 2011, *Market Microstructure Invariants*, Working Paper, University of Maryland.

[49] Lakonishok, J., Shleifer, A., Vishny, R., 1992. *The impact of institutional trading on stock prices*. Journal of Financial Economics 32, 23.43.

[50] Lee, C. M. C., Ready, M. J., 1991. *Inferring trade direction from intraday data.* Journal of Finance 46, 733.746.

[51] Lee, C. M. C., Radhakrishna, B., 2000. *Inferring investor behavior: evidence from TORQ data*. Journal of Financial Markets 3, 83.111.

[52] Madhavan, Ananth, 2000, *Market Microstructure: a Survey*, Journal of Financial Markets, 3, 205–258.

[53] Van der Wel, Michel, Albert J. Menkveld, and Asani Sarkar, 2009,*Are Market Makers Uninformed and Passive? Signing Trades in the Absence of Quotes*, *Federal Reserve Bank of New York Staff Reports* 395.

[54] Moallemi, Ciamac C., Benjamin Van Roy, and Beomsoo Park, 2009, *Strategic Execution in the Presence of an Uninformed Arbitrageur*, *Working Paper*.

[55] Moallemi, Ciamac C., and Mehmet Saglam, 2010, *The Cost of Latency*, Working Paper.

[56] Moulton, Pamela, 2005, *You can't always get what you want: Trade-size clustering and quantity choice in liquidity*, Journal of Financial Economics, 78, 89–119.

[57] Nelson, C. R., Siegel, A. F., 1987. *Parsimonious modeling of yield curves*. Journal of Business 60, 473-489.

[58] Newey, Whitney K., and Kenneth D. West, 1987, *A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, *Econometrica* 55, 703–708.

[59] Obizhaeva, Anna, and Jiang Wang, 2006,*Optimal Trading Strategy and Supply/Demand Dynamics*, NBER Working Paper.

[60] Parlour, Christine, 1998, *Price Dynamics in Limit Order Markets*, Review of Financial Studies, 11, 789–816.

[61] Peterson, Mitchell A., 2008 *Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches*, 22, 435-480.

[62] Rosu, Ioanid, 2009, *A Dynamic Model of the Limit Order Book*, Review of Financial Studies, 22, 4601-4641.

[63] Schied, Alexander, and Torsten Schoneborn, 2007, *Optimal Portfolio Liquidation for CARA Investors*, *MPRA Paper*.

[64] Sias, R., 2004 *Institutional herding. Review of Financial Studies*, 17, 165.206.

[65] Sias, R., Starks, L. T., Titman, S., 2006. *The price impact of institutional trading*. Journal of Business 79, 2869.2910.

[66] Wermers, R., 1999. *Mutual fund herding and the impact on stock prices*. Journal of Finance 54, 581.622.

[67] Wermers, R., 2000. *Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses*. Journal of Finance 55, 1655.1695.