

# Anticipatory Traders and Trading Speed

Raymond P. H. Fishe  
Richard Haynes  
and  
Esen Onur\*

This version: November 25, 2015

---

## ABSTRACT

We investigate a class of market participants who follow strategies that anticipate local price trends. These anticipatory traders can correctly process information prior to the overall market and systematically act before other participants. They use manual and automated order entry methods and exhibit varying processing speeds, but most are not fast enough to be high frequency traders. In certain cases, other participants are shown to gain by detecting such trading and reacting to avoid adverse selection costs. To identify these traders, we devise methods to isolate local price paths using order book data from the WTI crude oil futures market.

**Keywords:** Algorithmic traders, Manual traders, Cancellation rates, Order book, WTI crude oil futures, Local price trends

**JEL classification:** G10, G13

---

\***Fishe:** Patricia A. and George W. Wellde, Jr. Distinguished Professor of Finance, Department of Finance, Robins School of Business, University of Richmond, Richmond, VA 23173. Tel: (+1) 804-287-1269. Email: [pfishe@richmond.edu](mailto:pfishe@richmond.edu). **Haynes:** U.S. Commodity Futures Trading Commission, Washington, D.C. 20581. Tel: (+1) 202-418-5000. Email: [rhaynes@cftc.gov](mailto:rhaynes@cftc.gov). **Onur:** U.S. Commodity Futures Trading Commission, Washington, D.C. 20581. Tel: (+1) 202-418-5000. Email: [eonur@cftc.gov](mailto:eonur@cftc.gov). Tel: (+1) 202-418-5000. We thank a number of individuals including Jonathan Brogaard, Steve Kane, Albert Menkveld, Scott Mixon, David Reiffen, Michel Robe, John Roberts, and Sayee Srinivasan for comments on earlier versions of this research. The research presented in this paper was co-authored by Raymond Fishe, a CFTC limited term-consultant, and Richard Haynes and Esen Onur, who are both CFTC employees, in their official capacities with the CFTC. The Office of the Chief Economist and CFTC economists produce original research on a broad range of topics relevant to the CFTC's mandate to regulate commodity future markets, commodity options markets, and the expanded mandate to regulate the swaps markets pursuant to the Dodd-Frank Wall Street Reform and Consumer Protection Act. These papers are often presented at conferences and many of these papers are later published by peer-review and other scholarly outlets. The analyses and conclusions expressed in this paper are those of the authors and do not reflect the views of other members of the Office of Chief Economist, other Commission staff, or the Commission itself. All errors and omissions, if any, are the authors' own responsibility. Corresponding author: Raymond P. H. Fishe.

# Anticipatory Traders and Trading Speed

## I. INTRODUCTION

The growth of algorithmic and high frequency trading (HFT) in financial markets has led regulators and some industry participants to express concerns that such traders may use processing speed to disadvantage other participants.<sup>1</sup> One claim is that algorithmic traders can anticipate future order flows because they process intraday order information more quickly than other participants. Some researchers also claim that such anticipatory trading generates negative externalities, such as reducing liquidity provision, incentivizing behavior like quote stuffing, inducing slower traders to depart, or facilitating overinvestments in technology (e.g., Ye, Yao, and Gai, 2013; Biais, Foucault, and Moinas, 2014; Foucault, Kozhan and Tham, 2014; Han, Khapko, and Kyle, 2014; Menkveld, and Zoican, 2014).<sup>2</sup>

Market structure changes have been suggested to reduce the value of minimal speed advantages, such as batch auctions or the slowing down of algorithms by trading venues (Budish, Cramton, and Shim, 2013).<sup>3</sup> However, before regulators act broadly on market structure changes, it is important to understand whether the problems commonly associated

---

<sup>1</sup> The popular press has questioned whether diverting resources to gain a speed advantage as small as a millisecond improves welfare (Matthew O'Brien, "High-Speed Trading isn't about Efficiency—It's about Cheating," *The Atlantic*, February 8, 2014. Stiglitz (2014) provides an economic rationale behind such claims. A counter-point view argues that supply and demand for liquidity explains what critics claim to be anticipatory trading (Renee Cruthers, "High-Frequency Trading's Manoj Narang Fires Back at Critics," in *Traders Magazine Online News*, May 14, 2014).

<sup>2</sup> Biais, Foucault, and Moinas (2014) develop a model in which firms over-invest in speedy technologies because they ignore the negative externality of a "technological arms race." Empirically, such technologies may also have diminishing returns. Specifically, after message technology upgrades by the Nasdaq in 2010, the number of cancelled orders increased, but trading volume and bid-ask spreads were not affected (Gai, Yao, and Ye, 2012). Similarly, latency reductions at the London Stock Exchange between 2007 and 2010 resulted in increased HFT market share, but little change in execution quality for institutional investors (Brogaard, Hendershott, Hunt, and Ysusi, 2014).

<sup>3</sup> See Mark Buchanan, "High-Frequency Traders Need a Speed Limit" on Bloombergview.com. Also, the IEX exchange provides a test of a slowdown approach by introducing 350 microseconds of latency for any order message by a participant (e.g., entry, cancellation, or modification). See <http://www.iextrading.com/about/>.

with trading speed are due to high frequency participants or whether similar behavior, and a similar ability to process information, is found in other groups who may include both high- and low-speed traders.

This study focuses on the anticipatory trading claim noted above. Our primary goal is to identify participants who can repeatedly execute trades consistent with subsequent intra-day price changes and then examine whether speed is a determining characteristic of these participants. We consider two measures of speed, one based on the average time between order entry and cancellation and the other based on the average time between order entry and trade execution. The former is based solely on the participant's decision-making, while the latter involves both the participant and the matching engine actions of other participants.

Previous studies that address this question include Brogaard, Hendershott, and Riordan (2014), Hirschey (2013), and Jiang, Lo, and Valente (2013). These papers often find that HFTs, as a group, can identify and trade before future price movements. A few of these papers also note that the correlation between non-HFT activity and prices is similarly positive, so the relationship between speed and the ability to predict price movements may be unclear. Our approach is different because we do not condition our sample on a group of speedy traders, which avoids the criticism of data filters used in HFT research (Biais and Foucault, 2014). Rather, we analyze all of the traders in our data and identify who, if anyone has an anticipatory trading ability. As such, we can infer, without particular conditioning adjustments, whether speed is an important population characteristic of traders who correctly anticipate subsequent price changes. In short, we find that high speed is not a defining population characteristic.

The sample data we study are unique to academic research. These detailed, anonymous account-level data are part of the overall information collected from exchanges by the Commodity Futures Trading Commission (CFTC). We have both order book and trade files for the WTI crude oil futures contract traded on the CME/Nymex exchange. These data are for the December 2011 expiration and contain 48 trading days beginning on September 12, 2011. The WTI contract is one of the most active futures contracts and the December expiration is typically the month with the greatest open interest during a given year. There are a total of 20,977 unique accounts with trades in the sample data.

Our methods examine thousands of trader histories and generate a large number of test statistics related to speed and predictive ability. To limit the size of the error rate of our identification method, we control for the *false discovery rate* that is exacerbated by these numerous tests (Benjamini and Hochberg, 1995). Our sample provides evidence that a small number of traders—308 out of 7,871 tested—act to consistently anticipate price changes or price reversals.<sup>4</sup>

Participants whose trades are found to consistently forecast local price changes are henceforth defined as “*anticipatory*” traders.<sup>5</sup> The subset of accounts that act early during a directional price change are labeled as “*Type E*” traders, and those accounts that forecast an up-coming price reversal are labeled as “*Type R*” traders. Algorithmic and manual traders are found in our sample within both of these groups.<sup>6</sup> On any given day, algorithmic traders are

---

<sup>4</sup> To ensure statistical power of the FDR method, we limit our testing to those traders who had 30 or more trades in our sample. This reduces the number of analyzed accounts from the initial 20,977 to 7,871.

<sup>5</sup> The theoretical basis for anticipatory trading ability may follow from proprietary representations of dynamic limit-order book models, such as those by Cont, Stoikov and Talreja (2010), Avellaneda, Reed, and Stoikov (2011), and Huang and Kercheval (2012).

<sup>6</sup> Whether a trader is manual or algorithmic is determined using a specific variable found in our exchanged-created dataset.

about 13% of sample participants and about 20% of the Type E and Type R trader types. Some of these traders are speedy, but only a few appear to trade at high frequencies.

A key component of our analysis is the identification of local price paths. We use data from trade histories to define these local paths.<sup>7</sup> Our approach first considers how a trader with perfect foresight acts. This trader would try to initiate trades in the correct direction at or around price reversals. We build on this idea to identify local patterns of generally increasing or decreasing trade prices.

We use a statistical rule to define local neighborhoods along the time series of trade prices and to check for non-random, trending behavior. The statistical rule is based on the *Sequential Probability Ratio Test* (SPRT) by Wald (1945).<sup>8</sup> The requirements of the SPRT specify the number of trades in a local neighborhood. By selecting an appropriate alternative hypothesis and a Type II error rate, we find that a local neighborhood size of 17 trades is sufficient to locate non-random price sequences.

The SPRT reveals which sub-sequences of the price series have candidate prices indicative of a change in the down or up trend in the price path. From these candidate prices, we select the best (i.e., lowest or highest) price as the turning point. We define a price path as the set of trades between these two extreme points, which are periods in which prices generally trend upward or downward. Within this framework, our methods of finding Type E and Type R traders use information that these traders do not have at the time of their trades, so

---

<sup>7</sup> Our approach makes results conditional on local price paths in the same way that research on the “Flash Crash” or other market events is conditional on trade paths around those events (e.g., Kirilenko, Kyle, Samadi, and Tuzun, 2011; Menkveld and Yueshen, 2013).

<sup>8</sup> An alternative is to use a preset condition to define whether prices have changed sufficiently to identify turning points (cf., Hautsch (2012, p.36). Instead, our method is one that participants might use to test for non-randomness in short-term price changes. This approach is consistent with a momentum strategy that seeks confirmation on the underlying price direction.

this conservative approach is biased against finding such systematically successful participants.

The traders we study are considered anticipatory if they can consistently buy (sell) at the beginning of an upward (downward) price path or if they can consistently sell (buy) near the end of a path and immediately before the reversal of an upward (downward) price path. Our analysis uses the first 10% of path volume to define the “beginning” and the last 10% of path volume to define the “end” of a path.

We use standard and inverse regression techniques to investigate the characteristics of Type E and Type R traders. Type E participants generally are no faster than what would be predicted from the sample data. However, Type E traders are found to be faster than the execution speed of the overall sample when trading in the first 10% of path volume. Type R traders, in contrast, are distinguished by slower execution speeds than the overall sample and are slower still in the last 10% of path volume, indicating an inverse relationship between speed and predictive abilities. Also, being an algorithmic trader does not distinguish the Type E participant from the overall sample, but it does help identify a Type R trader. As execution speed is not a distinguishing characteristic of Type R participants in the last 10% of path volume, being algorithmic appears to matter only for their trades elsewhere on a local price path. In effect, we find that execution speed is less important than critics suggest, as many slower manual-entry traders are able to anticipate local price trends. As such, speed alone does not identify who acts as an anticipatory trader.

Our secondary goal is to examine whether these Type E and Type R participants bring new information to the market and/or affect the behavior of other participants. We investigate price changes in the first and last 10 percent of path volume with and without these traders,

and also in contiguous bins with the next 10 percent of path volume. We find no significant differences in price changes for Type E traders, but there is a difference for Type R traders in down paths for one case. However, this difference is in the wrong direction, indicating smaller absolute price changes when they are trading than otherwise. These results suggest that the Type E and Type R traders are not informed in the classical sense; that is, bringing new, permanent price information to the market.

We examine two avenues in which Type E and Type R traders may affect other participants. First, we investigate whether these traders may influence the trading volume of other participants. For example, they may incite other participants to trade—a so-called “momentum-ignition” effect—and thereby they are not exactly “anticipatory” because they have caused others to trade.<sup>9</sup> Our results show no such effects for Type R traders, but some possible effects for Type E traders, although not temporally contiguous with their trades.

We also investigate whether other traders can detect and react to these anticipatory traders in a manner that reduces adverse selection costs. Because Type E and Type R traders are not found in every local price path, we observe the behavior of other traders when anticipatory traders are present and contrast this to when they are absent. Participants who can identify when anticipatory traders are in the market and react accordingly may lower their own adverse selection costs, most simply by canceling or modifying resting orders on the order book. To the extent that other traders make such adjustments on paths with Type E or Type R traders, then such traders may offer a *positive externality* to other market participants.

---

<sup>9</sup> Momentum ignition and other follower-inducing strategies require that other participants believe there is a positive probability of very recent informed trades (cf. Allen and Gale, 1992). These strategies are among those thought to be used by HFT firms (U. S. Securities and Exchange Commission, 2014).

We find that market participants cancel standing orders at higher rates when Type R traders are present. Specifically, an increase in net buying (selling) by Type R traders in the last 10% of path volume is followed by an increase in sellside (buyside) cancellation rates in the next price path. As net buying (selling) by Type R traders signals an upcoming increase (decrease) in market price, cancelling sell (buy) orders avoids adverse selection costs. A one standard deviation change in net buying by Type R traders is expected to increase sellside cancellation rates by 0.42%. Interestingly, the market reacts to the net buying behavior of other participants in the opposite way. A one standard deviation change in net buying by all non-anticipatory traders is expected to decrease sellside cancellation rates by 0.51%, which will increase adverse selection costs.

In contrast, other market participants do not seem to gain from Type E traders. When Type E traders are present, other market participants cancel orders by a small amount in the wrong direction: a 0.20% decrease in sellside cancellation rates from a one standard deviation increase in net buying by Type E traders. The reaction to Type E traders appears larger for the modification of standing orders. A one standard deviation increase in net buying by Type E traders is expected to increase buyside modification rates by 0.49%. This modification result suggests that other market participants alter limit prices to chase a new price trend; this behavior seems to outweigh reactions which avoid adverse selection, though this behavior may still be present.<sup>10</sup> Because of these mixed findings, we can only conclude that Type R participants appear to offer a positive externality to the market as a whole.<sup>11</sup>

---

<sup>10</sup> The responses observed give support to the “influential” order component in the limit-order book model by Cartea, Jaimungal, and Ricci (2012).

<sup>11</sup> Type R participants may also be interpreted as liquidity providers using the definition offered by Albert Menkveld (<http://albertjmenkveld.org/2015/02/24/who-supplies-liquidity-we-need-a-new-definition/>): “He who trades against price pressures, supplies liquidity.”



This paper proceeds as follows. Section II discusses the related literature and how our work differs from previous research on anticipatory trading. Section III describes the methods used to identify anticipatory traders, along with a general discussion of the local price path approach. Section IV discusses the data and how we measure the speed of trading. Section V provides our analyses and results. Finally, Section VI offers a few conclusions.

## **II. RELATED LITERATURE**

Several researchers have evidence related to our primary goal of assessing whether some participants anticipate subsequent price behavior. A group of these papers combine anticipatory and HFT behavior and are similar to our paper in terms of addressing the speed of trading. They include Brogaard, Hendershott, and Riordan (2014), Clark-Joseph (2012), Hirschey (2013), and Jiang, Lo, and Valente (2013). These papers often find that HFTs, as a group, can identify and trade before future price movements.

To give more detail, Brogaard, Hendershott, and Riordan (2014) and Hirschey (2013) both use a Nasdaq sample with HFT participation noted in the trade data. Brogaard et al. find that the correlation between net order flow for all sample HFTs and subsequent returns is positive, but it is short lived and quite low—less than 4% at one second and near zero at two seconds. Interestingly, they find that non-HFTs demanding liquidity show higher, longer lived correlations with subsequent returns than HFTs demanding liquidity, implying that sub-groups excluding HFTs also appear informed of future returns. This implication is developed more fully by our methods, which focus on finding all members of a group that follow a given strategy without conditioning on exogenous characteristics.

Hirschey's (2013) also finds that liquidity-demanding trades by HFTs precede liquidity-demanding trades by non-HFTs. He examines whether serial correlation in non-HFT order

flow, momentum strategies by non-HFTs, or a faster reaction to news by HFTs explains these results. On net, he suggests that his results are best explained by HFTs anticipating price pressure from non-HFTs. There are some sample differences between Hirschey, who uses 2009 data, and Brogaard et al. (2014) who include data from 2008 and 2009. However, both studies include randomly selected stocks, so the question that arises is why are liquidity-demanding non-HFTs predictive of returns if HFTs have anticipated their net orders? In other words, why do HFTs leave “money on the table” for non-HFTs? A possible explanation that we offer is that successful anticipatory strategies are found among both HFT and non-HFT groups, so the HFT filter in the Nasdaq dataset does not sufficiently separate anticipatory vs. non-anticipatory behavior.

Jiang, Lo and Valente (2013) analyze how often transactions are in the “right” direction compared to subsequent price changes. Their sample consists of trade and order data for the U.S. Treasury market on the *BrokerTec* platform operated by ICAP plc. They specifically study price responses around major macroeconomic announcements and find that non-HFT limit orders are vastly more predictive of subsequent price changes, but that HFT trades are often more predictive than non-HFT trades. Again, these results suggest that those who can anticipate subsequent prices are not singly defined by an HFT label, and that there may be a broad range of strategies that can be considered anticipatory.<sup>12</sup>

While we examine anticipatory traders in our paper, we realize that trading patterns of anticipatory traders might be similar to those of momentum trading strategies. There is a rich literature analyzing momentum trading strategies and what may be motivating such

---

<sup>12</sup> For example, Clark-Joseph (2012) examines order and transaction data for the e-Mini S&P 500 futures contract during 30 days in 2010. He suggests that aggressive HFTs execute multiple smaller size orders, which are generally unprofitable, for the purpose of obtaining order book information. When this information indicates a high propensity for a future price movement, the firm then profits from well-timed larger orders. However, only eight out of the 30 HFTs identified in his sample follow this “exploratory” strategy.

behavior.<sup>13</sup> Those with strategies that may coincide with anticipatory behavior include Bikhchandani, Hirshleifer, and Welch (1992) and Allen and Gale (1992). Bikhchandani et al. develop a model in which informational cascades cause momentum reactions. Allen and Gale show that traders are affected by big traders because they may be informed. Thus, smaller traders have an incentive to trade in the same direction as a big trader. Both of these environments may create “following trades” by investors, which may imply that the initiating trader(s) is misclassified as “anticipatory” in our structure. To sort out the degree to which such traders vitiate our results, we examine the volume behavior following trades by those identified as anticipatory.

Our findings also straddle other research on automated or high-frequency trading behavior and informed trading. This literature has grown rapidly in the past decade. Jones (2013), Biais and Foucault (2014), and the U. S. Securities and Exchange Commission (2014) provide recent reviews. Much of this research is aimed at determining how such trading affects the overall quality of the market. Important papers that help to frame our results include Hendershott, Jones, and Menkveld (2011), Menkveld (2012), Hasbrouck and Saar (2013), Jovanovic and Menkveld (2013), and Van Kervel and Menkveld (2015).

In the HFT literature, we are not the first to identify sample filters—inventory turnover, trading volume, cancellations, etc.—as a potential problem for inference. Biais and Foucault (2014) discuss several of the filter methods used to classify data as algorithmic- or HFT-

---

<sup>13</sup> For examples, Chan, Jegadeesh, and Lakonishok (1996) suggest that momentum trading may be an under-reaction to good or bad earnings news, while De Long, Shleifer, Summers, and Waldman (1990) show an over-reaction response in the market when positive feedback investors are trading. Hirshleifer, Subrahmanyam, and Titman (1994) argue that the timing of information is important. Dong, Polk and Skouras (2014) find that profits for momentum strategies are realized overnight rather than intraday. However, Lei, Han, Li, and Zhou (2015) analyze the S&P 500 ETF and find an intraday momentum pattern where the first half-hour return predicts the last half-hour return. Lei, et al. suggest that an increase in the first-half hour return (possibly due to news) causes day-traders expect a price reversion and to short the S&P 500 ETF. Because trading options are valuable, they will wait until the last half-hour to close the short position, which creates an intraday pattern.

related. They warn that “One problem with this approach is that it may select HFTs with a specific trading style...while excluding others (p. 10).” Thus, rather than pre-condition our analysis on a subset of trader characteristics, we seek to infer those population characteristics from all successful traders associated with a specific strategy.

### III. METHODS

Our analysis involves statistical methods to identify local price paths in intraday data. In the discussion below we explain how such paths are identified. We then discuss how the FDR method is used to determine whether any participant can systematically execute trades during selected segments on these paths. Finally, we show how the characteristics of anticipatory traders may be inferred using regression and inverse regression techniques (Li, 1991).

#### A. Local Price Paths

Our approach to identify price paths uses statistical methods to locate sequences where a participant may believe that price changes exhibit some degree of short-term predictability. Specifically, we seek to isolate periods during which prices tend to move in one direction or another in a non-random manner. We use the SPRT to define the local neighborhood size and to test for non-randomness. We then search for local price extrema in the identified sequences.

We start with a trade price series and then remove  $P_t$  if  $P_t = P_{t-1}$ , keeping the first price of each such sequence to preserve any contiguous, unique price levels. Then, we remove all sequences of contiguous bid-ask bounce. Specifically, we identify all cases in which  $\Delta P_t = -\Delta P_{t-1}$ . We retain the prices into and out of such sequences, but remove the intermediate implied bid-ask trades. For example, consider the price sequence,  $P_t \equiv \{3,4,5,5,4,4,5,4,5,6\}$ . After the first filter, the sequence becomes  $P_t^* \equiv \{3,4,5,4,5,4,5,6\}$ , from which we retain

$P_1^*, P_2^*, P_7^*$ , and  $P_8^*$  after the second filter. The purpose behind removing bid-ask bounce sequences is to exclude periods in which liquidity replenishment is sufficient to satisfy liquidity demand at existing prices. These sequences may provide information to market participants, but arguably provide little information on local price trends.

The above procedure produces a sequence in which  $\Delta P_t \neq 0 \forall t$  so that the new set of all contiguous prices show non-zero price changes for each observation. We then define a set of candidate prices based on the SPRT test results. Within a group of  $K$  prices, the price  $P_{t^*}$  is a candidate for a local *minimum* at trade  $t^*$  if

- (i) the count of *previous* price changes,  $n^- \geq \delta K$ , where  $n^- = 1(\Delta P_t < 0)$  for  $t \in \{t^*, t^* - 1, t^* - 2, \dots, t^* - K\}$ , and
- (ii) the count of *subsequent* price changes,  $n^+ \geq \delta K$ , where  $n^+ = 1(\Delta P_t > 0)$  for  $t \in \{t^* + 1, t^* + 2, \dots, t^* + K\}$ .

The parameter  $\delta \in [0,1]$  creates a consistency condition and is used to assign confidence to our selection mechanism based on the power of the SPRT. The parameter  $K$  defines the local neighborhood of trades. For example, if  $K$  is large and  $\delta = 1$ , then every price change before the candidate local minimum will decrement the previous price towards the minimum and every price change after the local minimum will increment the previous price away from the minimum. We use the same approach, but reverse the inequalities to define a candidate local *maximum* price.<sup>14</sup>

The basic statistical properties of error rates guide the selection of the consistency parameter and the size of the local neighborhood. Consider the null hypothesis that the binary

---

<sup>14</sup> This method may produce cases in which multiple minimums or maximums are contiguous on a price path. We remove such cases by selecting a global maximum or minimum as appropriate in such sequences. The final price paths alternate in the sign of  $\Delta P_t$ , where this price change is from the beginning to the end of a path.

variable tracking the sign of any price change is binomial with null parameter,  $H_0: q = 0.5$ , which normalizes the null distribution of sign changes to a random sequence. A participant attempting to detect a price change is most concerned about rejecting this null in the neighborhood of a candidate price extrema; otherwise there is really no temporary trend. Thus, it is useful to establish confidence that the null is rejected. The SPRT offers an answer to the size of the local neighborhood necessary to reject this null as this test is uniformly most powerful against any other test in its expected stopping time (Wald, 1948).

The SPRT computes the likelihood ratio for each successive observation in the trade sequence given a null and alternative hypothesis. It uses type I ( $\alpha$ ) and type II ( $\beta$ ) errors rates for these hypotheses to establish bounds for rejecting one hypothesis versus another. In our calculations we set both of these error rates equal to 10 percent, which then feeds back to the neighborhood size and consistency parameter.

To determine the neighborhood size, we simulate the number of trades necessary to reject the null ( $H_0: q = 0.5$ ) against the alternative ( $H_a: q = 0.8$ ). We use a strongly convincing alternative versus one closer to the null as participants would not rely on a testing method for local trends if it required a large number of trades, perhaps more than might be observed in a local trend. Using small differences between the null and alternative hypotheses creates longer required sampling sequences. With 1,500 simulations, we found that if participants selected 17 observations, then in only 10 percent of the cases would they require more observations before the test signaled a rejection of the null. As this choice equals the required  $\beta = 10\%$  Type II error rate of the test, we use 17 observations on both sides of a candidate price extrema to define the local neighborhood.

To define the consistency parameter, we use a choice that follows from the  $\alpha = 10\%$  Type I error rate used in the SPRT. Under the null hypothesis, the error rate is  $\Pr[\sum x \geq x' \mid K = K^*] \leq 0.10$ , where  $x \in \{0,1\}$  to indicate either a negative or positive price change. As the neighborhood size is set by the SPRT such that  $K^* = 17$ , using the binomial distribution, we find that the cutoff for 10 percent arises when  $10 < x' < 11$ . We experimented on randomly chosen days with both choices and found that  $x' = 10$  gave somewhat more paths, but the overlap was near 100% with  $x' = 11$ . As more paths are expected to make it more difficult to consistently trade in the correct direction of path prices, we used  $x' = 10$  in our analysis. Thus if at least 10 out of 17 price changes are observed with the appropriate sign—positive (negative) before a candidate price maximum (minimum) and negative (positive) after the candidate price—then we define that price as a valid “candidate” for a local extrema. This approach produces a consistency parameter  $\delta$  approximately equal to 60%.<sup>15</sup> To choose among the set of candidate prices within the same neighborhood, we select based on the conditions:

(iii)  $P_{t^*}^{min} = \min\{P_{t^*}, P_t\} \forall_{t \neq t^*}$ , where  $t \in \{t^* - K, t^* - K - 1, \dots, t^* + K\}$  for a minimum, and

(iv)  $P_{t^*}^{max} = \max\{P_{t^*}, P_t\} \forall_{t \neq t^*}$ , where  $t \in \{t^* - K, t^* - K - 1, \dots, t^* + K\}$  for a maximum.

Figure 1 illustrates how we conceptualize the working of the price path algorithm and the relative location of anticipatory traders. The figure shows a sequence of trades (the “ $x$ ’s”) for

---

<sup>15</sup> We also simulated our results with Type I and Type II error rates equal to 5%. These simulations gave a neighborhood size of 24 observations to maintain the type II error rate and a consistency parameter of approximately 63%.

a portion of the day's trading. There are periods with market-making activity in which non-price moving liquidity-based trading creates a bid-ask bounce sequence (shaded areas). When new information about the value of the asset arrives or liquidity demand changes, the market price reacts until the information is impounded in the price or new liquidity arrives to resolve the imbalance. Local price reversals occur at the specified price extrema (the circled trades).

In the figure, Type E traders possess the skills to process order flow and trade information to systematically forecast the short-term direction of trade prices. These participants react quickly *after* a price reversal occurs. Type R traders may use strategies that analyze order book liquidity or possess new information to place limit order prices near upcoming price reversals. The trades of these participants occur *before* but close to the local price extrema.

### *B. Finding Anticipatory Traders*

The WTI crude oil futures contract is among the most active contracts in futures markets, with a diversity of participants in any given expiration month. To identify which of these participants may be making use of an anticipatory strategy, we use the *False Discovery Rate* (FDR) method of Benjamini and Hochberg (1995) as applied by Storey (2002) and Fishe and Smith (2012) to adjust for multiple testing problems.<sup>16</sup> The FDR controls for the expected proportion of false discoveries in our sample. Specifically, as we increase the number of test statistics, we expect to increase the number of false rejections of the null. By effectively adjusting the critical levels for the appropriate test statistic, the FDR method limits these expected mistakes to a pre-specified proportion of successful statistics. We use a 5 percent

---

<sup>16</sup> Recent applications of FDR include Barras, Scaillet, and Wermers (2010) who sought to identify fund managers with positive alpha performance, Bajgrowicz and Scaillet (2012) who examine the success of technical trading rules, Fishe and Smith (2012) who identified the number of informed traders in several futures contracts, and Harvey, Liu, and Zhu (2014) who examine threshold critical values necessary to claim a new risk factor after hundreds of asset pricing tests by previous researchers.



control rate for this fraction. The FDR method increases the hurdle level and gives greater confidence that the participants we identify as anticipatory traders are truly either Type E or Type R traders.

We use a volume metric to determine whether a trader may be classified as participating early or late in a price path. Let  $x_{i,t}^j \in \mathbf{X}_j$  be the quantity traded by participant  $i$  at time  $t$  in the vector of all trades ( $\mathbf{X}_j$ ) on price path  $j$ . If participant  $i$  is a buyer (seller) at time  $t$  in path  $j$  then  $x_{i,t}^j > 0$  ( $x_{i,t}^j < 0$ ). If participant  $i$  does not trade at time  $t$  in path  $j$  then  $x_{i,t}^j = 0$ . The heaviside function,  $s_{i,t}^j = h(x_{i,t}^j \in \mathbf{X}_j; d)$ , defines whether trade  $x_{i,t}^j$  arises in the first  $d^{\text{th}}$  percentile of path  $j$ 's volume. The heaviside function equals 1 if the trade is in the first  $d^{\text{th}}$  percentile and equals zero otherwise. The price direction along path  $j$  is defined to be increasing (decreasing) if  $\Delta P_j > 0$  ( $\Delta P_j < 0$ ), where the first and last trade prices on the path are used to compute this difference.

Participants are identified as successful *Type E* traders on a given path if their trades occur in the first  $d^{\text{th}}$  percentile of path volume and their trades are on the correct side for the path's price change. For a given value of  $d$ , we compute the sample frequency of successes for each participant:

$$F_i^{\text{Type E}} = \frac{\sum_{j=1}^J \sum_{t=1}^{T_j} \mathbf{1}(x_{i,t}^j > 0, \Delta P_j > 0 | s_{i,t}^j = 1) + \mathbf{1}(x_{i,t}^j < 0, \Delta P_j < 0 | s_{i,t}^j = 1)}{\sum_{j=1}^J \sum_{t=1}^{T_j} \mathbf{1}(x_{i,t}^j \neq 0)}, \quad (1)$$

where  $\mathbf{1}(\cdot)$  is the binary indicator function based on the given expression,  $T_j$  is the number of trades on path  $j$ , and  $J$  is the number of price paths in the sample.

To determine the null hypothesis, consider what may arise for traders who are not attempting to compute turning points for intraday prices. If a trader is randomly placing both

buys and sells during the day in small sizes, then across all paths we might expect to find about 10% of these trades in the first 10% of path volume with  $d = 10$ . But how many of these are expected to be successful, meaning that they are aligned with the price path direction? The answer depends on how volume is distributed across up and down paths as well as how a trader mixes order size and side (e.g., buy or sell sides). If volume is approximately equally distributed between up and down paths, order sizes are small, and order sides are about equal in number during the day, then a null of 5% may be appropriate for our tests. However, volume is on average higher in down paths, traders often vary order sizes, and many traders end up with an unequal numbers of buys and sells. Such differences will alter the relevant null hypothesis.

Rather than seek a general solution for such nuances, we back up a step and impose a more restrictive condition in our tests. The measure in equation (1) is a statistic indicating the proportion of participant  $i$ 's trades that were executed in the first  $d^{th}$  percentile of volume *and* were in the correct price direction. This proportion is conditional on our perfect foresight calculation of local price paths. If  $d = 10$ , then it is clear from how the price paths are created that a trader has a 10% chance of executing (a buy or sell) within the first 10% of path volume assuming trades are randomly placed during the day. Any adjustments for order size or order side will lower this fraction. Thus, *to make it more difficult* to find successful anticipatory traders we use 10% as our null hypothesis. For each trader we test the null hypothesis,  $H_0: F_i^{type E} = \frac{d}{100}$ . This is a binomial test and will have statistical power if a participant trades a sufficient number of times.

For our empirical work, we set  $d = 10$  to identify Type E traders. To identify Type R traders, we consider the *last* 10<sup>th</sup> percentile of trading volume to be indicative of whether a

trader uses information or foresight to anticipate the coming reversal of the price path. To measure success for Type R traders, we compute the proportion analogous to equation (1) using  $d = 90$  to define the heaviside function:

$$F_i^{type R} = \frac{\sum_{j=1}^J \sum_{t=1}^T \mathbf{1}(x_{i,t}^j > 0, \Delta P_j < 0 | s_{i,t}^j = 0) + \mathbf{1}(x_{i,t}^j < 0, \Delta P_j > 0 | s_{i,t}^j = 0)}{\sum_{j=1}^J \sum_{t=1}^T \mathbf{1}(x_{i,t}^j \neq 0)}. \quad (2)$$

The null hypothesis that we test to identify Type R traders is  $H_0: F_i^{type R} = \frac{100-d}{100}$ . Note that to ensure statistical power, we confine our investigation to participants with more than 30 trades in our sample.

### C. Inverse Regression and Anticipatory Trader Characteristics

We ask whether anticipatory traders are different from other participants in characteristics other than their trade placement along local price paths. Because the FDR method makes subsequent analyses conditional on the Type E and Type R groups, we use the (sliced) inverse regression method to extract these characteristics.<sup>17</sup> Consider the effects of a vector of exogenous variables ( $X_i$ ) on a binary dependent variable ( $y_i$ ) summarized as  $E(y_i | X_i)$ , where the subscript  $i$  denotes an observation index. Here the binary dependent variable indicates membership in either the Type E or the Type R group. We recognize the initial conditioning from the FDR method and seek to solve for  $E(X_i | y_i)$ , which requires a dimensional reduction. Fische and Smith (2012) provide a detailed discussion of this reduction for the case of a binary dependent variable. Following their approach, assume a regression model:

$$y = X\beta + \varepsilon, \quad (3)$$

---

<sup>17</sup> Examples of the (sliced) inverse regression technique are found in Härdle and Simar (2003).

where  $E(X'\varepsilon) = 0$  and  $\beta_j$  is a coefficient in the parameter vector  $\beta$  that belongs to a particular variable of interest ( $x_{ji}$ ), with observations indexed by  $i$ . Then, the effects of  $x_{ji}$  for the identified participants may be measured by:

$$\tilde{\beta}_j \equiv \beta_j \frac{E(x_{ji} - \hat{x}_{ji})^2}{Pr(y_i=1)} = E(x_{ji}|y_i = 1) - E(\hat{x}_{ji}|y_i = 1), \quad (4)$$

where  $\hat{x}_{ji} = X_{\sim j}\hat{\gamma}$  is a linear projection of the  $i^{th}$  observation of the variable of interest using the least squares estimator of  $\gamma$ ; the latter is computed by regressing the variable of interest on all other exogenous variables in  $X_i$ , which is labelled as  $X_{\sim j}$ . In this formulation, the least squares method serves to reduce the dimensionality of the problem.

For example, if the characteristic of interest is the average speed of trading and there is only an intercept term in the remaining vector ( $X_{\sim j}$ ), then  $\hat{x}_{ji}$  reduces to the average speed of trading in the sample. The estimate of the net effects of being a type of anticipatory trader ( $y_i = 1$ ) is  $E(x_{ji}|y_i = 1) - E(x_{ji})$  from equation (4), which is the average trading speed of that group of anticipatory participants net of the average trading speed of all participants. In effect, the least squares projection serves as a reference point, so that we are measuring Type E or Type R characteristics relative to what would be predicted given the overall incidence of those characteristics in the sample.

#### IV. DATA

The data we examine are derived from audit trail files for the CME/Nymex WTI light sweet crude oil futures contract. The WTI contract is traded worldwide on the Globex and ClearPort electronic platforms. A trading session commences at 6:00 p.m. and concludes at 5:15 p.m. (EST) the next day. However, the majority of a session's volume occurs during the open outcry period, which is from 9:00 a.m. to 2:30 p.m. (EST) on Monday through Friday. For the

WTI contract each one cent move in price represents a \$10 change in contract value, which provides leveraged returns even for relatively small changes in price.

#### *A. Sample Information*

Our sample covers the period from September 12, 2011 to November 18, 2011, the latter of which was the last day of trading for the December 2011 expiration. This period is selected based on the trading and open interest activity in the December 2011 contract. The December 2011 contract is traditionally the first or second most active month in the year. Beginning on September 12<sup>th</sup>, the December expiration becomes the 1<sup>st</sup> or 2<sup>nd</sup> largest contract by open interest and is the 1<sup>st</sup>, 2<sup>nd</sup>, or 3<sup>rd</sup> largest contract by volume going forward. On September 21<sup>st</sup>, this expiration becomes the second highest volume contract. On October 7<sup>th</sup> it is the highest open interest contract, and on Oct 18<sup>th</sup> it is the highest volume contract going forward. On November 16<sup>th</sup> to 18<sup>th</sup> the volume rank falls from 1<sup>st</sup> to 2<sup>nd</sup>, and then 5<sup>th</sup> on the last day of trading in the expiration.

These data contain all trades and orders posted, modified, and/or cancelled on the Nymex exchange. Because we use order book data, we limit our sessions to all trades and orders between 6:00 a.m. and 4:15 p.m. (EST), which is the time range provided for the order book information in the CFTC database. There are a total 48 trading days and 20,977 unique participant accounts. Note that all account data are anonymous, so we do not know names or locations of these market participants.

In order to determine price paths using the SPRT method described above, we remove non-price forming trades from the sample, which are mainly transfers and offsets. We filter out spread trades where both sides are holding the spread. If one side of a spread trade is an

outright, we keep that side's price if it is for the December 2011 contract. After applying these filters, there are 6,736,520 buy and sell trades in the sample.

Table 1 provides sample statistics on the trading volume, number of participants, and order book data. The information is calculated across days in the sample. The WTI contract is quite active during this period with an average of 356,645 contracts traded each day. There are in total 20,977 unique participants with trades in our data, with an average of 2,939 of them active in any given day. Out of the many participants in this market, the majority of them are using manual entry methods to place their orders. There are only 399 algorithmic traders on average each day or about 13.6% of the daily average. Participants will modify about 60.5% of the daily average new orders and eventually cancel an average of 85.5% of those orders. These data also show that the WTI crude oil contract is traded in a nearly pure limit order market, with market orders on average only 0.5% of daily orders. In our analysis, we do not examine stop orders, offsets, transfer messages, or special order trade types, such as TAS (Trade at settlement) trades.

### *B. Measuring and Modeling Speed*

There are several ways to measure and model speed. The autoregressive conditional durations models of Engle and Russell (1998) and the multi-fractal Markov models of Chen, Diebold, and Schorfheide (2013) focus on inter-trade durations. These empirical models often study the high persistence and over-dispersion of duration data. These models inherently capture the flow of bids and offers off of the order book without special regard to who is trading.

As our focus is on the participants' characteristics, we seek to identify and measure individual durations. To compute individual durations, we identify the initial order submission

time for every order in the sample and document the exit accounting for those orders. Specifically, there are three ways an order may exit the order book: (1) cancellation, (2) matching a counterparty for an execution, and (3) administrative action. We exclude orders canceled by administrative action. If an individual cancels, we measure duration as the difference between the time the Nymex received the cancel message and the initial confirmation time of the order. This duration may be considered the *participant decision speed* because it is based on the speed at which the participant acts and does not explicitly depend on flows onto the order book to find a match.

If a trade occurs, we measure duration as the difference between the time the Nymex confirms the trade message and the initial confirmation time of the order. This duration may be considered the *execution or matching engine speed* because it depends on a host of factors that affect the order book, such as liquidity flows and new information about price, as well as the initial and subsequent decisions of the participant placing the order message, such as whether to modify the limit price or quantity.

Figure 2 shows a comparison between participant cancellation and execution speeds. In this figure we have plotted the average values of these speeds for all participants in the sample. The plot is heavily populated near the origin, but there are many participants spread over the entire quadrant. In particular, there is a cluster of data points with low average execution speeds matched to higher average cancellation speeds and to some extent vice versa. In effect, some participants may use market orders to demand immediacy, but otherwise their strategy does not require fast actions. Also, the mass of points near the 45 degree line between axes suggests that average cancellation and execution speeds are similar for most participants.

## V. ANALYSIS

Our first task is to estimate local price paths in the WTI crude oil contract data. Then we use the FDR method to assess whether any participants can systematically trade in the correct direction either at the beginning of a price path or just before the next price reversal. After identifying such traders, we examine their characteristics relative to other traders, specifically decision speed and matching engine speed. Finally, we determine if the other traders react differently to the trades of anticipatory traders compared to how they react when such traders are not in the market.

### A. *Local Price Paths*

We apply the SPRT method to find price paths for each day in the sample. Table 2 reports summary statistics derived from the calculation of these daily price paths. This table summarizes information by month and path direction and reports the average and median path returns in percent, average path duration in seconds, average path volume, average number of trades, and the average number of unique participants in a local price path. These statistics show that September had fewer paths and lower trading volume, which is expected as December was not the front month contract at this time. Trading activity and the number of paths increase markedly in October and November when the December expiration becomes the front month contract. These data do not show any strong patterns except that the “up” paths have lower average volume and somewhat shorter average path durations.

### B. *Identifying Anticipatory Participants*

To provide statistical power to the FDR approach, we limit our analysis to a sub-sample of 7,871 participants who had 30 or more trades in the sample. This restriction removed 13,106 accounts, some of which may be anticipatory traders. Using the binomial statistic given by



equation (1) in the FDR method, we found 112 participant accounts that indicated a systematic ability to trade in the first 10 percent of a given local path's volume and to trade in the correct direction on that price path. These are defined as the Type E group. Using equation (2) in the FDR method, we identified 196 participant accounts that systematically executed trades during the last 10 percent of volume in the correct direction based on the subsequent price path, the Type R group. Within these two groups, we find that a total of 198 are algorithmic traders; 70 in the Type E group and 128 in the Type R group. The remaining traders use manual order entry methods.

Table 3 illustrates the effectiveness of the FDR method in identifying Type E and Type R traders. The table shows the fraction of all trades on the buy side by Type E, Type R, and other traders in four volume segments within the local price paths, and also by the months in the sample. The four segments correspond to the first 10 percent of path volume, the next 40 percent to the volume midpoint, then the next 40 percent to the 90 percent level, and finally the last 10 percent of path volume. Panel A shows results for upward trending price paths and Panel B shows the same results for downward trending paths. In the first 10 percent of volume, Type E anticipatory traders are expected to disproportionately buy in upward trending paths and sell in downward trending paths. The data show an overwhelming tendency for this result with no less than an average of 82% of the trades on the buy side in the first volume segment for upward trending prices and between only 10% and 38% on the buy side in downward trending prices.

For type R traders, a similar effectiveness is found. In the last 10 percent of volume, Type R traders are expected to sell in upward trending price paths and buy in downward trending paths, thereby anticipating the change in local price direction in the next path. Table 3 shows

that between 87% and 92% of the trades are on the buy side in the last volume segment for downward trending price paths. Similarly, between 5% and 12% of trades are on the buy side in the last segment for upward trending paths. These are compelling results given that the buy side percentages for other participants show no pattern different than a 50-50 split between buys and sells in these same volume segments.

Table 3 also shows trade count information for the different types of participants. The Type E and Type R traders execute only a small number of trades in September consistent with the smaller path counts found during these months. These statistics suggest that anticipatory traders may be focused on the most active contract for implementing their strategies.

We are also interested in whether these anticipatory traders tend to be low latency or high frequency traders in the usual sense of the term. The Securities and Exchange Commission (2014) offers five criteria to identify high frequency traders: (1) high speed in routing and executing orders, (2) use of co-location services, (3) short time between establishing and liquidating positions, (4) using a submit and cancel approach to orders, and (5) ending trading sessions with near zero inventories. We have information on items (1), (3), and (4) from our calculation of average decision speeds from cancellation times and average matching speeds from execution times. A comparison of these speeds for Type E and Type R participants is shown in Figure 3.

From Figure 3, we see that several of the identified anticipatory traders are reasonably speedy, but the average times shown here are not on the order of 1-2 seconds as examined in previous studies. Visually, these traders appear to be a microcosm of the plot in Figure 2. The log-log trend lines suggest an elasticity of 0.54 between average execution speed and average

cancellation speed for Type E traders, and a slightly higher 0.72 elasticity for Type R traders. For all other traders in Figure 2, this elasticity is estimated to be 0.6, about midway between Type E and Type R traders. In effect, we observe (approximately) that a one percent change in cancellation speed results in a smaller percentage execution speed response for Type E (0.54% than Type R (0.72%) traders. These findings suggest that execution speeds and the related strategy are less connected to cancellation actions for Type E than for Type R traders.

Table 4 provides additional evidence that almost no Type E traders, but possibly a few Type R traders, can be considered analogous to what the literature calls HFTs. This table reports summary statistics from the CME/Nymex order book. Since July 2011, the CME has required a binary flag on all orders submitted to its platform to identify whether an order is entered manually or electronically. A negative response means that the order is entered electronically, which then carries an automated trading system (ATS) label for that account. We report counts as well as the ATS order ratios for Type E, Type R, and other traders. The ATS order ratio is defined as the number of ATS orders divided by the total number of orders for a given trader group.

Panel A shows that a slight majority of the orders submitted by Type E traders are manual, with only 48.6% of their orders submitted electronically. In contrast, a much higher ratio of Type R traders orders are submitted electronically, 92.6%. Similar to Type R traders, the messages of other participants are dominated by electronic order submission with 92.8% of their orders marked as using an ATS. This strengthens the result that a few of the Type R traders can be HFTs.

Panel B of Table 4 reports the distribution of messages by participant type. We observe that cancellation, execution, and modification rates differ significantly between Type E, Type

R, and the rest of the market participants. Type E traders cancel very few orders when compared to everybody else. They also have a significantly higher order execution ratio, suggesting that they place their orders with the intention of execution, with few cancellations. Type R participants, on the other hand, have much higher modification rates compared to everyone else and lower execution rates. This suggests that they modify their orders a lot and few of those orders get executed.

Importantly, Panel C shows counts across trader types for those whose average decision and execution speeds are less than 0.5 seconds. Although we report counts in the teens for Type E and Type R execution speeds, these are driven by cases in which certain participants use more market orders. For average decision speeds, there are no Type E participants and only three Type R participants revealing actions faster than 0.5 seconds. In contrast we find 75 high-speed participants among the remaining traders.<sup>18</sup> These results confirm that the anticipatory trading we identify is not primarily a HFT phenomenon.

### *C. Speed Characteristics*

An important goal is to characterize the anticipatory participants in terms of speed. We follow two steps in this task. First, we specify a regression model with characteristics partitioned by a set of dummy variables, some of which identify the anticipatory traders. With the regression model, the effects of the anticipatory participants are measured relative to other omitted groups. In the second step, we implement the inverse regression approach to show the net effects of selected characteristics on anticipatory traders. Figure 2 and Table 4 have shown that the population includes a heterogeneous mix of manual and ATS traders with varying

---

<sup>18</sup> Similar results are found if we compute decision speeds from order entry to an order modification, although the counts are lower because not all participants use modification strategies.

speeds and types of messages. Thus, we choose several variables to characterize the Type E and Type R traders relative to other groups.

Table 5 shows regression results using participant decision speeds—measured from order cancellations—as the dependent variable. The transformation, natural logarithm of one plus the cancellation speed (in seconds with fractional milliseconds), is used in the regression.<sup>19</sup> The variables, "First 10% of Path Volume", "Last 10% of Path Volume", and "Between 10% to 50% of Path Volume", indicate whether the cancellation message occurs during the first 10% of path volume, last 10%, or during the first half of volume that excludes the first 10%, respectively. The "Type E Trader" and "Type R Trader" variables indicate whether the trade is made by a Type E or Type R trader, respectively. Interaction terms are included in selected models: (1) Type E trader in the first 10% volume bin; (2) Type R trader in the last 10% volume bin; (3) Type E trader that is also algorithmic (ATS); and (4) Type R trader that is also algorithmic. Additional binary variables are "Modified Order" if there are any modification messages to the original order, "Proprietary trader" if the trade is made from a proprietary account, "Buy-Sell Indicator" which is one if this is a buy order, and the "Algorithmic Trader" variable which is one if this side of the trade was submitted by a computer algorithm. As all variables are binary, the intercept captures the omitted categories.

The p-values of these estimates are shown in parentheses below each coefficient. These are computed using heteroscedasticity-consistent standard errors. The adjusted R-squared for each model is also shown at the bottom of the table. The sample size is 30,870,516 observations.

---

<sup>19</sup> We adjust by (1+time) because we use order confirmation time as our initial time. In the CME/Nymex data the confirmation time of a market order equals its execution time, so duration will be zero on these orders.

The intercept term captures the average decision speed of the omitted group, which in the most general case (Model VI) is a non-anticipatory, non-proprietary, manual participant, cancelling during the 50-90% of path volume with no order modifications before the cancellation message. The average cancellation speed of the omitted group is 29.858 seconds, which decreases to an average of 4.388 seconds if these are algorithmic participants. The estimates in Models II-IV show that Type E and Type R participants are faster than the omitted group until we control for other order characteristics: proprietary trades, modifications, order side, and algorithmic variables. Specifically, Model VI shows that the average cancellation speed of manual Type E (Type R) traders is 91.579 (76.663) seconds. The interaction terms with the algorithmic variable show that Type E (Type R) traders are marginally faster than the omitted algorithmic group with an average cancellation speed of 4.352 (4.090) seconds. From these estimates, the algorithmic trader coefficient is found to have the largest impact on average decision speed.

We reported above that Type E and Type R groups are composed of both manual and algorithmic traders and, as Table 5 shows, the average decision speeds of these traders are significantly different. Thus, it stands to reason that the capacity for speedy actions is *per se* insufficient to define anticipatory trading. At least as measured against our definition of local price paths.

Table 5 also includes regression estimates under the "Bootstrap" column heading. These estimates are averages of coefficient results from 1,500 random samples (with replacement) in which each trader account is chosen *only once* per sample. The purpose here is to equally weight each account in the regression so that the numerous orders from higher frequency participants do not give them greater influence on the resulting coefficients. The 95%

confidence intervals ("C.I.") from these simulations are shown below the average coefficient value. The sample size for each bootstrap sample is 11,869 observations.

The average coefficient signs and the confidence interval results generally confirm the comparisons using the full sample of data. However, the most notable difference is that the biggest effect on whether a participant's decisions are speedy compared to other participants is the proprietary trading variable. That is, removing the volume influence of higher frequency traders, the algorithmic coefficient reduces its relative impact by 75%. As many manual traders may also be proprietary, these results suggest that some manual traders may operate at relatively fast decision speeds.

In fact, these data reveal that 32 manual traders have average cancellation speeds less than one second, with eleven of these less than one-half a second. More broadly, there are 100 manual traders with average cancellation speeds less than three seconds. Certainly, there are more algorithmic traders with these speeds as the comparable counts are 130 and 268, respectively. However, this does highlight that some manual participants may use strategies that require quick actions, and some of these participants have that capacity.<sup>20</sup>

Table 6 presents similar speed results using the individual execution speed measure. In addition to the binary variables in Table 5, we include a dummy variable for whether this participant was on the aggressive side of the trade. Model VI coefficient estimates show that manual Type E (9.50 sec) and Type R (24.71 sec) traders are on average slower than the base speed of the omitted group (4.88 sec), which now includes non-aggressive traders. Execution speed increases significantly in the first 10% of path volume for Type E traders, falling to 6.70 seconds for manual traders and from 3.95 to 2.63 seconds for algorithmic traders. These

---

<sup>20</sup> For an example of fast manual trading see: <http://www.wsj.com/articles/how-spoofing-traders-dupe-markets-1424662202>

results suggest that when Type E traders detect that a path has a new local trend, they act quickly to execute directional positions. In contrast, average speeds decrease for Type R traders in the last 10% of path volume suggesting that strategies for these participants are not based on a fast execution. Given their relatively high use of modifications (see Table 6) Type R traders appear to reposition their limit prices until they execute in the order flow near the end of a local price path.

The bootstrap results are also computed for individual execution speeds and shown in the last column in Table 6. These results confirm the execution speed profile of the Type R participants found when using the full sample. However, they suggest that the Type E participants may not be significantly faster than non-anticipatory participants, particularly in the first 10% of path volume when the data are not influenced by the volume from frequent traders. In effect, those that act relatively fast in the first 10% of path volume are likely those who trade more than the average among the Type E participants.

We also estimated a logistic model (not shown) using the aggressor indicator as the dependent variable. The full sample results show that Type E traders (manual and algorithmic) are more aggressive than other participants and the probability of aggressive trades increases in the first 10% of path volume. The bootstrap estimates show that trader volume affects these observations as the previous claim holds only for algorithmic Type E participants with manual participants no more aggressive than other traders. Also, there is no aggressiveness effect in the first 10% of path volume in the bootstrap estimates. In contrast, the Type R traders are less likely to be aggressive than other participants, both for manual and algorithmic cases and this holds for the full sample and the bootstrap estimates. In short,



higher volume, algorithmic Type E participants are more aggressive than other traders, but Type R participants are uniformly less aggressive.

The inverse regression results are presented in Table 7. As noted in the discussion of our methods above, these results show how the characteristics of Type E and Type R participants are different from what we would predict from the general distribution of those characteristics across the sample. The data in this table are for the sample of participants tested using the FDR method, so they all have 30 or more trades. As Type E and Type R findings are drawn from this group, their characteristics are defined similarly.

Table 7 shows results for the following variables: average speed during the first 10% of path volume; average speed during the last 10% of path volume; average speed over all trades; average trade size; percentage of aggressive trades; percentage of trades on the buy side; binary indicator for proprietary trader; and a binary indicator for algorithmic trader. When a participant has no data in either the first or last 10% of path volume, they are excluded from these regressions, so sample sizes vary between Type E and Type R results.

The speed results in Table 7 are similar to the bootstrap findings for Table 6.<sup>21</sup> Type E traders are found to be no different in speed than the general speed of the sample, except when trading in the first 10% of path volume. Type R traders are slower than the sample average, but are not significantly slower during the last 10% of path volume. Thus, speed of execution is a distinguishing characteristic of Type E traders only when they detect a new price trend, otherwise they are similar to the sample average. Execution speed is a defining characteristic of Type R participants in that they are *slower* than would be expected from the sample data. These speed results suggest that Type E traders act quickly to execute orders when a new

---

<sup>21</sup> The bootstrap findings are based on equal-weights for all participants, so they are most similar to these inverse regression results.

local price trend is detected, while Type R traders appear to modify their limit orders (see Table 4) in order to time their executions near the end of a price path. In the former case speed is important, but not significantly for the latter case.

Related to execution speed are the results for the algorithmic flag. Table 7 shows that being algorithmic does not distinguish the Type E participant from the overall sample. This finding is consistent with the bootstrap results for the interaction between algorithmic trading and the first 10% of path volume in Table 6. However, algorithmic traders have significantly greater representation versus the overall sample for Type R traders, approximately 5.5% greater based on these estimates. We also isolated the algorithmic trader flag by excluding other variables and repeated the inverse regression procedure. The coefficient on this variable for the Type R group is 0.249 or 24.9%, which is significantly greater than the 18.3% sample average. Interestingly, because trading speed is not a distinguishing characteristic of Type R participants in the last 10% of path volume, being algorithmic appears to matter only for their trades elsewhere on a local path.

Other coefficient estimates in Table 7 show characteristics that distinguish the Type E and Type R groups from the remaining participants. Both Type E and Type R traders have relatively larger trade sizes, which may explain why market participants can detect their activity as reported in the next section. Both groups also tend to be more represented on the buy side of the market. Lastly, Type E participants are more likely to be aggressive with their trades and Type R participants are more likely to be proprietary compared to the overall sample.

#### *D. Measuring Information Effects and Actions of Other Market Participants*

Our secondary goal is to examine whether these Type E and Type R participants bring new information to the market and/or affect the behavior of other participants.<sup>22</sup> We first examine whether new information is provided by these anticipatory traders. Dividing each price path into volume deciles, we examine the first and second volume bins for Type E traders and the last bin in the previous path as well as the first bin in the current path for Type R traders. We compute the price change in each bin relative to the absolute price change in the path. Table 8 reports these average relative price changes for bins with and without anticipatory traders and for “up” and “down” price paths separately. The averages using all bins are also shown for comparison.

For Type E comparisons, we examine the bin averages conditional on whether Type E traders are transacting in Bin#1. Similarly, for Type R comparisons, we condition on whether Type R traders are transacting in Bin#10 on the previous path. We expect that if these traders bring new price information to the market, then in the current or next bin we will detect a difference in relative prices when they are participating in the current path (Type E) or previous path (Type R). We find no significant differences in relative price changes for bins with and without Type E traders, but one case of a significant difference for Type R traders in down paths. However, this difference is in the wrong direction, indicating smaller absolute price changes when they are trading than otherwise. These results suggest that the Type E and Type R traders are not informed in the classical sense; that is, bringing new, permanent price information to the market.

---

<sup>22</sup> In a similar idea, Bernales (2014) presents a model of dynamic limit order markets with algorithmic traders. In his model, slow traders observe the fundamental value of an asset with a time lag and they can learn from market trading activity and improve the accuracy of their beliefs.

Our second set of results address whether anticipatory traders affect the orders and/or trades by other market participants. First, we investigate whether these traders may influence the trading volume of other participants. For example, if they increase the volume by other participants, then price moves may be “caused” by their trading activity, a “momentum-ignition” effect. In this case, they are not literally “anticipatory” because they may have caused others to trade. We use a VAR model of price and volume to examine this question. This model is estimated using bin data across the 15,765 price paths. Table 9 shows reduced form results for the VAR model.

In Table 9, first differences of price and volume for all other participants are used to improve series stationarity.<sup>23</sup> As the data are bin specific, the difference for price and volume are relative to the previous bin. End-of-bin prices are used for price deltas. Both price and the volume of all other participants (i.e., not Type E or Type R) are endogenous, while the trading volume of Type E or Type R participants is exogenous, which implies that these variables enter the model contemporaneously. We also include a dummy exogenous variable for whether Type E (Type R) participants trade in Bin#1 (Bin#10 of the last path). This variable is the focus here because we want to know whether the volume of trading by other participants is affected by anticipatory traders, especially in those bins where they are likely implementing their anticipatory strategy. We consider effects for participants in the full sample and for Type E in Bin#2, and for Type R in Bin#1.<sup>24</sup>

The results in Table 9 suggest that contemporaneous trading by Type E participants is associated with reduced volume by other traders, but the reverse holds for Type R participants

---

<sup>23</sup> Adjusted Dickey-Fuller tests indicate that bin prices are stationary in first differences. Volume by trader type is generally stationary in levels and always stationary in first differences.

<sup>24</sup> We also estimated the VAR model without overlaps between Type E and Type R traders on a given path as discussed in Table 8. This filter did not change our conclusions.

(Model A). Important for momentum claims, we observe that trading by Type E participants in Bin#1 is associated with higher average changes in volume for all other participants (Model B). This effect appears to be driven by actions later in the price path because the Bin#2 results show no significant effects of Type E participants in Bin#1. In contrast, Type R participants in Bin#10 on the last path have a significant negative effect on the change in volume in the next path, both on average across bins and in Bin#1. In short, Type R participants are not associated with a “momentum ignition” effect, but Type E participants may incite some volume effect, but this effect is not specifically related to their volume in Bin#1.<sup>25</sup>

Because we observe that other market participants are acting as if they may detect trading by Type E and Type R participants, we also investigate whether other participants may use this information to reduce adverse selection costs. Commonly, adverse selection costs are measured as a component of the bid-ask spread (Van Ness, Van Ness, and Warr, 2001; Barclay and Hendershott, 2004). Because we have order book data, we examine whether order cancellations and modifications by other participants are different in local price paths with anticipatory traders versus in paths without such traders. An obvious way to avoid or lower adverse selection costs if you can react quickly is to cancel or modify your order on the order book.

Thus, we examine whether the cancellations and order modifications of other participants later in a price path are correlated with the earlier trades of Type E and Type R participants. For Type E traders, we examine the activity of other participants in the path volume intervals: 10<sup>th</sup> to 50<sup>th</sup> percentile and the 10<sup>th</sup> to 100<sup>th</sup> percentile. For the Type R traders, we examine the 0 to 50<sup>th</sup> percentile and the 0 to 90<sup>th</sup> percentile of volume in the subsequent price path. We use

---

<sup>25</sup> The specific bin results on Type E participants were also estimated for Bin#3. There is a significant negative effect from contemporaneous volume and an insignificant effect for the Bin#1 dummy variable.

regression methods adjusted appropriately for the time series characteristics of these data when analyzing cancellation and modification variables.

In Table 10 we measure the effects of Type E traders using their relative net buying behavior in the first 10% of current path volume and the effects of Type R traders using their relative net buying behavior in the last 10% of previous path volume. The relative net buying variable equals the buy minus sell volume of anticipatory traders divided by the total buy and sell volume of both anticipatory and other traders. These variables have zero values when there is no trading by Type E or Type R traders in these deciles. To correspond correctly, the dependent variables are all measured in the upper 90% of path volume in the Type E models and in the lower 90% of path volume in the next path following the path that measures Type R trading.

The dependent variable in Table 10 is the cancellation fraction for buy-side orders for all other participants. This variable equals the volume of buy orders cancelled divided by the volume of both buy and sell orders cancelled for other participants. All of these variables are measured over the time corresponding to the relevant volume percentile. The other independent variables are the rate of price change over the current path and its lag value, the rate of trading over the current path and its lag value, and a dummy variable for path direction (one is for upward paths). The regressions are estimated after controlling for an AR(1) process using Durbin's method, which normalizes the resulting residuals. Corrected p-values are shown below each coefficient and R-squared is shown at the bottom of each estimated model.

Cancellations are the clearest action a participant can take to avoid adverse selection costs if they detect directionally informed trading. Because of the accuracy of Type E and Type R traders (see Table III), detecting positive (negative) net buying by Type E traders strongly

suggests the path's prices will increase (decrease). Similarly, detecting positive (negative) net buying by Type R traders strongly suggests the next path's prices will increase (decrease). Thus, to avoid adverse selection costs from Type E traders, other traders are expected to act based on their own position in the book and the signal received about the direction of prices on the current path.

If the limit orders of other traders are on the buy side (sell side), then they are less affected by upward (downward) paths, but if they are on the sell side (buy side) then they risk trading a stale order and thereby incurring adverse selection costs. As such, we expect the coefficient on the net buys from Type E traders to have a negative effect on buy side and a positive effect on sell side cancellation rates. The negative effect on the buy side is because a negative rate of net buys for Type E traders and a negative coefficient implies an increase in buy side cancellation rates. By a similar argument, we expect the coefficient on the rate of net buys from Type R traders at the end of the current path to have a negative effect on buy side and a positive effect on sell side cancellation rates in the next path.<sup>26</sup>

Table 10 shows that the expected effects arise only for Type R traders, with significant coefficients in Models VII, IX, X, and XII for buy side reactions to the relative net buying volume of these traders. In this last 10% of path volume we also find that relative net buying of other traders sends the wrong signal to the rest of the market. There is also an incorrect signal from Type E traders, although it is insignificant in the 10% to 100% of path volume. In the 10% to 50% of path volume, the coefficient on relative net buying by Type E participants is positive, which implies an increase (decrease) in buy side (sell side) cancellations by other

---

<sup>26</sup> By definition, the coefficient on net buys for anticipatory traders is minus the coefficient that would arise if the dependent variable were defined as the sell side cancellation fraction.

traders. In terms of cancellation rates there is an apparent asymmetry in how other market participants respond to Type E and Type R traders.

This asymmetry may be resolved if modification rates also respond to these anticipatory participants. That is, it is possible there is a trade-off between modification and cancellation actions by other market participants. Other market participants may choose to modify orders instead of using outright cancellations in the presence of anticipatory traders. Table 11 investigates whether this possibility is supported in our sample using a definition for the modification fraction analogous to the cancellation fraction.

The coefficient signs expected for modification rates are the same as those expected for cancellation rates. As Table 11 shows, we estimate significant effects on buy-side order modifications for Type E participant volumes, but not for Type R. The sign of the effect for Type E participants is positive, which is the wrong direction if other participants seek to avoid adverse selection costs by modifying their limit prices. These results suggest that other participants may be chasing the trend with their modified orders.

The size of the effects in Tables 10 and 11 may be shown by examining how a one standard deviation change in net buying by Type E or Type R traders affects cancellation or modification rates. When Type E traders are present, we find that other market participants cancel orders by a small amount in the wrong direction: a 0.20% decrease in sell-side cancellation rates in the 10% to 50% of path volume from a one standard deviation increase in net buying by Type E traders. When Type R traders are present, other market participants are expected to increase sell-side cancellation rates by 0.42% in the 0% to 50% of next path's volume. The reaction to Type R traders is expected to reduce adverse selection costs; in contrast, the market reacts in the wrong direction to the net buying behavior of other



participants. A one standard deviation change in net buying by all other traders is expected to decrease sellside cancellation rates by 0.51%.

The reaction to Type E traders is of a similar magnitude for the modification of orders. A one standard deviation increase in net buying by Type E traders is expected to increase buy-side modification rates by 0.49% in the 10% to 50% of current path volume. Again, this appears to be buy-side limit prices adjusting to pursue the price trend and not to avoid adverse selection, which would be found if the sellside had increased modification rates.

Overall, the results Tables 10 and 11 show that other market participants are reacting to the trading of Type E and Type R participants. However, only the cancellation reaction to Type R participants serves to unambiguously reduce adverse selection costs.

## **VI. CONCLUSIONS**

We analyze whether there are traders who can consistently trade early when prices begin to trend locally in one direction or are about to reverse direction. This is arguably anticipatory trading behavior. We developed methods to identify local price paths and to identify traders (Type E) whose executions in the first 10% of path volume make them prescient for the remaining direction of prices on the local path. We also found traders (Type R) whose executions in the last 10% of path volume predicted a subsequent price reversal.

For characteristics, we found that these anticipatory traders were both manual and algorithmic, but that being algorithmic was not distinctive for the Type E group. It did help identify the Type R group relative to the overall sample, but we also found that participants in the Type R group were slower than would be expected from the overall sample. In addition, both Type E and Type R groups contain speedy traders, but they were not broadly characterized as speedy compared to the overall sample. We found that speed was only a

meaningful characteristic of Type E participants in the first 10% of path volume, where fast actions allow such participants to trade early in a price path. Otherwise, such participants were no speedier than the overall sample. Importantly, both Type E and Type R participants are populated with both algorithmic and manual entry traders, thus high frequency systems are not a necessary condition for an anticipatory trading strategy.

We also investigated whether other market participants may reduce adverse selection costs by reacting to the trading of these anticipatory traders. The results here are mixed with strong support for the view that other traders cancel orders on the appropriate side when Type R participants are present. There is also an increase in order cancellations and modifications when Type E traders are present, but these occur on the wrong side of the market, so they may exacerbate adverse selection costs for other market participants.

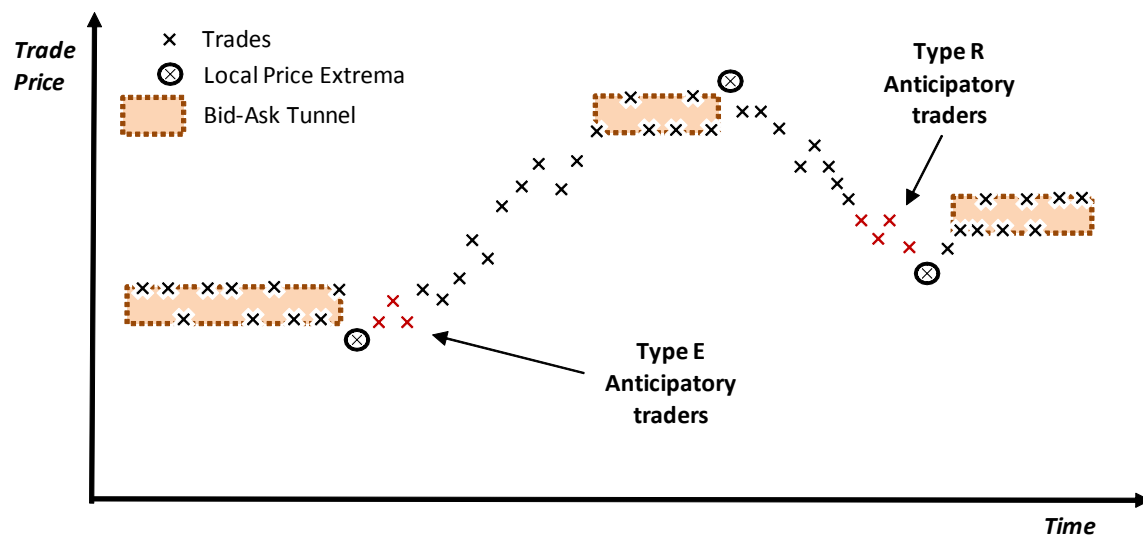
## REFERENCES

- Allen, Franklin, and Gale, Douglas, 1992, Stock-Price Manipulation, *The Review of Financial Studies*, 5, 503-529.
- Allen, Franklin, Morris Stephen, and Shin S. Hyun, 2006, Beauty Contests, Bubbles, and Iterated Expectations in Asset Markets, *The Review of Financial Studies*, 19, 719-752.
- Avellaneda, M., Reed, J., and Sasha Stoikov, 2011, Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity, *Algorithmic Finance* 1, 35-43.
- Bajgrowicz, P. and Scaillet, O., 2012, Technical Trading Revisited: False Discoveries, Persistence Tests, and Transaction Costs, *Journal of Financial Economics* 106, 447-684.
- Barras, L., Scaillet, O. and Wermers, R., 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *Journal of Finance* 65, 179-216
- Barclay, M. and Hendershott, T. 2004, Liquidity Externalities and Adverse Selection: Evidence from Trading After Hours, *Journal of Finance* 59, 681-710.
- Baron, M., Brogaard, J., and A. Kirilenko, 2014, Risk and Return in High Frequency Trading, Working paper, University of Washington.
- Benjamini, Y., and Y. Hochberg, 1995, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B* 57, 289-300.
- Bernales, Alejandro, 2014, Algorithmic and High Frequency Trading in Dynamic Limit Order Markets, working paper, Universidad de Chile.
- Boehmer, E., Fong, K., and J. Wu, 2012, International Evidence on Algorithmic Trading, Working paper, EDHEC Business School.
- Biais, B., T. Foucault, and S. Moinas, 2014, Equilibrium Fast Trading, Working paper, IDEI Toulouse, forthcoming in the *Journal of Financial Economics*.
- Biais, B., and T. Foucault, 2014, HFT and Market Quality, *Bankers, Markets & Investors* 128, (January-February), 5-19.
- Bikhchandani Sushil, Hirshleifer David, and Welch Ivo, 1992, A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades, *Journal of Political Economy* 992-1026

- Brogaard J., Hendershott, T., Hunt, S., and C. Ysusi, 2014, High Frequency Trading and the Execution Costs of Institutional Investors, *The Financial Review* 49, 345-369.
- Brogaard, J., Hendershott, T., and R. Riordan, 2014, High Frequency Trading and Price Discovery, *Review of Financial Studies* 27, 2267-2306.
- Budish, Eric, Cramton, P., and J. Shim, 2013, The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response, working paper, University of Chicago, Booth School of Business.
- Cartea, A. and J. Penalva, 2010, Where is the value in high frequency trading?, working paper, Universidad Carlos III, Madrid.
- Chan, K., C. Louis, Jegadeesh Narasimhan, and Lakonishok Josef, 1996, Momentum Strategies, *Journal of Finance* 51, 1681 – 1713.
- Chen, Fei, Diebold, Francis X., Schorfheide, Frank, 2013, A Markov-switching multifractal inter-trade duration model with application to US equities, *Journal of Econometrics*, forthcoming.
- Clark-Joseph, Adam, 2012, Exploratory trading, working paper, University of Illinois at Urbana-Champaign.
- Cont, R., Stoikov, S., and R. Talreja, 2010, A stochastic model for order book dynamics, *Operations Research* 58, 549-563.
- De Long, J., Bradford, Shleifer, Andrei, Summers H. Lawrence, and Waldman J. Robert, 1990, Positive Feedback Investment Strategies and Destabilizing Rational Speculation, *Journal of Finance* 45, 379-395
- Engle, Robert, and J. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, 66, 1127–1162.
- Fishe, Raymond P.H. and Aaron D. Smith, 2012, Identifying informed traders in futures markets, *Journal of Financial Markets* 15, 329-359.
- Foucault, Thierry, Roman Kozhan, and Wing Wah Tham, 2014, Toxic arbitrage, working paper, HEC Paris.
- Gai, J., Yao, C., and Mao Ye, 2012, The externalities of high frequency trading, working paper, University of Illinois at Urbana-Champaign.
- Gao Lei, Han Yufeng, Li, Z., Sophia, and Zhou Guofu, 2015, Intraday Momentum: The First Half-Hour Return Predicts the Last Half-Hour Return, working paper, Washington University, St. Louis

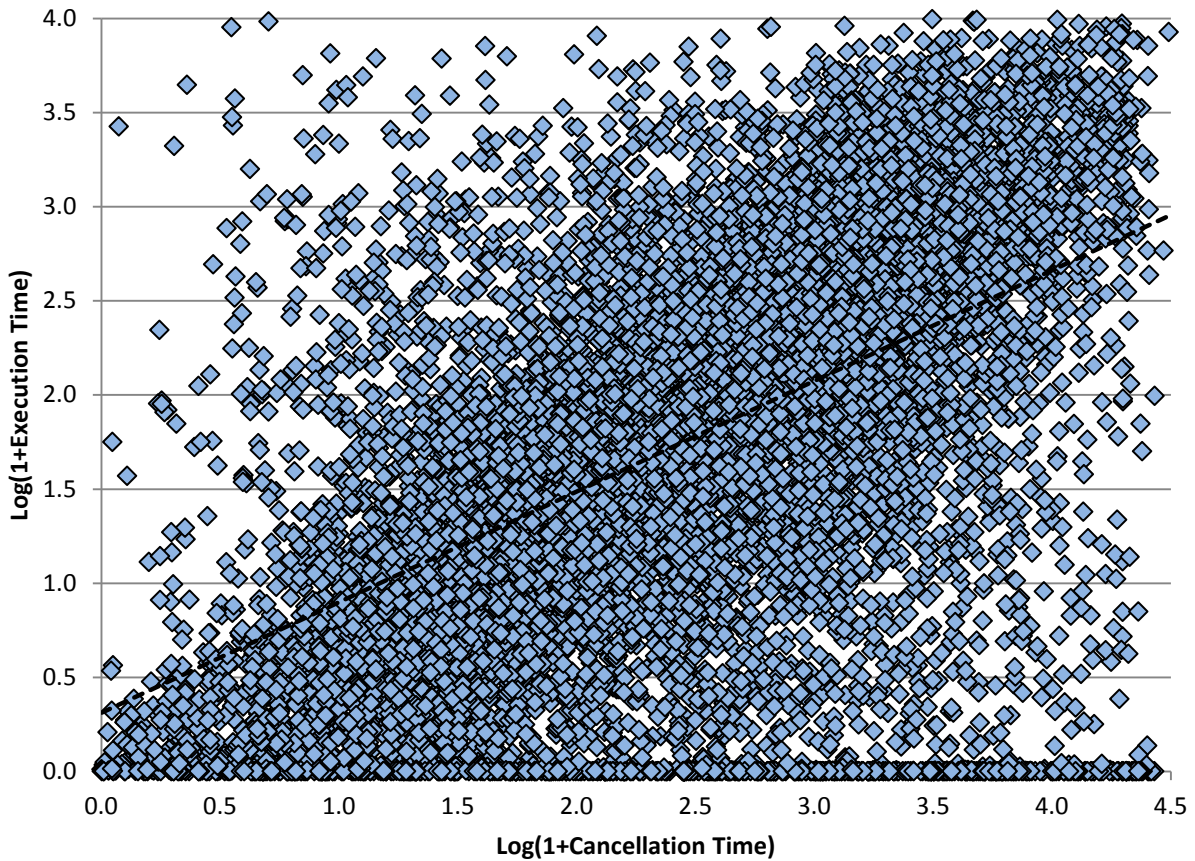
- Han, Jungsuk, Mariana Khapko, and Albert Kyle, 2014, Liquidity with high-frequency market making, working paper, Stockholm School of Economics.
- Härdle, W. and Simar, L., 2003, Applied Multivariate Statistical Analysis, Springer-Verlag: Heidelberg.
- Harvey, Campbell R., Liu, Yan, and Heqing Zhu, 2014, ...And the cross-section of asset returns, working paper, Fuqua School of Business, Duke University.
- Hasbrouck, J. and G. Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16, 646-679.
- Hautsch, Nikolaus, 2012, *Econometrics of Financial High-Frequency Data*, Springer, New York.
- Hendershott, T., Jones, C., and A. Menkveld, 2011, Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1-33.
- Hendershott, T. and A. Menkveld, 2014, Price pressures, *Journal of Financial Economics* 114(3), 405-620.
- Hirschey, Nicholas H., 2013, Do high-frequency traders anticipate buying and selling pressure?, IFA Working paper, London Business School.
- Hirshleifer David, Subrahmanyam Avanidhar, and Titman Sheridan, 1994, Security Analysis and Trading Patterns When Some Investors Receive Information Before Others, *Journal of Finance*, 49, 1665-1698.
- Huang, H., and Alec Kercheval, 2012, A generalized birth-death stochastic model for high-frequency order book dynamics, *Quantitative Finance* 12(4), 547-557.
- Jiang, J. G., Lo, I., Valente, G., 2013, High-frequency trading around macroeconomic news announcements: Evidence from the U.S. treasury market, working paper, The Bank of Canada, Ottawa, Canada.
- Jones, Charles M., 2013, What do we know about high-frequency trading?, working paper, Columbia Business School.
- Jovanovic, B., and A. Menkveld, 2013, Middlemen in limit-order markets, Working paper, VU University, Amsterdam.
- Kirilenko, A., Kyle, A., Samadi, M., and T. Tuzun, 2010, The flash crash: The Impact of High Frequency Trading on an Electronic Market, Working paper, Commodity Futures Trading Commission.

- Li, Ker-Chau, 1991, Sliced Inverse Regression for Dimension Reduction, *Journal of the American Statistical Association* 86, 316-327
- Lou Dong, Polk Christopher, and Skouras Spyros, 2014, A Tug of War: Overnight Versus Intraday Expected Returns, working paper, London School of Economics
- Menkveld, Albert J., 2012, High frequency trading and the new-market makers, working paper, VU University of Amsterdam.
- Menkveld, Albert J., and Bart Z. Yueshen, 2013, Anatomy of the flash crash, working paper, VU University of Amsterdam.
- Menkveld, Albert J., and Marius Zoican, 2014, Need for speed? Exchange latency and liquidity, working paper, VU University Amsterdam.
- Stiglitz, Joseph E., 2014, Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy?, Presentation at the Federal Reserve Bank of Atlanta, 2014 Financial Markets Conference, Atlanta, GA.
- Storey, John D., 2002, A direct approach to false discovery rates, *Journal of the Royal Statistical Society* 64, 479-498.
- U. S. Securities and Exchange Commission, 2014, Equity Market Structure Literature Review - Part II: High Frequency Trading, Staff of the Division of Trading and Markets, March 18, 2014.
- Van Kervel, Vincent, and Albert J. Menkveld, 2015, High-Frequency Trading around Large Institutional Orders, Working paper, VU University, Amsterdam.
- Van Ness B., Van Ness R., and Warr, R., 2001, How Well do Adverse Selection Components Measure Adverse Selection?, *Financial Management* 10, 77-98.
- Wald, Abraham, 1945, Sequential Tests of Statistical Hypotheses, *Annals of Mathematical Statistics* 16 (2), 117-186.
- \_\_\_\_\_, 1948, Optimum Character of the Sequential Probability Ratio Test, *Annals of Mathematical Statistics* 19 (3), 326-339.
- Ye, Mao, Yao, Chen, and Jiading Gai, 2013, The Externalities of High frequency Trading, working paper, University of Illinois - Urbana-Champaign.



**Figure 1. Price paths and anticipatory traders**

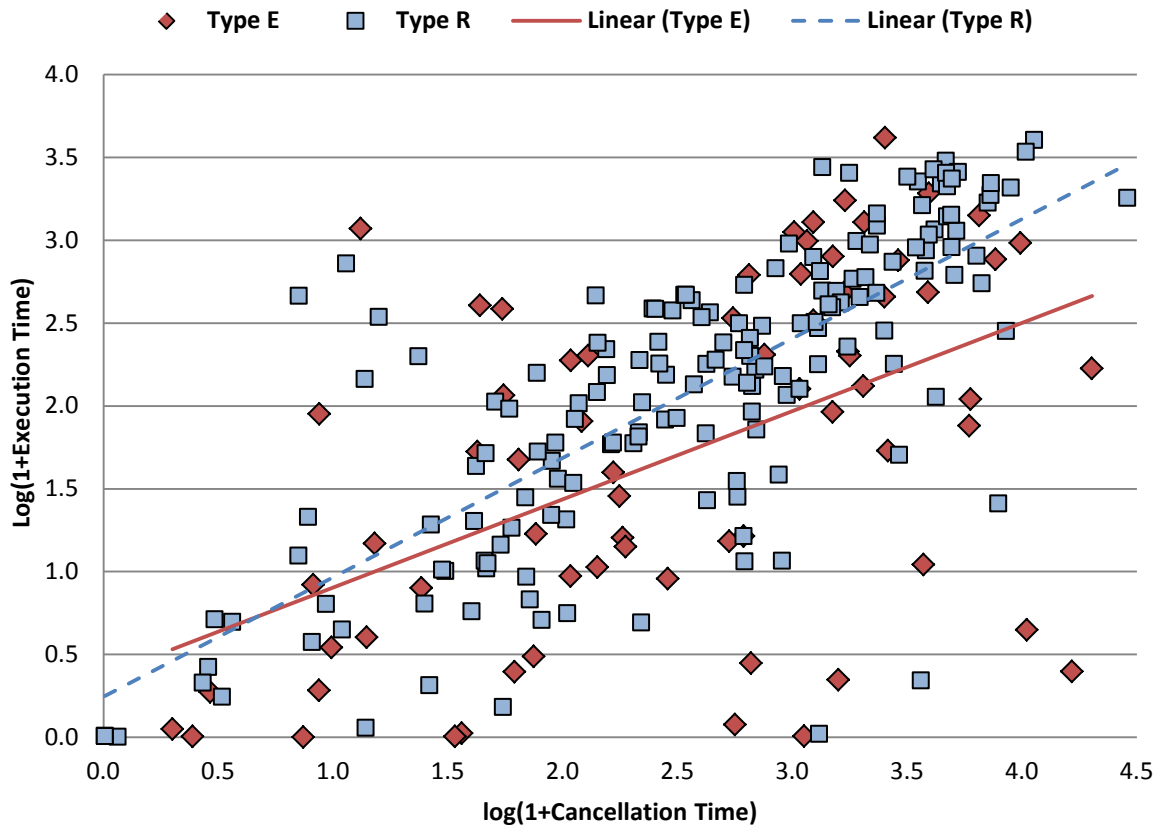
This figure illustrates a sequence of trades (the “x’s”) from participant order flow. When new information about the value of the asset arrives or liquidity demand changes, the market price reacts until the information is impounded in the price structure or new liquidity arrives to resolve the imbalance. Non-price moving liquidity based trading is shown as trades along a bid-ask bounce sequence (shaded areas). Price reversals occur at the specified local price extrema (the circled trades). A participant that processes the skills to process order flow and trade information to systematically forecast the short-term direction of trade prices is known as a Type E anticipatory traders. A participant who processes new information or market flow signals to systematic forecast the subsequent reversal in local trade price trend is known as a Type R anticipatory trader.



**Figure 2. Average Execution and Cancellation Speeds**

Average execution and cancellation speeds are computed for all sample participants using order book data. Cancellation time equals the difference between the time the Nymex received a cancel message and the initial order confirmation time. Execution time equals the difference between the time the Nymex confirms the trade message and the initial confirmation time of the order. These measures are computed in seconds (and milliseconds) and averaged for each participant. The data are plotted in the figure after transforming by the logarithm of one plus the average. A linear trend line ( $y = 0.600x + 0.288$ ) with an R-squared of 31.4% and a constant term and slope coefficient significant with p-values  $<0.001$  is shown for reference.





**Figure 3. Type E and Type R Participants' Average Execution and Cancellation Speeds**

Average execution and cancellation speeds are computed for the Type E (diamond shapes) and Type R (square shapes) participants. The method of calculation is the same as in Figure 2. The data are plotted in the figure after transforming by the logarithm of one plus the average. Linear trend lines from log-log regressions are also shown for reference. The Type E trend line is  $y = 0.544x + 0.348$  with an R-squared of 26.4%, insignificant constant term, and a p-value less than 0.001 for the slope coefficient. The Type R trend line is  $y = 0.724x + 0.239$  with an R-squared of 56.7%, insignificant constant term, and a slope coefficient with a p-value of  $<0.001$ .

**Table 1**  
**Daily Characteristics of Sample Data**

Summary statistics are shown for all messages and orders in the WTI crude oil contract expiring in December 2011. The sample covers all trading days from September 12 to November 18, 2011. The table provides average, median, standard deviation, minimum and maximum statistics across sample days for volume, number of traders (algorithmic or manual order entry methods), message types (entry, modification, or cancellation), and order types (market, limit, or stop). The sample contains 20,977 unique participants with trades during the period.

Characteristics	Average	Median	Standard Deviation	Minimum	Maximum
Volume	356,645	276,986	207,236	58,718	860,810
Number of Participants	<u>2,939</u>	<u>1,672</u>	<u>1,984</u>	<u>587</u>	<u>6,442</u>
<i>Algorithmic Traders</i>	399	263	252	62	779
<i>Manual Traders</i>	2,540	1,444	1,733	525	5,663
Order Message Type	<u>2,091,518</u>	<u>1,252,017</u>	<u>1,575,044</u>	<u>24,880</u>	<u>5,728,985</u>
<i>Entry Message</i>	744,905	428,782	569,240	2,906	2,031,921
<i>Modification Message</i>	450,684	265,522	401,582	1,611	1,471,517
<i>Cancellation Message</i>	636,707	399,296	454,444	1,516	1,698,800
Order Type	<u>2,091,518</u>	<u>1,252,017</u>	<u>1,575,044</u>	<u>24,880</u>	<u>5,728,985</u>
<i>Market Order</i>	10,858	3,405	11,540	411	35,042
<i>Limit Order</i>	2,027,329	1,194,546	1,555,986	23,853	5,674,858
<i>Stop Order</i>	10,603	779	12,550	77	30,695

**Table 2**  
**Local Price Path Characteristics**

Summary statistics are shown for all local price paths in the December 2011 WTI crude oil contract. The paths are computed from intraday local price path data. The sample covers all trading days from September 12 to November 18, 2011. The table summarizes information by month and path direction. The table shows average and median path returns, average path duration in seconds, average path volume, average number of trades, and the average number of unique participants in a path.

Month	Path Direction	Total Path Count	Avg. Path Return (%)	Median Path Return (%)	Average			
					Path Duration (sec)	Path Volume	Number of Trades	Number of Unique Participants
September	down	611	-0.346	-0.291	481.2	153.1	102.7	27.4
	up	608	0.337	0.279	390.6	138.7	92.1	26.2
October	down	4,085	-0.145	-0.114	96.0	302.3	236.6	90.9
	up	4,090	0.148	0.113	89.6	269.3	210.5	83.6
November	down	3,207	-0.114	-0.094	70.9	294.1	237.0	97.8
	up	3,218	0.115	0.094	65.3	269.6	215.2	90.6

**Table 3**  
**Buy-Sell Frequencies on Local Price Paths**

The table shows the fraction of all trades on the buy side by Type E, Type R, and all other participants in four volume segments along the local price paths and by months in the sample. The four segments correspond to the first 10 percent of path volume, the next 40 percent to the volume midpoint, then the next 40 percent to the 90 percent level, and finally the last 10 percent of path volume. Panel A shows results for upward trending price paths and Panel B shows the same results for downward trending price paths. In the first 10 percent of volume, Type E traders are expected to disproportionately buy in upward trending paths and sell in downward trending paths. In the last 10 percent of volume, Type R traders are expected to sell in upward trending paths and buy in downward trending paths, anticipating the change in local price direction in the next path. Trade counts by month and participant type over the sample are shown at the bottom of the table. These counts are by participant, so they will include both the buy and sell sides in the totals.

Volume Bins	Type E Participants			Type R Participants			All Other Participants		
	Sept.	Oct.	Nov.	Sept.	Oct.	Nov.	Sept.	Oct.	Nov.
<i>Panel A: Price Paths Trending Upward</i>									
First 10%	0.94	0.85	0.82	0.61	0.72	0.75	0.48	0.50	0.50
10% to 50%	0.41	0.21	0.20	0.26	0.26	0.31	0.50	0.49	0.49
50% to 90%	0.05	0.07	0.07	0.11	0.11	0.12	0.51	0.51	0.51
Last 10%	0.15	0.05	0.04	0.10	0.05	0.05	0.48	0.53	0.55
<i>Panel B: Price Paths Trending Downward</i>									
First 10%	0.38	0.10	0.12	0.32	0.19	0.20	0.50	0.50	0.50
10% to 50%	0.37	0.75	0.75	0.69	0.62	0.64	0.48	0.51	0.51
50% to 90%	0.49	0.91	0.90	0.82	0.84	0.86	0.48	0.49	0.49
Last 10%	0.65	0.94	0.87	0.87	0.89	0.92	0.51	0.47	0.47
Sample Trade Counts	431	22,218	32,812	9,899	88,212	69,628	286,422	3,590,797	2,843,233

**Table 4**  
**Electronic Order Submission and Speed by Participant Type**

The table shows order book statistics for Type E, Type R, and all other participants. Panel A shows how many messages are flagged as being entered by an automated trading system (ATS) and reports the overall ratio of such messages within each group. Panel B shows the distribution of different kinds of messages entered by Type E, Type R, and all other participants. Panel C shows counts based on two measures of trading speed. The first is execution speed, measured as the difference between the time the Nymex confirms a trade execution and the time it receives the initial order from the participant (in seconds with fractional milliseconds). The second is individual decision speed measured by cancellations. Decision speed equals the difference between the time the NYMEX receives a cancellation message and the confirmation time of the original order (in seconds with fractional milliseconds). Counts are shown for the number of participants whose average speed is less than 0.5 seconds.

Characteristic	Type E Participants	Type R Participants	All Other Participants
<i>Panel A: Message Statistics</i>			
Number of ATS Messages	30,196	3,990,180	89,107,276
Number of Manual Messages	31,961	319,010	6,914,740
Total Number of Messages	62,157	4,309,190	96,022,010
ATS Order Ratio	48.58%	92.60%	92.80%
<i>Panel B: Distribution of Messages as Percentage of All Messages</i>			
Order Entry	5.86%	19.43%	37.14%
Modification Message	27.86%	59.31%	20.32%
Cancellation Message by Trader	0.98%	16.54%	31.75%
Order Execution Message	65.07%	4.70%	9.74%
- ATS Executions	18.37%	2.83%	5.81%
- Manual Executions	46.70%	1.87%	3.93%
Other Messages	0.22%	0.01%	1.06%
<i>Panel C: Execution and Decision Speeds</i>			
# with Avg. Execution Speed < 0.5 sec	18	19	5,601
# with Avg. Decision Speed < 0.5 sec	0	3	75





**Table 7**  
**Characteristics of Anticipatory Traders**

This table uses the inverse regression technique to infer characteristics of the Type E and Type R traders relative to the entire sample of participants. In the Type E (Type R) results, the dependent variable equals 1 if the trader is identified by the FDR methods as significant in the first 10% (last 10%) of the local price paths. The independent variables are averages by participant across all trades in the sample. Execution speed is measured as the difference between the time the Nymex confirms a trade execution and the time it receives the initial order from the participant (in seconds with fractional milliseconds). The speed variable is the natural logarithm of one plus the execution speed. The coefficients reported are transformed from initial least squares estimates following Fishe and Smith (2012). These coefficients measure the difference between the characteristic of the Type E (or Type R) trader and what would be expected given all of the other characteristics in the sample as defined by equation (4). The sample is defined to be those participants tested using the FDR method, so all traders have 30 or more transactions. Sample sizes are reduced when participants are not observed in the first or last 10% of path volume. The p-values of the underlying coefficient are shown in parentheses beneath each estimate.

Characteristics	Sample Average over Participants	Type E Participants	Type R Participants
Average Speed during the First 10% of Path Volume	1.246 sec	-0.264 (0.001)	
Average Speed during the Last 10% of Path Volume	1.348 sec		0.069 (0.161)
Average Speed over All Trades	1.255 sec	0.014 (0.818)	0.182 (0.000)
Average Trade Size	1.275	0.285 (0.000)	0.051 (0.120)
Percentage of Aggressive Trades	55.8%	-0.052 (0.026)	-0.072 (0.000)
Percentage of Trades on the Buyside	50.2%	-0.007 (0.647)	0.014 (0.250)
Binary Indicator for Proprietary Trader	9.8%	0.033 (0.214)	0.069 (0.001)
Binary Indicator for Algorithmic Trader	18.3%	0.004 (0.911)	0.055 (0.037)
R-Squared		2.1%	4.2%
Sample Size		6,886	7,208



**Table 8**  
**Do Anticipatory Traders Bring New Information to the Market?**

To test whether Type E or Type R anticipatory traders appear to bring new price information to the market, the price change in each bin relative to the absolute price change in the path is computed specific volume bins, and for all volume bins combined. Volume is equally divided into ten bins for each price path. Averages of these data are reported for different cases based on whether Type E or Type R participants are trading in a given volume bin in a price path. Panel A computes averages without regard to whether the other type of anticipatory trader is in the path, too. Panel B excludes paths in which the other trader is found in the bin#1 for Type R averages or Bin#10 in the last path for Type E averages. The first two columns report averages across all bins separating the averages for relative price changes by "upward" and "downward" trending paths, respectively. The remaining columns also average the data by upward and downward paths, but only use specific volume bins in the calculation. The figures shown in parentheses beneath these averages are p-values for the t-test of comparing two means (unequal variances).

Path Characteristics	Means of Sample Data - Relative Price Changes by Bin							
	Up	Down	Up Price Paths			Down Price Paths		
	All Volume Bins		Bin#10	Bin#1	Bin#2	Bin#10	Bin#1	Bin#2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: All paths including any with Type E and Type R traders overlapping								
No Type E in Current Path Volume Bin#1	0.095	-0.096		0.262	0.116		-0.245	-0.110
Type E in Current Path Volume Bin#1	0.094	-0.109		0.282	0.106		-0.250	-0.088
p-value for t-Test of Diff.	(0.916)	(0.063)		(0.564)	(0.656)		(0.753)	(0.183)
No Type R in Previous Path Volume Bin#10	0.094	-0.098	0.154	0.261		-0.125	-0.241	
Type R in Previous Path Volume Bin#10	0.096	-0.092	0.150	0.265		-0.138	-0.251	
p-value for t-Test of Diff.	(0.363)	(0.004)	(0.588)	(0.593)		(0.057)	(0.159)	
Panel B: Type E test excludes cases with Type R in Bin#10 and Type R test excludes cases with Type E in Bin#1 on the next path								
No Type E in Current Path Volume Bin#1	0.088	-0.090		0.259	0.112		-0.244	-0.105
Type E in Current Path Volume Bin#1	0.084	-0.102		0.272	0.116		-0.236	-0.074
p-value for t-Test of Diff.	(0.759)	(0.110)		(0.629)	(0.917)		(0.740)	(0.125)
No Type R in Previous Path Volume Bin#10	0.094	-0.098	0.155	0.267		-0.125	-0.255	
Type R in Previous Path Volume Bin#10	0.096	-0.091	0.150	0.256		-0.136	-0.232	
p-value for t-Test of Diff.	(0.348)	(0.005)	(0.526)	(0.136)		(0.093)	(0.001)	

**Table 9**  
**Volume Effects of Anticipatory Traders**

This table estimates VAR models of volume and price using bin data derived from the sample price paths. Prices and volume are differenced for stationarity, although volume is generally stationary without differencing. Delta price uses end-of-bin trade prices and delta volume is the change in bin volume for all traders excluding Type E and Type R traders. These VAR models are estimated with one lag, but ten lags produced similar results. The exogenous variables are delta volume for either Type E or Type R traders and dummies variables for whether Type E participants traded in Bin#1 of the current path or Type R participants traded in Bin#10 of the previous path. Results are shown using all sample bins and for Bin#2 only for Type E tests and Bin#1 only for Type R tests. The cumulative impulse responses are the coefficients on the associated variable in each model. The significance of these estimated coefficients is indicated by a sign for the 5% level and a double sign for the 1% level. An 'ns' indicates that the coefficient was not significant. The Akaike information criterion (AIC) is shown for model comparison.

Exogenous Variables	'X' = Type E Participant and 'Z' = #1				'X' = Type R Participant and 'Z' = #10 Last Path			
	All Bins		Bin#2 Only		All Bins		Bin#1 Only	
	$\Delta P_t$	$\Delta V_t$	$\Delta P_t$	$\Delta V_t$	$\Delta P_t$	$\Delta V_t$	$\Delta P_t$	$\Delta V_t$
<u>Model A:</u>								
Delta(Type 'X' Volume in Bin)	-	--	--	--	ns	++	ns	++
- Cumulative Impulse Response	-0.00	-0.29	-0.00	-1.08	ns	0.33	ns	3.64
- Akaike Info. Criterion	-0.315		-3.185		-0.316		2.096	
<u>Model B:</u>								
Delta(Type 'X' Volume in Bin)	-	--	--	--	ns	++	ns	++
- Cumulative Impulse Response	-0.00	-0.29	-0.00	-1.03	ns	0.30	ns	3.36
Dummy for Type 'X' in Bin 'Z'	ns	++	--	ns	ns	--	ns	--
- Cumulative Impulse Response	ns	4.98	-0.01	ns	ns	-4.57	ns	-10.96
- Akaike Info. Criterion	-0.315		-3.187		-0.323		2.091	



