ABSTRACT

Title of dissertation: ESSAYS ON MARKET MICROSTRUCTURE

Kyoung Hun Bae, Doctor of Philosophy, 2015

Dissertation directed by: Professor Albert "Pete" Kyle
Department of Finance

This dissertation includes two essays on topics related to market microstructure. In the first essay, we analyze algorithmic trading in the Korean Index Futures market. We document that short-term traders consistently anticipate the order flow of large traders that build large positions within a short period of time. We study trade-by-trade data around 36,164 trades by large traders among the largest 1% of all active trades during 66 trading days in 2009 from the Korean Index Futures market. We find that large traders manage their orders first by executing small, positively correlated trades, which are followed by a single large trade. While the small trades are executed, short-term traders gradually increase their inventories in the direction of the forthcoming large trade. After the execution of the large trade, short-term traders unload their inventories to other traders. We find that short-term traders correctly anticipate the direction of large trades 56.06% of the time. Furthermore,

the aggregate positions of short-term traders are statistically significant predictors for the direction of large trades that will arrive within 120 seconds.

In the second essay, we explore market microstructure invariance in the Korean stock market. We define the number of buy-sell switching points based on the number of times that individual traders change the direction of their trading. Based on the hypothesis that switching points take place in business time, market microstructure invariance predicts that the aggregate number of switching points is proportional to the 2/3 power of the product of dollar volume and volatility. Using trading data from the Korea Exchange (KRX) from 2008 to 2010, we estimate the exponent to be 0.675 with standard error of 0.005. Invariance explains about 93% of the variation in the logarithm of the number of switching points each month across stocks.

ESSAYS ON MARKET MICROSTRUCTURE

by

Kyoung Hun Bae

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
2015

Advisory Committee:
Professor Albert "Pete" Kyle, Chair/Advisor
Professor Steven L. Heston
Professor Mark Loewenstein
Professor Anna Obizhaeva
Professor John C. Chao

Acknowledgments

I am indebted to my dissertation advisors, Albert S. "Pete" Kyle, Steve Heston, Mark Loewenstein, Anna A. Obizhaeva, and John Chao for their invaluable guidance. I am very thankful to Don Bowen, Joon Chae, Peter Dixon, Eun Jung Lee, Yoon Jung Lee, Wei Li, Richmond Mathews, Alberto Rossi, Shrihari Santosh, Austin Starkweather, Tonia Wang and participants at the University of Maryland seminars for their comments and discussions.

Table of Contents

List of Figures

Chapter 1: Can Short-term Trading Algorithms Anticipate Large Order Executions?

1.1 Overview

How trading algorithms affect financial markets has been largely debated. Current literature assesses the market influence of algorithms as a whole, overlooking the differences among the various types of algorithms and the complexity of their interactions. Taking this into account, we disaggregate the algorithms by their characteristics as the first step in the attempt to answer the question. Using account-level data, in which dynamics among all algorithms are presented, this paper asks whether short-term trading algorithms can anticipate the order flow of large order execution algorithms, and then examines the influence of this anticipation on each trading entity.

When building or unloading large positions within a short time, a large trader has an incentive to reduce price impact cost by hiding her order flow. To minimize the price impact cost within a set time constraint, theory might suggest that the large trader should split her order into smaller orders of similar-size. Such an execution strategy would not be optimal under a time constraint, because the early trades in the order flow sequence would be large enough to reveal the entire order flow to short-term traders trying to profit by trading ahead of the large trader. Furthermore, the large trader is not better off pursuing a strategy of executing just one large order to hide her order flow, as this would incur even larger price impact, increasing overall transaction costs.[1]

To examine the trade-off faced by large traders, we develop a simple two-period model based on Bertsimas and Lo (1998), in which a large trader decides how to split a large purchase or sale of a risky asset over two periods. In the model, there exists a short-term trader who receives a signal based on the order flow in period one and becomes informed about the large trader's second trade with a probability that is endogenous to the large trader's first trade size. In equilibrium, the large trader reveals her order flow to the short-term trader to the extent that the marginal benefits of smoothing out her trades offset the marginal costs of revealing her second trade to the short-term trader.

In the model, there is one risky asset, whose price follows an arithmetic random walk with a constant linear price impact. The large trader's objective is to minimize the transaction costs of demanding a large number of shares of the risky asset within two periods. The short-term trader anticipates the second trade of the large trader

[1]Reference Harris (1997) for the concept of order exposure.

and uses this knowledge to maximize his profits from the price impact caused by the large trader. This model produces the following order-anticipation dynamics between the large trader and the short-term trader: (1) Short-term traders anticipate the autocorrelated trades of large traders with a probability greater than random chance based on order flow information. (2) In the presence of short-term traders, when building large positions, it is optimal for large traders to trade small, positively correlated “child orders”, which are followed by a large trade. (3) Short-term traders exit their positions when large traders initiate a large trade. (4) If large traders trade with bigger “child orders” before initiating a large trade, their large trade following the “child orders” is more likely to be anticipated by short-term traders.

We document order-anticipation dynamics between large traders and short-term traders with comprehensive data that contains the complete dynamics among all traders at an account level in the Korea index futures market for 66 consecutive trading days beginning in March 26, 2009. In our empirical analysis, short-term traders are identified as traders whose inventories are strongly mean-reverting. This is defined as (1) having an average holding time per position of less than 3 minutes, (2) having an average daily ratio of overnight inventories to their own contracts traded of less than 0.01%, and (3) trading with an active order more often than with a passive order. Large traders are identified as traders who are not short-term traders and initiate at least one trade among the largest 1% of all trades during the sample period. All other traders are classified as "small traders". During the sample period, large traders initiate 36,164 such large trades. We find that, conditional of a large trade being executed, short-term traders correctly anticipate the direction of the large trades 56.06% of the time by taking a long position in advance of a large buy order or a short position in advance of a large sell order. This result suggests that short-term traders are informed about the direction of the forthcoming large trades with a probability greater than random chance and that they profit from the price impact caused by large traders.

A simple event study is used around the execution of large trades to describe order-anticipation dynamics among the large, short-term, and small traders. We find that, prior to initiating a large trade, large traders smooth out their large demand with small "child orders" in the direction of the forthcoming large trade. During this period of "child orders", short-term traders gradually increase their inventories in the direction of the large trade as if they are informed about the direction of the large trade. When large traders initiate a large trade, the price jumps because of the price impact of the large trade, and short-term traders are likely to be on the right side, profiting from the price impact caused by the large trade. As the large order arrives, short-term traders begin to liquidate their positions by trading against small traders who want to actively trade to respond to new information learned from the large trade.

We examine the order-anticipation horizon of short-term traders, using predictive regressions with 5 to 240 seconds time intervals. We find that in all predictive regressions, the aggregate positions of short-term traders are statistically sufficient and significant predictors for the direction of large trades that will arrive in a short period of time. Furthermore, in the predictive regressions, the coefficient of the aggregate positions of short-term traders monotonically increases as the time intervals become longer. This suggests that their order-anticipation strategy is not based on an extremely low latency.

Our empirical results support the order-anticipation dynamics implied by the model. When large traders build or unload large positions, they try to slowly accumulate their desired positions by splitting their large orders to minimize price impact by hiding their order flow from short-term traders. Because of limited liquidity and time constraints to fill their large orders, large traders encounter a trade-off between trading faster and not revealing their order flow to short-term traders. When scheduling their orders, large traders rationally expect that short-term traders extract some order flow information from their "child orders". Therefore, large traders build or unload their inventory with small "child orders" early on, and at the very last moment on their time constraint, they initiate a large trade to finish filling their large order. In equilibrium, large traders reveal their order flow information to short-term traders to the extent that the marginal benefits of smoothing out their trades offset the marginal costs of revealing their order flow to short-term traders. Therefore, before a large trade, short-term traders anticipate a sequence of autocorrelated trades initiated by large traders. The large trade then provides an exit point to short-term traders, who rationally expect that large traders finish executing their large orders with a large trade and that small traders will want to actively trade in the direction of the large trade to respond to new information learned from large traders.

In addition to the literature on order execution and order exposure, this paper is connected with multiple other strands of literature including the market microstructure invariance hypothesis proposed by Kyle and Obizhaeva (2013) and high frequency trading. Using an event study approach to analyze repetitive large trades, we can clearly describe order-anticipation dynamics among large, short-term, and small traders. When applying the event study methodology, we use a time invariant trading sequence instead of physical time based on the invariance hypothesis suggesting that "market microstructure properties become constant when measured in units of business time". This intuition from the invariance hypothesis is essential in our analysis since market microstructure noise is substantial in the physical time domain, and we can mitigate the noise by aggregating the data with time invariant trading sequence in the event study.

We use vector autoregression analysis to show that short-term traders predict short-term price changes by anticipating large trades initiated by large traders. In our vector autoregression, we partition the trade-by-trade data with large trades instead of a regular time interval. This approach is less subject to market microstructure noise than a general method since trading activity between large trades is controlled by the invariance hypothesis. According to the invariance hypothesis, "business time" runs differently from physical time. Therefore, trading activity in regular time intervals is not comparable because the level of risk transferred in regular time intervals is not homogeneous. However, the time domain partitioned by large trades is endogenous to trading activity since, if "business time" runs faster, large trades are more likely to arrive in the market, therefore the time span between large trades endogenously becomes smaller.

Our study is related to the high frequency trading (HFT) literature. There is a controversial debate on whether HFT firms can use their faster trading speed to trade ahead of institutions, thereby raising transaction costs for institutional investors; Brogaard et al. (2014), Clark-Joseph (2013), Hirschey (2013) and Li (2014). Although we do not take a stand on whether short-term traders identified in this paper are HFT firms, we see that short-term traders have trading patterns similar to those of HFT firms such as quickly mean-reverting inventories and low overnight inventories. We also find that short-term traders make consistent profits and that the size of "child orders" of large traders is the first order reason for the consistent order-anticipation trading of short-term traders. This finding raises an important issue in the HFT literature as "speed" may not be the first order reason for consistent HFT firms profits. We need to distinguish between HFT profitability resulting from a speed advantage versus order-anticipation trading based on public order flow information.

The next section proceeds as follows. A simple two-period model is introduced to provide intuition on the trade-off faced by large traders and to derive order-anticipation dynamics between large and short-term traders. We then describe the data and institutional background. Finally, using large trades as repetitive random experiments, we document the order-anticipation dynamics implied by the model.

1.2 Two-Period Model: Large Trader's Problem

When scheduling a large order for a risky asset, large traders face a trade-off between order execution speed and order execution cost. Faster order execution makes orders more expensive by making them easier to anticipate. To examine this trade-off faced by large traders, we introduce a short-term trader anticipating the order flow of large trader to the two-period model developed by Bertsimas and Lo (1998). In our economy, there is one riskless asset with zero interest rate and one risky asset, whose price follows an arithmetic random walk with a constant linear price impact. There exists one large trader and one short-term trader. The large trader demands a large number of shares of the risky asset that have to be executed within two periods, and her objective is to minimize transaction costs. The short-term trader is informed about the large trader's second trade with probability β, which is assumed to be proportional to the large trader's first trade size: $\beta = \alpha \cdot |y_1|$, where y_1 is the large trader's first trade, and α is a positive constant.

Let p_t, x_t and y_t denote price, short-term trader's trade and large trader's trade, respectively at time $t = 1, 2$.

For $t = 1, 2$, the price motion is

$$p_t = p_{t-1} + \lambda z_t + \varepsilon_t, \tag{1.1}$$

where $z_t = x_t + y_t$, and λ is a linear price impact factor, which is assumed to be a positive constant.

The short-term trader's problem is

$$\min_{x_1, x_2} \beta \mathbb{E}\left[p_1 x_1 + p_2 x_2\right], \qquad s.t. \quad x_1 + x_2 = 0. \tag{1.2}$$

We assume that the short-term trader exits his entire positions at $t = 2$. Although the model is fixed to end in two periods, in a real financial market, large trades would repeatedly arrive in the market, and short-term traders would consistently anticipate the direction of large trades, if possible. Since large traders may demand a risky asset in a different direction, whether long or short, short-term traders need to exit their positions to be ready for anticipating the next large trader.

Given the price motion p_t, we can rewrite the short-term trader's problem:

$$\min_{x_1} \beta\lambda \left(-x_1 + y_2\right) x_1.$$

When the short-term trader is informed about the second trade of large trader, his optimal trading is

$$x_1^* = -x_2^* = \frac{y_2}{2}. \tag{1.3}$$

The large trader's problem is

$$\min_{y_1, y_2} \left\{(1-\beta)\mathbb{E}\left[p_1 y_1 + p_2 y_2 | x_1 = 0\right] + \beta\mathbb{E}\left[p_1 y_1 + p_2 y_2 | x_1 = y_2/2\right]\right\}, \tag{1.4}$$

$$s.t. \quad y_1 + y_2 = \overline{Y}.$$

When deciding a demand schedule, the large trader should consider the possibility that the short-term trader anticipates her second trade. This is because order-anticipation trading by the short-term trader increases the total transaction costs of the large trader. Furthermore, the large trader should consider how her first trade affects the probability of her second trade being anticipated.

The large trader faces a trade-off between smoothing out her trades and hiding her second trade from the short-term trader. Without the order-anticipation of the short-term trader, it is optimal for the large trader to evenly split her large trades over two periods. With the order-anticipation of the short-term trader, it is optimal for the large trader to reduce her first trade size to decrease the probability of her second trade being anticipated. The expected transaction costs with and without the order-anticipation by the short-term trader are

$$\mathbb{E}\left[p_1 y_1 + p_2 y_2 \middle| x_1 = y_2/2\right] = \left(p_0 + \left(\overline{Y} - y_1\right)/2 + \lambda y_1\right) y_1 + \left(p_0 + \lambda \overline{Y}\right)\left(\overline{Y} - y_1\right) \tag{1.5}$$

$$\mathbb{E}\left[p_1 y_1 + p_2 y_2 \middle| x_1 = 0\right] = \left(p_0 + \lambda y_1\right) y_1 + \left(p_0 + \lambda \overline{Y}\right)\left(\overline{Y} - y_1\right). \tag{1.6}$$

We can rewrite the large trader's problem of minimizing expected transaction costs with and without the order-anticipation by the short-term trader such that

$$\min_{y_1} \left\{ \overline{Y}\left(p_0 + \lambda \overline{Y}\right) - \lambda y_1 \left(\overline{Y} - y_1\right) + \frac{\alpha y_1^2}{2}\left(\overline{Y} - y_1\right) \right\}, \tag{1.7}$$

where $\overline{Y}\left(p_0 + \lambda \overline{Y}\right)$ are the expected transaction costs when the large trader trades all shares at either time $t = 1$ or 2, $-\lambda y_1 \left(\overline{Y} - y_1\right)$ are the benefits of smoothing out her trades, and $\alpha y_1^2 \left(\overline{Y} - y_1\right)/2$ are the costs of revealing her second trade to the short-term trader. In equilibrium, the large trader chooses y_1 such that the marginal benefits of smoothing out her trades offset the marginal costs of revealing her second trade to the short-term trader. At the optimum, the large trader optimally chooses her demand schedule such that [2]

$$y_1^* = \frac{\overline{Y}}{3} + \frac{2\lambda - \sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}}{3\alpha}, \quad y_2^* = \frac{2\overline{Y}}{3} - \frac{2\lambda - \sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}}{3\alpha}. \tag{1.8}$$

The equilibrium is defined as the price motion $\{p_1^*, p_2^*\}$, the large trader's demand schedule for the risky asset $\{y_1^*, y_2^*\}$ and the short-term trader's order-anticipation strategy $\{x_1^*, x_2^*\}$ when he is informed about the large trader's second trade, y_2^*, and the probability of order-anticipation by the short-term trader $\beta^* = \alpha \cdot |y_1^*|$.

When the short-term trader is not informed about y_2^*, he cannot strategically trade to extract more information about the large trader's second trade because both the short-term trader and the larger trader are risk-neutral. If we introduce small amounts of trading fees proportional to the trading volume to the short-term trader, it is optimal for the short-term trader not to trade when he is not informed about the large trader's second trade.

The total demand for the risky asset of the large trader, $\overline{Y}$, is assumed to be a large number, which would determine the signs in the comparative statics analysis below. A few important comparative statics results are noted: (1) In equilibrium,

[2] There exist two solutions that satisfy the first order condition. Since the large trader minimizes the transaction costs, the solution chosen is the unique one with the second order condition being positive.

the short-term trader anticipates the large trader's second trade with a probability greater than random chance since $\beta > 0$ for all $\overline{Y}$. (2) In the presence of the short-term trader, the large trader trades a small trade, which is followed by a large trade: $y_2^* > y_1^*$. (3) If the short-term trader can more accurately extract order flow information from the large trader's first trade (i.e. α increases), the marginal costs of revealing the large trader's second trade increase. Therefore, the large trader reduces her first trade size while increasing her second trade size:

$$\frac{\partial y_1^*}{\partial \alpha} = -\lambda \cdot \frac{\overline{Y}\alpha - 4\lambda + 2\sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}}{3\alpha^2\sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}} < 0, \quad \frac{\partial y_2^*}{\partial \alpha} > 0.$$

(4) If the price impact factor becomes larger, the marginal benefits of smoothing out trades increase. Then the large trader balances more evenly her trade size between her first and second trade:

$$\frac{\partial\left(y_2^* - y_1^*\right)}{\partial \lambda} = -\frac{4}{3\alpha} + \frac{-2\overline{Y}\alpha + 8\lambda}{3\alpha\sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}} < 0.$$

(5) If the large trader demands larger liquidity, her demands are more concentrated on the second trade:

$$\frac{\partial\left(y_1^*/\overline{Y}\right)}{\partial \overline{Y}} = -\lambda \cdot \frac{\overline{Y}\alpha - 4\lambda + 2\sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}}{3\overline{Y}^2\alpha\sqrt{\overline{Y}^2\alpha^2 - 2\overline{Y}\alpha\lambda + 4\lambda^2}} < 0.$$

The model produces the following order-anticipation dynamics: (1) In equilibrium, when large traders build or unload large positions within a short period of time, short-term traders can anticipate the autocorrelated trades of large traders with a probability greater than random chance based on order flow information. (2) In the presence of short-term traders, when building large positions within a short period of time, it is optimal for large traders to trade small “child orders”, which are followed by a large trade. (3) Short-term traders exit their positions when large traders initiate a large trade. (4) If large traders trade with bigger “child orders” before initiating a large trade, their large trade is more likely to be anticipated by short-term traders.

The model has a few limitations. First, the total demand $\overline{Y}$ of the large trader is not endogenous to the price. In theory, large traders should endogenously adjust their total demand, depending on the price changes. In a real financial market, there are many cases in which large traders cannot change their demand, especially in a short time. For instance, large traders may delegate the execution of large demands to an execution agent, which will execute the requested large order within a time constraint. Also, within an investment bank, deciding total demand could be separated from executing it in that portfolio managers decide total demand of the risky asset, and an execution department manages details of small orders which fill the large order. Moreover, since we are modeling an execution decision over one to two minutes, large traders have limited ability to react to a price change. Another limitation is that the model forces the large trader to trade over two periods. In a real financial market, large traders trade over a multiple periods. To apply the model to a real financial market, we may consider large trader's first trade as all "child orders" before large traders initiate a large trade, and her second trade as a large trade belonging to the largest 1% of all trades during the sample period. Third, there is no competition among large traders and short-term traders in the model. It would be an interesting extension if we introduce multiple agents of large traders and short-term traders competing on liquidity with correlated information.

1.3 Institutional Background and Data Description

The data is from the Korea Exchange (KRX hereafter). This section discusses the market conditions and rules in the KRX as well as the data descriptions that give us a unique opportunity to examine the complete dynamics among large, short-term, and small traders.

The KRX is an automated centralized electronic market based on a limit order book; it is the sole exchange that houses both the stock and derivatives market in South Korea. Compared to the U.S. markets, its size is small but its intraday trading is very active relative to the size. According the 2009 statistics published by the World Federation of Exchanges, the market capitalization of the KRX reaches one trillion USD, which is nearly 7% of NYSE's, while the KRX has relatively high daily turnover of approximately 10% of NYSE's. The intraday trading in the index futures market is 5 times more active than in the stock market; traders in the KRX hold 106,151 contracts of daily open interest, which mark 3% of those of the CME; however, the notional value of daily trading volume takes nearly 20% of CME's.

The data contains the complete records of trades and quotes time-stamped at one millisecond with an encrypted account identification for 66 consecutive trading days beginning in March 26, 2009.[3] When several events occur in the limit order book during the same millisecond, the order of the events is recorded in the proper sequence in the data. Therefore, we can observe the complete dynamics of the limit order book at an individual trader level. For instance, the data records a time-stamp of the times of when a message is submitted and when an order is matched. Therefore, when a trader submits a limit order, we can determine when this limit order is matched or canceled at a millisecond precision. Furthermore, by comparing the times of when a buyer and a seller sent their messages for each trade, it is possible to accurately identify the trader that initiated the trade. That is, given a trade, if the buyer sent a message later than the seller, then such a trade is a buyer-initiated trade, which implies that the buyer crossed the bid-ask spread, and bought at the ask price.

Another unique feature of this data is that we can identify whether a trader

[3]During this period, the KRX did not provide a collocation service to any trader.

is an institution or a retail investor and also determine if it is a foreign or domestic investor from the perspective of South Korean. Although we cannot rule out the possibility that a domestic investor opens an account outside of South Korea so that he is classified as a foreign investor, it is likely that foreign investors are foreign investment banks, mutual funds or hedge funds that are actively trading in the KRX.

Our analysis focuses on the KOSPI 200 index futures market for the following reasons. First, the KOSPI 200 index futures contract is one of the most liquid index futures contracts in the World. Second, the underlying asset is a well-diversified index, which is the KOSPI 200 index, a basket of two hundred major stocks listed in the KRX. Therefore, there is not much idiosyncratic risk involved in trading the index futures contract. This implies that an idiosyncratic shock from an individual stock does not affect the index futures price to a large extent. Instead of idiosyncratic risk in an individual stock, macro news such as interest rate changes, Chinese economic growth forecast, etc. are major determinants of significant price changes in the KOSPI 200 index futures contract.

An open outcry market does not exist for the KOSPI 200 index futures contract; all contracts are traded electronically. In 2009, 83 million KOSPI 200 index futures contracts were traded. Unlike the E-mini S&P 500 index futures contracts, which are traded mostly by institutions, retail investors provide substantial liquidity to the KOSPI 200 index futures market. The substantial trading volume from retail investors allows us to examine the interactions among large, short-term traders and small traders who are mostly retail investors.

The notional value of one KOSPI 200 index futures contract is KOSPI 200 futures price times a multiplier of 500,000 Korean Won (KRW). The average notional value of one contract during our sample period is USD 67,779, which is higher than that of the E-mini S&P 500 index futures contract.[4] Its tick size is 0.05, which is about USD 19.37 or 2.86 basis points.

We analyze only the front month contracts. June 11, 2009, is the only expiration date. Therefore, until June 11, 2009, we use the data of the June 2009 contract, and after June 11, 2009, we use the September 2009 contract for our analysis. Since the back month contracts were illiquid except for a few days right before or on the expiration date, including the data on the back month contracts in our analysis does not qualitatively change the results in this paper.

The daily price limit on the KOSPI 200 futures contracts is plus and minus 10% of previous closing price. There are few market conditions severe enough to trigger a circuit breaker. During the sample period, the price fluctuated within the daily price limit and a circuit breaker never came into effect.

[4]The average closing price is 174.97, and the average exchange rate is USD/KRW 1,290.73 during our sample period.

1.4 Identifying Large, Short-term, and Small Traders

The data tracks all traders in the market of 25,172 traders. We identify short-term traders and large traders based on their trading records. Short-term traders are identified by the following criteria: (1) An average holding time for one position is less than 3 minutes.[5] (2) An average daily ratio of overnight inventories to their own contracts traded is less than 0.01%. (3) The number of contracts with a marketable order is greater than the number of contracts with a non-marketable order.

Among all traders, excluding short-term traders, we identify a "large trader" as a trader who initiated at least one large trade that belongs to the largest 1% of all trades during the sample period. All other traders are classified as "small traders".

During the sample period, the average daily volume was 348,114 contracts and the notional value was USD 24 billion. Based on the three criteria above, of the 25,172 traders, 3% are classified as either short-term traders or large traders, which initiate 72% of daily volume. Large traders initiate 14% of daily volume with a large trade. When a large trade is initiated, large traders trade against non-marketable orders of large, short-term, and small traders by 46.05%, 10.22%, and 43.73%, respectively.

We observe 32 short-term traders who switch their positions as frequently as

[5] We define one position as a sequence of trades that maintain the same sign of the inventories. For example, a long position starts from a zero-inventory, and retains the position as long as the inventories are positive, and ends with the next zero-inventory.

Panel A. Trader Entity			
	# Total	# Foreign	# Institution
Short-term Trader	32	5	24
Large Trader	737	179	557
Small Trader	24,403	391	3,057
Total	25,172	575	3,638

Panel B. Volume Ratio						
	Volume(%)		Large Trade(%)		Small Trade(%)	
Trader	Take	Make	Take	Make	Take	Make
Short-term Trader	30.48	14.51	0.00	10.22	34.76	15.12
Large Trader	42.13	43.18	100.00	46.05	34.01	42.77
Small Trader	27.39	42.31	0.00	43.73	31.23	42.11
Daily Volume	348,114 (100%)		49,671 (14%)		298,443 (86%)	

Panel C. Other Statistics			
	Mean	Median	Std.
# of Switch/Day			
Short-term Trader	91.10	58.53	92.83
Large Trader	5.09	1.45	18.67
Small Trader	4.32	2.16	10.00
Switch Time(sec)			
Short-term Trader	104.89	81.96	115.76
Large Trader	9,607.83	10,014.12	5,135.44
Small Trader	5,860.38	4,730.69	5,080.46
Overnight Ratio(%)			
Short-term Trader	0.00	0.00	0.02
Large Trader	40.33	29.80	36.47
Small Trader	21.40	4.44	32.86

Table 1.1: This table reports summary statistics for large, short-term, and small traders. The total volume consists of small trade (%) and large trade (%), which are expressed as the proportion to the total volume. Large trades are defined as active trades by large traders among the largest 1% of all active trades. We define one position as a sequence of trades that retain the same sign of the inventories. The number(#) of Switch/Day is daily average number of position changes such as changing from a long to short position or vice versa. Switch Time(sec) is mean holding time for one position. Overnight Ratio(%) is daily average ratio of overnight inventory to whole day trading volume by each trader.

91.10 times per day. Their average holding time for one position is 104.89 seconds. They liquidate most of their inventories at the end of day. This leads to the average low overnight inventory ratio of 0.004%.

We identify 737 traders as large traders. They are long-term investors compared to short-term traders, since they hold their positions for 2.7 hours on average. They tend to take large directional positions, keeping a high overnight inventory ratio compared to short-term traders. On average, they keep 40.33% of intraday trading volume as overnight inventory. Large traders may be an execution algorithm, index arbitrage, or portfolio insurance program that occasionally execute a large trade.

We conjecture that small traders are similar to noise traders in Kyle (1985) since their trading volumes are small compared to those of short-term traders and large traders, and 87% of small traders are domestic retail investors who may trade for exogenous reasons.

1.5 Large Order Executions

Executing large trades is an economically significant event in the market for the following reasons. First, large trades incur substantial price impact. Second, the direction of large trades is uncertain to market participants except for the one who initiates them. Traders are likely to be on the wrong side of large trades unless they can consistently anticipate the trading direction of large trades. Thus, large trades

Variable	Mean	Median	Std.	Max	Min
Large Trade Size	78.76	62.00	40.74	800.00	50.00
# Large Trade per Day	547.94	531.00	109.30	951.00	334.00
Time btwn Large Trades (sec)	39.97	16.31	65.17	1,239.29	0.01
Volume btwn Large Trades	1,074.79	638.00	1,287.64	24,523.00	1.00
# Trades btwn Large Trades	489.33	281.00	612.58	11,253.00	1.00
# Message btwn Large Trades	848.49	471.00	1,108.84	22,220.00	0.00

Table 1.2: This table reports summary statistics of large trades. Large trades are defined as active trades by large traders among the largest 1% of all active trades. We observe 36,164 of large trades, among which 48.99% are buyer-initiated and 51.01% are seller-initiated. Buyer-initiated trade is a trade that the buyer crossed the bid-ask spread and bought at the ask price. Similarly, seller-initiated trade is a trade that the seller sold at the bid price.

cause negative skewness in the profits distribution of traders who are on the wrong side.

Using the trade-by-trade data in the KOSPI 200 index futures market, we define a "large trade" as an active trade by large traders among the largest 1% of all active trades in the sample period. We observe 36,164 of such large trades during the sample period. This paper analyzes 200 trades before and after large trades to document the order-anticipation dynamics of large, short-term, and small traders. Two hundred trades occur in approximately 1 minute.

The total number of large trades during the sample period is 36,164. The minimum trade size to be considered as a large trade is 50 index futures contracts, which have a notional value of 3.4 million USD. The mean size of large trades is 78.76 contracts. Of the 36,164 large trades, 48.99% are buyer-initiated, meaning that a large trader crossed the bid-ask spread and bought at least 50 contracts at the ask price. Large trades occur 547.94 times per day on average. The average time between two consecutive large trades is 39.97 seconds, during which 1,074.79 contracts are traded while 489.33 trades and 848.49 messages occur on average.

The number of large trades is distributed over the intraday trading hours in a U-shape. The directions of large trades are positively autocorrelated. Panel A of figure 1.1 plots the total number of large trades during the sample period in 10 minute intervals between 9:00 a.m. and 3:00 p.m. Panel B shows that the directions of large trades are positively autocorrelated with the directions of up to 10 previous large trades.

We use an event study to examine buying pressure, selling pressure, and price changes around the execution of large trades. We calculate the mean active trades and the mean relative price that have the same trading sequence around large trades. For example, denote the price as $p\,(i,j)$, where i indexes large trades, and j indexes the trading sequence around the i^{th} large trade. Let's define the relative price such that

$$p'\,(i,j) = \{lnp\,(i,j) - lnp\,(i,0)\} \times 10^4,$$

where $p\,(i,0)$ is price at the i^{th} large trade. The mean relative price $p'\,(j)$ can be

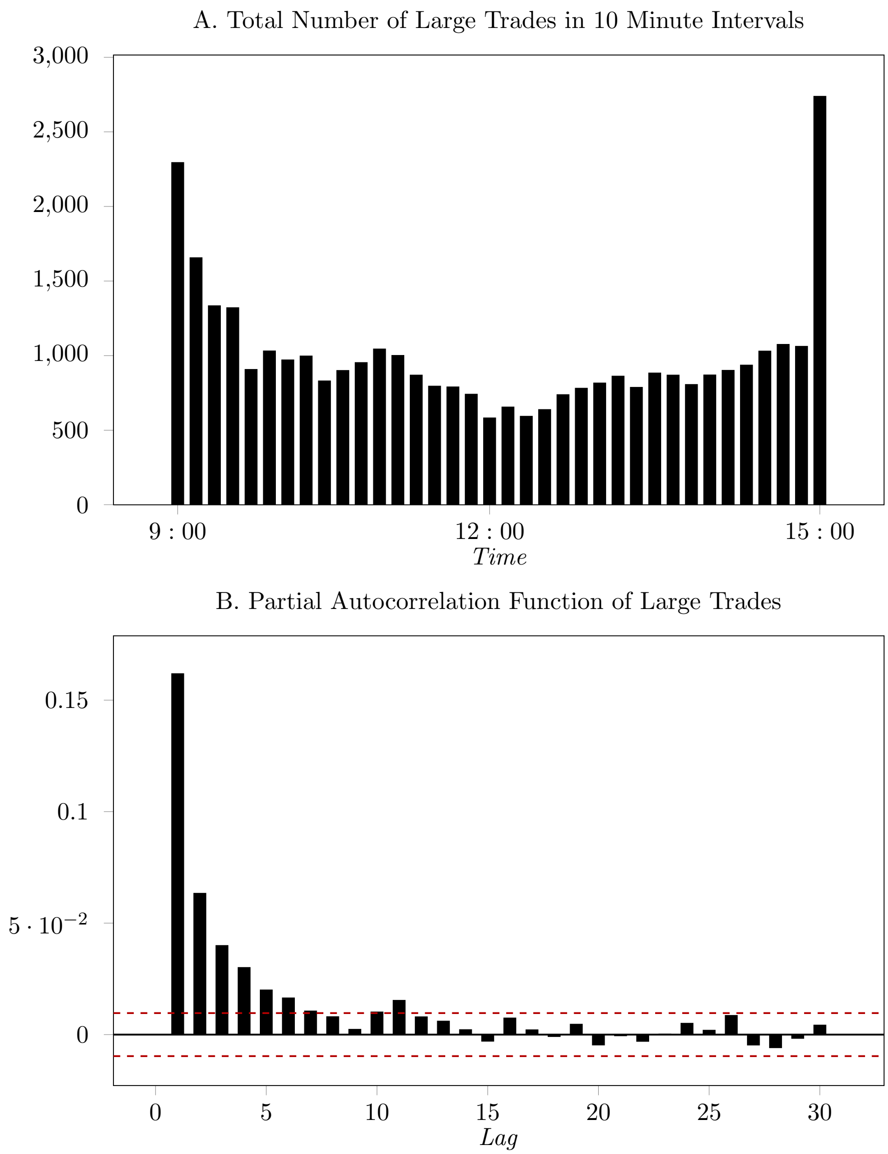

Figure 1.1: Large trades are defined as active trades by large traders among the largest 1% of all active trades. We observe 36,164 large trades in the KOSPI 200 index futures market from March 26 to June 29, 2009. Panel A plots the total number of large trades during the sample period in 10 minute intervals between 9:00 a.m. and 3:00 p.m. Panel B shows the partial autocorrelation function of large trades. The sign and size of large trades represent the trading direction and volume of large traders at each time when large traders initiate a trade.

computed as

$$p'(j) = \frac{1}{T}\sum_{i=1}^{T} p'(i,j),$$

where j indexes the trading sequence around large trades. Note that a negative value j indicates a trade prior to large trades and a positive value j indicates a trade after large trades. Due to difference in their dynamics, we aggregate buyer-initiated large trades and seller-initiated large trades separately. The mean active trades and the mean relative price are calculated by their trading sequence around large trades.

Figure 1.2 presents the mean active trades and the mean relative price around large trades. The x-axis is the time invariant trading sequence centered at large trades and the y-axis is either the mean active trades or the mean relative price.

In panel B of figure 1.2, the relative price increases slowly and monotonically as the trading sequence approaches the execution of the large buyer-initiated trades. The price jumps instantly at the execution, and the price is maintained for up to 200 trades after the execution.

Current literature provides a partial explanation for this price pattern. Gradually increasing price pattern before the execution point simulates the pattern generated by the order splitting of the informed trader (Kyle, 1985). Namely, the trader with monopolistic information splits orders in building a large position while hiding his information, resulting in gradual price increase. Therefore, the model of informed

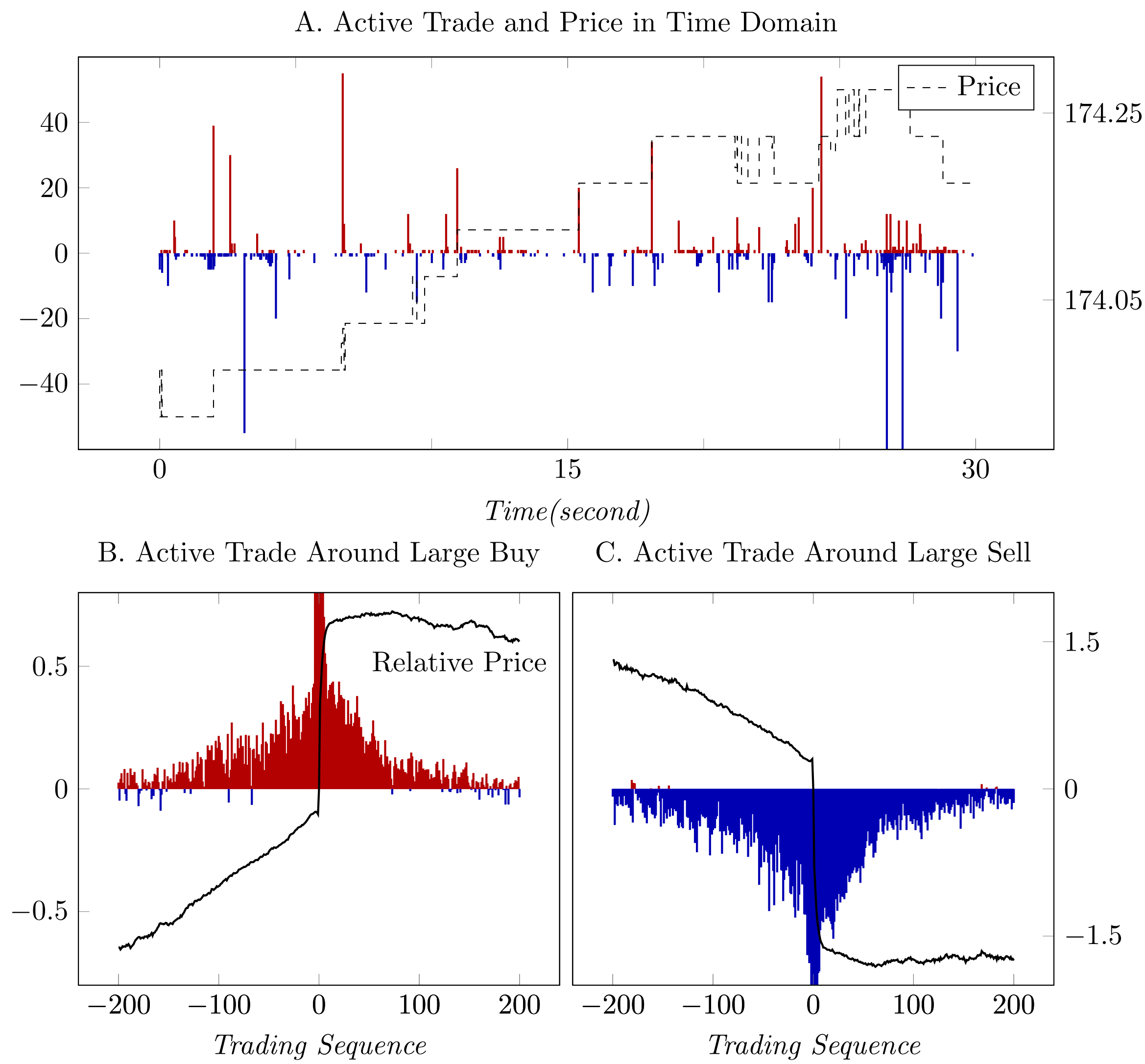

Figure 1.2: Large trades are defined as active trades by large traders among the largest 1% of all active trades. We observe 36,164 large trades in the KOSPI 200 index futures market from March 26 to June 29, 2009. The buyer-initiated trades and seller-initiated trades account for 48.99% and 51.01% of large trades, respectively. Panel A plots the signed active trades and prices on the time domain for 30 seconds. Panels B and C plot the mean active trades and the mean relative price that have the same trading sequence around large trades. Panels B and C aggregate the buyer-initiated trades and seller-initiated trades, respectively. The relative price $p'(i,j)$ is defined as $\{\ln p(i,j) - \ln p(i,0)\} \times 10^4$, where i indexes large trades and j indexes the trading sequence around the i^{th} large trade.

trader does not anticipate neither the large order execution nor the price jump.

In order to explain the price jump occurred by large order executions, we separate the traders that initiated the large orders from other traders, and track their trading behavior around large order executions. In addition, we decompose remaining traders by short-term trader and small traders, and compare their behaviors with that of large traders.

1.6 Order-Anticipation Dynamics

We analyze order-anticipation dynamics around the execution of large trades using a simple event study. The traders are decomposed into three groups: short-term traders, small traders and large traders who initiated large trades. We calculate the average inventories of the 3 groups of traders around large trades by the direction of large trades and that of aggregate positions of short-term traders.

Figure 1.3 presents the average inventories of large, short-term, and small traders as a function of trading sequence centered at large trades. Figure 1.4 plots the average relative prices as a function of trading sequences centered at large trades. In figure 1.3 and 1.4, the panels for the seller-initiated large trades are symmetric with those for the buyer-initiated large trades.

Figure 1.3 compares 4 combinations generated by two directions of large trades and two aggregate positions of short-term traders. The aggregate positions of short-

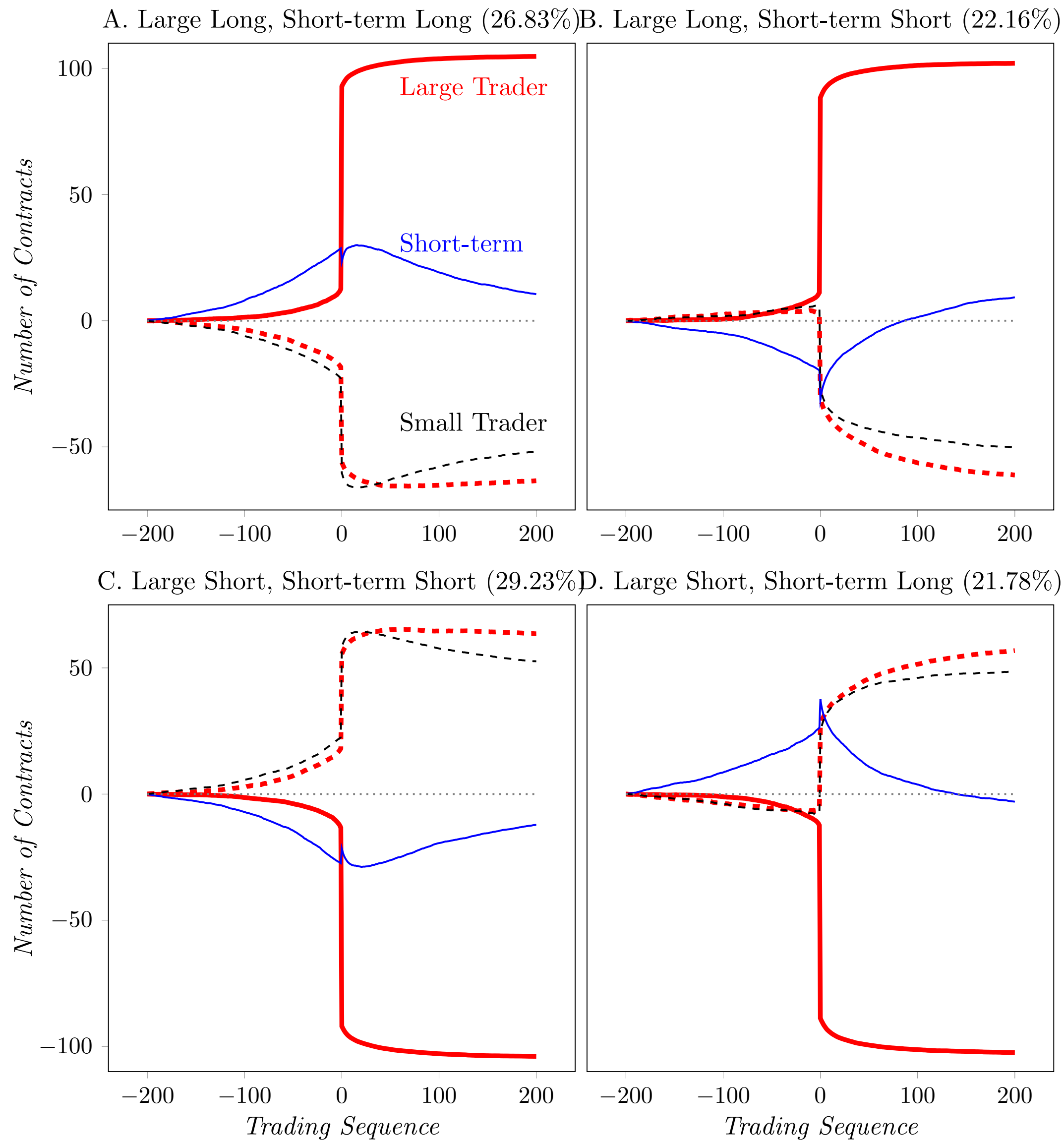

Figure 1.3: Using an event study, this figure plots the average inventories of large, short-term, and small traders around large trades. We analyze 36,164 large trades during the sample period. The x-axis is the time invariant trading sequence centered at large trades, and the y-axis is the number of contracts. Two hundred trades occur in approximately 1 minute. The thick dashed line represents the average inventories of passively traded large traders. Panels A and C plot the average inventories when short-term traders correctly anticipate the direction of large trades. Panels B and D plot the average inventories when short-term traders incorrectly anticipate the direction of large trades. The ratio in the title indicates the proportion of each panel among total large trades.

term traders are likely to be on the right side of large trades. When large traders initiate a large buy order, the aggregate positions of short-term traders are long 54.77% of the time. Similarly, when large traders initiated a large sell order, the aggregate positions of short-term traders are short 57.3% of the time.

When short-term traders are on the right side of large trades, the aggregate positions of short-term traders gradually increase as if they are informed about the direction of the forthcoming large trade approximately 200^{th} trade prior to large trades (Panels A and C of figure 1.3). At the execution of large trades, short-term traders trade against large traders instantly, and their positions shrink. Short-term traders recover some of their positions after the immediate response. They slowly exit their positions entirely by unloading the positions to small traders and other large traders who demand liquidity to respond to new information from large trades.

Figure 1.3 also details how large traders build their positions. Large traders manage their orders by executing small, positively correlated trades, which are followed by a single large trade. This order execution pattern is consistent with the strategy of large traders trying to reduce the price impact of their large demand in the presence of short-term traders, thereby minimizing overall transaction cost.

Figure 1.4 tracks relative prices around large trades. When short-term traders are on the right side of large trades, the price slowly moves toward the price at

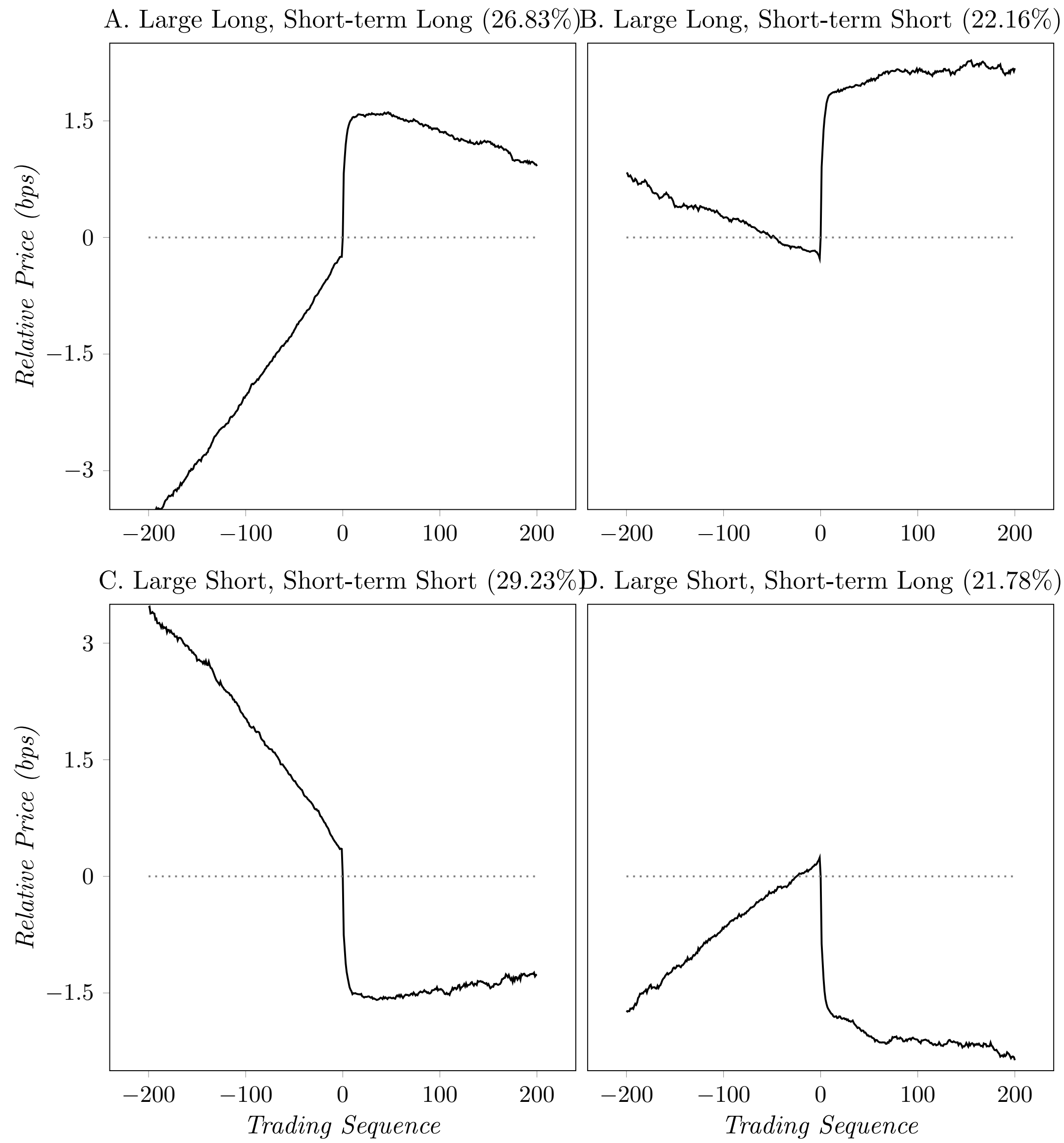

Figure 1.4: Using an event study, this figure plots the average relative prices around large trades. We analyze 36,164 of large trades during the sample period. The relative price $p'(i,j)$ is defined as $\{\ln p(i,j) - \ln p(i,0)\} \times 10^4$, where i indexes large trades and j indexes the trading sequence around the i^{th} large trade. The x-axis is the time invariant trading sequence centered at large trades. The y-axis is the relative price(bps). Two hundred trades occur in approximately 1 minute. Panels A and C plot the average relative prices when short-term traders correctly anticipate the direction of large trades. Panels B and D plot the average relative prices when short-term traders incorrectly anticipate the direction of large trades. The ratio in the title indicates the proportion of each panel among total large trades.

the execution of large trades (Panels A and C). This price pattern is generated by demand pressure from both large traders splitting their trades and short-term traders anticipating the order flow of large traders. Their competition for liquidity contributes to the price discovery by making the price converge to the fundamental value at the execution of large trades.

Panels B and D in Figure 1.3 and 1.4 track three groups of traders when short-term traders are on the wrong side of large trades. The average inventories in Figure 1.3 shows that short-term traders quickly reverse their inventories when they are on the wrong side of large trades. We conjecture that such behavior is the strategy taken by short-term traders to minimize their losses, because they expect that small traders will trade actively to respond to large trades.

Such behavior is distinguished from the behavior of short-term traders when they are on the right side of large trades. Short-term traders do not unload their positions quickly since they expect that small traders will actively trade in the direction of large trades. They are better off exiting their positions with a passive order instead of an active order to maximize their profits.

Short-term traders face trade-off between favorable price and order information accuracy when they build positions. They are better off building their positions quickly before their competitors push the price. However, early position building increases the risk of being on the wrong side of large trades, because they are less informed about the order flow. Therefore, trading speed prior to large trades is not as critical as that after large trades, especially when short-term traders are on the wrong side.

In summary, when large traders build or unload large positions within a short period of time, they want to slowly accumulate their desired positions by smoothing out their large order to minimize price impact and to hide their order flow from short-term traders. Theoretically, such as the informed trader in Kyle (1985), large traders want to perfectly smooth out their trades, but in a real financial market, because of competition with other large traders who have correlated information, and because of limited liquidity, large traders have a time constraint on their order, within which they have to fill their large order. Under this condition, large traders have a trade-off between smoothing out their trades and hiding their order flow from short-term traders. Large traders rationally expect that short-term traders extract their order flow information from their early trades. Therefore, they shift some demand in their earlier trades to later trades as shown in the model, in which the large trader shifts some shares of the first trade to the second trade. At the very last moment on their time constraint, large traders have to initiate a large trade to fill their large order. In equilibrium, large traders reveal their order flow information to short-term traders to the extent that the marginal benefits of smoothing out trades offset the marginal costs of revealing their order flow to short-term traders. Therefore, near the execution of large trades, there are short-term traders anticipating a sequence of autocorrelated trades of large traders. Large trades provide an exit point to short-term traders as short-term traders rationally expect that large traders fill their large order with a large trade.

To formally test whether short-term traders anticipate large trades, we run a simple regression to test the null hypothesis that the probability of short-term traders being on the right side of large trades is less than or equal to 50%. Let y_t be the sign of the t^{th} large trade: Plus 1 is a buyer-initiated trade, and minus 1 is a seller-initiated trade. Let $x_{i,t}$ be the sign of positions of short-term trader i at the t^{th} large trade: Plus 1 is a long position, and minus 1 is a short position.[6]

$$y_t = \beta_0 \cdot x_{i,t} + \sum_{j=1}^{15} \gamma_j \cdot y_{t-j} + \alpha_i + \varepsilon_t, \tag{1.9}$$

where α_i represents a fixed effect of short-term traders. The null hypothesis is that $\beta_0 \leq 0$. To control for the autocorrelation of large trades, we include lagged large trades. We choose the number of lags based on the partial autocorrelation function of large trades in figure 1.1.

[6]The subscript i indexes short-term traders.

Model		Estimate	S.E.	t value	Pr(≤ |t|)	Adj. R^2
Panel A. Direction of Trade and Position						
Simple OLS	(Intercept)	−0.021	0.002	−10.005	0.000	0.004
	$x_{i,t}$	0.063	0.002	29.493	0.000	
Fixed Effect	$x_{i,t}$	0.064	0.002	29.617	0.000	0.004
Fixed Effect	$x_{i,t}$	0.036	0.002	16.766	0.000	0.041
+ Lag y	$\sum_{j=1}^{15} y_{t-j}$	0.365	0.006	61.817	0.000	
Panel B. Direction × Size of Trade and Position						
Simple OLS	(Intercept)	−1.023	0.019	−5.371	0.000	0.004
	$x_{i,t}$	0.085	0.003	29.112	0.000	
Fixed Effect	$x_{i,t}$	0.085	0.003	29.092	0.000	0.004
Fixed Effect	$x_{i,t}$	0.050	0.003	17.150	0.000	0.040
+ Lag y	$\sum_{j=1}^{15} y_{t-j}$	0.350	0.006	58.651	0.000	

Table 1.3: This table tests the null hypothesis that the probability of short-term traders being on the right side of large trades is less than or equal to 50%. The dependent variable is y_t, which is the direction of the t^{th} large trade: Plus 1 is a buyer-initiated trade and minus 1 is a seller-initiated trade. The subscript t indexes large trades. The independent variable is $x_{i,t}$, which is the position of the i^{th} short-term trader at the t^{th} large trade: Plus 1 is a long position and minus 1 is a short position. There are 32 of short-term traders and 36,164 of large trades. The number of observation is 217,583.

The regressions in panel A in table 1.3 reject the null hypothesis $\beta_0 \leq 0$. This implies that the probability of short-term traders being on the right side of large trades is statistically strictly greater than 50%, and that short-term traders can anticipate large trades initiated by large traders.

Regression (1.9) is not a spurious regression because both the dependent variable and the independent variables are a stationary time series and there is no cointegration between them. That is, y_t is the sign of the t^{th} large trade, which is a stationary time series, and $x_{i,t}$ is the sign of the positions of short-term trader i at the t^{th} large trade, which is a strongly mean-reverting process. This variable indicates whether short-term traders actively trade and take a long or short position prior to the t^{th} large trade. Since short-term traders switch their positions as frequently as 91 times per day, either from a long to short position or vice versa, the sign of positions of short-term traders at the t^{th} large trade is a stationary process.

Regression (1.9) is a predictive regression. If $x_{i,t}$ indicates a direction of short-term traders, β_0 implies the likelihood that short-term traders correctly anticipate the direction of large trades.

If we use the actual size and the direction of large trades along with the actual positions of short-term traders instead of $+1$ or -1 for y_t and $x_{i,t}$ in regression (1.9), the signs and their statistical significance are the same with those from the regression with +1 and −1. See panel B in table 1.3. However, the regression with +1 and −1 is more meaningful in that β_0 has a simple interpretation associated with the probability of short-term traders being on the right side of large trades such that

$$\begin{aligned} \mathbb{P}\left[y_t = x_{i,t}\right] &= \mathbb{E}\left[y_t \cdot x_{i,t}\right]/2 + 1/2 \\ &\approx \beta_0/2 + 1/2 = 52.05\%. \end{aligned}$$

Based on the regression result in table 1.3, short-term traders are on the right side of large trades with a probability 53.50% when not controlling for the autocorrelation of large trades, and 52.05% when controlling for the autocorrelation.

We do not drop any large trades that were initiated by large traders during trading hours except the beginning and closing times. As long as a short-term trader has an open position when a large trader initiate a large trade, this is included as an observation in regression (1.9)

Short-term traders consistently anticipate the direction of large trades with a probability greater than 50%. We run regression (1.9) with every two consecutive trading days of the data, and plot the time series of β_0 in figure 1.5.

Figure 1.5 demonstrates that the probability of short-term traders being on the right side of large trades is consistently greater than 50% over the sample period.

To illustrate the economic significance of order-anticipation strategy by short-term traders, let's simply assume that short-term traders lose one tick price ($20)

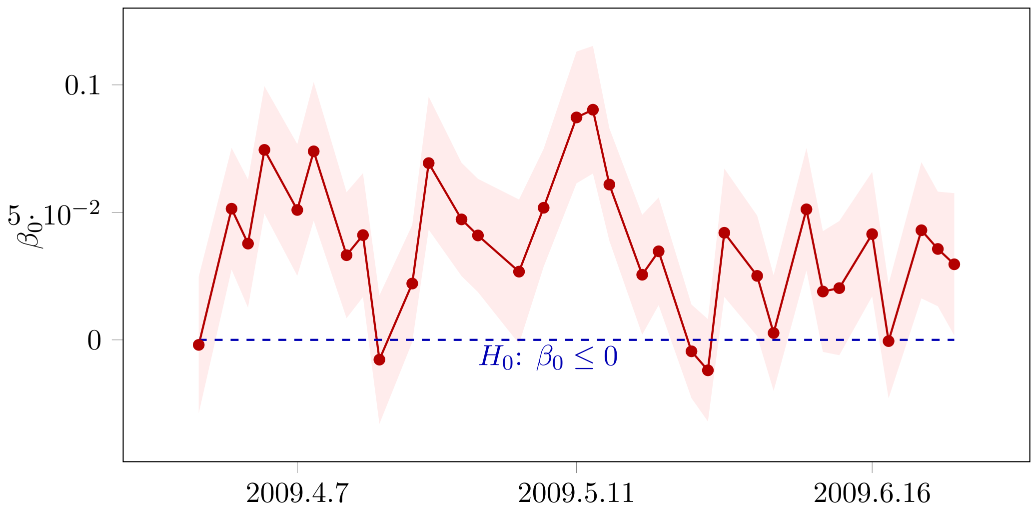

Figure 1.5: This figure plots the time series of β_0 estimated in two-day interval during the sample period: $y_t = \beta_0 \cdot x_{i,t} + \sum_{j=1}^{15} \gamma_j \cdot y_{t-j} + \alpha_i + \varepsilon_t$, where the subscript t indexes large trades, and y_t is the direction of the t^{th} large trade: Plus 1 is a buyer-initiated trade and minus 1 is a seller-initiated trade. The variable $x_{i,t}$ is the position of the i^{th} short-term trader at the t^{th} large trade: Plus 1 is a long position and minus 1 is a short position. Our null hypothesis is $\beta_0 \leq 0$, which implies that the probability of short-term traders being on the right side of large trades is less than or equal to 50%. The shade area represents a 95% confidence interval. The coefficient β_0 is related to the probability of short-term traders being on the right side of large trades such that $\mathbb{P}[y_t = x_{i,t}] \approx \beta_0/2 + 1/2$

if they are on the wrong side, and gain one tick price if they are on the right side of large trades. Also, let's assume that short-term traders trade just one futures contract in their order-anticipation trading strategy. There are 548 large trades per day on average. Short-term trader's net profits per day would be $548 \times (0.52 - 0.48) \times 1$ tick $\times$ 1 contract $\times$ multiplier $= 548 \times 0.04 \times 20\,\text{USD} = 438\,\text{USD}$. These profits would be consistent across trading days. If short-term traders can replicate their order-anticipation strategy with 50 contracts instead of one contract, the total profits of all 32 short-term traders would be $438 \times 50 \times 32 = 700{,}800$ USD per day. Short-term traders may attemp similar strategies not only in futures markets but also in options markets, and they may use their strategies in other international markets. The profits can become economically significant.

Short-term traders make consistent profits with a positive skewness. We aggregate the mark-to-market profits of all short-term traders in one hour intervals during our sample period, and normalize them to have a standard deviation of one. Figure 1.6 is the histogram of the mark-to-market profits aggregated across all short-term traders.

1.6.1 "Child Order Size" and Order Exposure Probability

The order-anticipation dynamics in the model is mainly driven by the assumption that, if the large trader increases her first trade size, the short-term trader is

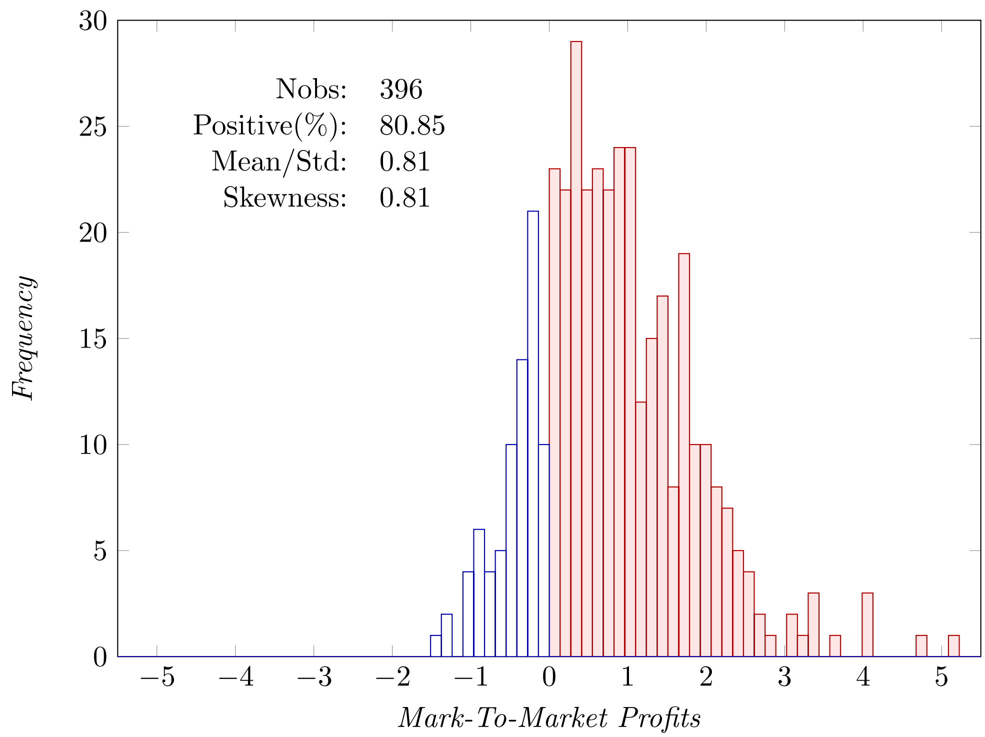

Figure 1.6: This figure plots the histogram of mark-to-market profits aggregated across all short-term traders in one hour interval during the sample period. The mark-to-market profits are normalized to a standard deviation of one.

more likely to be informed about her second trade. To validate this assumption, we take the data to the model as follows. We consider the large trader's second trade as large trades belonging to the largest 1% of all trades during the sample period, and consider the large trader's first trade as all small "child orders". For each large trade, we compute the mean size of small "child orders", and test whether short-term traders are more likely to anticipate a large trade initiated by a large trader who traded with bigger "child orders" before initiating the large trade.

As a competing hypothesis, we test whether the trading speed of large traders affects the probability of large trades being anticipated by short-term traders. The trading speed of large traders is measured with order matching and submission times in all passive trades of large traders. If large traders are faster, they are more likely to be in the front of queue in the limit order book by responding to market conditions faster than others. Therefore, the difference between order matching and submission times would be smaller for faster large traders. To measure large trader's speed, we take the mean differences between order matching and submission times for all matched passive orders of the large trader. The reason to use only the passive orders is that the difference between order matching and submission times in an active trade is subject to the latency within the exchange servers instead of the latency between the exchange servers and large traders.

The idea that "child order" size affects the probability of order-anticipation is not new. Harris (1997) argues that "the aggregate order size may then attract a costly response from other traders". The market makers in Kyle (1985) respond to the trade size of the informed trader by making the market thin when the market makers expect that the informed trader will demand large liquidity. The idea that the trading speed matters in order-anticipation by short-term traders is not new, either. Clark-Josep (2012) and Hirschey (2013) argue that HFT firms can trade ahead of others with their speed advantage. Li (2014) develops a theoretical model in which HFT firms can anticipate the order flow of all other traders due to their speed advantage. What we are testing in this section is whether "child order size" or "speed" affects the probability of short-term traders being on the right side of large trades.

The ideal experiment to test whether "child order size" or "speed" affects the probability of order-anticipation by short-term traders would be a random experiment, in which "child order size" and "speed" are randomly assigned to otherwise identical large traders (or their characteristics except for "child order size" and "speed" are randomly distributed), and compare the order exposure probabilities of the trader groups that have various "child order size" and "speed". The first issue to implement the ideal experiment is that "child order size" and "speed" may be positively correlated if traders with more capital are more likely to invest in increasing their trading speed. To address the correlation between "child order size" and "speed", we test the null hypotheses associated with "child order size" and "speed", both separately and jointly. The second issue is that characteristics of large traders except for "child order size" and "speed" may not be identical or not be randomly distributed. However, we argue that most characteristics of large traders except for "child order size" and "speed" are unobservable to short-term traders. Therefore, from the perspective of short-term traders, large traders are almost identical, expect for "child order size" and "speed". Furthermore, since large traders should randomize their trades and try to find the best time to manage their large orders, such trading behaviors would make it hard for us to reject the two null hypotheses.

Let $x^c_{i,t}$ be the mean trade size between the $(t-3)^{\text{th}}$ and the t^{th} large trade traded by the large trader who initiated the t^{th} large trade. Let D^B_t be a dummy variable indicating whether $x^c_{i,t}$ is greater than or equal to the median of $x^c_{i,t}$ during the sample period. The dummy variable $D^B_t = +1$ indicates that a large trader uses a relatively big child order compared to other large traders before initiating the t^{th} large trade. Similarly, let D^S_t be a dummy variable indicating whether the speed of large trader who initiates the t^{th} large trade is slower than the median speed of larger traders. The dummy variable $D^S_t = +1$ indicates that the large trader who initiates the t^{th} large trade is more likely to be in the back of queue in the limit order book than other large traders, and the average time difference between order matching and submission times in her passive trades is longer than the median difference of all large traders.

Using the dummy variables, D_t^B and D_t^S indicating the "child order" size and the relative speed of large traders, the two null hypotheses can be formally stated as follows:

$$H_0^a : \mathbb{P}\left[y_t = x_{i,t} | D_t^B = 1\right] - \mathbb{P}\left[y_t = x_{i,t} | D_t^B = 0\right] \leq 0$$

$$H_0^b : \mathbb{P}\left[y_t = x_{i,t} | D_t^S = 1\right] - \mathbb{P}\left[y_t = x_{i,t} | D_t^S = 0\right] \leq 0.$$

Given the two null hypotheses above, we design a simple regression to test two hypotheses:

$$\begin{aligned} y_t =& \beta_1 \cdot D_t^S \cdot D_t^B \cdot x_{i,t} + \beta_2 \cdot D_t^S \cdot x_{i,t} + \beta_3 \cdot D_t^B \cdot x_{i,t} \\ &+ \beta_4 \cdot x_{i,t} + \sum_{j=1}^{15} \gamma_j \cdot y_{t-j} + D_t^S \cdot D_t^B + D_t^S + D_t^B + \alpha_i + \varepsilon_t. \end{aligned}$$

Based on the regression, we can rewrite our null hypotheses as follows:

$$H_0^a : \beta_3 \leq 0,\ \beta_1 + \beta_3 \leq 0$$

$$H_0^b : \beta_2 \leq 0,\ \beta_1 + \beta_2 \leq 0.$$

A linear hypothesis test rejects H_0^a with a p-value less than 0.001 since $\beta_3 > 0$ and $\beta_1 + \beta_3 > 0$, and rejects H_0^b with a p-value less than 0.005 since $\beta_2 > 0$ and

Model	Covariate	Estimate	S.E.	t value	Pr(≤ |t|)	Adj. R^2
Child Size	$x_{i,t}$	−0.019	0.003	−6.392	0.000	0.043
	$D_t^B \cdot x_{i,t}$	0.110	0.004	26.183	0.000	
	$\sum_{j=1}^{15} y_{t-j}$	0.363	0.006	61.634	0.000	
Speed	$x_{i,t}$	0.021	0.003	6.684	0.000	0.041
	$D_t^S \cdot x_{i,t}$	0.026	0.004	6.253	0.000	
	$\sum_{j=1}^{15} y_{t-j}$	0.364	0.006	61.583	0.000	
Child Size + Speed	$x_{i,t}$	−0.031	0.005	−6.103	0.000	0.044
	$D_t^S \cdot x_{i,t}$	0.018	0.006	2.879	0.004	
	$D_t^B \cdot x_{i,t}$	0.086	0.006	13.318	0.000	
	$D_t^S \cdot D_t^B \cdot x_{i,t}$	0.059	0.009	6.909	0.000	
	$\sum_{j=1}^{15} y_{t-j}$	0.362	0.006	61.513	0.000	

Table 1.4: This table tests the "child order size" and "speed" hypotheses, separately and jointly. Let $x_{i,t}^c$ be the mean size of child orders between the $(t-3)^{\text{th}}$ and the t^{th} large trade traded by the large trader who initiated the t^{th} large trade. The speed of large traders is measured by the mean differences between order matching and submission times for all matched passive orders of large traders. The dummy variable D_t^B indicates whether $x_{i,t}^c$ is greater than or equal to the median of $x_{i,t}^c$ during the sample period. The dummy variable D_t^S indicates the speed of the large trader who initiates the t^{th} large trade is slower than the median speed of large traders.

$\beta_1 + \beta_2 > 0$. This result implies that if large traders use bigger "child orders" with slower trading speed before initiating a large trade, their large trade is more likely to be anticipated by short-term traders. Although large traders can reduce the probability of order-anticipation by being faster, they cannot completely avoid order-anticipation by short-term traders if their "child order" size before initiating a large trade is relatively bigger than that of other large traders.

If we interpret D_t^S as a dummy variable indicating whether large traders use a trading algorithm to execute their large orders, the result implies that although an execution algorithm helps large traders to hide their order flow, large traders cannot completely avoid order-anticipation by short-term traders if they demand large liquidity within a short period of time.

Our tests have a few limitations. First, the regressions in table 1.4 cannot test whether an extreme low latency affects the probability of order-anticipation since the proxy for speed is not a good measure for the extreme low latency. It is possible that HFT firms exploit their extreme low latency to anticipate the order flow of large traders, but these tests cannot reveal whether HFT firms have the capability to do so. Second, the assumption that other characteristics of large traders are randomly distributed or not observable to short-term traders may not be valid. Large traders may have trading characteristics that are observable and correlated with "child order size", and it is possible that such characteristics are driving the results in table 1.4.

The results in table 1.4 do not conflict with the findings in Clark-Josep (2012), Hirschey (2012) and Li (2014). It is perfectly possible that HFT firms can trade ahead of institutions or informed traders based on their extreme low latency. What we are arguing in this paper is that in addition to the "trading speed" of large traders, "child order size" is an important factor that affects the probability of order-anticipation since "child orders" of large traders may reveal information about the forthcoming large trade to short-term traders.

The results in table 1.4 along with order-anticipation dynamics imply that the trading speed is important to short-term traders. The "speed" of large traders affects the probability that short-term traders are on the right side of large trades, and the "speed" of short-term traders is valuable when they are on the wrong side of large trades. See panels B and D in figure 1.3. When large traders initiate a large trade, and if short-term traders are on the wrong side of it, short-term traders need to get out of their positions as quickly as possible since they expect small traders to respond to the large trade by trading actively against their positions. Therefore, "speed" is valuable to short-term traders in the sense that if short-term traders have low latency, they can reduce negative skewed profits when they are on the wrong side of large trades.

In summary, the two main objectives of short-term traders are (1) to be on the right side when large traders initiate a large trade and (2) to get out of their positions as quickly as possible when they are on the wrong side of large trades. Bigger "child order size" of large traders increases the probability of short-term traders to be on the right side, and faster "speed" of short-term traders reduces their negative skewed profits when they are on the wrong side of large trades.

1.6.2 Large Trade Size and Order Exposure Probability

The model predicts that larger trades among the largest 1% of all trades are more likely to be anticipated by short-term traders because the size of the large trade is positively correlated with the size of its "child order", and the bigger "child orders" increase the probability of large trades being anticipated. By estimating the probability of order-anticipation for different size groups, we test whether the larger trades among the largest 1% of all trades are more likely to be anticipated.

The estimated β_0 in table 1.5 does not monotonically increase as the size of large trades increases. However, the extreme large trades among the largest 1% of all trades are more likely to be anticipated by short-term traders than relatively small large trades.

The model is inconsistent with the result in table 1.5 for a few possible reasons. First, the model forces the large trader to schedule her order over only two periods. If

Group(Largest%)	Size Bound	β_0	S.E.	t value	Pr($\leq$ |t|)	Nobs.
$1.00 \sim 0.60\%$	$[50, 62)$	0.032	0.003	10.531	0.000	108,001
$0.60 \sim 0.20\%$	$[62, 100)$	0.031	0.004	7.373	0.000	55,481
$0.20 \sim 0.10\%$	$[100, 128)$	0.022	0.006	3.923	0.000	31,812
$0.10 \sim 0.05\%$	$[128, 154)$	0.040	0.010	4.078	0.000	10,494
$0.05 \sim 0.01\%$	$[154, 228)$	0.120	0.010	12.048	0.000	9,068
$0.01 \sim 0.00\%$	$[228, \max)$	0.120	0.018	6.577	0.000	2,727
$1.00 \sim 0.00\%$	$[50, \max)$	0.036	0.002	16.766	0.000	217,583

Table 1.5: This table tests whether the larger trades among large trades are more likely to be anticipated. Large trades are defined as active trades by large traders among the largest 1% of all active trades. We divide large trades into 6 groups based on their trade size, and estimate the probability of order-anticipation, β_0 for each group. The variable y_t is the direction of the t^{th} large trade: Plus 1 is a buyer-initiated trade and minus 1 is a seller-initiated trade. The subscript t indexes large trades. The variable $x_{i,t}$ is the position of the i^{th} short-term trader at the t^{th} large trade: Plus 1 is a long position and minus 1 is a short position.

large traders have a longer time horizon to manage their large order and if short-term traders extract more accurate order flow information with a longer time series, the larger trades among the largest 1% of all trades may be less likely to be anticipated than the smaller trades. Second, the model implicitly assumes that liquidity provision from noise traders is constant since we assume a constant linear price impact factor. Since large traders have large demands, it is optimal for them to work their large orders when the market is more liquid so that they can efficiently hide their order flow while smoothing out their trades. In order to properly test the model prediction, we need to measure the size of large trades relative to liquidity.

1.7 Order-Anticipation Horizon

To show the order-anticipation horizon of short-term traders, we run predictive regressions, unconditional of the execution of large trades, with 120 second time intervals:

$$y_{(t,t+1]} = \beta_0 x_t + \sum_{i=1}^{5} \beta_i x_{(t-i,t-i+1]} + \varepsilon_{(t,t+1]},$$

where $y_{(t,t+1]}$ is the number of buyer-initiated large trades minus the number of seller-initiated large trades between time t and $t+1$. The variable x_t is the sign of the aggregate inventories of short-term traders at time t. The variable $x_{(t-i,t-i+1]}$ is the sign of short-term traders' trades between time $t-i$ and $t-i+1$. Note that the subscript $(t', t'+1]$ implies that t' is not included, but $t'+1$ is included and that the 120 second time intervals cover the entire trading hours during the sample period.

The predictive regressions in table 1.6 show that the aggregate positions of short-term traders are statistically significant predictors for the direction of large trades that will arrive within 120 seconds. The positions of short-term traders are a still significant predictor, even after controlling for the lagged trades of short-term traders. This implies that the positions of short-term traders are sufficient statistics to evaluate the predictability of short-term traders.

In panel B of table 1.6, we use the actual size of large trades for the dependent variable and use the aggregate inventories and trades of short-term traders for the

Covariate	Panel A. Direction (1)	(2)	(3)	Panel B. Direction × Size (1)	(2)	(3)
Intercept	−0.06	−0.06	−0.06	−2.94	−3.15	−2.94
	(−2.57)	(−2.44)	(−2.54)	(−1.34)	(−1.40)	(−1.34)
x_t	0.17	–	0.16	0.21	–	0.22
	(7.49)	–	(6.04)	(7.50)	–	(3.70)
$x_{(t-1,t]}$	–	0.09	0.01	–	0.16	−0.02
	–	(3.90)	(0.34)	–	(6.42)	(−0.28)
$x_{(t-2,t-1]}$	–	0.09	0.05	–	0.17	0.03
	–	(3.30)	(1.83)	–	(5.47)	(0.51)
$x_{(t-3,t-2]}$	–	0.05	0.03	–	0.14	0.03
	–	(1.83)	(1.12)	–	(4.23)	(0.60)
$x_{(t-4,t-3]}$	–	0.03	0.02	–	0.06	−0.02
	–	(1.18)	(0.82)	–	(1.88)	(−0.60)
$x_{(t-5,t-4]}$	–	−0.02	−0.02	–	0.01	−0.03
	–	(−0.75)	(−0.97)	–	(0.48)	(−1.11)
Nobs	11,674	11,674	11,674	11,674	11,674	11,674
Adj. R^2	0.0049	0.0017	0.0050	0.0056	0.0043	0.0057

Table 1.6: This table tests whether the aggregate inventories of short-term traders predict the direction of large trades. We run predictive regressions with data points extracted in 120 second intervals between 9:00 a.m. and 3:00 p.m. for 66 consecutive trading days of the sample period: $y_{(t,t+1]} = \beta_0 x_t + \sum_{i=1}^{5} \beta_i x_{(t-i,t-i+1]} + \varepsilon_{(t,t+1]}$, where $y_{(t,t+1]}$ is the number of buyer-initiated large trades minus the number of seller-initiated large trades between time t and $t+1$. The variable x_t is the sign of the aggregate inventories of short-term traders at time t, and $x_{(t-i,t-i+1]}$ is the sign of short-term traders' net trades between time $t-i$ and $t-i+1$. Note that the subscript $(t', t'+1]$ implies that t' is not included, but $t'+1$ is included. In panel B, we use the actual size of large trades for the dependent variable, and use the aggregate inventories and trades of short-term traders for the independent variables. Newey and West (1994) t-statistics with 30 lags are reported in parentheses.

independent variables. In these specifications, we still find the predictability of short-term traders on the direction of large trades. As a robustness check, we run the same regressions in table 1.6 with 5, 20, 40, 60, 120 and 240 second time intervals, and report the results in table 1.7. In all specifications, we find that the aggregate positions of short-term traders are statistically significant predictors for the direction of large trades that will arrive in a short period of time. Furthermore, the coefficient of x_t monotonically increases as the time intervals become longer. This implies that the order-anticipation strategy of short-term traders is not subject to their trading speed.

1.7.1 Order-Anticipation Horizon across Large Trades

Order-anticipation by short-term traders can occur over a few large trades. Using vector autoregression with contemporaneous trades, this section analyzes order-anticipation dynamics over ten large trades among large, short-term, and small traders.

When taking the data to a vector autoregression model, large trades are used as a partition of time domain to aggregate the active trades of large, short-term, and small traders. Based on the market microstructure invariance hypothesis proposed by Kyle and Obizhaeva (2013), the partition of time domain by large trades is more appropriate than regular time intervals for analyzing order-anticipation dynamics.

	Panel A. Direction: $y_{(t,t+1]}$					
Covariate	5 sec	20 sec	40 sec	60 sec	120 sec	240 sec
Intercept	0.00	−0.01	−0.02	−0.03	−0.06	−0.13
	(−0.16)	(−2.45)	(−2.47)	(−2.49)	(−2.54)	(−2.61)
x_t	0.02	0.04	0.08	0.10	0.16	0.26
	(8.16)	(11.45)	(9.42)	(8.01)	(6.04)	(4.42)
$x_{(t-1,t]}$	0.00	0.00	0.00	−0.02	0.01	0.00
	(0.78)	(−0.47)	(0.04)	(−1.90)	(0.34)	(0.04)
$x_{(t-2,t-1]}$	0.00	0.01	0.00	0.01	0.05	−0.03
	(0.93)	(1.92)	(0.28)	(0.45)	(1.83)	(−0.61)
$x_{(t-3,t-2]}$	0.00	0.01	0.00	0.02	0.03	−0.01
	(−0.22)	(2.57)	(−0.27)	(1.63)	(1.12)	(−0.24)
$x_{(t-4,t-3]}$	0.00	0.00	0.01	0.01	0.02	−0.04
	(0.50)	(0.73)	(1.54)	(0.47)	(0.82)	(−0.64)
$x_{(t-5,t-4]}$	0.00	0.01	0.01	0.01	−0.02	0.03
	(1.37)	(1.80)	(1.95)	(1.13)	(−0.97)	(0.53)
Nobs	71,421	71,819	35,749	23,683	11,674	5,672
Adj. R^2	0.0013	0.0026	0.0036	0.0034	0.0050	0.0045

Table 1.7: This table reports the order-anticipation horizon of short-term traders. We run the predictive regressions in table 6 with various time intervals: 5, 20, 40, 60, 120 and 240 seconds: $y_{(t,t+1]} = \beta_0 x_t + \sum_{i=1}^{5} \beta_i x_{(t-i,t-i+1]} + \varepsilon_{(t,t+1]}$, where $y_{(t,t+1]}$ is the number of buyer-initiated large trades minus the number of seller-initiated large trades between time t and $t+1$, where t is not included, but $t+1$ is included. The variable x_t is the aggregate inventories of short-term traders at time t, and $x_{(t-i,t-i+1]}$ is the net trades between time $t-i$ and $t-i+1$, where $t-i$ is not included, but $t-i+1$ is included. For panel B, we use the actual size of large trades instead of the number of large trades. Newey and West (1994) t-statistics with 30 lags are reported in parentheses.

The reasons are that (1) large trades provide a natural partition of time domain that is endogenous to trading activity, and (2) large trades can be considered as an exit point to short-term traders. If the market runs fast in business time, the level of risk transferred per unit of calendar time increases and a large trade is more likely to occur. Therefore, the time spans between two consecutive large trades would become smaller when business time runs faster, controlling for trading activity across the time spans between large trades.

Let $x^l_{\{t,t+1\}}$, $x^s_{\{t,t+1\}}$ and $x^m_{\{t,t+1\}}$ denote the net signed active trades of large, short-term, and small traders, respectively, between the t^{th} and the $(t+1)^{\text{th}}$ large trade. Let's define short-term price change $\Delta p_{\{t,t+1\}}$ as $(\ln p_{t+1} - \ln p_t) \times 10^4$, where p_t is the price at the t^{th} large trade. Let's define a 4×1 vector $\Lambda_{\{t,t+1\}}$ and $\mathbb{Y}_{\{t,t+1\}}$ as:

$$\Lambda_{\{t,t+1\}} = \left[\, \lambda_1 x^l_{\{t,t+1\}} + \lambda_2 x^s_{\{t,t+1\}} + \lambda_3 x^m_{\{t,t+1\}} \;\; 0 \;\; 0 \;\; 0 \,\right]'$$

$$\mathbb{Y}_{\{t,t+1\}} = \left[\, \Delta p_{\{t,t+1\}} \;\; x^l_{\{t,t+1\}} \;\; x^s_{\{t,t+1\}} \;\; x^m_{\{t,t+1\}} \,\right]'.$$

Our vector autoregression model with contemporaneous trades is

$$\mathbb{Y}_{\{t,t+1\}} = \Lambda_{\{t,t+1\}} + \sum_{i=1}^{5} \Theta_i \mathbb{Y}_{\{t-i,t-i+1\}} + \varepsilon_{\{t,t+1\}}, \tag{1.10}$$

where $\Lambda_{\{t,t+1\}}$ is a vector controlling for the price impact of contemporaneous trades, and Θ_i is a 4×4 vector of coefficients capturing the relationship among the active trades of each trader group and the short-term price change.

Based on the model, short-term traders anticipate the order flow of large traders instead of predicting the short-term price change directly. Consistent with the model, when controlling for the contemporaneous active trades of $x^l_{\{t,t+1\}}$, $x^s_{\{t,t+1\}}$ and $x^m_{\{t,t+1\}}$, the lagged active trades of short-term traders cannot predict the short-term price change, $\Delta p_{\{t,t+1\}}$ directly.

The variable $x^s_{\{t,t+1\}}$ is strongly negatively associated with $x^s_{\{t-i,t-i+1\}}$ for $1 \leq i \leq 10$. This is consistent with the summary statistics that short-term traders switch their positions, either from a long to short position or vice versa, as frequently as 91 times per day. This negative correlation is also consistent with the model, in which the short-term trader exits his positions when the large trader initiates a large trade at her second trade. The variable $x^l_{\{t,t+1\}}$ is strongly positively associated with both $x^l_{\{t-i,t-i+1\}}$ and $x^s_{\{t-i,t-i+1\}}$ for $1 \leq i \leq 10$. The positive autocorrelation of $x^l_{\{t-i,t-i+1\}}$ is due to large traders smoothing their trades. The positive correlation between $x^l_{\{t,t+1\}}$ and $x^s_{\{t-i,t-i+1\}}$ implies that short-term traders anticipate the order flow of large traders.

Short-term traders profit from the price impact caused by large traders. To show this, the short-term price change $\Delta p_{\{t,t+1\}}$ is regressed on the active trades of

	(1) $\Delta p_{\{t,t+1\}}$		(2) $x^s_{\{t,t+1\}}$		(3) $x^l_{\{t,t+1\}}$		(4) $x^m_{\{t,t+1\}}$	
Covariate x	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
$x^s_{\{t,t+1\}}$	−0.015	(−3.36)	–	–	–	–	–	–
$x^l_{\{t,t+1\}}$	0.095	(50.52)	–	–	–	–	–	–
$x^m_{\{t,t+1\}}$	−0.035	(−8.59)	–	–	–	–	–	–
$x^s_{\{t-1,t\}}$	0.000	(−0.09)	−0.115	(−20.75)	0.244	(18.35)	0.157	(25.03)
$x^s_{\{t-2,t-1\}}$	−0.002	(−0.45)	−0.092	(−16.18)	0.185	(13.66)	0.075	(11.80)
$x^s_{\{t-3,t-2\}}$	−0.008	(−1.71)	−0.074	(−13.02)	0.155	(11.42)	0.068	(10.64)
$x^s_{\{t-4,t-3\}}$	−0.006	(−1.43)	−0.060	(−10.69)	0.090	(6.70)	0.026	(4.15)
$x^s_{\{t-5,t-4\}}$	−0.004	(−0.93)	−0.038	(−6.96)	0.064	(4.89)	0.014	(2.33)
$x^l_{\{t-1,t\}}$	0.003	(1.64)	0.010	(4.04)	0.112	(19.00)	0.038	(13.67)
$x^l_{\{t-2,t-1\}}$	0.005	(2.71)	0.003	(1.04)	0.067	(11.30)	0.024	(8.51)
$x^l_{\{t-3,t-2\}}$	0.000	(−0.15)	−0.001	(−0.28)	0.024	(4.09)	0.009	(3.34)
$x^l_{\{t-4,t-3\}}$	0.001	(0.66)	−0.004	(−1.60)	0.021	(3.59)	0.008	(2.85)
$x^l_{\{t-5,t-4\}}$	0.002	(0.82)	−0.004	(−1.44)	0.033	(5.51)	0.010	(3.73)
$x^m_{\{t-1,t\}}$	0.004	(1.02)	−0.019	(−3.65)	0.003	(0.24)	0.078	(13.43)
$x^m_{\{t-2,t-1\}}$	0.003	(0.68)	−0.005	(−1.00)	0.041	(3.36)	0.066	(11.40)
$x^m_{\{t-3,t-2\}}$	0.002	(0.55)	−0.022	(−4.33)	0.050	(4.10)	0.048	(8.19)
$x^m_{\{t-4,t-3\}}$	0.009	(2.17)	−0.015	(−2.95)	0.055	(4.45)	0.047	(8.11)
$x^m_{\{t-5,t-4\}}$	−0.005	(−1.12)	−0.018	(−3.55)	0.037	(3.03)	0.042	(7.35)
$\Delta p_{\{t-1,t\}}$	0.027	(5.12)	0.188	(0.28)	2.102	(1.32)	0.013	(0.02)
$\Delta p_{\{t-2,t-1\}}$	−0.005	(−0.90)	−1.332	(−2.00)	−0.624	(−0.39)	−0.445	(−0.59)
$\Delta p_{\{t-3,t-2\}}$	0.018	(3.41)	−0.180	(−0.27)	3.049	(1.92)	−0.161	(−0.22)
$\Delta p_{\{t-4,t-3\}}$	0.018	(3.37)	−0.041	(−0.06)	2.026	(1.28)	0.325	(0.43)
$\Delta p_{\{t-5,t-4\}}$	0.004	(0.74)	−0.061	(−0.09)	−2.139	(−1.35)	−0.366	(−0.49)
Nobs	35,438	–	35,438	–	35,438	–	35,438	–
Adj. R^2	0.080	–	0.030	–	0.060	–	0.090	–

Table 1.8: This table analyzes order-anticipation dynamics among large, short-term, and small traders across large trades. Let $\mathbb{Y}_{\{t,t+1\}} = \left[\Delta p_{\{t,t+1\}}\ x^l_{\{t,t+1\}}\ x^c_{\{t,t+1\}}\ x^m_{\{t,t+1\}}\right]'$, where $\Delta p_{\{t,t+1\}}$ is the log return (bps) of midpoint of the bid-ask prices between the t^{th} and the $(t+1)^{\text{th}}$ large trade. Let $x^l_{\{t,t+1\}}$, $x^s_{\{t,t+1\}}$ and $x^m_{\{t,t+1\}}$ be the net signed active trades of large, short-term, and small traders, respectively, between the t^{th} and the $(t+1)^{\text{th}}$ large trade. Define $\Lambda_{\{t,t+1\}} = \left[\lambda_1 x^l_{\{t,t+1\}} + \lambda_2 x^s_{\{t,t+1\}} + \lambda_3 x^m_{\{t,t+1\}}\ 0\ 0\ 0\right]'$. Our VAR model is $\mathbb{Y}_{\{t,t+1\}} = \Lambda_{\{t,t+1\}} + \sum_{i=1}^{5} \Theta_i \mathbb{Y}_{\{t-i,t-i+1\}} + \varepsilon_t$, where $\Theta_i = \left\{\theta_i^{m,l}\right\}$ is a 4 by 4 coefficient matrix, and the coefficients λ_1, λ_2 and λ_3 in $\Lambda_{\{t,t+1\}}$ capture the contemporaneous price impact of $x^l_{\{t,t+1\}}$, $x^s_{\{t,t+1\}}$ and $x^m_{\{t,t+1\}}$ on the price change $\Delta p_{\{t,t+1\}}$, respectively. Every first 5 large trades in a day are dropped since their lagged variables are missing. The unit on the coefficients in the first column is 10^{-2} and the unit on the independent variables is the number of contracts actively traded by the corresponding trader group.

large, short-term, and small traders along with the lagged short-term price changes:

$$\Delta p_{\{t,t+1\}} = \lambda_1 x^s_{\{t,t+1\}} + \lambda_2 x^l_{\{t,t+1\}} + \lambda_3 x^m_{\{t,t+1\}} + \sum_{i=1}^{5} \Delta p_{\{t-i,t-i+1\}} + \sum_{i=1}^{5} \left(\theta_{1,i} \cdot x^s_{\{t-i,t-i+1\}} + \theta_{2,i} \cdot x^l_{\{t-i,t-i+1\}} + \theta_{3,i} \cdot x^m_{\{t-i,t-i+1\}}\right) + \varepsilon_{\{t,t+1\}}. \tag{1.11}$$

We intentionally omit $x^l_{\{t,t+1\}}$ or $x^m_{\{t,t+1\}}$ to examine which trader group contributes to the profits of short-term traders when short-term traders exit their positions. When omitting $x^l_{\{t,t+1\}}$ or $x^m_{\{t,t+1\}}$, the dependent variable in regression (1.11) becomes the price impact from the contemporaneous trades of the omitted trader group plus the short-term price change that cannot be explained with contemporaneous trades. If short-term traders profit from the price impact caused by a certain group, its omitted active trades would result in positively biased estimates between $x^s_{\{t-i,t-i+1\}}$ and $\Delta p_{\{t,t+1\}}$.

When omitting $x^l_{\{t,t+1\}}$ in regression (1.11), the short-term price change $\Delta p_{\{t,t+1\}}$ is strongly positively associated with $x^s_{\{t-i,t-i+1\}}$. We interpret this positive correlation with the negative autocorrelation of $x^s_{\{t,t+1\}}$ such that short-term traders anticipate a sequence of trades of large traders by accumulating their positions in the direction of the forthcoming large trade, and when large traders initiate a large trade, short-term traders consider this as an exit point, in which they start to realize their profits by liquidating their positions to other traders who respond to large trades. Such order-anticipation dynamics show up in predictive regression (1.11) as if short-

Panel A. Short-term Traders Active Trades at Large Trades t								
	(1) $\Delta p_{\{t,t+1\}}$		(2) $\Delta p_{\{t,t+1\}}$		(3) $\Delta p_{\{t,t+1\}}$		(4) $\Delta p_{\{t,t+1\}}$	
Covariate x	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
$x^s_{\{t,t+1\}}$	−0.015	(−3.36)	0.019	(4.36)	0.010	(2.25)	−0.024	(−5.52)
$x^l_{\{t,t+1\}}$	0.095	(50.52)	–	–	–	–	0.091	(50.24)
$x^m_{\{t,t+1\}}$	−0.035	(−8.59)	–	–	0.028	(6.83)	–	–
$x^s_{\{t-1,t\}}$	0.000	(−0.09)	0.021	(4.63)	0.016	(3.42)	−0.006	(−1.28)
$x^s_{\{t-2,t-1\}}$	−0.002	(−0.45)	0.016	(3.44)	0.013	(2.81)	−0.005	(−1.02)
$x^s_{\{t-3,t-2\}}$	−0.008	(−1.71)	0.007	(1.53)	0.005	(0.99)	−0.010	(−2.21)
$x^s_{\{t-4,t-3\}}$	−0.006	(−1.43)	0.003	(0.72)	0.002	(0.45)	−0.007	(−1.65)
$x^s_{\{t-5,t-4\}}$	−0.004	(−0.93)	0.003	(0.64)	0.002	(0.47)	−0.005	(−1.05)
$x^l_{\{t-1,t\}}$	0.003	(1.64)	0.012	(6.04)	0.011	(5.55)	0.003	(1.28)
$x^l_{\{t-2,t-1\}}$	0.005	(2.71)	0.011	(5.31)	0.010	(4.99)	0.005	(2.47)
$x^l_{\{t-3,t-2\}}$	0.000	(−0.15)	0.002	(0.84)	0.001	(0.71)	−0.001	(−0.26)
$x^l_{\{t-4,t-3\}}$	0.001	(0.66)	0.003	(1.56)	0.003	(1.44)	0.001	(0.55)
$x^l_{\{t-5,t-4\}}$	0.002	(0.82)	0.004	(2.20)	0.004	(2.05)	0.001	(0.70)
$x^m_{\{t-1,t\}}$	0.004	(1.02)	0.002	(0.56)	0.000	(0.01)	0.001	(0.31)
$x^m_{\{t-2,t-1\}}$	0.003	(0.68)	0.005	(1.08)	0.003	(0.64)	0.001	(0.15)
$x^m_{\{t-3,t-2\}}$	0.002	(0.55)	0.006	(1.45)	0.005	(1.09)	0.001	(0.15)
$x^m_{\{t-4,t-3\}}$	0.009	(2.17)	0.013	(3.06)	0.011	(2.72)	0.007	(1.80)
$x^m_{\{t-5,t-4\}}$	−0.005	(−1.12)	−0.002	(−0.45)	−0.003	(−0.77)	−0.006	(−1.48)
$\Delta p_{\{t-1,t\}}$	0.027	(5.12)	0.029	(5.29)	0.029	(5.30)	0.027	(5.13)
$\Delta p_{\{t-2,t-1\}}$	−0.005	(−0.90)	−0.005	(−0.86)	−0.005	(−0.86)	−0.005	(−0.89)
$\Delta p_{\{t-3,t-2\}}$	0.018	(3.41)	0.021	(3.85)	0.021	(3.86)	0.018	(3.45)
$\Delta p_{\{t-4,t-3\}}$	0.018	(3.37)	0.020	(3.59)	0.020	(3.58)	0.018	(3.37)
$\Delta p_{\{t-5,t-4\}}$	0.004	(0.74)	0.002	(0.37)	0.002	(0.38)	0.004	(0.74)
Nobs	35,438	–	35,438	–	35,438	–	35,438	–
Adj. R^2	0.080	–	0.010	–	0.010	–	0.080	–

Table 1.9: This table identifies a channel through which short-term traders predict the short-term price change. The short-term price change $\Delta p_{\{t,t+1\}}$ is regressed on the lagged short-term price changes and the active trades of large, short-term, small traders. The variable $x^l_{\{t,t+1\}}$ or $x^m_{\{t,t+1\}}$ are intentionally omitted in the regression. The omitted variables result in biased estimates for the coefficients on $x^s_{\{t-i,t-i+1\}}$, which identify the indirect channel, through which short-term traders predict the short-term price change $\Delta p_{\{t,t+1\}}$. Panel B replaces $x^s_{\{t,t+1\}}$ with $x_{\{t\}}$, which are the aggregate positions of short-term traders at the t^{th} large trade. As a counterfactual analysis, panel C replicates panel B with randomly chosen trades. Every first 5 large trades in a day are dropped since their lagged variables are missing. The unit on the coefficients is 10^{-2} and the unit on the independent variables is the number of contracts actively traded by the corresponding trader group.

term traders predict the short-term price change between two consecutive large trades when omitting $x^l_{\{t,t+1\}}$. What actually happens is that short-term traders anticipate the order flow of large traders, and they profit from the price impact caused by large traders by liquidating their positions when large traders initiate a large trade.

Even if $x^s_{\{t-i,t-i+1\}}$ is positively associated with $\Delta p_{\{t,t+1\}}$, if the short-term trader's position is on the wrong side of large trades, short-term traders cannot profit from the price impact incurred by large traders. To address this concern, $x^s_{\{t,t+1\}}$ is replaced with the aggregate positions of short-term traders at the t^{th} large trade, $x_{\{t\}}$ in regression (1.11), and the results are presented in panel B in table 1.7.1.

The results in panel B in table 1.7.1 imply that short-term traders' position at the t^{th} large trade is a strong predictor for the short-term price change between the t^{th} and the $(t+1)^{\text{th}}$ large trade. The variable $x^s_{\{t-i,t-i+1\}}$ is no longer positively associated with $\Delta p_{\{t,t+1\}}$ when controlling for the positions of short-term traders, $x_{\{t\}}$. This is because the active trades of short-term traders, $x^s_{\{t-i,t-i+1\}}$, are highly correlated with their positions, $x_{\{t\}}$, which is a sufficient predictor for $\Delta p_{\{t,t+1\}}$. The underlying mechanism in panel B in table 1.7.1 is that short-term traders take positions in the direction of large trades before large traders initiate large trades, and they profit from the price impact caused by large traders.

As a counterfactual analysis, panel C replicates panel B in table 1.7.1 with randomly chosen trades instead of large trades. Between randomly chosen two consecutive trades, short-term traders' position, $x_{\{t\}}$ is negatively associated with the short-term price change $\Delta p_{\{t,t+1\}}$. The results in panel B and C in table 1.7.1 imply that except for large trades, the short-term price is likely to move against the position of short-term traders and large trades are the main source of profits of short-term traders.

1.8 Conclusion

This is the first paper to document that short-term traders anticipate the direction of large trades, and they profit from the price impact caused by large traders. When large traders initiate a trade, short-term traders correctly anticipate the direction of the large trade 56.06% of the time. By either taking a long position in advance of a large buy order or a short position in advance of a large sell order, short-term traders profit from the price impact caused by the large trade. Furthermore, we find that the aggregate positions of short-term traders are statistically sufficient and significant predictors for the direction of large trades that will arrive within 120 seconds.

Based on the findings, we argue that order-anticipation trading occurs because of order flow revealed by large traders rather than because of their speed slower than short-term traders. Large traders inevitably reveal their order flow to reduce the price impact of large positions. This acts as a trading signal to short-term traders

Although this paper analyzes the index futures market of South Korea, the main findings are driven by the foreign trading algorithms that would actively trade in other international markets. Thus, we conjecture that the results can be replicated in other major markets such as E-mini S&P 500 index futures market.

This paper does not evaluate overall effect of order-anticipation by short-term traders on financial markets. Short-term traders clearly increase the transaction cost of large traders, however they also contribute to the price discovery process by competing for liquidity with large traders prior to large trades. After large trades are made, short-term traders provide liquidity to other market participants who want to actively trade to respond to new information revealed by large trades.

In our next paper, we ask how exogenous order flow frequency change affects the order-anticipation of short-term traders. The KRX went through a structural change on March 23, 2009, when the exchange server capacity was doubled from the former system. Since the upgrade, market participants have received more frequent limit order book information from once per 10 millisecond to once per 1 millisecond. This structural change would exogenously give an advantage to short-term traders as they receive more refined information about large traders who do not want to reveal their order flow. Using this exogenous change as a natural experiment, we examine how this structural change affects order-anticipation dynamics, and how order-anticipation by short-term traders affects price informativeness, short and long-term volatility, market depth and liquidity.

Chapter 2: An Invariance Relationship in the Number of Buy-Sell Switching Points

2.1 Overview

Financial markets generate a voluminous amount of data on order placements and quote updates. These data leave little doubt that trading patterns vary significantly across securities. The market microstructure invariance hypothesis developed by Kyle and Obizhaeva (2013) nevertheless claims that trading patterns in different markets look similar when viewed from the perspective of an appropriate "business" time clock. Market microstructure invariance predicts similarities in the dollar amounts expected to be at stake, the scale of risk transferred, the magnitude of transaction costs, and the size of profits.

In this paper, we test the market microstructure invariance hypothesis by examining variation in the aggregate number of buy-sell switching points across stocks. We define the number of buy-sell "switching" points based on the number of times that individual traders change the direction of their trading. We hypothesize that the number of switching points is proportional to the rate at which business time passes. Under this hypothesis, market microstructure invariance predicts that the aggregate number of switching points is proportional to the 2/3 power of the product of dollar volume and volatility.

Using account-level data from the Korea Exchange (KRX) from 2008 to 2010, we estimate the exponent to be 0.675 with standard error of 0.005. Invariance explains about 93% of the variation in the number of switching points each month across stocks. Invariance patterns are especially pronounced for the subset of domestic retail investors. A decomposition into the number of unique accounts and the average number of switching points per account shows that it is the cross-sectional variation in the number of accounts that exhibits the invariance patterns, while the number of switching points per account is relatively stable.

2.2 Market Microstructure Invariance, Business Time, and Switching Points

According to the market microstructure invariance hypothesis of Kyle and Obizhaeva (2013), the business time clock is governed by the frequency at which independent ideas—referred to as "bets"—are expected to arrive into the marketplace. In more active markets, bets arrive more frequently as the time clock runs faster. As bets are placed at a faster rate, trading costs decrease and the average distance between the market price and unobserved fundamental value decreases by an amount proportional to the square root of the arrival rate of bets. For informed traders to make the same expected dollar profits per bet when trading costs fall and price efficiency increases, they scale up the dollar size of their bets proportionally. Holding volatility constant, invariance implies that the dollar size of bets increases at a rate proportional to the square root of the number of bets per day; this implies that, as trading volume varies across securities, the number of bets increases twice as fast as the size of bets. Thus, if trading volume increases by a factor of 8, the number of bets increases by a factor of 4 and the dollar size of bets increases by a factor of 2. Since the business time clock—which ticks at the rate bets arrive—effectively speeds up by a factor of 4, the speed with which business time passes is proportional to the 2/3 power of trading volume.

To adjust for differences in percentage returns volatility, Kyle and Obizhaeva (2013) introduce the concept of "trading activity" denoted W. Trading activity is defined as the product of daily dollar volume $P \cdot V$ (dollar share price times share volume per day) and daily percentage returns volatility σ,

$$W := P \cdot V \cdot \sigma. \tag{2.1}$$

It is a measure of aggregate risk transfer per calendar day. The expected number of bets per calendar day—and thus the rate at which trading unfolds—is proportional to $W^{2/3}$. Invariance implies that specific exponents of 1/3 and 2/3 govern relationships between trading activity W and various market characteristics such as bet size, bid-ask spreads, market impact costs, speed of mean reversion, and price efficiency. These relationships should be present in data on trading in financial markets.

The variable of interest in this paper is the aggregate number of buy-sell "switching" points. For each month and each security, we count how many times individual traders change their trading direction from buying to selling or from selling to buying and then aggregate those numbers across all accounts to find an aggregate number of switching points for all traders in a given stock in a given month. If an account trades a given stock in a given month but not in the previous month, then we count its number of switching point as at least one. Each time an individual account changes the direction of its trading from buying to selling or from selling to buying, the number of switching points is increased by one. We denote the aggregate number of switching points, summed across all accounts which traded stock i during month t, as S_{it}.

We expect to find an invariance relationships in the cross-sectional patterns of switching points. More precisely, consistent with the invariance hypothesis, we hypothesize that S_{it} is proportional to $W_{it}^{2/3}$,

$$S_{it} = a \cdot \left(\frac{W_{it}}{W^*}\right)^{2/3}, \tag{2.2}$$

where a is the same "invariant" constant for all stocks i and all months t. The constant a is scaled by W^* so that it quantifies the expected number of switching points per calendar day for a hypothetical benchmark stock with trading activity W^*. To match the benchmark stock of Kyle and Obizhaeva (2013), we define the benchmark stock to have a daily volume of one million shares, daily volatility of 2%, and price of $47,440$ KRW per share (approximately equal to $40 per share given the average exchange rate of 1,186 KRW per USD between 2008 to 2010). This hypothetical stock would be at the bottom of the top 50 stocks in the Korean Composite Stock Price Index (KOSPI). In this paper, we present evidence supporting our hypothesis.

Our tests have a number of advantages over other tests for invariance relationships in trading data. Kyle and Obizhaeva (2013) document invariance relationships for the size distributions of portfolio transition orders. These tests require the identifying assumption that portfolio transition orders of institutional investors be proportional to bets. Kyle, Obizhaeva and Tuzun (2012) document invariance relationships for the size distributions of "prints" of quantities traded in the Trade and Quote dataset (TAQ). These tests rely on the even stronger assumption that print sizes are proportional to bets. This assumption broke down after the 2001 reduction of tick size to one cent and electronic order handling algorithms motivated traders in the earlier 2000s to shred their larger "meta-orders" into trades equal in size to the minimum lot size of 100 shares or even smaller odd lots.

Although the aggregate number of switching points is hardly of any economic interest in itself, it is a convenient tool for testing invariance. The tests based on the aggregate number of switching points do not require strong assumptions about bets.

Kyle and Obizhaeva (2013) develop invariance hypotheses using the concept of "bets." In theory, a portfolio manager places a bet when he makes a statistically independent decision to accumulate a position of a particular size. In practice, the concept of a bet is difficult to map into data. Bets do not map easily into orders, since one bet might be broken into many orders or spread across different accounts; thus, bets do not necessarily show up in an obvious way in consolidated audit trail data. Bets map even less easily into public data on trades, such as TAQ prints.

In contrast to the concept of a bet, the concept of a switching point can be given a more unambiguous definition which maps into data in a straightforward manner, provided trading data is available by individual account. There is some ambiguity concerning the possibility that bets are spread across multiple accounts or multiple bets are merged together; these possibilities may affect the number of switching points, but the effect is likely to be proportional across stocks. Empirical tests of cross-sectional variation based on the number of switching points only require the structure of trading to be approximately preserved across securities, regardless of the specifics of how the flow of bets in the marketplace is expressed as a flow of trades. Of course, switching point results may be affected by various market frictions and institutional features such as minimum tick size, minimum lot size, the level of cross-market arbitrage, and the industrial organization of entities participating in trading financial securities. We examine these issues in later sections of the paper.

2.3 The South Korean Stock Market Data

Our study is based on trade-level and account-level data provided by the Korea Exchange (KRX) for the period from February 2008 through November 2010. The Korea Exchange was created after the integration of the Korea Stock Exchange, the KOSDAQ Stock Exchange, and the Korea Derivatives Market in 2005. According to the World Federation of Exchanges, the South Korean stock market is ranked 17^{th} in terms of market capitalization (about $1 trillion). Our sample includes only the stocks listed in the KOSPI Market division at the Korea Exchange.

The KRX operates a single central limit order book for each KOSPI stock. The dataset contains records of all orders placed, canceled, or modified as well as all transactions executed. Records include blocking trading codes, short-sale codes, trading system codes, and time stamps to the millisecond. Each message is linked to the specific accounts involved and some additional information on account types is collected, such as whether accounts belong to domestic retail investors, domestic institutional investors (financial investment companies, insurance companies, private equity funds, etc.), or foreign investors. The KRX database has about 2.69 billion messages and 1.29 billion distinct trade records during our sample period.

For our analysis, one observation is associated with each stock for each period of 20 trading days from February 2008 through November 2010. In this paper, we refer informally to each period of 20 trading days as a "month" (even though the 20-trading-day period do not correspond precisely to calendar months). Using this definition, our dataset covers 36 months. We begin with 24,441 observations, one observation for each KOSPI stock and each month from February 2008 through November 2010. We drop $2,506$ stock-month observations, because trading of some stocks was discontinued during particular months, thus biasing downwards the number of switching points calculated for those observations. Our final sample has $21,935$ observations of stock-month pairs. There are on average 609 KOSPI stocks traded during each month.

Using these data, we calculate for each stock i and for each month t the aggregate number of accounts which trade N_{it} and the aggregate number of buy-sell switching points S_{it} (summed across accounts). For each observation, we calculate the dollar share price P_{it} as the product of the exchange rate between the South Korean won and the U.S. dollar (KRW-USD exchange rate) and the closing KRW stock price. We obtain share volume V_{it} from the official daily public share volume report. We calculate daily returns volatility σ_{it} as the sample standard deviation of daily percentage returns during the same month. Trading activity W_{it} is defined as $W_{it} := P_{it} \cdot V_{it} \cdot \sigma_{it}$. We define market capitalization based on the number of shares outstanding at the end of each year. We calculate the annualized turnover rate ν_{it} based on share volume in month t and shares outstanding at the end of the previous year.

The dataset identifies three broad categories of traders: domestic retail investors, domestic institutional investors, and foreign investors. The number of accounts N_{it} and number of switching points S_{it} represent sums across these three investor types. We let α_{it} denote the fraction of share volume due to domestic retail investors.

There are in total $425,440,260$ switching points in the sample, on average $19,395$ switching points per month per stock in the KOSPI universe: 94.2% from accounts of domestic retail investors, 4.7% from accounts of domestic institutions, and 1.1% from accounts of foreign investors. There are $5,886,557$ distinct accounts in the sample: 94% domestic retail investors, 5.1% domestic institutions, and 0.8% foreign investors.

Table 2.1 shows summary statistics for the entire sample as well as the six

Variable	All	30^{th}	60^{th}	75^{th}	85^{th}	95^{th}	100^{th}
	Volume Group						
Price	36,839	13,957	26,530	37,815	47,599	77,869	119,947
Daily Volume (1B)	8.50	0.08	0.50	2.08	6.68	23.55	94.88
Volatility (%)	2.79	2.22	2.88	3.34	3.21	2.97	2.74
Capitalization (1T)	1.32	0.07	0.15	0.33	0.99	3.39	14.62
Annual Turnover (%)	263.70	49.23	193.23	429.22	553.08	495.40	363.91
Tick Size (BPS)	22.10	21.53	22.25	22.44	22.80	22.30	21.69
# Trades/Day	5,659	255	1,170	3,574	7,893	17,033	41,400
Avg Trade Size (1M)	2.87	1.21	2.06	2.51	3.75	5.90	10.37
Trades at Min Lot Size (%)	23.25	28.78	23.42	20.27	19.21	18.46	17.51
DR Volume (%)	78.32	86.63	81.57	79.09	71.95	62.31	54.71
DI Volume (%)	13.93	10.11	12.08	13.78	17.87	21.45	23.91
FI Volume (%)	7.75	3.27	6.35	7.13	10.18	16.24	21.38
Avg # Switches	19,395	930	4,072	13,353	28,501	57,567	136,710
Avg # Stock	609	176	185	93	62	62	32
# Observations	21,935	6,330	6,669	3,341	2,220	2,235	1,140

Table 2.1: The Summary Statistics: The table shows the price (KRW), daily volume (1 billion KRW), volatility (%), market capitalization (1 trillion KRW), annual turnover (%), percentage tick size (bps), number of trades, average trade size (1 million KRW), percentage of trades of minimum lot size, the fraction of double-sided volume of domestic retail investors, the fraction of double-sided volume for domestic institutional investors, the fraction of double-sided volume of foreign investors, average number of switches per month, average number of stocks, and number of month-stock observations. The average exchange rate is $1,186$ KRX/USD during the sample period.

volume subgroups defined by the 30^{th}, 60^{th}, 75^{th}, 85^{th}, 95^{th} 100^{th} percentiles of average daily volume. The largest volume group is dominated by Samsung Electronics, the largest stock in the Korea Exchange, which accounts for about 5% of the total trading volume in KRW.

The average number of switching points per month increases by a factor of 147 from 930 for the lowest volume group to $136,710$ for the highest volume group.

Trading activity $W_{it} = P_{it} \cdot V_{it} \cdot \sigma_{it}$ increases by a factor $1,464$ from the lowest to the highest group. Most of the variation in trading activity is due to variation in daily volume, which increases from 0.08 billion KRW to 94.88 billion KRW. Volatility does not change much across groups and the modest changes are not monotonic across groups; volatility is 2.22 percent in the lowest group, 3.34 percent in the 75$^{\text{th}}$ percentile group, and 2.74 percent in the highest group. These patterns are consistent with invariance predictions, since 147 is approximately equal to 2/3 power of $1,464$.

The minimum lot size is equal to ten shares if the share price is below $50,000$ KRW and one share if share price is above $50,000$ KRW. In our sample, the median size of trades is equal to 38 shares, implying that the minimum lot size constraint is often binding. Indeed, about 23.25% of trades are executed in the minimum size allowed; the fraction decreases from 28.78% for the low volume group to 17.51% for the high volume group. As in the U.S. market, extensive order shredding makes it difficult to test directly the invariance hypothesis by identifying bets in market data.

The tick size is determined according to a schedule.[1] The average tick size is about 22.10 basis points, approximately ten times larger than the typical tick size in the U.S. stock market (e.g., one penny on $40 stock or 2.5 basis points). The average

[1]The tick size is equal to 1 KRW if share price is below 1,000 KRW; 5 KRW if share price is between 1,000 KRW and 5,000 KRW; 10 KRW if share price is between 5,000 KRW and 10,000 KRW; 50 KRW if share price is between 10,000 KRW and 50,000 KRW; 100 KRW if share price is between 50,000 KRW and 100,000 KRW; 500 KRW if share price is between 100,000 KRW and 500,000 KRW; and 1,000 KRW if share price is above 500,000 KRW.

tick size is relatively stable across volume groups, ranging from 21.53 basis points for low volume group to 22.83 basis points for high volume group. The large tick size is likely to influence the trading behavior of market participants and have an effect on the aggregate number of switching points.

Let Δ_{it} denote the tick size in units of KRW for stock i in month t (e.g., Δ_{it} is 1 KRW if the share price is below 1,000 KRW). Following Kyle, Obizhaeva and Tuzun (2012), we define effective relative tick size e_{it}/e^* as the ratio of tick size in basis points Δ_{it}/P_{it} to the standard deviation of returns over one unit of business time (which is proportional to $\sigma_{it}/W_{it}^{1/3}$), scaled so that this ratio is equal to one for the benchmark stock, i.e., we have

$$\frac{e_{it}}{e^*} := \frac{\Delta_{it}}{P_{it}} \cdot \frac{P^*}{\Delta^*} \cdot \frac{W_{it}^{1/3}}{\sigma_{it}} \cdot \frac{\sigma^*}{W^{*1/3}}. \tag{2.3}$$

Another possibly important market friction is South Korea's transactions tax. The exchange collects a tax of about 30 basis points on the sale of securities, paid by the seller. Trading fees of about 1.50 basis points are paid to on-line brokers on executed orders.

Several stock indices are used as reference values for actively traded derivatives contracts. The Korea Composite Stock Price Index (KOSPI) includes all common stocks traded on the Korea Exchange, with weights proportional to market capitalization. The KOSPI includes about 688 stocks. The KOSPI 50 index includes the

50 largest companies listed on the Korea Exchange, approximately corresponding to the 95$^{\text{th}}$ percentile and the 100$^{\text{th}}$ percentile volume groups in table 2.1). The KOSPI 200 index includes the 200 largest companies listed on the Korea Exchange, approximately corresponding to the 75$^{\text{th}}$ percentile to 100$^{\text{th}}$ percentile volume groups in table 2.1). The largest 200 stocks are often traded by investors engaging in cross-market and index arbitrage strategies. The resulting basket trades will tend to affect the number of switching points across stocks in the KOSPI 50 and KOSPI 200 universes. The identification of basket trades in the dataset is complicated, because the dataset does not link accounts trading in the stock market to accounts trading in the derivatives market.

2.4 Trading Activity and Switching Points

The main result of this paper concerns the empirical relationship between the logarithm of the aggregate number of buy-sell switching points $\ln(S_{it})$ and the logarithm of scaled trading activity $\ln(W_{it}/W^*)$ in the same month.

Figure 2.1 shows that all $21,935$ observations line up along a straight line whose fitted slope of 0.675 (from an OLS regression) is very close to the predicted slope of 2/3. Observations for stocks included in the KOSPI 50 universe (black points) and KOSPI 200 universe (blue points) are close to the fitted line as well. At the far right corner of figure 2.1, the observations for the largest South Korean stock, Samsung

Electronics, do not deviate much from that line. When Samsung Electronics is compared to the stock with the least amount of trading activity, the difference in trading activity is a factor of about exp(10), or approximately $22,000$. It is apparent from visual observation that the data is relatively homoskedastic. For a given level of the logarithm of trading activity, the logarithm of the number of switching points for the less actively traded stocks deviates from the fitted line only slightly more than for the more actively traded stocks. This slightly higher deviation may indicate a larger estimation error in the estimates of expected trading activity for smaller stocks.

A similar conclusion can be drawn from a regression analysis of the logarithm of the aggregate number of buy-sell switching points $\ln(S_{it})$ on the logarithm of scaled trading activity $\ln(W_{it}/W^*)$, clustering standard errors in the panel data regression at monthly levels:

$$\ln(S_{it}) = 11.156 + 0.675 \cdot \ln(W_{it}/W^*) + \epsilon_{it}. \tag{2.4}$$

The estimated coefficient of 0.675 has a clustered standard error of 0.005, implying that the hypothesis that the coefficient is equal to the predicted value of 2/3 is not rejected ($t = 1.67$). The non-clustered standard error is 0.0012. The constant term of 11.156 implies that benchmark stock has on average about $53,000$ buy-sell switching points per month. The R^2 of the regression is equal to 0.935.

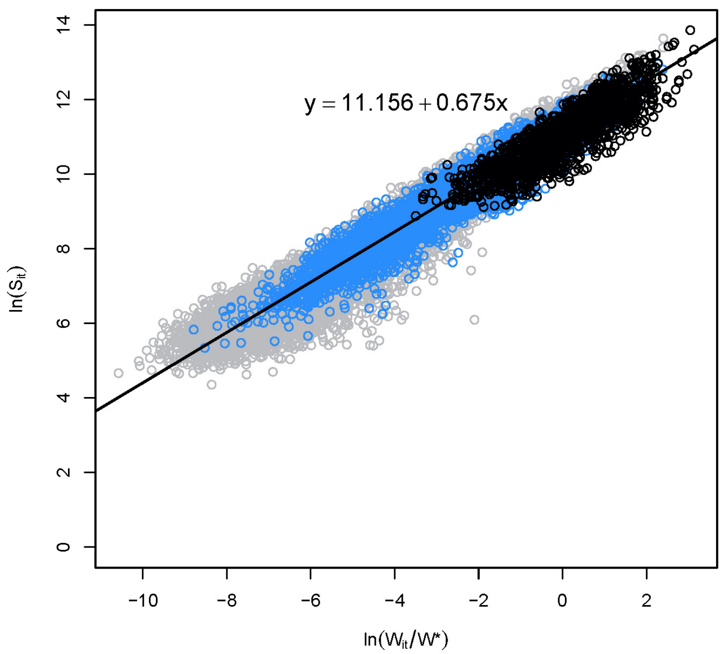

Figure 2.1: Aggregate Number of Switching Points $\ln(S_{it})$ against Trading Activity $\ln(W_{it}/W^*)$: The vertical axis is $\ln(S_{it})$. The horizontal axis is $\ln(W_{it}/W^*)$, where $W^* = 10^6 \cdot 40 \cdot 1186 \cdot 0.02$ and $W_{it} = V_{it} \cdot P_{it} \cdot \sigma_{it}$. The fitted line is $11.156 + 0.675 \cdot \ln(W_{it}/W^*)$. The invariance-implied slope is 2/3

Figure 2.2 presents estimates from monthly regressions of the logarithm of the aggregate number of switching points $\ln(S_{it})$ on the logarithm of scaled trading activity $\ln(W_{it}/W^{*})$. To make interpretation of results easier, the figure also contains a horizontal line indicating the regression coefficient of 2/3 predicted by invariance. All 36 point estimates of monthly regression coefficients are very close to 2/3. Only 15 out of 36 point estimates lie slightly outside of 95%-confidence bounds. Most of these 15 months occur between October 2008 and November 2009, when the South Korean market was most affected by the 2008 financial crisis. The estimated coefficients exhibit persistence across months, fluctuating over time between 0.64 and 0.72.

We conclude that even though there is enough variation in the time series of regression coefficients to reject the hypothesis that the coefficient is 2/3 every month, the coefficient estimates are economically close to this predicted value.

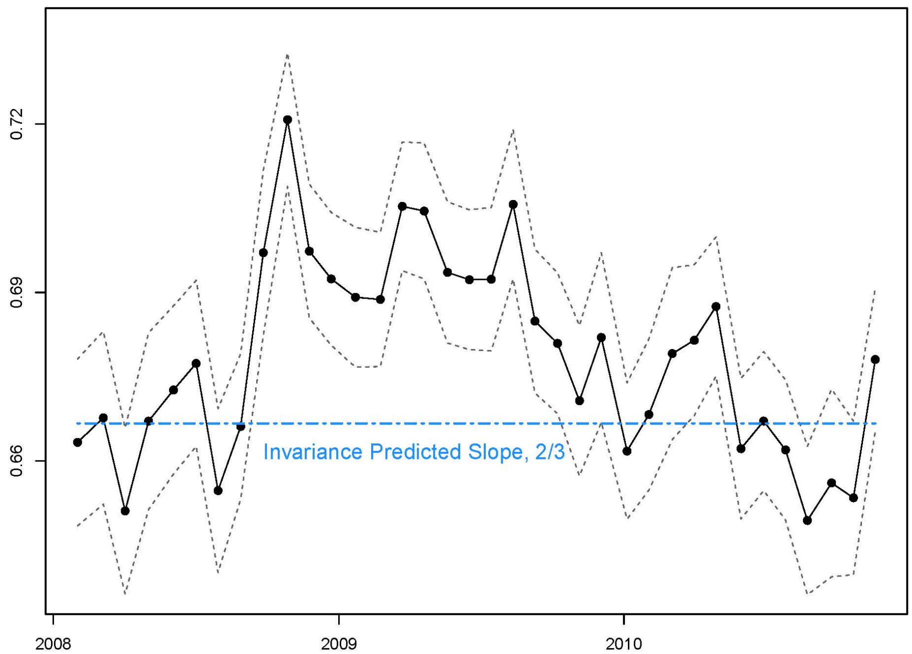

Figure 2.2: Time Series of Monthly Regression Coefficients: The time series of estimates β_s and their 95%-confidence intervals from 36 cross-sectional regressions $\ln(S_{it}) = \ln(a) + \beta_s \cdot \ln(W_{it}/W^*) + \epsilon_{it}$, where S_{it} is the aggregate number of switching points and W_{it} is expected trading activity for stock i and month t. The time period is from February 2008 to November 2010. The invariance predicted slope is 2/3.

2.5 Number of Switching Points and Different Types of Traders

Figure 2.3 shows the relationship between the logarithm of buy-sell switching points and the logarithm of scaled trading activity for different types of traders: domestic retail investors, domestic institutional investors, and foreign investors..

Panel A of figure 2.3 shows results for the subset of domestic retail investors.

These observations reveal a striking invariance relationship. The slope of the fitted line 0.669 (t=0.4630 using clustered standard error, t=1.7903 using non-clustered standard error) does not reject the hypothesis of equality to the predicted value of 2/3. Trades by retail investors dominate the results in figure 2.1, since domestic retail investors account for about 94.7% of switching points in the entire sample.

Panel B of figure 2.3 shows results for the subset of domestic institutional investors. These observations account only for about 4.7% of switching points of the entire sample. They satisfy the invariance relationship less closely. The slope of the fitted line 0.82 is higher than predicted coefficient of 2/3. The number of switching points for stocks included in the KOSPI 50 universe is flatter than predicted by invariance; the estimated slope for these observations is 0.332. The number of switching points for stocks in the KOSPI 200 universe but outside of the KOSPI 50 universe is slightly steeper; the estimated slope for these observations is 0.532. The flatness of the empirical distribution on the right side of the graph suggests that cross-market arbitrage plays an important role in trading patterns of domestic institution, especially for stocks in the KOSPI 50 universe. The small counts for less actively traded securities (as revealed by horizontal lines corresponding to one through ten switching points per month) introduces further distortions.

Panel C of figure 2.3 shows results for the subset of foreign investors. The slope of the fitted line 0.639 is lower than the predicted slope of 2/3, but not by much. The points representing stocks included in KOSPI 50 and KOSPI 200 indices have much flatter slopes; the slopes of the fitted lines are 0.451 for the stocks in the KOSPI 50 universe and 0.35 for the stocks in the KOSPI 200 universe but outside of the KOSPI 50 universe. These slopes are similar in magnitude to the slopes for domestic institutions, suggesting that cross-market arbitrage affects trading patterns of both domestic institutions and foreign investors in a similar way. Since these observations account for about 0.6% of all switching points, these patterns are also influenced by small counts for less actively traded stocks, but this issue is less important for this subset than for the subset of domestic institutions.

The main lesson from these results is that trading by retail investors, as measured by the rate at which switching points occur, reflects the passage of business time in a manner strikingly close to the predictions of market microstructure invariance. A conceptual issue raised by this result concerns whether invariance results from the trading behavior of institutional investors or retail traders. As developed by Kyle and Obizhaeva (2013), the invariance hypothesis is based on the idea that institutional investors choose their strategies for placing bets and professional intermediaries respond to these bets in a manner which leads to invariance relationships. The powerful results in this paper suggests that the trading behavior of individual

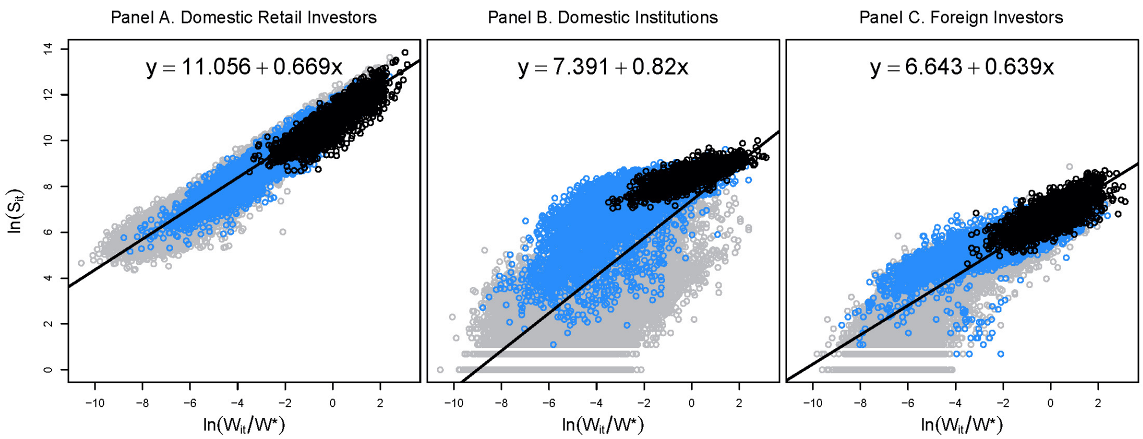

Figure 2.3: Aggregate Number of Switching Points $\ln(S_{it})$ against Trading Activity $\ln(W_{it}/W^{*})$ for different types of investors: The vertical axis is $\ln(S_{it})$. The horizontal axis is $\ln(W_{it}/W^{*})$, where $W^{*} = 10^{6}{\cdot}40{\cdot}1186{\cdot}0.02$ and $W_{it} = V_{it}{\cdot}P_{it}{\cdot}\sigma_{it}$. Panel A presents results for domestic retail investors; the fitted line is $11.056+0.669\cdot\ln(W_{it}/W^{*})$. Panel B presents results for domestic institutional investors; the fitted line is $7.391+0.82\cdot\ln(W_{it}/W^{*})$. Panel C presents results for foreign investors; the fitted line is $6.643 + 0.639 \cdot \ln(W_{it}/W^{*})$.

investors leads to invariance relationships as well.

The importance of retail investors may be an institutional characteristic specific to the South Korean stock market. In the South Korean stock market, retail investors account for a much larger share of trading than in most other countries, about 78.32% of double-counted trading volume, i.e., about 39.16% of buys and 39.15% of sells. Many large traders are classified as retail investors in the data, but they trade in a manner similar to institutional investors; South Koreans often refer to large retail investors as "super-ants".

2.6 Effective Relative Tick Size, Index Inclusion, and Other Explanatory Variables

When the slope is fixed at the predicted value of 2/3 and only a constant term is estimated, we obtain $\ln(S_{it}) = 11.123 + 2/3 \cdot \ln(W_{it}/W^*) + \epsilon_{it}$; the mean squared error is 0.190 and the R^2 is 0.927 (where $1 - R^2$ is defined as the variance of residuals divided by the variance of the demeaned data, i.e., 0.190/2.60). Neither the mean squared error nor the R^2 are different from the regression equation (2.4) in an economically significant way, since the data closely fit the invariance relationship to begin with. Thus, invariance explains about 93% of the variations in the logarithm of the number of buy-sell switching points. We next study what explains the remaining variation in the aggregate number of switching points.

Table 2.2 presents results of panel data regressions of the logarithm of the number of switching points by month and stock on five sets of explanatory variables:

1. a constant term;

2. a constant term and the logarithm of trading activity $\ln(W_{it}/W^*)$;

3. a constant term; the logarithm of trading activity $\ln(W_{it}/W^*)$; and the logarithm of effective relative tick size $\ln(e_{it}/e^*)$;

4. a constant term; the logarithm of the three separate components of trading activity, share volume $\ln(V_{it}/V^*)$, share price $\ln(P_{it}/P^*)$, volatility $\ln(\sigma_{it}/\sigma^*)$); the logarithm of effective relative tick size $\ln(e_{it}/e^*)$; the logarithm of the stock's turnover rate $\ln(\nu_{it}/\nu^*)$; the logarithm of a fraction of volume executed by domestic retail investors $\ln(\alpha_{it}/\alpha^*)$; dummy variables for stocks in the KOSPI 50 and the KOSPI 200 universes; and month fixed effects;

5. the logarithm of trading activity $\ln(W_{it}/W^*)$ and stock fixed effects;

6. the logarithm of effective relative tick size $\ln(e_{it}/e^*)$; the logarithm of the components of trading activity (share volume $\ln(V_{it}/V^*)$, share price $\ln(P_{it}/P^*)$, volatility $\ln(\sigma_{it}/\sigma^*)$); the logarithm of the turnover rate $\ln(\nu_{it}/\nu^*)$; the logarithm of the fraction of volume executed by domestic retail investors $\ln(\alpha_{it}/\alpha^*)$;

dummy variables for the stocks in the KOSPI 50 and the KOSPI 200 universes; month and stock fixed effects.

All explanatory variables are scaled so that the estimated coefficients correspond to the benchmark stock with $V^* = 10^6$, $P^* = 40 \cdot 1,186$, $\sigma^* = 0.02$, $\alpha^* = 1$, $\nu^* = 1/12$, and $W^* = V^* \cdot P^* \cdot \sigma^*$. The standard errors are clustered at the monthly level.

The most important results are the R^2 and the mean squared errors of each specification. The coefficients themselves are less important, because they are heavily affected by multi-collinearity.

The main lesson of table 2.2 is that the addition of other explanatory variables, including month and stock fixed effects, improves the R^2 in a statistically significant manner but nevertheless leaves some economically significant variation unexplained. The initial variation of the dependent variable is equal to 2.60 $(21,395$ observations). In comparison with the R^2 of 0.927 for in the first column (where only a constant term is estimated), the remaining five specifications have R^2 of 0.935, 0.936, 0.973, 0.969, and 0.984, respectively. The highest value of 0.984 is achieved in the fifth specification which has 8 estimated parameters, 36 month fixed effects, and 686 stock fixed effects $(20,665$ degrees of freedom). The mean squared errors of the regressions show similar variation across different specifications.

Covariate	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	8.500	11.156	11.358	–	–	–
	(0.059)	(0.022)	(0.046)	–	–	–
$\ln(W/W^*)$	–	0.675	0.659	–	0.679	–
	–	(0.005)	(0.005)	–	(0.005)	–
$\ln(e/e^*)$	–	–	0.066	–	–	−0.047
	–	–	(0.012)	–	–	(0.007)
$\ln(P/P^*)$	–	–	–	0.539	–	0.617
	–	–	–	(0.012)	–	(0.014)
$\ln(V/V^*)$	–	–	–	0.727	–	0.802
	–	–	–	(0.016)	–	(0.018)
$\ln(\sigma/\sigma^*)$	–	–	–	0.245	–	0.228
	–	–	–	(0.008)	–	(0.011)
$\ln(\nu/\nu^*)$	–	–	–	0.049	–	−0.023
	–	–	–	(0.018)	–	(0.020)
$\ln(\alpha/\alpha^*)$	–	–	–	0.590	–	0.562
	–	–	–	(0.025)	–	(0.025)
KOSPI50	–	–	–	−0.028	–	−0.030
	–	–	–	(0.020)	–	(0.017)
KOSPI200	–	–	–	0.120	–	0.127
	–	–	–	(0.026)	–	(0.027)
F.E. Month	No	No	No	Yes	No	Yes
F.E. Stock	No	No	No	No	Yes	Yes
Nobs	21,935	21,935	21,935	21,935	21,935	21,935
Adj. R^2	0.927	0.935	0.936	0.973	0.969	0.984
MSE	2.926	0.190	0.188	0.078	0.091	0.047

Table 2.2: Explanatory Power of Other Variables: The explanatory variables are trading activity $\ln(W_{it}/W^*)$, share volume $\ln(V_{it}/V^*)$, share price $\ln(P_{it}/P^*)$, volatility $\ln(\sigma_{it}/\sigma^*)$, effective relative tick size $\ln(e_{it}/e^*)$, turnover rate $\ln(\nu_{it}/\nu^*)$, the fraction of volume executed by domestic retail investors $\ln(\alpha_{it}/\alpha^*)$, and dummy variables for stocks in the KOSPI 50 and the KOSPI 200 universes. Some specifications have month and stock fixed effect. In the first column, $1 - R^2$ is defined as the variance of residuals divided by the variance of the demeaned data, i.e., 0.190/2.60

2.7 Decomposition into the Number of Accounts and the Number of Switching Points per Account

By definition, the aggregate number of switching points is equal to the product of the number of unique accounts traded in a given month and the average number of switching points per account. The cross-sectional variation in those two factors is the question we examine next.

Figure 2.4 shows the relationship between the logarithm of the number of unique accounts $\ln(N_{it})$ trading a given security i during a given month t and the logarithm of trading activity $\ln(W_{it})$. The slopes of 0.625, 0.666, and 0.595 for domestic retail investors, domestic institutions, and foreign investors, respectively, are slightly lower than the value of $2/3$ implied by invariance if the number of switching points per account is constant. The higher intercept for domestic retail investors reveals the exceptionally high level of retail participation in the South Korean stock market. Domestic institutions and foreign investors are less active than retail investors. Many stocks were traded by only a few domestic institutions or foreign investors during a particular month, as reflected by clustering of data points around horizontal lines of $\ln(1)$, $\ln(2)$, $\ln(3)$, and $\ln(4)$.

Figure 2.5 shows the analogous relationship for the average number of switching points per account, $\ln(S_{it}/N_{it})$. The clouds of data points for all three categories of

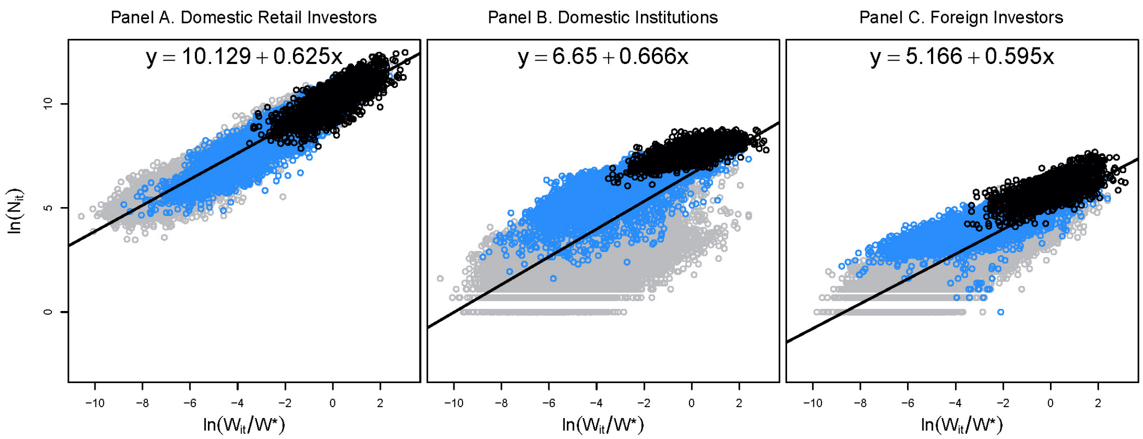

Figure 2.4: The Number of Unique Accounts $\ln(N_{it})$ against Trading Activity $\ln(W_{it}/W^*)$ for different types of investors: The vertical axis is $\ln(N_{it})$. The horizontal axis is $\ln(W_{it}/W^*)$, where $W^* = 10^6 \cdot 40 \cdot 1186 \cdot 0.02$ and $W_{it} = V_{it} \cdot P_{it} \cdot \sigma_{it}$. Panel A presents results for domestic retail investors; the fitted line is $10.129 + 0.625 \cdot \ln(W_{it}/W^*)$. Panel B presents results for domestic institutional investors; the fitted line is $6.65 + 0.666 \cdot \ln(W_{it}/W^*)$. Panel C presents results for foreign investors; the fitted line is $5.166 + 0.595 \cdot \ln(W_{it}/W^*)$.

traders—domestic retail, domestic institutions, foreign investors—are almost flat. The slopes of 0.044, 0.154, and 0.043 for the three investor categories are close to zero. The sums of the slopes in figure 2.4 and figure 2.5 are by construction equal to the corresponding slopes in figure 2.3. There are more data points on the left side of the subplot for domestic retail investors rather than the other subplots, since domestic institutions and foreign investors avoid trading South Korean stocks with low trading activity.

The clustering patterns along horizontal lines are less distinct than before because the horizontal lines correspond to both integers (such as one switch for one account, two switches for one account, two switches for two accounts) and fractions (one switch for two accounts, one switch for three accounts, two switches for three accounts, etc.); nevertheless, the horizontal clustering is still visible on panel B and panel C. Also, the data points in those two panels are somewhat symmetric relative to each other.

We conclude that the invariance relationship arises mostly from cross-sectional variation in the number of unique accounts, not from number of switching points per account. This empirical fact is consistent with the spirit of the theoretical model in Kyle and Obizhaeva (2013), where the endogenously determined number of traders—each of whom makes decision to participate in the trading game, buy

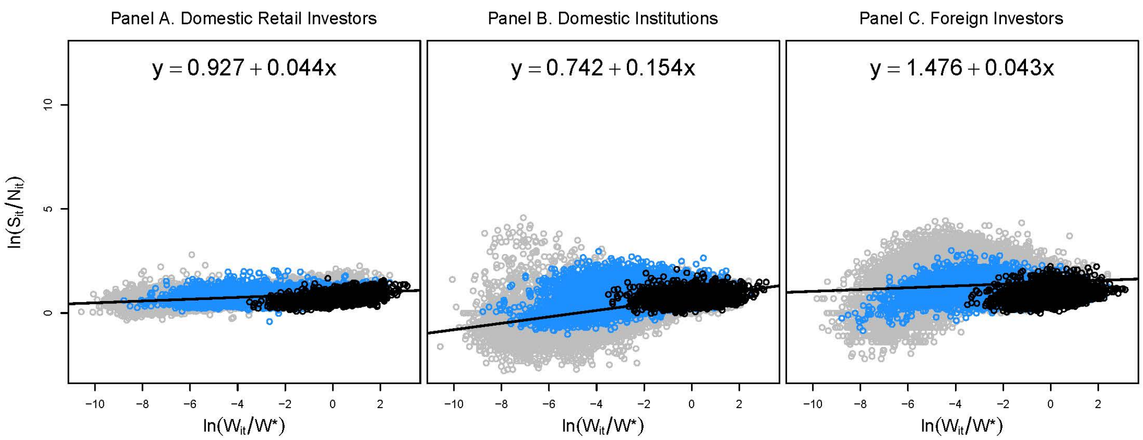

Figure 2.5: The Average Number of Switching Points per Account $\ln(S_{it}/N_{it})$ against Trading Activity $\ln(W_{it}/W^{*})$ for different types of investors: The vertical axis is $\ln(S_{it}/N_{it})$. The horizontal axis is $\ln(W_{it}/W^{*})$, where $W^{*} = 10^{6} \cdot 40 \cdot 1186 \cdot 0.02$ and $W_{it} = V_{it} \cdot P_{it} \cdot \sigma_{it}$. Panel A presents results for domestic retail investors; the fitted line is $0.927 + 0.044 \cdot \ln(W_{it}/W^{*})$. Panel B presents results for domestic institutional investors; the fitted line is $0.742 + 0.154 \cdot \ln(W_{it}/W^{*})$. Panel C presents results for foreign investors; the fitted line is $1.476 + 0.043 \cdot \ln(W_{it}/W^{*})$.

a signal of the same precision, and place exactly one bet—is shown to satisfy the invariance relationship.

Yet, this similarity should be taken with a word of caution. A slope slightly lower than 2/3 for the number of accounts may indicate that financial firms devote more resources, generate better signals, and place bigger bets when trading more active stocks. For example, domestic institutions and foreign investors may restrict their trading to stocks present in relevant benchmark indices such as the MSCI Emerging Markets Index, of which South Korea is one of the largest components. The empirical patterns may also be influenced by trades of cross-market arbitrageurs that tend to flatten the average number of switching points across stocks in indices.

2.8 Conclusion

The patterns documented in this paper strongly support the predictions of market microstructure invariance. This evidence complements the evidence on the invariance relationships in the U.S. market data documented by Kyle and Obizhaeva (2013), Kyle, Obizhaeva and Tuzun (2012), and Kyle et al. (2014). It suggests that invariance relationships hold in all markets, not just the U.S. markets. It also suggests that the trading of individual traders, not just institutions, exhibits invariance relationships.

The results in this paper are so precise that they look like empirical evidence from a physics journal rather than from an economics or finance journal. Yet, the empirical patterns reported in this paper are not regularities which have an explanation based on a mechanical interdependence among variables. If there is an alternative to the market microstructure invariance hypothesis which better explains how the number of buy-sell switching points varies across stocks, we leave it to other researchers to discover it.

Bibliography

[1] Matthew Baron, Jonathan Brogaard, and Andrei Kirilenko. The trading profits of high frequency traders. *Working Paper*, 2012.

[2] Dan Bernhardt and Bart Taub. Front-running dynamics. *Journal of Economic Theory*, 138(1):288–296, 2008.

[3] Dimitris Bertsimas and Andrew W Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[4] Hendrik Bessembinder. Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis*, 38(04):747–777, 2003.

[5] Hendrik Bessembinder and Paul J Seguin. Price volatility, trading volume, and market depth: Evidence from futures markets. *Journal of financial and Quantitative Analysis*, 28(01):21–39, 1993.

[6] Bruno Biais, Thierry Foucault, and Sophie Moinas. Equilibrium fast trading. *Working Paper*, 7, 2013.

[7] Johannes H Breckenfelder. Competition between high-frequency traders, and market quality. *Available at SSRN 2336028*, 2013.

[8] Jonathan Brogaard, Björn Hagströmer, Lars L Norden, and Ryan Riordan. Trading fast and slow: Colocation and market quality. *Available at SSRN 2317797*, 2013.

[9] Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High-frequency trading and price discovery. *Review of Financial Studies*, page hhu032, 2014.

[10] Markus K Brunnermeier and Lasse Heje Pedersen. Predatory trading. *The Journal of Finance*, 60(4):1825–1863, 2005.

[11] Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *Working Paper*, 6, 2013.

[12] Louis KC Chan and Josef Lakonishok. Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, 33(2):173–199, 1993.

[13] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Order imbalance, liquidity, and market returns. *Journal of Financial economics*, 65(1):111–130, 2002.

[14] Adam D Clark-Joseph. Exploratory trading. *Working Paper*, 2013.

[15] Carole Comerton-Forde, Terrence Hendershott, Charles M Jones, Pamela C Moulton, and Mark S Seasholes. Time variation in liquidity: The role of market-maker inventories and revenues. *The Journal of Finance*, 65(1):295–331, 2010.

[16] Jennifer Conrad, Kevin M Johnson, and Sunil Wahal. Institutional trading and alternative trading systems. *Journal of Financial Economics*, 70(1):99–134, 2003.

[17] F Douglas Foster and S Viswanathan. Strategic trading when agents forecast the forecasts of others. *The Journal of Finance*, 51(4):1437–1478, 1996.

[18] Thierry Foucault and Albert J Menkveld. Competition for order flow and smart order routing systems. *The Journal of Finance*, 63(1):119–158, 2008.

[19] Simon Gervais, Ron Kaniel, and Dan H Mingelgrin. The high-volume return premium. *The Journal of Finance*, 56(3):877–919, 2001.

[20] John M Griffin, Jeffrey H Harris, and Selim Topaloglu. The dynamics of institutional and individual trading. *The Journal of Finance*, 58(6):2285–2320, 2003.

[21] Lawrence Harris. Order exposure and parasitic traders. *Working Paper*, 1997.

[22] Lawrence Harris and Joel Hasbrouck. Market vs. limit orders: the superdot evidence on order submission strategy. *Journal of Financial and Quantitative analysis*, 31(02):213–231, 1996.

[23] Joel Hasbrouck. Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207, 1991.

[24] Joel Hasbrouck and Gideon Saar. Low-latency trading. *Journal of Financial Markets*, 16(4):646–679, 2013.

[25] Joel Hasbrouck and George Sofianos. The trades of market makers: An empirical analysis of nyse specialists. *The Journal of Finance*, 48(5):1565–1593, 1993.

[26] Terrence Hendershott, Charles M Jones, and Albert J Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33, 2011.

[27] Terrence Hendershott and Mark S. Seasholes. Market maker inventories and stock prices. *The American Economic Review*, 97(2):pp. 210–214, 2007.

[28] Nicholas Hirschey. Do high-frequency traders anticipate buying and selling pressure? *Working Paper*, 2013.

[29] Craig W Holden, Stacey E Jacobsen, and Avanidhar Subrahmanyam. The empirical analysis of liquidity. *Available at SSRN*, 2014.

[30] Craig W Holden and Avanidhar Subrahmanyam. Long-lived private information and imperfect competition. *The Journal of Finance*, 47(1):247–270, 1992.

[31] Andrei A Kirilenko, Albert P Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: The impact of high frequency trading on an electronic market. *Working Paper*, 2010.

[32] Alan Kraus and Hans R Stoll. Price impacts of block trading on the new york stock exchange. *The Journal of Finance*, 27(3):569–588, 1972.

[33] Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.

[34] Albert S Kyle and Anna A Obizhaeva. Market microstructure invariance: Theory and empirical tests. *Working Paper*, 2013.

[35] Albert S. Kyle, Anna A. Obizhaeva, Nitish R. Sinha, and Tugkan Tuzun. News Articles and the Invariance Hypothesis. *Working Paper*, 2014.

[36] Albert S. Kyle, Anna A. Obizhaeva, and Tugkan Tuzun. Trading Game Invariance in the TAQ Dataset. *Working Paper*, 2012.

[37] Albert S Kyle, Anna A Obizhaeva, and Yajun Wang. Smooth trading with overconfidence and market power. *Available at SSRN 2423207*, 2013.

[38] Josef Lakonishok, Andrei Shleifer, and Robert W Vishny. The impact of institutional trading on stock prices. *Journal of financial economics*, 32(1):23–43, 1992.

[39] Wei Li. High frequency trading with speed hierarchies. *Available at SSRN 2365121*, 2013.

[40] Ananth Madhavan and Seymour Smidt. An analysis of changes in specialist inventories and quotations. *The Journal of Finance*, 48(5):1595–1628, 1993.

[41] Steven Manaster and Steven C Mann. Life in the pits: Competitive market making and inventory control. *Review of Financial Studies*, 9(3):953–975, 1996.

[42] Sarah Zhang. Need for speed: An empirical analysis of hard and soft information in a high frequency world. *Manuscript, Karlsruhe Institute of Technology*, 2012.