Large Trade Anticipation

Kyoung-Hun Bae*   Peter Dixon**   Eun-Jung Lee*

We study whether a subset of high frequency traders, which we refer to as opportunistic high frequency traders, anticipate and trade opportunistically around individual large trades to profit from their price impact. The trading patterns we document using account level transaction data from the Korean Stock Exchange are consistent with opportunistic high frequency traders anticipating large trades and trading opportunistically around them. Our findings are difficult to reconcile with alternative hypotheses that opportunistic high frequency traders and large traders are trading on a common price signal, or that the observed trading behavior is the byproduct of market making strategies.

JEL classification: G12, G14

Keywords: High Frequency Trading, Order Execution, Order Anticipation

***** Hanyang University
** The University of Tennessee-Knoxville
A prior version of this paper was circulated under the title "An Empirical Analysis of Predatory Trading" and "Predatory High Frequency Trading". The authors would like to thank Don Bowen, John Chao, Kaitlyn Dixon, Steve Heston, Eric Kelley, Hyeng Keun Koo, Albert S. "Pete" Kyle, Wei Li, Mark Loewenstein, David Maslar, Richmond Mathews, Tom McInish, Anna A. Obizhaeva, Andy Puckett, Albert Rossi, Shrihari Santosh, Austin Starkweather, Tom Steffen, Antonio Mello, and participants at the 2016 FMA annual meeting for their help and useful comments.

***** Hanyang University
** The University of Tennessee-Knoxville
A prior version of this paper was circulated under the title "An Empirical Analysis of Predatory Trading" and "Predatory High Frequency Trading". The authors would like to thank Don Bowen, John Chao, Kaitlyn Dixon, Steve Heston, Eric Kelley, Hyeng Keun Koo, Albert S. "Pete" Kyle, Wei Li, Mark Loewenstein, David Maslar, Richmond Mathews, Tom McInish, Anna A. Obizhaeva, Andy Puckett, Albert Rossi, Shrihari Santosh, Austin Starkweather, Tom Steffen, Antonio Mello, and participants at the 2016 FMA annual meeting for their help and useful comments.

1. Introduction

Large traders face a complex decision of when and how to expose their trading interests. If they underexpose their intent to trade they risk being unable to find a trading partner. If they overexpose their intent they may face increased transaction costs as other market participants opportunistically trade around their trade (Harris (1997), Brunnermeier and Pedersen (2005), Schöneborn and Schied (2009)).

Over the past decade or so, high frequency trading has transformed the environment in which large traders place orders and signal their intent to trade. One concern expressed by large traders is that in this new trading environment, some high frequency traders use their speed and technological sophistication to identify a large trader's intent to transact and then opportunistically trade around their large trades.

When seeking to transact a large trade, large traders attempt to signal their intent to trade to potential counterparties while preventing that information from leaking to opportunistic traders. If opportunistic traders become informed of an incoming large trade, it can be costly to large traders for two reasons. The first is if other market participants learn of the upcoming large trade and take away liquidity prior to its arrival. This behavior increases the transaction costs of the large trade (Brunnermeier and Pedersen, 2005). Second, it decreases the value of information. If other traders become aware of an incoming trade, they may incorporate some of the information of that trade prior to its arrival – effectively increasing the cost of becoming informed (Brunnermeier, 2005).

Understanding if large trades can be predicted by opportunistic traders may also have implications for market design. If large traders fear that too much of their intent to trade will leak to opportunistic traders they may take their trades off of lit exchanges and into venues such as dark pools.[1]

In this study, we explore whether a subset of high frequency traders, which we refer to as opportunistic high frequency traders (hereafter OHFTs), anticipate individual large trades and trade opportunistically around them to

[1] Many dark pools openly court traders with such concerns by advertising less information leakage and better terms for large traders. For example, the ITG POSIT dark pool, one of the largest dark pools in the US, advertises on its homepage "minimized information leakage" and states that it "rewards size over speed" https://www.itg.com/solutions/liquidity/ accessed on May 12, 2017, see appendix A for a screenshot

benefit from their price impact. The evidence presented in this study is consistent with the hypothesis that OHFTs anticipate individual large trades and trade opportunistically around them, and are difficult to reconcile with alternative hypotheses that OHFTs and large traders transact on a common price signal or common trading strategies – e.g. momentum – or that the results are a byproduct of market making strategies.

To perform our analysis, we obtain account level data from the Korean Stock Exchange for trading in Korean KOSPI 200 futures for 66 consecutive trading days beginning March 26, 2009. We define large trades as those trades in the top 1% of trades by size. We then use trading behavior to classify each of the 25,172 accounts in our data as being either a large trader, a small trader, or an OHFT.

We first study how OHFTs trade around the arrival of large trades. For an OHFT to profit from trade anticipation, they must be able to both identify when a large trade is imminent, and the sign of the large trade. If OHFTs can anticipate a large trade, then we should observe empirically that they trade actively in one direction prior to the arrival of the large trade, and that they unwind their trades immediately after the large trade. If OHFTs can sign large trades with greater than random chance, then their inventory run-up will be in the same direction as the incoming large trade statistically more than 50% of the time.

Using an event study of OHFT trading behavior around large trades, we observe behavior that is consistent with order anticipation and opportunistic trading. OHFTs aggressively build positions during the 200 trades prior to the arrival of a large trade and they immediately begin to unwind these positions after the large trade arrives. Also, consistent with OHFTs being able to sign large trades with greater than 50% probability, we observe that their inventory accumulation is in the same direction as the large trade 56% of the time. This fraction is statistically greater than 50%.

This trading behavior is hard to reconcile with the hypothesis that OHFTs are targeting common price signals. If this were true, then to generate the trading patterns we observe, it would need to be the case that the OHFTs and the large trader executing the large trade (but not other large traders) consistently receive simultaneous signals about price that are strong enough to generate significant amounts of trading. However, these signals would be in the same direction only 56% of the time. The other 44% of the time the OHFTs would need to receive a signal that is simultaneous with the arrival of the incoming large trade, but opposite to the information that the large trade is acting on. Further, if OHFTs and large traders do receive simultaneous signals that are in the opposite direction 44% of the time, it is difficult to understand why the arrival of a large trade would induce the OHFTs to systematically give up on their accumulated positions.

Although suggestive, the event study by itself is not conclusive evidence that large trades are systematically anticipated by OHFTs. The event study is conditional on the arrival of a large trade and cannot speak directly to whether OHFTs can predict the arrival of large trades. To address this limitation, we perform an analysis of large trade anticipation that is unconditional of the arrival of a large trade. In this unconditional analysis, we use the inventories of OHFTs to predict the net size and direction of future large trades. In these regressions, we observe that the size and direction of OHFTs' inventories successfully predict the size and direction of future large trades – consistent with order anticipation.

A drawback to the unconditional analysis is that it cannot explicitly differentiate between whether the OHFTs are targeting the large trade itself or whether they are simply trading on a common signal.[2] In essence, the regressions may suffer from omitted variable bias. In this case, the potential omitted variable – in both the same signal and the momentum trading cases – is the price change that both the OHFTs and the large traders may be responding to.

If this is the case, and OHFTs are targeting a common price signal, then after controlling for price changes in our regressions, the predictive power of OHFT inventory to predict the size and direction of large trades should disappear. To this end we re-estimate our unconditional regressions which determine if OHFT inventory predicts the size and direction of future large trades with the added control variables of contemporaneous and lagged returns.

We find in these regressions that the coefficients indicating the magnitude of the correlation between OHFT inventories and future large trades remains positive and significant at the 1% level after controlling for lagged and

[2] These could be external signals not yet incorporated into prices, or they could be prior price changes i.e. momentum.

contemporaneous price changes. These results are also difficult to reconcile with the idea that OHFTs and large traders are trading on common price signals. For this to be true, the price signal observed by the OHFT and the large trader would have to be uncorrelated with the returns on the underlying asset leading up to and contemporaneous with the large trade. If our results are the byproduct of OHFTs and large traders both trading momentum strategies, then their trades should be correlated with prior price changes. Consequently, after controlling for prior and contemporaneous price changes, the relation between OHFT inventory and large trades should disappear – but it does not.

We next seek insight into how OHFTs anticipate large trades. We explore two hypotheses related to the mechanisms which may allow OHFTs to anticipate large trades. First, we speculate that large trades associated with larger than average child orders may be easier to identify and thus would be more likely to be anticipated. Second, we conjecture that large traders that are 'slower' may not be as able to respond to changes in the trading environment, and thus would be more likely to have their large orders anticipated.

We test both conjectures via regression analysis and find support for both the child order size and the speed conjecture. Although, in the context of our study, it is child order size that plays a larger role in determining whether OHFTs correctly accumulate positions in the same direction as the large trade prior to its arrival.

Lastly, we explore the profitability of the OHFTs in our sample. We find that OHFTs earn an average profit (loss) of 2.1 (1.6) basis points for each successfully (unsuccessfully) signed large trade. The asymmetry between profit and loss for successful and unsuccessful attempts by OHFTs to sign large trades suggests that OHFTs trade in a manner that mitigates losses to failed attempts, and maximizes gains to successful attempts to sign trades and trade opportunistically around them.

We estimate hourly mark-to-market profits for the OHFTs in our sample and find that OHFT profits are positive 81% of the time and that the distribution of hourly mark-to-market profits exhibits positive skewness. These findings are consistent with Baron, Brogaard, and Kirilenko (2012) in suggesting that OHFTs are on average profitable. They also suggest that trade anticipation may be a viable strategy for OHFTs to pursue.

Our study makes several contributions to the literature. First, our analysis would appear to validate some of the concerns of large traders that their large trades are systematically anticipated in financial markets. This result suggests a possible additional explanation for recent growth in off exchange trading. Also, the finding that both child order size and speed of the parent trader influence the likelihood that a large trade is anticipated have implications for optimal order execution strategies. Lastly, this study adds to the growing literature seeking to understand the impact of high frequency trading on financial markets by studying individual HFT strategies.

2. Related Literature

In the study perhaps most relevant to ours, Clark-Joseph (2014) how high frequency traders may gather information about the state of the market from small exploratory trades. He then then tests the implications of his model with data on high frequency trading in the E-mini S&P 500 futures market. In his model, a high frequency trader uses small exploratory trades to try and discern the current state of price impact. This information is valuable to the high frequency trader because the profit earned when an HFT successfully anticipates a large trade is contingent on the price impact of the large trade. Consequently, small exploratory trades help the HFT to trade around predictable trades only when price impact is sufficient to make trade anticipation profitable. Clark-Joseph (2014) finds empirical support for his model by documenting that the market response to the HFT's most recent exploratory trade is related to the profit earned on subsequent larger trades executed by the HFT.

Our study compliments Clark-Joseph (2014) because, while in his study, he suggests that HFTs can anticipate large trades and thus profit from their price impact, he never explicitly tests this assertion. Consequently, the findings in our study help to support and motivate the analysis in Clark-Joseph (2014). Clark-Joseph (2014) also provides motivation for studying high frequency trading around individual trades in addition to analyzing high frequency trading across the life of a large order as in Tong (2015), Kervel and Menkveld (2016), and Korajczyk and Murphy (2016).

Additional motivation for studying high frequency trading around individual trades comes from Bernhardt and Taub (2008) who analyze how traders with advance knowledge of an impending trade can trade in a dynamic setting to maximize profit. Other studies which analyze order anticipation include Yang and Zhu (2015) who use a two period Kyle (1985) model with a 'back runner' who observes the period one order flow and places an order in period two. From period one order flow the 'back runner' extracts a signal about the informed trader's intention to trade in period two and trades accordingly.

Their study suggests that careful analysis of order flow may allow a trader to anticipate and opportunistically trade around another trader's trade. Li (2014) also uses a variation of a Kyle (1985) model to study the trading dynamics when one trader has a speed advantage suggesting that speed may also play a role in allowing a trade to be anticipated. Attari, Mello, and Ruckes (2005) show that when a large trader is capital constrained their trades may become predictable, and that this predictability attracts other traders who opportunistically trade around the predictable trades of the capital constrained trader. In this model, the financial state of the large trader may provide information about future trades which makes the individual trades predictable.

Other studies related to ours include Hirschey (2013), van Kervel and Menkveld (2016) and Korajczyk and Murphy (2016). The key difference between our study and these is that we analyze high frequency trading around individual orders. Hirschey (2013) provides evidence that high frequency traders tend to trade in advance of periods of large buying or selling pressure. However, it is unclear in Hirschey's (2013) study whether the high frequency traders are anticipating order flow, or are simply quicker to act on a common signal than other traders. This is because the high frequency traders in Hirschey's (2013) study do not liquidate their holdings after the buying or selling pressure abates as would be expected if the high frequency traders were targeting the order flow.

Van Kervel and Menkveld (2016) and Korajczyk and Murphy (2016) use Swedish and Canadian data respectively to analyze the effect of high frequency trading on implementation shortfall across the life of a large order that is executed via smaller trades over a long time period. Their analysis suggests that high frequency traders tend to 'back run' large orders that are spread over a long period. That is, when an order is shredded and implemented over the course of multiple hours, high frequency traders tend to provide liquidity to the initial trades, but then take away liquidity for the later trades.

The analysis performed in Hirschey (2013), van Kervel and Menkveld (2016), and Korajczyk and Murphy (2016) helps to shed light on the complex relationship between HFTs and large institutional traders, however none of these studies directly address the fundamental question analyzed in this study, which is whether or not individual large trades are anticipated by high frequency traders as is asserted by large traders.

Our study also contributes to the growing literature which sheds light on the impact of high frequency trading on financial markets by studying their individual strategies. As discussed in O'Hara (2015), high frequency traders engage in a variety of strategies. Many of which are liquidity providing and market quality enhancing, while others may be less so. Studies documenting the significant amount of liquidity provision provided by HFTs include Brogaard (2010), Carrion (2013), Menkveld (2013), and Hagstromer & Norden (2013). Other studies which analyze individual HFT strategies include O'Hara (2011) who examines potential spoofing by high frequency traders. Egginton, Van Ness, and Van Ness (2011) and Ye, Yao, and Jiading (2013) examine potential quote stuffing by high frequency traders.[3] Brogaard, Hendershott, and Riordan (2017) study specifically short selling by high frequency traders, and as previously discussed Clark-Joseph (2014) studies the use by high frequency traders of small exploratory trades to acquire information about market conditions. We add to this growing literature examining the effects of high frequency trading by studying HFT trading dynamics around individual large trades.

Lastly, to the extent that trade anticipation on exchanges induces large traders to move their trading off of lit exchanges and into off exchange venues such as dark pools, the findings in this study are relevant to the literature studying off exchange trading venues such as dark pools. The theoretical literature studying the interaction of dark pools (or crossing networks) and lit exchanges tends to model dark pools as providing a tradeoff of low cost transactions for immediacy (Degryse, Van Achter, and Wuyts (2009), Ye (2012), Zhu (2014), Buti, Rindi, and Werner (2017)). These lower costs occur because dark pools often transact at the prevailing NBBO midpoint at the time of the trade, and thus minimize the spreads paid by investors. However, dark pools cannot guarantee that there will be a counterparty for your order. Thus, traders in dark pools face the tradeoff of cost for immediacy.

[3] See also Biais and Wooley (2011) and Jarrow and Protter (2012) for a further discussion of high frequency trading strategies.

In addition to lower cost execution, Kwan, Masulis, and McInish (2015) suggest that dark pools in the US provide investors with the opportunity to circumvent trading rules that require the tick size on lit exchanges to be equal to a penny. To this growing literature our findings suggest that trade anticipation on lit exchanges may decrease the appeal of trading on lit exchanges and may induce large traders to move into off exchange venues such as dark pools.

3 Data

We obtain data from the Korean Stock Exchange (hereafter KRX). Our data contains the complete record of account level trades and quotes for KOSPI 200 index futures time-stamped at the millisecond level. If multiple events occur in the same millisecond, the correct order of events is preserved in the data. Our sample begins on March 26, 2009 and continues for 66 consecutive trading days.

The KOSPI 200 Index consists of 200 large cap stocks listed on the KRX. Options and futures derived from the KOSPI 200 index are among the most liquid and highly traded derivatives in the world[4]. As a well-diversified basket of stocks, the KOSPI 200 index is less susceptible to stock specific idiosyncratic risk, providing a relatively clean environment with which to study trading activity. All KOSPI 200 contracts are traded electronically. The notional value of one KOSPI 200 index futures contract is the KOSPI 200 futures price times a multiplier of 500,000 Korean Won. This translates to an average notional value of one contract of approximately USD 68,000[5] during our sample. The tick size for KOSPI 200 index futures is 0.05 or 2.86 basis points (approximately USD 19.37).

Because of illiquidity and low trading volumes (less than 2% of total trading volume) in back month contracts we limit our analysis to trading in front month contracts.[6] As June 11, 2009 is the only expiration day in our sample, prior to June 11 we study trading in contracts with a June 11 expiration date and after June 11 we study trading in

[4] See Ahn, Kang, and Ryu (2008) (2010)
[5] The average closing price for the KOSPI 200 index during our sample is 174.97 and the average USD/KRW exchange rate is 1,290.73.
[6] Including back month contracts does not qualitatively change our results.

contracts which expire on September 10, 2009. KOSPI 200 futures contracts operate with a daily price limit of 10%. During our sample, there were not any days with fluctuations severe enough to trigger the circuit breaker.

The granularity of our data allows us to sign order flow very accurately without relying on algorithms such as the Lee and Ready (1991) algorithm. We do this by simply ordering the events in a given transaction. When two accounts complete a transaction, we observe both the time that the transaction was completed as well as the times that both traders submitted the orders which resulted in the transaction. If the buyer in a given trade submits his order after the seller does, then the buyer is determined to be the active party and the trade is identified as a buyer-initiated trade.

To facilitate our analysis, we define large trades as those active trades which are in the top 1% of trades by size across our whole sample. We exclude large trades which occur in the first five minutes of trading, because OHFTs lack sufficient order flow to attempt to anticipate large trades at the very start of trading. We also exclude from our analysis large trades which are time stamped at exactly the market open or close (9:00am or 3:00pm). This process identifies 36,164 qualifying large trades. The smallest trade that qualifies as a large trade is for 50 contracts, or a notional value of approximately 3.4 million USD. As presented in Table 1, the mean size of a large trade is 78.76 contracts or approximately 5.4 million USD. Of the 36,164 large trades, 48.99% are buyer-initiated trades. Large trades occur an average of 548 times per day with an average time of 40 seconds between consecutive large trades.

<Insert Figure 1>

<Insert Table 1>

In panel A of Figure 1, we present the distribution of large trades across the trading day in ten minute increments from 9:00 am until the close of the trading day at 3:00 pm. We observe, consistent with Jain and Joh (1988) and McInish and Wood (1990) that large trading volume has a U shape with more trading in the opening and closing of each day. We also plot the ratio of large trades to small trades for each ten-minute increment; in panel A we find that the ratio of large trades to small trades remains fairly constant throughout the day except for during the last 10 minutes of trading where the ratio nearly doubles. This finding suggests that large traders rush at the end of the day to complete their remaining large trades before the trading day concludes. In panel B of Figure 1, we plot the partial autocorrelation function for the sign of large trades, and we observe that the sign of large trades exhibits positive autocorrelation for up to six lags.

There are 25,172 unique accounts in our date, and we classify each account as being either a large trader, a small trader, or an OHFT based on trading behavior. Our procedure identifying OHFT accounts follows closely the procedure employed by Baron, Brogaard, and Kirilenko (2012). In this process, our goal is to eliminate accounts that trade primarily market making strategies, and then to isolate those accounts which actively trade short term price changes rather than long term fundamentals. To do so, we define accounts as OHFTs if they meet the following three criteria: (1) trade with an active trade more often than with a passive trade, (2) have an average holding time for a position of less than three minutes, and (3) have an average daily ratio of overnight inventory to daily trading volume that is less than 0.01%. This procedure is similar to the procedure employed by Baron, Brogaard, and Kirilenko (2012) and yields 32 OHFT accounts.

There is significant empirical research documenting that high frequency traders are important sources of liquidity (e.g. Brogaard (2010), Carrion (2013), Menkveld (2013), Hagstromer and Norden (2013)). Liquidity provision is primarily a passive trading strategy (Menkveld, 2013), and so the requirement that the account trade actively more often than passively is designed to filter out accounts that primarily trade market making strategies, and mitigate the concern that the behavior we observe is a byproduct of market making strategies. The last two criteria are designed to isolate accounts that specialize in trading short term price changes rather than long term fundamentals. To classify the remaining accounts, we define large traders as accounts that are not OHFTs, and that have at least one trade in the top 1% of trades by size. All other accounts are classified as small traders. This process yields 737 large traders and 24,403 small traders.

<Insert Table 2 >

In Table 2 we present descriptive statistics for each of the three account types identified in our sample. In panel A we present statistics for the number of traders in each of our three trader classifications as well as additional statistics about whether these accounts are foreign (non-Korean) or domestic (Korean), or retail or institutional accounts. In panel B of Table 2 we present volume statistics by trader group. OHFTs in our data account for 30.5% (14.5%) of all active (passive) trading volume. Large traders account for 42.1% (43.2%) of active (passive) trading volume, and small traders comprise the remaining 27.4% (42.3%) of active (passive) trading volume. Consistent with our definition, OHFTs are the initiating counterparty party more frequently than they are the passive counterparty. The volume statistics reveal that the OHFTs in our sample trade actively (i.e. demand liquidity) more than twice as often as they trade passively (i.e. provide liquidity). This statistic makes it unlikely that the accounts identified as OHFTs in our sample are predominately market makers. Finally, large trades comprise 14% of trading volume while small trades comprise the remaining 86% of trading volume.

In panel C of Table 2 we present descriptive statistics for the trading behavior of each of the three classes of investor. The first statistic that we calculate is the number of times an account's position switches from long to short or from short to long in a day. We find that the inventory for the median OHFT switches direction 27 times more than the median small trader and 40 times more frequently than the median large trader. Consistent with our definition, OHFTs hold essentially zero overnight inventory whereas large and small traders carry a significant portion of their daily trading volume overnight.

4 Empirical Analysis

4.1 OHFT Trading Around the Arrival of Large Trades

Trading opportunistically around the arrival of large trades is profitable only if those trades are associated with some measure of price impact. If there were no price impact associated with a large trade then opportunistic traders could not profit from the arrival of the large trade. We therefore first study the price dynamics around large trades using an event study. Price is denoted as $p(i,j)$ where i indexes the large trade and j indexes the trading sequence around the i^{th} large trade where $j = 0$ indicates the price at the large trade. Since we are concerned with the effect a large trade has on price changes, we define the relative price for each trade position and each large trade as presented in equation 1.

$$p'(i,j) = \ln\left(\frac{p(i,j)}{p(i,0)}\right) * 10\text{^}4 \tag{1}$$

We compute the average relative price across all large trades for each relative trading position *j* beginning at $j = -200$ and ending at $j = 200$ as $\overline{p'(j)}$ as presented in equation 2.

$$\overline{p'(j)} = \frac{1}{N}\sum_{i=1}^{N} p'(i,j) \tag{2}$$

In equation 2, *N* is the number of large trades in the sample. Because large buyer-initiated trades are likely to be associated with price increases and seller-initiated trades are likely to be associated with price declines, we calculate $\overline{p'(j)}$ separately for buyer and seller-initiated large trades. This process yields two series displaying the average price dynamics around large trades beginning 200 trades before the large trade and ending 200 trades after the arrival of the large trade. Our results are presented in Figure 2. We observe that for both large buys and large sells, the price tends to drift in the direction of the large trade in the 200 trades prior to the execution of the large trade. Then at the moment of the large trade, the price jumps in the direction of the large trade, and then remains fairly stable for the 200 trades following. When converted to basis points, the average price jump up for large buys is approximately .7 basis points and the price jump down for large sell trades is approximately 1.5 basis points. The analysis presented in Figure 2 demonstrates that large trades do generate price impact and thus OHFTs can potentially profit if they can successfully build up positions before the large trade and unwind them afterward.

<Insert Figure 2 >

We next study the behavior of the different classes of traders around large trades using an event study in Figure 3. We analyze large buys and large sells separately and we aggregate trading activity around large trades into four groups of traders including: large traders engaging in the given large trade, large traders not engaging in the large trade, OHFTs, and small traders. Since we are concerned with changes in inventory around a large trade we normalize the inventory of all groups to be equal to zero 200 trades before the arrival of the large trade.[7] For each trading position $j \epsilon[-200,200]$ we compute the average holdings for each investor class across all large trades in the given sample. This process yields four sequences representing the average inventory accumulations for each of the four groups around the arrival of a large trade. For illustration purposes, we perform this analysis for large buys, large sells, and for when OHFTs accumulate positions prior to the large trade that are in the same direction of the large trade and when they are not. Analyzing trading behavior for large trades when OHFTs are on the correct and incorrect side of the market when the large trade arrives allows us to study how OHFTs trade differently when their inventories stand to benefit or lose from the price impact generated by the arrival of a large trade.

<Insert Figure 3>

In Figure 3 we observe that, on average, OHFTs trade very actively and in one direction prior to the arrival of large trades suggesting that they as a group are responding to some signal that is correlated with the arrival of the large trade. This is true both when OHFTs build positions in the same and opposite direction as the impending large trade. We also observe in Figure 3 that after the arrival of the large trade the OHFTs begin to unwind their accumulated positions. We also measure how often OHFTs use passive (liquidity providing) versus active (liquidity taking) trades in the vicinity of large trades. We observe that that the OHFTs trade very actively as we find that in the vicinity of large trades OHFTs trade take liquidity in 68% of their trades and provide liquidity in 32% of their trades. The active nature of OHFT trading around large trades makes it unlikely that the behavior that we observe is the byproduct of market making strategies provided by the OHFTs.

Of the large trades that arrive, 26.83% (22.16%) are large buys where OHFTs accumulate positions in the same (opposite) direction as the large trade before its arrival, and that 29.23% (21.78%) of large trades are large sells where OHFTs accumulate positions in the same (opposite) direction as the large trade before its arrival. These

[7] In our data approximately 100 trades arrive every minute and thus 200 trades equals approximately two minutes of real time elapsing.

numbers indicate that OHFTs accumulate positions in the same direction as large trades prior to the arrival of the large trade for 54.77% of large buy trades and 57.39% of large sell trades for a total of 56.06% of all large trades.

Figure 3 also documents that OHFTs trade in economically meaningful magnitudes before the arrival of the large trade, as we find that the average OHFT inventory accumulation at the time of the arrival of the large trade is equal to approximately 30% of the size of the average large trade. OHFTs also trade somewhat differently after the arrival of the large trade when their inventories are on the same side as compared to when their inventories are on the opposite side of the large trade. We find that when OHFTs' inventories are on the opposite side of a large trade the OHFTs exit their positions much more quickly than they do when on the same side as a large trade.

Figure 3 shows that the OHFTs are actively accumulating inventory immediately prior to large trades, both when their inventory is in the same and opposite directions as the large trade. Given the pattern of trading behavior documented in Figure 3 it seems unlikely that the OHFTs are trading on a common price signal, and given the liquidity taking nature of OHFT trading around large traders it also seems unlikely that the trading patterns we observe are the byproduct of market making strategies. If OHFTs are trading on common price signals, then to generate our results, the OHFTs and the large trader executing the large trade (but not other large traders) would need to consistently receive simultaneous signals about price that are strong enough to generate significant amounts of trading by both large traders and OHFTs. However, these signals would be in the same direction only 56% of the time. The other 44% of the time the OHFTs would need to be receiving a signal that, although it arrives at the same time as the signal received by the large trader, is opposite to the information that the large trade is acting on, and the remaining large traders would need to not receive any signal. Further if the OHFTs are trading on an external signal that is opposite to what would be implied by the large trade 44% of the time it is difficult to explain why the OHFTs abandon their inventory, and hence their signal, upon the arrival of the large trade.

An alternative explanation for the observed trading behavior is that the OHFTs are able to anticipate the arrival of large trades, but are less successful at signing the large trade. In this case, the OHFTs act on two signals relating to a large trade, one about when the large trade is likely to occur, and a second weaker signal about the direction of the large trade. In this case, the OHFTs will consistently predict the arrival of a large trade and will thus accumulate inventory in anticipation, however the weaker signal about the direction of the large trade means that their inventory accumulation will be in the same direction as the large trade less frequently but still more than 50% of the time. Even if OHFTs in this scenario incorrectly sign and lose money on just under half of the large trades, so long as the OHFTs can correctly anticipate the sign of the large trade more than 50% of the time an application of the Kelly criteria (Kelly, 1956) would allow OHFTs to earn consistent profits from anticipating large trades and trading opportunistically around them.

We also observe in Figure 3 across all panels that large traders executing the large trade begin to accumulate inventory well in advance of the large trade suggesting that the large traders transacting the large trade engage in order shredding. Small traders and large traders not executing the large trade serve as counterparties trading against the OHFTs and the large traders who are seeking to execute a large trade. In panels A and C of Figure 3, where we plot the trading behavior when OHFTs accumulate positions in the same direction as large trades, we observe that after the arrival of the large trade small traders switch from trading in the opposite direction of the large trade, to the same direction immediately after. This behavior is convenient for the OHFTs because it provides a counterparty willing to accept the positions of the OHFTs allowing the OHFTs to unwind their positions with reduced price impact.

In Figure 4 we use the same methodology as was employed in Figure 2 to plot the average relative price around large buy and sell trades for trades where the OHFTs accumulate inventory positions in the same direction as the incoming large trade and when their inventory accumulation is in the opposite direction. In panels A and B we present the price dynamics for large buy trades when OHFTs' inventories at the time of the large trade are in the same and opposite direction as the large trade respectively. In panels C and D we present the price dynamics for large sell trades when OHFTs inventories at the time of the large trade are in the same and opposite direction as the large trade respectively.

In panels A and C, where we plot the price dynamics around large trades where the OHFTs accumulate positions in the same direction as the large trade, we observe that prior to the arrival of the large trade there is a run-up in price, followed by a jump in price corresponding to the arrival of the large trade, and after the arrival of the large trade, the price reverts somewhat as the OHFTs exit their positions. The price dynamics are somewhat different when OHFTs accumulate positions in the opposite direction as the large trade. We observe a milder run-up in price, but the run-up is in the opposite direction to the incoming large trade. After the arrival of the large trade, we do not observe a price reversal as we do in the cases where OHFTs accumulate positions on the same side as the large trade. When OHFTs' inventory is in the opposite direction as the large trade the price after the large trade appears to continue to drift in the same direction as the large trade. This drift could occur as OHFTs quickly unwind their positions adding to the price pressure of the large trade.

<Insert Figure 4 >

The analysis suggests that OHFTs may be able to anticipate large trades and trade opportunistically in response to them, however it has heretofore been essentially descriptive in nature. To formally test the null hypothesis that the probability that OHFTs accumulate inventory in the same direction as large trades is less than or equal to 50%, we use a simple regression. For this analysis, we define $Sign_i$ as the sign of the i^{th} large trade and is equal to -1 for large sells and 1 for large buys. We define $SignPT_{k,i}$ as the sign of the k^{th} OHFT account's inventory at the time of the i^{th}large trade. $SignPT_{k,i}$ is either 1 or -1 indicating whether the k^{th} OHFT is long or short at the time of the i^{th}large trade. Because large trades are autocorrelated, we control for the sign of the prior 15 large trades in our regressions. We also include fixed effects for each of the 32 OHFTs in our sample. Our base model for our hypothesis test is presented formally in equation 3.

$$Sign_i = \beta_0 SignPT_{k,i} + \sum_{j=1}^{15} \gamma_j Sign_{i-j} + \alpha_k + \varepsilon_i \tag{3}$$

From equation 3 rejecting the null hypothesis that $\beta_0 \leq 0$ implies that OHFTs accumulate inventory in the same direction as incoming large trades with a probability that is greater than random chance. In panel A of Table 3 we present the results from our regressions. We estimate three variations of equation 3. The first is a simple OLS model where we do not include the values for lagged $Sign_i$ or fixed effects for the 32 OHFTs in our data. In the second variation, we include OHFT fixed effects but we do not include the values for lagged $Sign_i$, and the last specification is the complete model presented in equation 3. We find that across all specifications the coefficient for β_0 is significantly greater than zero, leading us to reject the null hypothesis that OHFTs accumulate inventory in the same direction as the large trader due to random chance.

In panel B of Table 3 we estimate similar regressions to those in panel A except that we allow the coefficient β_0 to be influenced by the size of the large trade. If OHFTs have some notion of the expected size of the incoming large trade then we may expect them to place proportionally larger bets on larger trades and the coefficient β_0 from the regressions in panel B should be similar to those in panel A. However, if the magnitude of OHFT inventories is uncorrelated with the magnitude of the large trade, as would be expected if the OHFTs have no foreknowledge of the large trade, then the statistical significance of the coefficient β_0 should diminish significantly as the standard error increases. To this end in the regressions in panel B we estimate the same basic regression as in panel A except that we multiply $Sign_i$ by the size in contracts of the i^{th} large trade and $SignPT_{k,i}$ by the size of OHFT k's position at the time that the i^{th} large trade arrives. In these specifications, the coefficient β_0 remains positive and statistically significant in all specifications, and the t statistics for the hypothesis test that $\beta_0 \leq 0$ remain essentially unchanged relative to the corresponding t statistics from the regressions in panel A.

<Insert Table 3>

There is always a concern that our full sample results are generated by some abnormal period in the data. To determine if our results are driven by a certain time period within our sample, we divide our 66-day sample period into 33 two day sub-periods and estimate the full equation 3 for each of the two-day sub periods. In Figure 5 we plot the estimates of β_0 for each of the 33 sub-periods. From these regressions, we find that in none of the two-day sub-periods does the estimated value for β_0 become statistically negative, and in 25 of the 33 two-day sub-periods β_0 is statistically greater than zero. Further the 8 periods where the observed β_0 coefficients appear indistinguishable from zero are not clustered in any given time period. The first event occurs in the 1st two-day time period and the last occurs at the 30th two-day time period. These findings suggest that the phenomenon of OHFT inventories at the time of a large trade predicting the sign of large trades is fairly consistent throughout the sample.

<Insert Figure 5>

Our analysis up to this point finds that OHFT inventories at the time of a large trade positively predict the sign of the large trade with greater than random chance. This analysis, although suggestive, is not conclusive evidence that OHFTs can actually anticipate large trades. First, the analysis thus far has been conditional on the arrival of a large trade and thus does not test whether OHFTs can predict the arrival of large trades. Second, although this does not appear to be the case given the pattern of trading presented in Figure 3, these results could potentially be generated by an alternative scenario where both the large trader and the OHFT are acting on the same signal about price and thus their trading behavior will be correlated producing the observed results.

We first address the concern by studying whether the inventory of OHFTs can predict the sign of large trades in a setting that is unconditional on the arrival of a large trade. If OHFTs anticipate large trades and trade opportunistically with this information, then we should find that the inventories' of OHFTs at time t should be positively correlated with the sign of future large trades. To this end, we divide our sample into 2-minute segments and for each 2-minute segment we compute $y_{(t,t+1]}$ as the number of buyer-initiated large trades minus the number of seller-initiated large trades which arrive during time segment t. We calculate x_t as the sign of the aggregate inventory positions of OHFTs at time t, and $x_{(t-i,t-i+1]}$ is the sign of OHFTs net trades between time $t-i$ and $t-i+1$. We then use the predictive regression model presented in equation 4 to determine if OHFT inventories predict the arrival of future large trades.

$$y_{(t,t+1]} = \beta_0 + \beta_1 x_t + \sum_{i=1}^{5} \beta_{i+1} x_{(t-i,t-i+1]} + \varepsilon_{(t,t+1]} \tag{4}$$

In this regression, the key coefficient is β_1 which tests whether, after controlling for past trading, the aggregate inventory position of OHFTs at the beginning of time period t predicts the aggregate direction of large trades which will arrive between time period t and $t+1$. Since this is a time series regression we control for autocorrelation in the residuals by employing the Newey and West (1987) (1994) methodology with 30 lags. Our results for this specification are presented in panel A of Table 4. We find in these regressions that β_1 is positive and statistically significant implying that the sign of OHFT inventories at time t positively predicts the sign of aggregate large trading over the following 120 seconds. In panel B of Table 4 we use the actual size, in number of contracts, of the aggregate large trades which arrive during a given 2-minute interval along with the actual size of OHFT positions to perform the same analysis. Here we find the same results. The size and direction of OHFT positions at time t positively predicts the size and direction of aggregate large trading during the following 2-minutes. From this analysis we conclude that OHFTs trade in such a manner that their inventories consistently predict the size and direction of future large trades in such a way that allows the OHFT to profit from any upcoming price change engendered by the large trade.

<Insert Table 4>

We next apply ourselves to the question of whether the OHFTs are actively anticipating the large trade, or whether they are acting on – along with the large trader – some external price signal.[8] This concern essentially come down to an omitted variable problem. If the OHFTs are not trading on the large trade itself – but on some other price signal – then excluding the signal in our regressions will bias our analysis toward thinking that OHFTs are anticipating large trades because both the OHFTs' and large trader's trades are correlated with the omitted variable. If this is the case, then upon including the omitted variable as a control in our regressions we should observe the predictive power of OHFT inventories for future large trades disappear. In this case, the omitted variable is a signal about price changes. If the observed behavior is the byproduct of common momentum strategies, then by definition, prior price changes are the signal that momentum traders employ.

If large traders and OHFTs are trading on some external signal, and if this signal is correlated in an unbiased manner with the actual price changes, then including the contemporaneous and lagged returns in our regressions should ameliorate the concern that OHFTs' inventories predict large trades because they are both targeting a common signal about price.

[8] These could be external signals not yet incorporated into prices, or they could be prior price changes i.e. momentum.

By including contemporaneous price changes in our regression, the coefficient β_1 from equation 4 will only capture the variation in large trades explained by the inventory of OHFTs that is orthogonal to price changes. To this end we re-estimate equation 4, which tests whether OHFT inventories predict the net sign of upcoming large trades, with the added control variables of contemporaneous and lagged returns on the KOSPI 200 index and present our results in panels C and D of Table 4. These estimations reveal that after controlling for price changes, the OHFT inventories at time t continue to statistically significantly predict the net number and direction of large trades which occur over the following two minutes. This would seem to provide evidence that the OHFTs can anticipate large trades and that they trade opportunistically in response to this anticipation.

We explore the robustness of this result by repeating the analysis from Table 4 but with window lengths varied from 5 seconds to 240 seconds. In addition to providing a robustness check, varying the time window allows us to study the time horizon that OHFT inventories predict future large trades. If we observe that the coefficient for x_t is monotonically increasing in the window length up until a point at which the coefficient remains relatively stable, then this would indicate that OHFT positions predict large trade flow up to a given time and not afterward. The results from these regressions are presented in Table 5. We find that the coefficients from 5 seconds to 240 seconds are statistically significant and monotonically increasing with time. This pattern holds true both when we do and do not include lagged and contemporaneous returns. This finding suggests that the horizon at which OHFT inventories at time *t* predict the direction of upcoming large trades is at least 240 seconds. Put another way, the OHFTs can anticipate with greater than random chance large trades that will occur over the next 4 minutes.

<Insert Table 5>

The findings in this section are hard to reconcile with the alternative hypothesis that the OHFTs and the large traders are acting on a common signal, and are unlikely, given the predominately active nature of OHFT trading, to be the byproduct of market making strategies. They are, however, supportive of the hypothesis that OHFTs can anticipate individual large trades and that they trade opportunistically in response to this information. In the following section, we seek to shed light on the determinates of OHFT trade anticipation.

4.2 Determinates of OHFT Trade Anticipation

In this section, we turn our attention to analyzing the mechanisms which may allow OHFTs to anticipate large trades. We focus on two mechanisms described in the theoretical literature relating to order flow analysis and speed. We will first discuss the order flow hypothesis. Harris (1997) argues that investors face a tradeoff between trading and exposing intent. If too much intent is exposed then the way is opened for traders to trade opportunistically around their trade, however if not enough intent is exposed then traders will fail to execute a trade. Likewise, market makers in Kyle (1985) respond to aggregate order flow by making the market thin when they expect that informed traders will demand large amounts of liquidity. Yang and Zhu (2015) use a two period Kyle (1985) model with a 'back runner' who observes the period one order flow and places an order in period two based on this observation. From period one order flow the 'back runner' extracts a signal about the informed trader's intention to trade in period two, and trades accordingly. This model suggests that careful analysis of order flow may allow a trader to anticipate and opportunistically trade around another trader's trade. We refer to this as the order flow hypothesis. This hypothesis suggests that the more that large traders expose intent to trade prior to the large trade, the easier it will be for opportunistic traders to extract a signal about the upcoming large trade and thus the more likely their large trade is to be anticipated.

The second idea is that OHFTs use their speed and technological advantages to trade opportunistically around large trades. Since the Korean Market is not fragmented speed does not allow an OHFT to observe a trade on one market, and then race to another market in advance of the incoming trade as might be argued in a fragmented market. Speed in the context of our study may allow an OHFT to quickly react to changes in the market environment and either submit or cancel orders before other investors. Li (2014) proposes a theoretical model in which OHFTs are able to opportunistically trade around large trades due to their speed advantage. We refer to this hypothesis as the speed hypothesis, and it suggests that large traders who are slower at transacting are more likely to have their trades anticipated by OHFTs because it takes more time for them to respond to changes in the trading environment.

To test these hypotheses, we compute simple metrics for large trader intent exposure, and large trader speed. To estimate the amount of intent to trade that a large trader exposes we measure the average child order size prior to the arrival of the large trade for each of the large trades. The assumption implicit in this metric is that large traders who use larger than average child orders in the run-up to the large trade are exposing more intent to trade than traders who use smaller child orders. Thus, we expect that it will be easier for an OHFT to identify an upcoming large trade if that large trade is preceded by large child orders. To compute average child order size for each large trade we calculate the average trade size transacted by large trader l beginning three large trades prior to the arrival of the given large trade. Since large trades arrive on average every 40 seconds, this amounts to the average trade size for large trader l approximately 120 seconds before large trader l executes large trade i. This process produces an average child order size for each large trade in our sample. We then sort all large trades by their corresponding average child order size and identify those large trades with child order sizes in the largest 50% as having large child orders. We then test to see if large trades associated with larger than average child orders are more likely to be anticipated by OHFTs.

To measure large trader speed, we calculate for each of the 737 large traders in our sample the average difference in time between order submission and order fulfillment for all passive orders that a large trader executes in our sample. We do not use active orders, because we only observe the time that the trade arrived at the exchange, not the time that the order left the originator's computer. Consequently, for active orders the time between the order arrival at the exchange and the time stamp on the trade that fulfills the order is a function of the matching engine at the KRX and not determined by any speed advantage that a trader may enjoy in processing market data and then sending an order to the exchange.

By using passive orders our assumption is that large traders with lower latency can more quickly submit competitive quotes and cancel obsolete quotes than can higher latency large traders. Thus, we expect that the faster large traders will on average have quotes nearer the top of the order book than their slower counterparts. This assumption is supported theoretically by Ye (2017). Since passive orders are fulfilled in the order of their priority, the time between order submission and order fulfilment for passive orders higher in the order book should be less than for those orders that lie deeper in the book. Since we expect large traders who are faster at submitting and canceling orders to on average have orders higher in the book than their slower counterparts, we also expect faster traders to have a lower average time between submission and fulfillment for their passive orders. Consequently, we use the average time between order submission and execution across all passive trades for each large trader as our measure of large trader speed. We then sort large traders by average execution speed and those large traders in the slowest 50% of large traders are identified as being slow traders i.e. those whose trades are most likely to be anticipated.

We use regression analysis to determine if large trades associated with large child orders or slow traders are more likely to be anticipated by OHFTs. We test the two hypotheses that child order size and speed play a role in OHFTs ability to anticipate trades separately and jointly using the regression model presented in equation 5.

$$Sign_i = \beta_0 + \beta_1 SignPT_{k,i} + \beta_2 SignPT_{k,i} * Child_i + \beta_3 SignPT_{k,i} * Speed_i + \beta_4 SignPT_{k,i} * Child_i * Speed_i + \sum_{j=1}^{15} \gamma_j Sign_{i-j} + \alpha_k + \varepsilon_i \tag{5}$$

In equation 5 $Sign_i$ is the sign of the i^{th} large trade, and $SignPT_{k,i}$ is the sign of the k^{th} OHFT's inventory at the time of the i^{th} large trade. The variable $Child_i$ is an indicator variable identifying whether large trade i is associated with child orders that are larger than average. The variable $Speed_i$ is an indicator variable identifying whether large trade *i* is executed by a large trader that is slower than average. In this specification, the coefficients β_2 and β_3 test the hypothesis that child order size and speed respectively affect the likelihood that a given large trade will be correctly anticipated by OHFTs. A positive coefficient for β_2 would indicate that large trades associated with larger than average child orders are more likely to be anticipated by OHFTs. Likewise, a positive coefficient for β_3 would indicate that slower large traders are more likely to have their large trades anticipated by OHFTs. The coefficient β_4 tests the joint hypothesis that the size of child orders and the speed of the large trader interact to affect the probability that a large trade will be successfully anticipated by OHFTs. A positive coefficient for β_4 would indicate that child order size and speed interact to make large trades executed by slow large traders and associated with large child orders additionally likely to be anticipated by OHFTs. We present our results for these regressions in Table 6.

<Insert Table 6>

Our analysis provides evidence that both child order size and speed play a role in OHFTs anticipating large trades, but that child order size appears to play a larger role in explaining the ability for OHFTs to anticipate large trades. In our first specification, we exclude variables relating to speed and test independently the size hypothesis. We find that the coefficient for the interaction between child size and the sign of OHFT positions is positive and is equal to 0.11 which means that large trades associated with larger than average child orders are 5.5% more likely to be anticipated. It is also interesting to note that after controlling for child order size, the coefficient for the sign of OHFT inventory at the time of large trade is negative. This result suggests that child order size is essential in OHFT trade anticipation, and that after controlling for child order size, OHFTs' inventories predict large trades in the wrong direction.

Our results for the effect of speed also support the idea that OHFTs are more likely to anticipate large trades associated with slower large traders, but the effect is not as large as that of child order size. We find that OHFTs are 1.3% more likely to anticipate large trades executed by slow large traders. We do not however find that the coefficient for the sign of OHFT inventory at the time of the large trade is negative, suggesting that the speed of the large trader has a lesser impact than does the size of the child order.

Lastly in our third specification we estimate the full model presented in equation 5. Our results from this specification confirm what our prior analysis indicates, which is that child order size appears to play a larger role in determining OHFTs ability to anticipate a large trade. The coefficient for the interaction of speed and child order size is positive and significant, indicating that large trades associated with large child orders and slow traders are additionally likely to be anticipated by OHFTs. These findings provide evidence that both child order size and speed play a role in OHFTs' ability to anticipate large trades, but that child order size prior to the arrival of the large trade plays a larger role in allowing the OHFTs in our sample to anticipate large trades.

4.3 Estimating OHFT Profitability

In this section, we estimate how profitable it is to anticipate large trades and attempt to trade opportunistically around them. Baron, Brogaard, and Kirilenko (2012) study the profitability of high frequency traders in E-mini S&P 500 futures contracts and find that opportunistic high frequency traders tend to be the most profitable. We likewise estimate the profitability of OHFTs in our sample. We first estimate the profitability of opportunistic trading around large trades; we then estimate the profitability of OHFTs in general. The first analysis is facilitated by the event study presented in section 4.1. Recall that in Figure 3 we observe the trading dynamics of OHFTs both before and after a large trade for when OHFTs accumulate inventory in the same and opposite direction as a large trade. By combining this analysis of trading behavior from Figure 3 with that of average price movements around large trades from Figure 4 it is straightforward to estimate the average relative price paid by an OHFT to accumulate a position prior to a large trade and what OHFTs earn unwinding their positions after the large trade arrives.

We compute the profit or loss to opportunistic trading around large trades by taking the average relative price at trade position j from Figure 4, and multiplying by the average number of contracts purchased at trade position j from Figure 3. This procedure allows us to compute the weighted average relative price that OHFTs pay to build a position prior to a large trade, and then likewise the weighted average price that OHFTs receive to unwind their positions after the large trade arrives. Since we compute OHFT trading and price dynamics for both when OHFTs build inventory in the same and opposite direction as the large trade we can estimate the average profit and loss due to correctly or incorrectly signing a given large buy or sell trade.

Using this methodology, we estimate that the profit for successfully signing and trading opportunistically around a large buy trade is about 2.05 bps, and 2.08 bps for a large sell trade. Losses to OHFTs when they incorrectly sign a large buy trade amount to approximately 1.69 bps, and are approximately 1.44 bps for a large sell trade. In our sample, OHFTs successfully sign 54.77% of large buy trades and 57.37% of large sell trades.

Although OHFTs appear to successfully sign and trade opportunistically around more than half of large trades, they also lose money on their attempts to trade opportunistically for approximately 44% of large trades which arrive. However, given the consistency of the predictive power of OHFTs to predict large trades with greater than random chance and the asymmetry of profits and losses, an application of the Kelly criterion (Kelly, 1956) may allow OHFTs to earn positive consistent profits suggesting that trade anticipation may be a profitable strategy for OHFTs to pursue.

We next investigate the consistency of OHFT profits by dividing our sample into one hour segments and calculating the aggregate mark-to-market profits for OHFTs for each hour separately. In computing these profits, we aggregate all OHFT positions together into one series. At the beginning of each one-hour segment we impose that OHFT inventories are equal to zero; we then track the profits and losses on all trades throughout the one-hour time period. At the end of the one-hour time period we liquidate any remaining inventory at the current market price to compute the mark-to-market profit or loss for the given one-hour segment. This process yields 396 observations (66 days multiplied by 6 one hour segments). We normalize mark-to-market profits by dividing by their standard deviation and present a histogram of normalized hourly mark-to-market OHFT profits in Figure 6. We observe in Figure 6 that hourly mark-to-market profits are positive 80.85% of the time, and that the distribution of hourly mark-to-market profits exhibits positive skewness. Taken as a whole, the findings in this section suggest that pursuing trade anticipation strategies may be profitable. They also are consistent with the findings of Baron, Brogaard, and Kirilenko (2012) that opportunistic high frequency traders tend to be consistently profitable.

<Insert Figure 6>

5 Conclusion

In this study, we investigate whether individual large trades are anticipated by a subset of high frequency traders which we refer to as opportunistic high frequency traders. We find evidence consistent with the hypothesis that this subset of high frequency traders can anticipate individual large trades. Using an event study methodology, we find that OHFTs begin actively building positions up to 200 trades in advance of the large trade and that they immediately begin to unwind their positions after the arrival of the large trade. The direction of the OHFT inventory accumulation is in the same direction as the incoming large trade statistically more than 50% of the time. We find that the profits that OHFTs earn when their inventories are on the same side as the large trade are greater than the losses incurred when on the opposite side to the large trade. We also document that large trades associated with larger than average child orders or slower large traders are additionally likely to be anticipated. These findings suggest that order anticipation strategies may be profitable for OHFTs to pursue.

The trading behavior documented herein is hard to reconcile with the hypothesis that OHFTs are targeting common price signals, or that the results are the byproduct of market making strategies. OHFTs in general, and in the vicinity of large trades tend to take liquidity more than twice as often as they provide liquidity casting doubt on the hypothesis that the trading behavior observed is part of market making strategies employed by the OHFTs. Also, if OHFTs and large traders are transacting on the same signal, then to generate the trading patterns we observe, it would need to be the case that the OHFTs and the large trader executing the large trade (but not other large traders) consistently receive simultaneous signals about price that are strong enough to generate significant amounts of trading. However, these signals would be in the same direction only 56% of the time. The other 44% of the time the OHFTs would need to receive a signal that is correlated with the arrival of the incoming large trade, but opposite to the information that the large trade is acting on. Further, if OHFTs and large traders do receive simultaneous signals that are in the opposite direction 44% of the time, it is difficult to understand why the OHFTs give up on their accumulated position, and hence their information, immediately following the arrival of a large trade. Also, the price signal that triggers the inventory buildup and wind down for the OHFT and the large trade would have to be uncorrelated with the returns on the underlying asset leading up to and contemporaneous with the large trade, as we find that after controlling for contemporaneous and lagged returns OHFT inventories continue to predict the arrival of large trades.

These findings documented in this study contribute to our understanding of financial markets in several ways. First, they shed light on the complex behavior of high frequency traders in financial markets. Second, they would appear to validate the concerns of large traders that there is information leakage on lit exchanges which may make their large trades predictable. Lastly, to the extent that information leakage drives large traders from lit exchanges into off exchange venues such as dark pools, our results provide an additional explanation for the recent rise in off exchange trading volume.

Works Cited

Ahn, H., Kang, J., & Ryu, D. (2008). Informed Trading in the Index Option Market: The Case of KOSPI 200 Options. *Journal of Futures Markets*, 1118-1146.

Ahn, H., Kang, J., & Ryu, D. (2010). Information Effects of Trade Size and Trade Direction: Evidence from the KOSPI 200 Index Options Market. *Asia-Pacific Journal of Financial Studies*, 301-339.

Attari, M., Mello, A., & Ruckes, M. (2005). Arbitraging Arbitrageurs. *Journal of Finance*, 2471-2511.

Baron, M., Brogaard, J., & Kirilenko, A. (2012). The trading profits of high frequency traders. *Working Paper*.

Bernhardt, D., & Taub, B. (2008). Front-running dynamics. *Journal of Economic Theory*, 288-296.

Biais, B., & Wooley, P. (2011). High frequency trading. *Working Paper*.

Boehmer, E., & Wu, J. (2013). Short selling and the price discovery process. *Review of Financial Studies*, 287-322.

Brogaard, J. (2010). High frequency trading and its impact on market quality. *Working Paper*.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High frequency trading and price discovery. *Review of Financial Studies*, 2267-2306.

Brogaard, J., Hendershott, T., & Riordan, R. (2017). High Frequency Trading and the 2008 Short Sale Ban. *Journal of Financial Economics*, 22-42.

Brunnermeier, M. (2005). Information leakage and market efficiency. *Review of Financial Studies*, 417-457.

Brunnermeier, M., & Pedersen, L. (2005). Predatory Trading. *Journal of Finance*, 1825-1863.

Buti, S., Rindi, B., & Werner, I. (2017). Dark pool trading strategies, market quality and welfare. *Journal of Financial Economics*, 244-265.

Carrion, A. (2013). Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets*, 680-711.

Clark-Joseph, A. (2014). Exploratory Trading. *Working Paper*.

Comerton-Forde, C., & Putniņš, T. (2015). Dark trading and price discovery. *Journal of Financial Economics*, 70-92.

Degryse, H., Van Achter, M., & Wuyts, G. (2009). Dynamic order submission strategies with competition between a dealer market and a crossing network. *Journal of Financial Economics*, 319-338.

Egginton, J., Van Ness, B., & Van Ness, R. (2011). Quote Stuffing. *Working paper*.

Grossman, S., & Stiglitz, J. (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review*, 393-408.

Hagstromer, B., & Norden, L. (2013). The diversity of high-frequency traders. *Journal of Financial Markets*, 741-770.

Hagstromer, B., & Norden, L. (2013). The diversity of high-frequency traders. *Journal of Financial Markets*.

Harris, L. (1997). Order exposure and parasitic traders. *Working Paper*.

Hirschey, N. (2013). Do high-frequency traders anticipate buying and selling pressure? *Working Paper*.

Jain, P., & Joh, G. (1988). The dependence between hourly prices and trading volume. *Journal of FInancial and Quantitative Analysis*, 269-283.

Jarrow, R., & Protter, P. (2012). A dysfunctional role of high frequency trading in electronic markets. *International Journal of Theoretical and Applied Finance*, 2-15.

Kelly, J. (1956). A new interpretation of information rate. *Information Theory, IRE Transactions on 2.3*, 185-189.

Korajczyk, R., & Murphy, D. (2016). High-Frequency Market Making to Large Institutional Trades. *Working Paper*.

Kwan, A., Masulis, R., & McInish, T. (2015). Trading rules, competition for order flow and market fragmentation. *Journal of Financial Economics*, 330-348.

Kyle, A. (1985). Continous Auctions and Insider Trading. *Econometrica*, 1315-1335.

Lee, C., & Ready, M. (1991). Inferring trade direction from intraday data. *Journal of Finance*, 733-746.

Li, W. (2014). High frequency trading with speed hierarchies. *Working Paper*.

McInish, T., & Wood, R. (1990). An analysis of transactions data for the Toronto stock exchange. *Journal of Banking and Finance*, 441-458.

Menkveld, A. (2013). High frequency trading and the new market makers. *Journal of Financial Markets*, 712-740.

Menkveld, A. (2013). High frequency trading and the new market makers. *Journal of Financial Markets*.

Menkveld, A., Yueshen, B., & Zhu, H. (2017). Shades of darkness: A pecking order of trading venues. *Journal of Financial Economics*, 503-534.

Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 777-787.

Newey, W., & West, K. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 631-653.

O'Hara, M. (2011). What is a quote? *Journal of Trading*, 10-16.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 257-270.

Schöneborn, T., & Schied, A. (2009). Liquidation in the Face of Adversity: Stealth Versus Sunshine Trading. *Working Paper*.

Tong, L. (2015). A blessing or a curse? The impact of high frequency trading on instituional investors. *Working Paper* .

van Kervel, V., & Menkveld, A. (2016). High-Frequency Trading around Large Institutional Orders. *Working Paper*.

Yang, L., & Zhu, H. (2015). Back-running: seeking and hiding fundamental information in order flows. *Working Paper*.

Ye, M. (2012). Price Manipulation, Price Discovery and Transaction Costs in the Crossing Network. *Working Paper*.

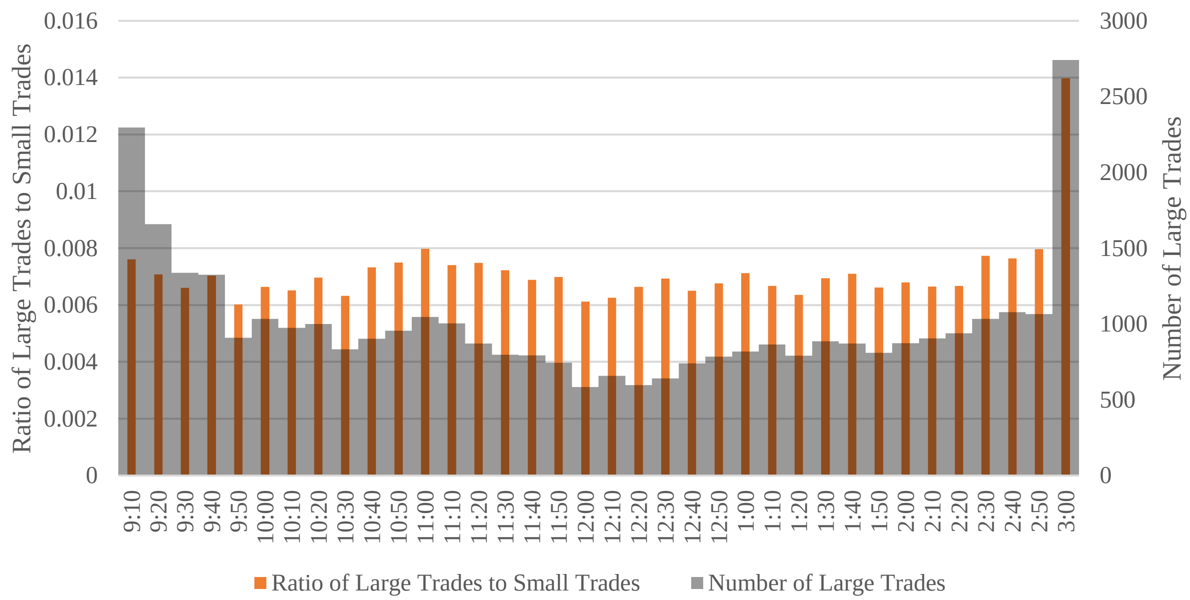
Ye, M. (2017). Who Provides Liquidity, and When: An Analysis of Price vs Speed Competition on Liquidity and Welfare. *Working Paper*.

Ye, M., Yao, C., & Jiading, G. (2013). The externalities of high frequency trading. *Working Paper*.

Zhang, X. (2010). High-frequency trading, stock volatility, and price discovery. *Working Paper*.

Zhu, H. (2014). Do Dark Pools Harm Price Discovery? *Review of Financial Studies*, 747-789.

Panel A: Distribution of Large Trades Throughout the Trading Day



Panel B: Partial Autocorrelation Function of Large Trades

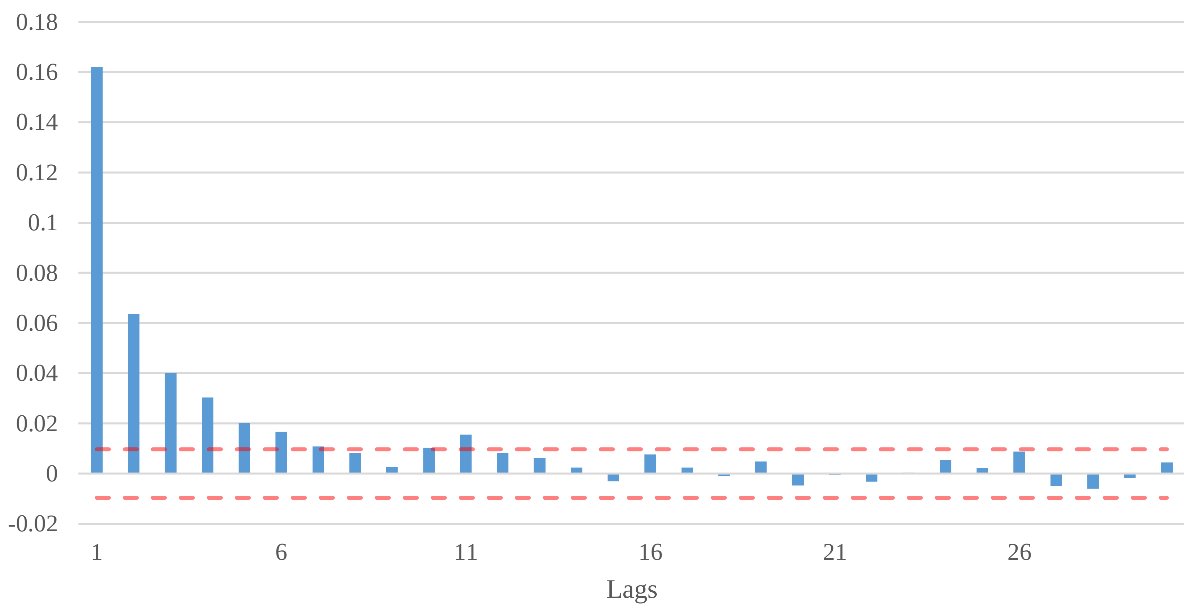
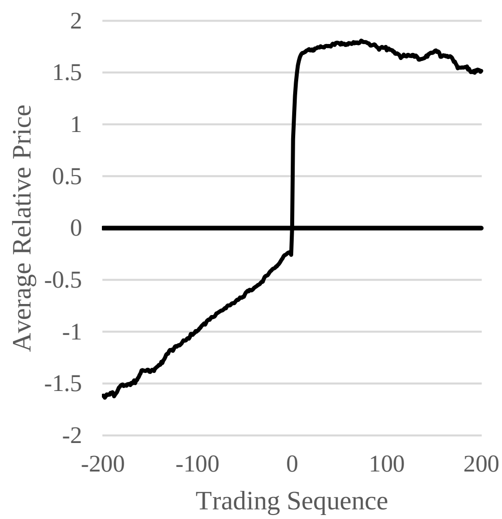


Figure 1 Characteristics of Large Trades: In Panel A we show the distribution of large trades throughout the trading day broken up by 10-minute increments beginning at 9:00 a.m. and ending at 3:00 p.m. Large trades are defined as those trades in the top 1% of all trades by trade size. We plot the ratio of large trades to small trades across all trading days for each 10-minute increment on the left hand axis, and the total number of large trades which occur in each 10-minute increment across all trading days on the right hand axis. In Panel B we present the partial autocorrelation function for large trades for 30 lags. Values outside of the dotted lines are significant at the 5% level.

Panel A: Relative Price Changes Around a Large Buy



Panel B: Relative Price Changes Around a Large Sell

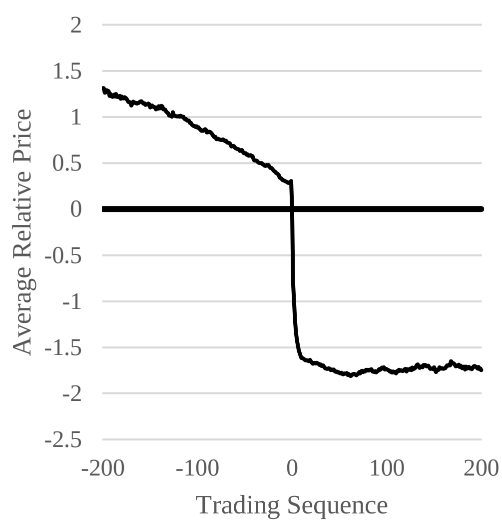


Figure 2: Relative Price Changes Around Large Trades. In this figure, we present the average relative price changes around large trades. For each of the 36,164 large trades we begin measuring the relative price 200 trades before the arrival of the large trade. For each trade position $j \epsilon [-200,200]$ and each trade $i \epsilon [1,\ 36{,}164]$ the relative price is computed as $p'(i,j) = \ln\left(\frac{p(i,j)}{p(i,0)}\right) * 10\text{^}4$ where $p(i,j)$ indicates the price at trade position j around large trade i and $p(i,0)$ indicates the price of the i^{th} large trade. For each trading position $j \epsilon [-200,200]$ we compute the average relative price across all trades as $\overline{p'(j)} = \frac{1}{N}\sum_{i=1}^{N} p'(i,j)$, where N is the total number of large buys or sells observed in the data. This process is performed for buyer and seller-initiated trades separately. In Panel A we present the average relative price around the arrival of large buyer-initiated trades and in Panel B we present the average relative price around the arrival of large seller- initiated trades.

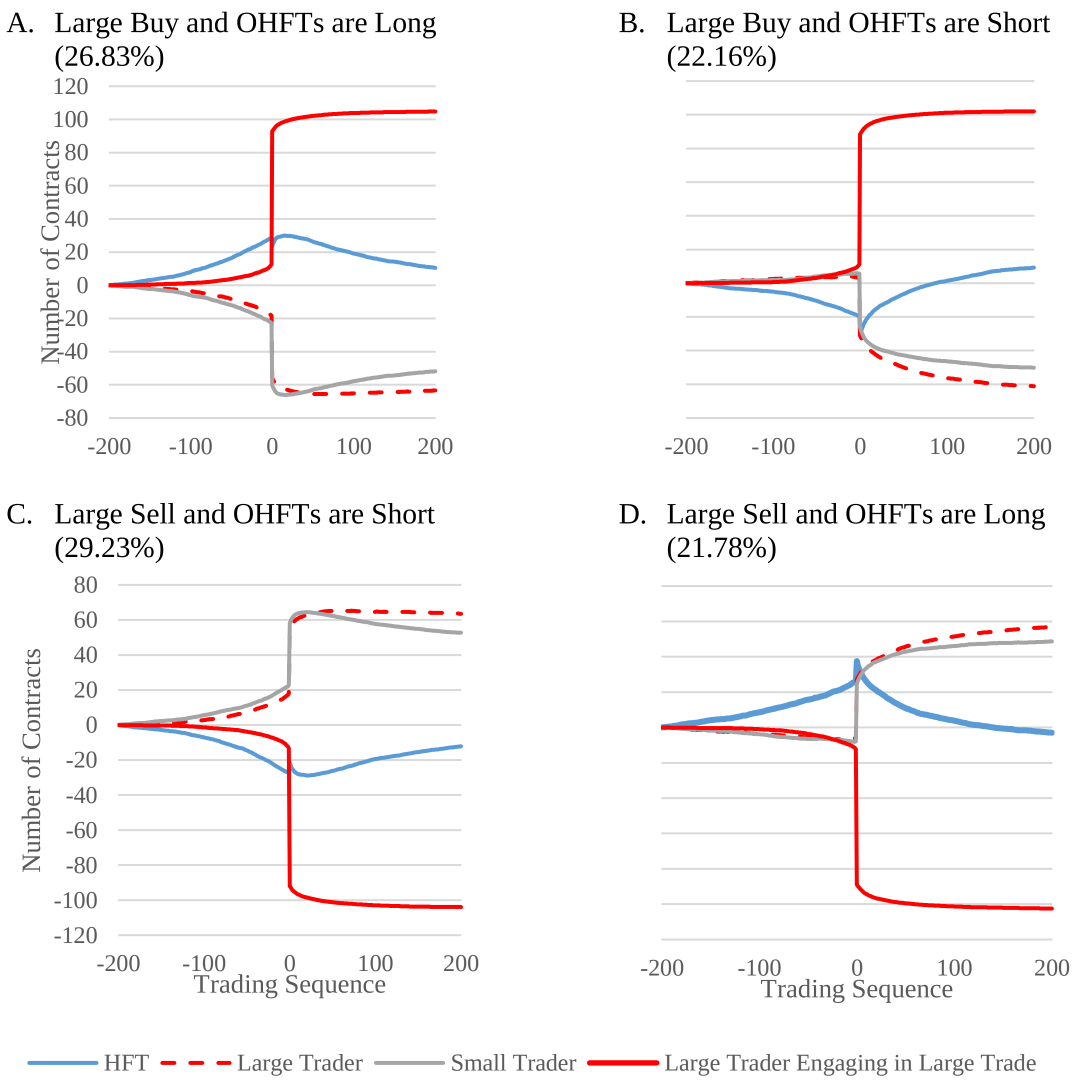


Figure 3 Trading Dynamics Around Large Trades. In this figure, we plot trading dynamics for 200 trades before and after the arrival of large trades. Traders are classified as either OHFTs, small traders, large traders engaging in a given large trade, or large traders not engaging in a given large trade. All positions begin at zero at trade position $j = -200$, and for each trade position $j \in [-200,200]$ and each trade $i \in [1,\ 36{,}164]$ we compute the aggregate position of the four trader types relative to their position at $j = -200$. For each relative trading position around the large trade $j \in [-200,200]$ we compute the average position in number of contracts of each of the four trader types across all large buyer and seller-initiated large trades separately. In Panel A we present the trading dynamics around the arrival of large buy trades when OHFTs are on the same side as the large trade. In Panel B we present the trading dynamics around the arrival of large buy trades when OHFTs are on the opposite side as the large trade. In Panel C we present the trading dynamics around the arrival of large sell trades when OHFTs are on the same side as the large trade. In Panel D we present the trading dynamics around the arrival of large sell trades when OHFTs are on the opposite side as the large trade.

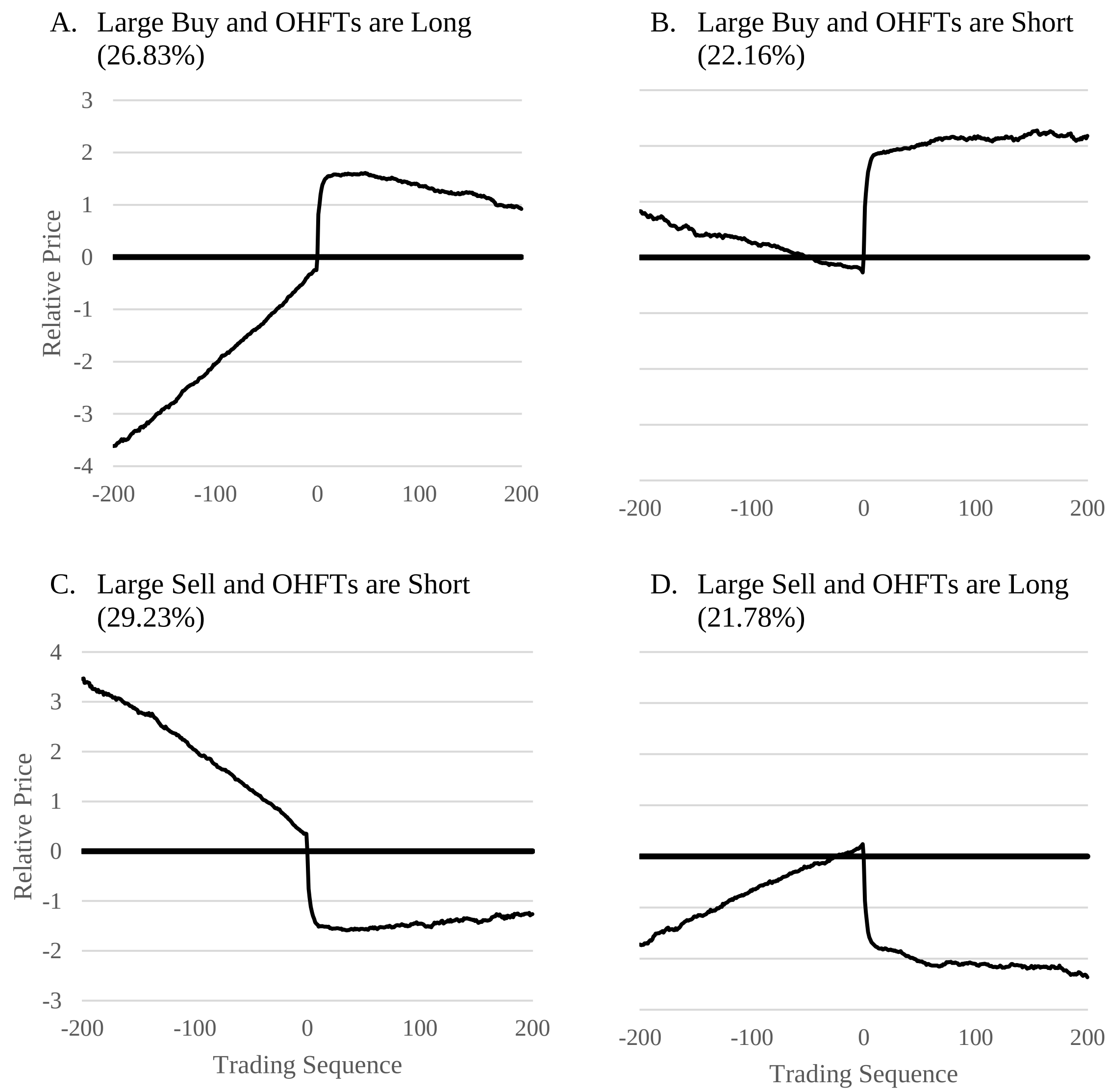


Figure 4: Relative Price Changes Around Large Trades. In this figure, we plot average relative prices around the arrival of large buy and sell trades both for when the aggregate positions of OHFTs are on the same side as the large trade and when the aggregate positions of OHFTs are on the opposite side as the large trade. In Panels A and B we present the average relative price around the arrival of large buy trades when OHFTs are on the same and opposite side as the large trade respectively. In Panels C and D we present the average relative price around the arrival of large sell trades when OHFTs are on the same and opposite side as the large trade respectively. For each trade position $j \epsilon [-200{,}200]$ and each trade $i \epsilon [1,\ 36{,}164]$ the relative price is computed as $p'(i,j) = \ln\left(\frac{p(i,j)}{p(i,0)}\right) * 10\text{^}4$ where $p(i,j)$ indicates the price at trade position j around large trade i and $p(i,0)$ indicates the price of the i^{th} large trade. For each trading position $j \epsilon [-200{,}200]$ we compute the average relative price across all trades as $\overline{p'(j)} = \frac{1}{N}\sum_{i=1}^{N} p'(i,j)$ where N is the number of large trades meeting a given criterion.

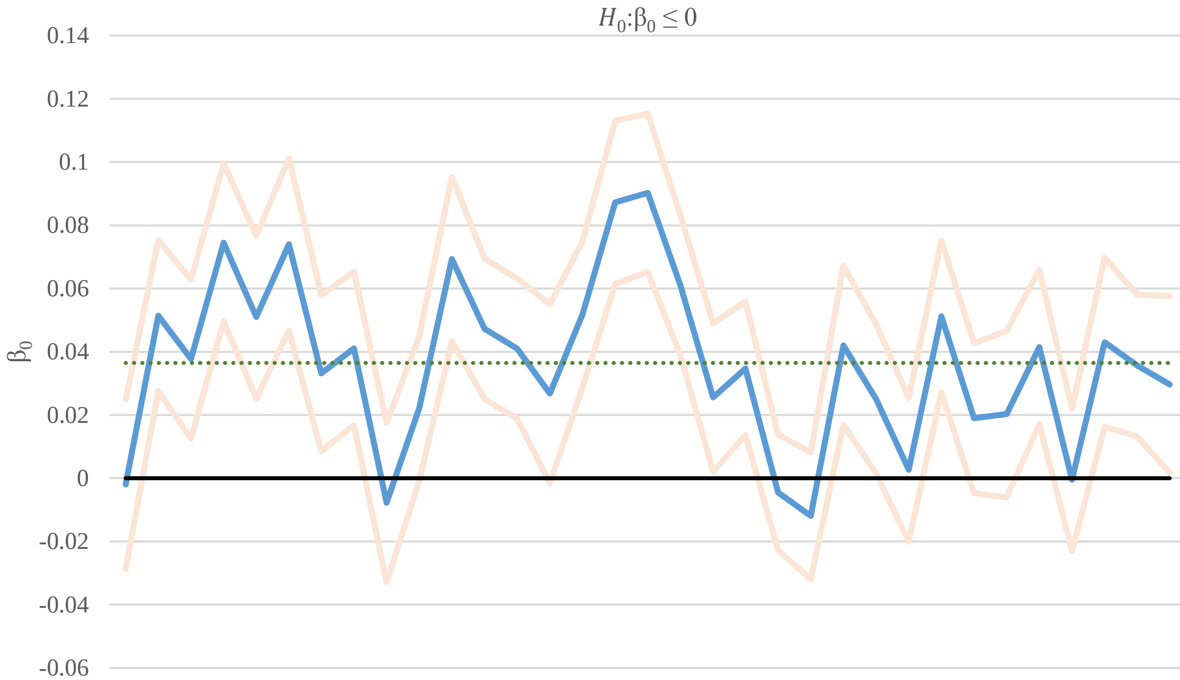


Figure 5 Time Series of β₀. This figure plots the time series of β_0 throughout the sample period. β_0 is estimated from the regression $Sign_i = \beta_0 SignPT_{k,i} + \sum_{j=1}^{15} \gamma_j Sign_{i-j} + \alpha_k + \varepsilon_i$ and indicates whether the 32 OHFTs in our sample build up positions on the same side as a large trade with a probability greater than random chance. $Sign_i$ is the sign of the i^{th}large trade and is equal to -1 for large seller-initiated trades and 1 for large buyer-initiated trades. We identify 32 accounts that trade in a manner consistent with the current literature associated with OHFTs and the variable $SignPT_{k,i}$ is the sign of the inventory accumulated by the k^{th} OHFT at the arrival of large trade i. Because large trades exhibit autocorrelation we include in our regression the sign of the prior 15 large trades. We also include fixed effects for each of the 32 OHFTs in our sample. We divide our 66-day sample into 33 two-day time periods and estimate the above regression for each two-day sub-period separately and plot the estimated value of β_0 for each of the 33 sub-periods. The dotted line represents the mean value of β_0 across the sample while the solid blue line represents the evolution of β_0. We also include 95% confidence intervals for β_0.

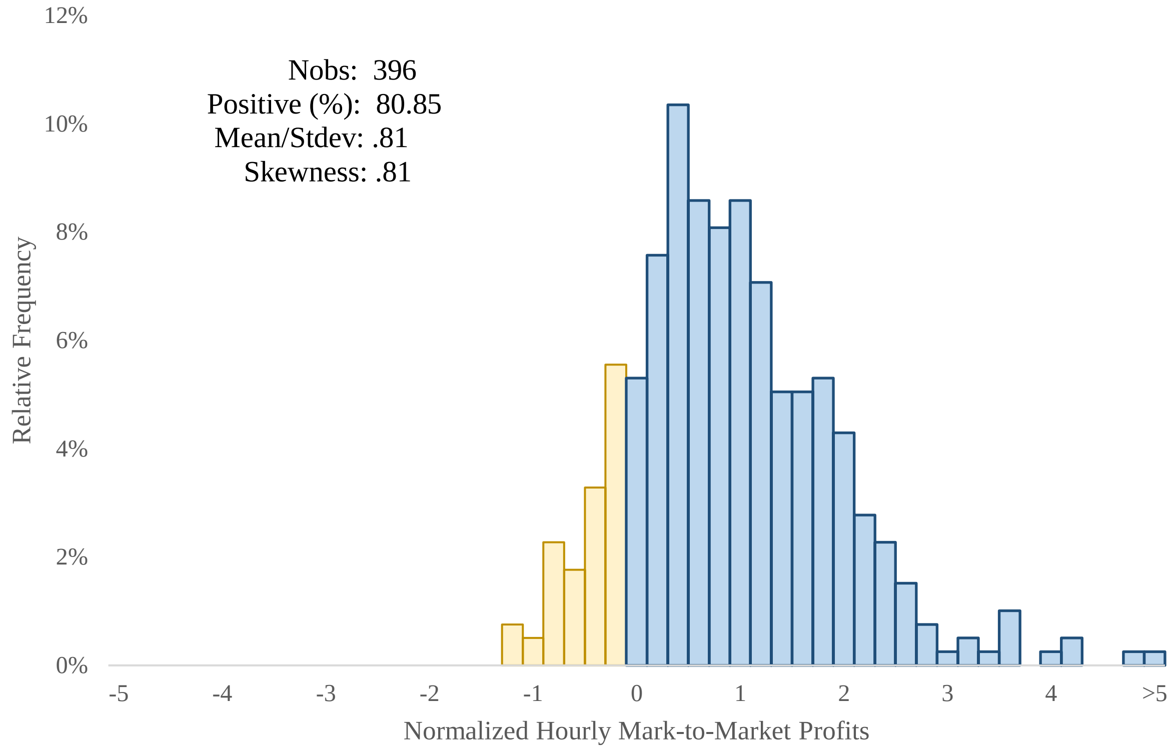


Figure 6: Normalized Hourly Mark-to-Market Profits. In this figure, we present the histogram of standardized aggregate hourly mark-to-market profits of the OHFTs in our sample. We divide our 66-day sample into 396 one hour segments representing the six trading hours of each of the 66 days in our sample. We impose that all OHFT positions are zero at the beginning of each 1-hour period. We then compute the profits and losses for OHFTs on all subsequent trades during that 1-hour period. At the end of the 1-hour period we liquidate any remaining OHFT positions at the current market price and we calculate the aggregate profits and losses for OHFTs during each 1-hour segment of our data. We normalize hourly profits by their standard deviation and plot a histogram of the relative frequencies of the 396 hourly mark-to-market profits in this figure.

Table 1: Characteristics of Large Trades

This table reports summary statistics for large trades. Large trades are defined as active trades by large traders which are among the largest 1% of all large trades. We identify 36,164 large trades in our sample. Large trades are defined as buyer-initiated if the large trader crossed the bid-ask spread and bought at the ask price or seller-initiated if the large trader crossed the bid-ask spread and sold at the bid price. In our sample 48.99% of large trades are buyer-initiated and 51.01% of large trades are seller-initiated trades.

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

Table 2: Trading Characteristics of Account Types

This table presents descriptive statistics for the trader types in our data. Accounts are determined to be OHFTs if 1) the number of opportunistic trades exceeds the number of passive trades 2) average holding time for one position is less than 3 minutes, and 3) average daily ratio of overnight inventories to contracts traded is less than 0.01%. Large traders are defined as accounts that are not OHFTs which execute at least one active trade that is among the top 1% of trades by size. Small traders are all other traders. In Panel A we provide statistics about the classifications of traders: foreign vs domestic and institutional vs retail. In Panel B we provide volume statistics for each of the three trader types. *Take* and *Make* indicates active and passive trading, and in Panel C we provide other statistics. These include *# of Switch/Day* which is defined as the number of times that a traders' inventory switches from positive to negative or from negative to positive, *Switch Time (sec)* which is measured as the time in seconds between when aggregate inventories switch in their direction, and *Overnight Ratio (%)* which is the ratio of inventory held overnight to daily trading volume.

Panel A: Trader Entity

	Total	Foreign	Institutional
OHFTs	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

	Total Volume		Large Trades		Small Trades	
	Take	Make	Take	Make	Take	Make
OHFTs	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			
OHFTs	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)			
OHFTs	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 (%)			
OHFTs	0.00	0.00	0.02
Large Trader	40.33	29.80	36.47
Small Trader	21.40	4.44	32.86

Table 3: OLS Regression Conditional of Large Trades

In Panel A of this table we use three variations of the model

$$Sign_i = \beta_0 SignPT_{k,i} + \sum_{j=1}^{15} \gamma_j Sign_{i-j} + \alpha_k + \varepsilon_i$$

to test the hypothesis that OHFTs are on the right side of large trades with greater than 50% probability. $Sign_i$ is the sign of the i^{th} large trade and is equal to -1 for large sells and 1 for large buys. $SignPT_{k,i}$ as the sign of the k^{th} OHFT's inventory at the time of the i^{th}large trade and is either 1 or -1 indicating whether the k^{th} OHFT was long or short at the time of the i^{th}large trade respectively, and α_k is a dummy variable indicating the k^{th} OHFT. In model (1) we do not include OHFT fixed effects or the values for the lagged sign of the large trades, in model (2) we include OHFT fixed effects, but not the lagged signs of the prior large trades, in model (3) we estimate the full model. In Panel B we present the results from a related regression presented below.

$$Size_i * Sign_i = \beta_0 Size_{k,i} * SignPT_{k,i} + \sum_{j=1}^{15} \gamma_j Size_{i-j} * Sign_{i-j} + \alpha_k + \varepsilon_i$$

The only modification from the model used in Panel A is that we multiply $Sign_i$ by the size in contracts of the i^{th} large trade and $SignPT_{k,i}$ by the size of OHFT k's position at the time that the i^{th} large trade arrives. Like Panel A, in model (4) we do not include OHFT fixed effects or the values for the lagged sign of the large trades, in model (5) we include OHFT fixed effects, but not the lagged signs of the prior large trades, in model (6) we estimate the full model. We do not include stars because all estimates are significant at the 1% level.

Model		Estimate	S.E.	t value	p value	Adj.R^2
		Panel A: Direction of Trade and Position				
(1) Simple OLS	$SignPT_{k,i}$	0.063	0.002	29.493	0.000	0.004
(2) Fixed Effect	$SignPT_{k,i}$	0.064	0.002	29.617	0.000	0.004
(3) Fixed Effect + Lag $Sign_i$	$SignPT_{k,i}$	0.036	0.002	16.766	0.000	0.041
		Panel B: Direction*Size of Trade and Position				
(4) Simple OLS	$Size_{k,i} * SignPT_{k,i}$	0.085	0.003	29.112	0.000	0.004
(5) Fixed Effect	$Size_{k,i} * SignPT_{k,i}$	0.085	0.003	29.092	0.000	0.004
(6) Fixed Effect +Lag $Size_i * Sign_i$	$Size_{k,i} * SignPT_{k,i}$	0.050	0.003	17.150	0.000	0.040

Table 4: Predictive Regression Unconditional of Large Trades

In this table, we test the unconditional hypothesis that OHFT inventories predict the aggregate direction of large trades with greater than random chance. For this analysis, we divide our sample into 120 second segments and for each 120 second segment we compute $y_{(t,t+1]}$ as the number of buyer-initiated large trades minus the number of seller-initiated large trades which arrive during time segment t to $t+1$. We also calculate x_t as the sign of the aggregate inventory position of OHFTs at time t, and $x_{(t-i,t-i+1]}$ is the sign of OHFT's net trades between time segment $t-i$ and $t-i+1$. In panel A we use the regression model

$$y_{(t,t+1]} = \beta_0 + \beta_1 x_t + \sum_{i=1}^{5} \beta_{i+1} x_{(t-i,t-i+1]} + \varepsilon_{(t,t+1]}$$

to determine if OHFT inventories at time *t* predict the net direction of large trades which arrive between time t and t+1. In Panel B we perform the same analysis as in Panel A except that we use the net size of large trades multiplied by the sign of aggregate large trades, and the aggregate size of OHFT positions at time *t* multiplied by the sign of OHFT positions. Panel C repeats the analysis of panel A but includes lagged and contemporaneous two minute returns on the KOSPI 200 index as control variables. Panel D repeats the analysis of panel B but includes lagged and contemporaneous two minute returns on the KOSPI 200 index as control variables. Autocorrelation in the residuals is controlled by employing the Newey and West (1994) (1987) methodology with 30 lags. One, two, and three stars represent significance at the 10, 5, and 1% levels respectively. *t* statistics are presented in parentheses.

	Panel A: Direction			Panel B: Direction * Size		
	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	-0.06**	-0.06**	-0.06**	-2.94	-3.15	-2.94
	(2.57)	(2.44)	(2.54)	(1.34)	(1.4)	(1.34)
x_t	0.17***		0.16***	0.21***		0.22***
	(7.49)		(6.04)	(7.5)		(3.7)
$x_{(t-1,t]}$		0.09***	0.01		0.16***	-0.02
		(3.9)	(0.34)		(6.42)	(0.28)
$x_{(t-2,t-1]}$		0.09***	0.05*		0.17***	0.03
		(3.3)	(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)	(0.6)
$x_{(t-4,t-3]}$		0.03	0.02		0.06*	-0.02
		(1.18)	(0.82)		(1.88)	(0.6)
$x_{(t-5,t-4]}$		-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

	Panel C: Direction			Panel D: Direction*Size		
	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	-0.06***	-0.06***	-0.058***	-2.646	-2.961	-2.555
	(-3.09)	(-2.69)	(-3.08)	(-1.6)	(-1.52)	(-1.6)
x_t	0.177***		0.112***	0.227***		0.159***
	(9.31)		(4.59)	(10.04)		(3.32)
$x_{(t-1,t]}$		0.037	-0.023		0.066**	-0.051
		(1.52)	(-0.97)		(2.5)	(-1.09)
$x_{(t-2,t-1]}$		0.05*	0.003		0.08**	-0.008
		(1.9)	(0.14)		(2.49)	(-0.2)
$x_{(t-3,t-2]}$		0.025	0.001		0.069**	0.023
		(0.99)	(0.03)		(2.11)	(0.64)
$x_{(t-4,t-3]}$		0.021	0.016		0.007	-0.005
		(0.83)	(0.7)		(0.25)	(-0.17)
$x_{(t-5,t-4]}$		-0.027	-0.022		-0.02	-0.026
		(-1.21)	(-1.11)		(-0.85)	(-1.23)
r_t	0.122***		0.121***	11.279***		11.226***
	(19.3)		(18.72)	(22.77)		(22.29)
r_{t-1}		0.032***	0.029***		2.958***	2.703***
		(8.72)	(6.86)		(11.87)	(9.95)
r_{t-2}		0.01***	0.009***		0.921***	0.833***
		(4.19)	(3.41)		(4.57)	(3.82)
r_{t-3}		0.002	0.002		0.304	0.249
		(1.02)	(1.15)		(1.46)	(1.48)
r_{t-4}		-0.001	0		0.238	0.243
		(-0.34)	(-0.09)		(1.29)	(1.47)
r_{t-5}		0.004*	0.002		0.51***	0.365**
		(1.83)	(1.34)		(2.9)	(2.22)
Nobs	11,674	11,674	11,674	11,674	11,674	11,674
Adj.R^2	0.2864	0.0230	0.3028	0.3188	0.0277	0.3381

Table 5: Predictive Regression with Varying Time Interval

In this table, we test the unconditional hypothesis that OHFT inventories predict the aggregate direction of large trades with greater than random chance. For this analysis, we divide our sample into segments of varying time length and for each segment we compute $y_{(t,t+1]}$ as number of buyer-initiated large trades minus the number of seller-initiated large trades which arrive during time segment t to $t+1$. We also calculate x_t as the sign of the aggregate inventory position of OHFTs at time t, and $x_{(t-i,t-i+1]}$ is the sign of OHFTs' net trades between time $t-i$ and $t-i+1$. W then use the regression model presented below to test the hypothesis that OHFT inventories at the beginning of time *t* predict the net direction of large trades over the following period.

$$y_{(t,t+1]} = \beta_0 + \beta_1 x_t + \sum_{i=1}^{5} \beta_{i+1} x_{(t-i,t-i+1]} + \varepsilon_{(t,t+1]}$$

The results from these regressions are presented in panel A. We vary the time segments from 5 seconds to 240 seconds in these regressions in columns one through six respectively. In Panel B we perform the same analysis as in Panel A except that we use the net size of large trades multiplied by the sign of aggregate large trades, and the aggregate size of OHFT positions at time *t* multiplied by the sign of OHFT positions. Panel C repeats the analysis of panel A but includes lagged and contemporaneous two minute returns on the KOSPI 200 index as control variables. Panel D repeats the analysis of panel D but includes lagged and contemporaneous two minute returns on the KOSPI 200 index as control variables. Autocorrelation in the residuals is controlled by employing the Newey and West (1994) methodology with 30 lags. One, two, and three stars represent significance at the 10, 5, and 1% levels respectively. *t* statistics are presented in parentheses.

Panel A: Regressions Predicting Direction of Large Trades

	5sec	20sec	40sec	60sec	120sec	240sec
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.00***	-0.01**	-0.02**	-0.03**	-0.06**	-0.13***
	(-2.68)	(-2.45)	(-2.47)	(-2.49)	(-2.54)	(-2.61)
x_t	0.01***	0.04***	0.08***	0.10***	0.16***	0.26***
	(12.89)	(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
	(0.13)	(0.47)	(0.04)	(-1.9)	(0.34)	(0.04)
$x_{(t-2,t-1]}$	0.00**	0.01*	0.00	0.01	0.05*	-0.03
	(2.42)	(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
	(1.91)	(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
	(0.78)	(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
	(3.73)	(1.8)	(1.95)	(1.13)	(-0.97)	(0.53)
Nobs	283,093	71,819	35,749	23,683	11,674	5,672
Adj.R^2	0.0009	0.0026	0.0036	0.0034	0.0050	0.0045

Panel B: Regressions Predicting Direction*Size of Large Trades						
	5sec	20sec	40sec	60sec	120sec	240sec
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.14	−0.36	−0.78	−1.28	−2.94	−6.72
	(-0.74)	(−1.07)	(−1.11)	(−1.20)	(−1.34)	(−1.52)
x_t	0.03***	0.08***	0.14***	0.16***	0.22***	0.46***
	(9.43)	(11.39)	(8.96)	(6.47)	(3.7)	(3.68)
$x_{(t-1,t]}$	−0.02***	−0.04***	−0.06***	−0.06**	−0.02	−0.16
	(−4.59)	(−5.51)	(−4.13)	(−2.37)	(−0.28)	(−1.40)
$x_{(t-2,t-1]}$	−0.01***	−0.03***	−0.04***	−0.02	0.03	−0.16
	(−3.25)	(−3.84)	(−2.63)	(−0.86)	(0.51)	(−1.47)
$x_{(t-3,t-2]}$	−0.01**	−0.01	−0.03**	0.01	0.03	−0.11
	(−2.06)	(−1.34)	(−2.32)	(0.57)	(0.6)	(−1.22)
$x_{(t-4,t-3]}$	−0.01	−0.01*	0.00	0.00	−0.02	−0.13
	(−1.57)	(−1.78)	(−0.27)	(0.25)	(−0.60)	(−1.55)
$x_{(t-5,t-4]}$	0.00	0.00	0.00	0.00	−0.03	−0.05
	(−1.31)	(−0.69)	(0.45)	(−0.16)	(−1.11)	(−0.81)
Nobs	71,421	71,819	35,749	23,683	11,674	5,672
Adj.R^2	0.0019	0.0033	0.0046	0.0044	0.0057	0.005

Panel C: Regressions Predicting Direction of Large Trades While Controlling for Returns

	5sec	20sec	40sec	60sec	120sec	240sec
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.002***	-0.009***	-0.019***	-0.028***	-0.058***	-0.122***
	(-2.75)	(-2.72)	(-2.93)	(-3.02)	(-3.08)	(-3.28)
x_t	0.006***	0.026***	0.043***	0.052***	0.112***	0.219***
	(7.56)	(6.94)	(6.08)	(5.09)	(4.59)	(4.55)
$x_{(t-1,t]}$	-0.003***	-0.009***	-0.007	-0.027**	-0.023	-0.072
	(-3.12)	(-2.63)	(-0.98)	(-2.48)	(-0.97)	(-1.41)
$x_{(t-2,t-1]}$	-0.002*	-0.002	-0.004	-0.009	0.003	-0.018
	(-1.78)	(-0.58)	(-0.59)	(-0.92)	(0.14)	(-0.39)
$x_{(t-3,t-2]}$	-0.001	0.001	-0.007	0.014	0.001	-0.027
	(-0.74)	(0.39)	(-1.11)	(1.43)	(0.03)	(-0.61)
$x_{(t-4,t-3]}$	-0.001	-0.003	0.007	-0.002	0.016	-0.04
	(-1.2)	(-0.78)	(0.99)	(-0.21)	(0.7)	(-0.87)
$x_{(t-5,t-4]}$	0.002***	0.003	0.01	0.005	-0.022	0.063
	(2.8)	(0.85)	(1.53)	(0.56)	(-1.11)	(1.54)
r_t	0.052***	0.074***	0.09***	0.092***	0.121***	0.137***
	(30.87)	(12.38)	(15.18)	(7.01)	(18.72)	(13.9)
r_{t-1}	0.01***	0.022***	0.026***	0.026***	0.029***	0.025***
	(14.37)	(9.76)	(10.37)	(8.08)	(6.86)	(5.29)
r_{t-2}	0.008***	0.01***	0.009***	0.014***	0.009***	0.002
	(14.69)	(7.02)	(3.78)	(9.92)	(3.41)	(0.62)
r_{t-3}	0.006***	0.006***	0.007***	0.008***	0.002	0.003
	(7.55)	(5.67)	(5.32)	(6.09)	(1.15)	(1.26)
r_{t-4}	0.005***	0.005***	0.006***	0.004***	0	0
	(12.14)	(6.18)	(4.37)	(3.41)	(-0.09)	(0.11)
r_{t-5}	0.003***	0.002**	0.005***	0.003**	0.002	-0.003
	(6.76)	(2.12)	(3.89)	(2.26)	(1.34)	(-1.26)
Nobs	283,093	71,823	35,749	23,683	11,674	5,672
Adj.R^2	0.1011	0.1802	0.2256	0.2399	0.3028	0.3359

Panel D: Regressions Predicting Direction*Size of Large Trades While Controlling for Returns

	5sec	20sec	40sec	60sec	120sec	240sec
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.074	-0.307	-0.725	-1.188	-2.555	-6.308**
	(-1.07)	(-1.12)	(-1.33)	(-1.49)	(-1.6)	(-2.02)
x_t	0.017***	0.062***	0.104***	0.109***	0.159***	0.294***
	(14.49)	(11.56)	(8.53)	(5.05)	(3.32)	(3.05)
$x_{(t-1,t]}$	-0.016***	-0.048***	-0.072***	-0.068***	-0.051	-0.116
	(-10.81)	(-8.15)	(-5.72)	(-3.3)	(-1.09)	(-1.36)
$x_{(t-2,t-1]}$	-0.012***	-0.04***	-0.055***	-0.045**	-0.008	-0.038
	(-8.05)	(-6.84)	(-4.56)	(-2.52)	(-0.2)	(-0.47)
$x_{(t-3,t-2]}$	-0.009***	-0.024***	-0.051***	-0.016	0.023	-0.036
	(-6.67)	(-4.32)	(-4.82)	(-0.98)	(0.64)	(-0.53)
$x_{(t-4,t-3]}$	-0.008***	-0.021***	-0.022**	-0.018	-0.005	-0.071
	(-6.33)	(-4.14)	(-2.45)	(-1.29)	(-0.17)	(-1.17)
$x_{(t-5,t-4]}$	-0.003**	-0.011***	-0.005	-0.015	-0.026	0.014
	(-2.47)	(-2.77)	(-0.64)	(-1.33)	(-1.23)	(0.3)
r_t	4.829***	6.822***	8.312***	8.506***	11.226***	12.65***
	(30.64)	(12.26)	(15.92)	(6.98)	(22.29)	(17.59)
r_{t-1}	0.921***	2.007***	2.432***	2.47***	2.703***	2.496***
	(14.83)	(10.97)	(14.24)	(14.81)	(9.95)	(7.24)
r_{t-2}	0.755***	0.973***	0.85***	1.288***	0.833***	0.356
	(17.01)	(10.67)	(5.18)	(9.24)	(3.82)	(1.39)
r_{t-3}	0.552***	0.546***	0.658***	0.795***	0.249	0.479**
	(7.87)	(7.87)	(6.09)	(7.14)	(1.48)	(2.22)
r_{t-4}	0.456***	0.499***	0.512***	0.336***	0.243	0.381
	(11.32)	(6.7)	(5.29)	(3.04)	(1.47)	(1.58)
r_{t-5}	0.322***	0.18**	0.29***	0.296***	0.365**	-0.119
	(9.41)	(2.41)	(2.64)	(2.78)	(2.22)	(-0.56)
Nobs	283,093	71,823	35,749	23,683	11,674	5,672
Adj.R^2	0.11	0.20	0.25	0.26	0.34	0.37

Table 6: Regression Results for Effect of Child Size and Speed

In this table, we test separately and jointly the hypothesis that the size of the child orders preceding the large trade and the speed of the large trader affect the probability of being anticipated by OHFTs, using the following model.

$$Sign_i = \beta_0 + \beta_1 SignPT_{k,i} + \beta_2 SignPT_{k,i} * Child_i + \beta_3 SignPT_{k,i} * \ Speed_i + \beta_4 SignPT_{k,i} * Child_i * Speed_i + \sum_{j=1}^{15} \gamma_j Sign_{i-j} + \alpha_k + \varepsilon_i$$

$Sign_i$ is the sign of the i^{th} large trade and is equal to -1 for large sells and 1 for large buys. $SignPT_{k,i}$ as the sign of the k^{th} OHFT's inventory at the time of the i^{th}large trade and is either 1 or -1, indicating whether the k^{th} OHFT was long or short at the time of the i^{th}large trade respectively, and α_k provides an OHFT specific intercept for each OHFT. For each large trade i we calculate the average trade size for all trades executed by the given large trader beginning at large trade $(i-3)$ and ending just before large trade i. We sort all large trades by their corresponding average child order size and identify those large trades with child order sizes in the top 50% of child order sizes as having large child orders. For large trades associated with larger than average child order sizes, the variable $Child_i = 1$ otherwise it equals 0. We calculate speed for each of the 737 large traders in our sample as the average difference in time between order submission and order fulfillment for all passive orders that a large trader executes in our sample. We then sort large traders by average execution speed and those large traders in the slowest 50% of large traders are identified as being slow large traders and for large trades associated with these slow traders the variable $Speed_i = 1$ otherwise it equals 0. In model (1) we omit the interaction of speed and OHFT position, and speed and child order size to test the hypothesis that large trades associated with large child orders are more likely to be anticipated. In model (2) we omit the interaction of child order size and OHFT positon, and speed and child order size to test the hypothesis that large trades executed by slow large traders are more likely to be anticipated. In model (3) we estimate the full model with tests the joint hypothesis that both child order size and speed impact the likelihood that a large trade will be successfully anticipated. We do not include stars because all estimates are significant at the 1% level.

		Estimate	S.E.	t value	Pr($\leq \|t\|$)	Adj.R^2
(1) Child Size	$SignPT_{k,i}$	-0.019	0.003	-6.392	0.000	
	$SignPT_{k,i} * Child_i$	0.110	0.004	26.183	0.000	0.043
(2) Speed	$SignPT_{k,i}$	0.021	0.003	6.684	0.000	
	$SignPT_{k,i} * \ Speed_i$	0.026	0.004	6.253	0.000	0.041
(3) Child Size + Speed	$SignPT_{k,i}$	-0.031	0.005	-6.103	0.000	
	$SignPT_{k,i} * \ Speed_i$	0.018	0.006	2.879	0.004	
	$SignPT_{k,i} * Child_i$	0.086	0.006	13.318	0.000	0.044
	$SignPT_{k,i} * Child_i * Speed_i$	0.059	0.009	6.909	0.000	

Appendix: Screenshot of ITG POSIT homepage May 12, 2017

ITG®

CAREERS INVESTOR RELATIONS CONTACT CLIENT LOGIN ENG

SOLUTIONS | GLOBAL REGIONS | ABOUT | THINKING | INCUBATOR | NEWS

Liquidity

Electronic block crossing and alternative trading systems in 35+ countries

POSIT*

Price improvement of $1bn+ in midpoint executions over 10 years

- Global dark ATS/MTF, operating in 33 countries
- Technology rejects trades that don't meet our standards; for example, Liquidity Guard filters out 1 million shares a day in the U.S.
- Rewards size over speed, using a size-based pro rata matching logic instead of the price-time method many other pools use
- Minimized information leakage: We do not communicate IOIs or route orders outside POSIT
- Access to unique institutional block liquidity, readily accessible at the BBO midpoint

	% POSIT Traded at Midpoint	Average Price Improvement for Midpoint Fills
APAC	96%	9.6 bps
CAN*	89%	9.9 bps
EMEA	93%	4.5 bps
US	62%	4.2 bps

*MATCHNow in Canada

ITG is at the forefront of the industry in our commitment to providing transparency. We have published a guide to POSIT as well as POSIT's Form ATS.

Source: ITG, Q1 2017

Above is a screenshot from the ITG POSIT Dark pool's home page accessed on May 12,2017
https://www.itg.com/solutions/liquidity/