Do Cancellations Increase Trading Cash Flows?

Raymond P. H. Fishe

and

Esen Onur*

First Draft: March 2025

ABSTRACT

Cancellations are the most common message type after new order submissions. Using E-mini futures data, we identify cases when participants reveal that adverse selection costs encourage cancellations. Using an algorithm to identify whether cancelled orders would execute if left on the book, we show that participants avoiding adverse selection increase cash flows on average between $1.23 and $4.19 per contract relative to executing a prior cancel order and not altering price. These results vary by whether new orders are placed before or after the cancellation of an existing order.

Keywords: Adverse Selection, Cancelled Orders, E-Mini Futures

JEL classification: G10, G13

* **Fishe**: Professor of Finance (emeritus), Department of Finance, Robins School of Business, University of Richmond, Richmond, VA 23173. Tel: (+1) 804-287-1269. Email: pfishe@richmond.edu. **Onur**: U.S. Commodity Futures Trading Commission, Washington, D.C. 20581. Tel: (+1) 202-418-5000. Email: eonur@cftc.gov. We thank members of the Office of the Chief Economist 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 Esen Onur, who is a CFTC employee, in their official capacities with the CFTC. The paper was cleared for public distribution by the Office of the Chief Economist. 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. The analyses and conclusions expressed in this paper are those of the author(s) and do not necessarily reflect the views of other Commission staff, the Office of the Chief Economist, or the Commission. All errors and omissions, if any, are the authors' own responsibility. Corresponding author: Esen Onur.

Do Cancellations Increase Trading Cash Flows?

I. INTRODUCTION

The presence of informed order flows exposes traders to adverse selection risk, which is a compelling reason for traders to reduce their time on the book with cancellations.[1] However, the Dodd-Frank legislation of 2010 made it illegal to submit an order without an 'intent' to trade. Thus, market participants with many cancel messages relative to actual executions are at risk of being charged with 'spoofing'.[2] In this new environment participants are expected to be more calculating about when and how often they cancel orders.

There are many good reasons for cancelling an order. In addition to adverse selection risks, Hasbrouck and Saar (2009) observe that traders may cancel to chase a price, seek an immediate execution, and/or search for hidden liquidity or other information. Each of these strategies has implications for what a participant will do after cancelling and hence what may happen to cash flows.

If adverse selection costs are the motive for a cancellation, then participants may reason that executing at the cancel price will lower cash flows relative to a subsequent execution. Alternatively, if participants are chasing a price—a momentum-type strategy—then actual executions are likely to result in lower cash flows than what would arise at the cancel price. Specifically, when chasing a price, they would be expected to submit another order at a higher or lower price depending on whether they are buying or selling. If that order executes, they would have bought at a higher price or sold at a lower price and be worse off relative to the cancel price. The "seeking an immediate execution" motive may produce a cash flow result like

[1] The growth of high frequency trading has resulted in faster cancel speeds. Using pre-Dodd-Frank data, Hasbrouck and Saar (2009) find that approximately 36% of limit orders are cancelled within two seconds of submission. Using data from 2018, McInish, et al. (2020) find that about one-half of new quotes are on the book for less than one-half a second on average.

[2] Recent legal cases involving spoofing include U.S.A v. Michael Coscia, 14 CR 551; CFTC v. Oystacher, et al., 15 CV 9196; and U.S.A. v. Navinder Singh Sarao, 15 CR 75. Michael Coscia was the first trader in the U.S. to go to trial for spoofing. He was found guilty and sentenced to three years in prison. Igor Oystacher settled his spoofing case with the CFTC and agreed to pay a fine of $2.5 million. Navinder Sarao pled guilty to spoofing after being extradited to the U.S. and was penalized over $500,000.

chasing a price as the participant pays the half-spread to whoever is at the top of the book. Lastly, if participants are searching for hidden liquidity or other information, they may continue the effort by submitting another order that is likely to be cancelled in the near term.

Using regulatory data on the S&P500 E-mini futures contract, which contain all messages submitted for this contract, we investigate how cancellations affect cash flows by examining the contrapositive situation. That is, the situation in which the cancelled order is allowed to stay on the book to determine if subsequent prices indicated it would execute.[3] With this information, we can calculate the cash flow from the cancelled order. To this end, we sort the actions of traders into three buckets based on their order flows. The first bucket includes traders who cancel an order and then submit another order on the same side (buy or sell) at either a higher sell price or lower buy price, which then executes. These traders are acting as if adverse selection costs have increased, so executing the cancelled order is expected to reduce cash flows when positions are settled at the end-of-session price.[4]

The second bucket includes cases for traders who cancel and then submit another order on the same side but at a lower sell price or higher buy price. These participants appear to be momentum-like traders who are chasing the market, so executing at the cancel price is expected to increase cash flows relative to the actual execution of the next order. This bucket may also include those who seek an immediate execution. The third bucket includes cases where an order is cancelled, and the subsequent order executes at the same price as the cancelled order. This bucket may include those who have been searching for liquidity and find it with their next order or simply those seeking information about order flows on the book. This last bucket acts as a control group for most of our analyses.

[3] Cancelled orders are only executed in this analysis when we can show that future execution prices imply that they would have, in fact, reached the top of the book and found an execution.

[4] This interpretation also overlaps with the submission strategy in Harris (1998), where participants may optimally select a limit price when prices follow a stochastic process. As time and prices evolve, the limit decision is revisited, and a new optimal limit price may arise, so the order is cancelled and a new order submitted. Similarly, Fishe, Haynes and Onur (2017) develop sequential probability methods to identify directional trends in intra-day prices for WTI futures contracts. Such methods identify hundreds of short-term directional changes in a typical trading day, encouraging short term traders to cancel-and-resubmit orders.

A feature of the bucket approach is that the participants' reveal by their actions when adverse selection may be a concern in the market. This contrasts with spread decomposition models or methods that estimate toxic order flows (Van Ness, et al. (2001) or Easley and O'Hara (1987)). In these models, estimates of the adverse selection risk applies to the entire market. In our approach, however, only some participants may have detected that risk and reacted, which is what is found by Fishe and Robe (2004) in a study using data from an actual insider trading case.

To create these buckets, we examine the order flow sequence of market participants for twenty-four days between October 22, 2018 – November 23, 2018, excluding the Thanksgiving holiday on November 22, 2018. These account-level data show all messages (orders, trades, cancels, and modifications) submitted by every participant in the market. We also know characteristics of an order—limit, marketable limit, or stop order. We identify if an account is an algorithm or a manual-entry participant.[5] There are over 218 million messages in these sample data, of which approximately 85 million are new order messages.

Our sample shows that participant order flow sequences may be idiosyncratic. Specifically, some participants submit orders as packet sequences. That is, they submit a new order, then either cancel, modify, or observe an execution for that order. After a resolution, they submit another order, possibly with a time delay, and the process repeats. We call these "in sequence" orders. The idiosyncratic participants often submit multiple new orders in contiguous time, then consider actions to take for each of these orders resting on the book—cancel, modify, wait for an execution, or submit another new order. We call these "out of sequence" orders. These out-of-sequence orders are often at the same price points or sides, but less often at different quantities. Such idiosyncratic behavior may serve the purpose of executing a larger number of contracts without revealing the total order size. However, such idiosyncratic behavior makes participant motives less transparent.

[5] The sample data provide an identifier (TAG50) that is unique to the individual or team submitting orders for an account. A team uses an automated trading system (ATS) while an individual may use manual submission or an ATS system. A flag is provided to identify if an order is submitted manually or via an ATS. See CME Group Advisory Notice RA1805-5 (2018).

To help isolate participant motives and to populate the buckets, we sacrifice information from selected message sequences. Instead of trying to determine what a contiguous sequence of new orders or submit-and-cancel orders means, particularly if they are on the same side, price and quantity, we focus on cases in which a cancelled order is followed by a new order that subsequently executes. The cancelled order may be part of a sequence or not, but it is the last action of the participant before an execution is observed for that participant's next order. This provides a subset of data that include executions closely tied to a cancel decision. We then connect the characteristics of both the cancelled order and the execution message to populate the buckets described above. The sorting into these buckets is not exhaustive, but it does contain approximately 20.5% of the orders that have executions, which amounts to over 5.86 million cancel-execution pairs, so they are expected to provide an indication of the level of support for the various cancellation strategies.[6]

Our sample shows that nearly 21% of cancels by manual-entry traders are indicative of chasing the market and approximately 13-15% are indicative of adverse selection fears. However, algorithmic traders behave differently. We track whether the next new order in the message sequence arises after or before the previous order is cancelled. For manual entry results, the sequencing makes only a minor difference, but for algorithms it is meaningful if the algorithm re-submits before the cancel message. We find that approximately 5.5% of cancels by algorithmic traders are indicative of chasing the market when the next order is in time sequence, but this jumps to 9.5% when the next order is placed before the cancel message. Similarly, approximately 3.0% of cancels by algorithmic traders are indicative of adverse selection when the next order is in time sequence, but this increases to 7.1% when the next order is placed before the cancel message. What these results suggest is that the sequencing of new orders and cancel messages provides additional information on the purpose of these cancels.

[6] Because of the pairing of cancellations and executions we lose many new orders that cancel and are followed by other new orders that also cancel. We also lose new orders that execute and are followed by other new orders that execute, or by cancels that are not followed by executions. Many of these cases appear to be participants who are acting as market makers.

We calculate the effects of cancels on cash flows across all cancels; cancels that are in sequence and out of sequence. Our approach is simplified because we avoid re-building the book after each message.[7] Instead, we ask whether the cancel price and quantity would have executed based on future minimum and maximum execution prices. Thus, we know whether leaving a cancelled order on the book would have resulted in an execution.[8] We call these *pseudo-cancel* executions or just *pseudo* executions. Any outstanding positions from trades or pseudo executions are settled at the end-of-session price and combined with trading cash flows to compute end-of-session net cash flows for cancel-execution pairs in the sample. The daily regressions use these cancel-execution pairs to analyze the effects of adverse selection and momentum buckets by comparing execution to cancel cash flows.

Our regressions show that across 24 days, cancellations motivated by adverse selection risks increase net cash flows per contract by $1.23 for the buy side and $4.19 for the sell side relative to the bucket in which the cancel-execution pair does not change price and the data are in time sequence. The signs here are as predicted. The momentum bucket results in an average increase of net cash flows per contract of $1.41 for the buy side and a decrease of $2.46 for the sell side relative to the bucket in which the cancel-execution pair does not change price. The former positive result is unexpected, but the latter negative value what we have predicted. The out-of-sequence results are similar for the adverse selection variable above but somewhat higher in absolute value for the momentum results. We also compute results using daily regressions, and these are significant and consistent for adverse selection effects but are more mixed for momentum variables.

The remainder of this paper proceeds as follows. Section II provides a discussion of the previous literature on adverse selection costs and cancellation decisions. Section III discusses how we compute net trading revenues and determine whether a cancelled order would have executed during the daily trading

[7] In theory, our approach may be applied using non-proprietary data, which usually identify book depth at active quote levels, best quotes and trades, so that future execution prices identify whether depth is exhausted at previous quote levels.

[8] Cancels are assumed to be atomistic in this analysis, so they do not affect the course of inflows or outflows on the book at any future price levels.

session. Section IV explains the data analyzed and provides details on the variables and filters used to create our sample. Section V explains our regression results and Section VI offers conclusions.

II. Previous Literature

This research is related to various areas in market microstructure. Our work connects to a large literature on limit-order strategies in markets. Copeland and Galai (1983) offer a characterization of the limit order option within a dealer model of market making. The option value arises because the security price may subsequently change to the disadvantage of the posted limit price, creating adverse selection. Participants may reduce adverse selection costs by cancelling (or modifying) limit orders and submitting new orders at a more preferred price. Parlour and Seppi (2008) provide an extensive review of the limit-order theory literature, which helped motivate empirical efforts to measure adverse selection costs.

Specifically, adverse selection costs are expected to be a component of the bid-ask spread, increasing the size of the spread when such costs increase. Spread decomposition models were developed to estimate such effects, although early models appeared to provide widely varying estimates of this component.[9] More recent models based on structural arguments offer possibly better methods of estimating the size of adverse selection costs (Dahlström, Henricson and Nordén, 2016).

An alternative to these component models is the efforts started by Easley and O'Hara (1987) and Easley, de Prado, and O'Hara (2012) to estimate the toxicity of order flow. The PIN and VPIN models have generated extensive critical research on how to detect informed trading, measure order flow toxicity, and determine whether it affects return volatility and liquidity. Similar to component models, there are disagreements on the implementation and interpretation of these models (e.g., Andersen and Bondarenko (2014), Lai, Ng, and Zhang (2014), Griffin, Oberoi, and Oduro (2021), and Yang and Xue (2021)).

Our approach generally complements these earlier efforts to measure adverse selection risk but uses the revealed preference of participants—that is, limit price changes—to indicate whether adverse selection

[9] Van Ness, Van Ness, and Warr (2001) compare five spread decomposition models. They find that adverse selection estimates are not consistent across models and are only weakly related to other proxies for informed trading.

costs may be a factor in their decisions. With this account-level approach we avoid methods that sweep the extant market into a single measure of adverse selection risk. We do not know if all market participants at a given time know about informed order flow, but we infer from the actions of selected participants that they may have detected informed trading.

Although intuitively correct, there is little evidence on whether cancellations avoid adverse selection because account-level order book data are proprietary. An exception is Fishe and Robe (2004), who examine illegal insider trades in both dealer and specialist markets. They report minimal spread impacts in dealer markets but find depth impacts in both market types. Their evidence suggests that dealers and specialists are fading order sizes when adverse selection risk appears to increase. A parallel finding in current order-driven markets would suggest that cancellations would be used when participants detect greater adverse selection risk.

Other efforts study the decision to cancel an order. Using data on limit orders, executions and cancellations for a sample of 116 stocks traded on Nasdaq in 2011, Subrahmanyam and Zheng (2016) find that HFT firms use cancellations to manage their limit orders ahead of short-term price movements versus non-HFT firms. These results suggest that HFT firms may may cancel orders prior to a short-term adverse price movement. In addition to Hasbrouck and Saar (2009), the research by Liu (2009) and Dahlström, Hagströmer, and Nordén (2024) investigate the costs and benefits of retaining a limit order on the book. The model by Liu (2009) creates cancellations when information arrives to affect the fundamental value, which favors either the buy- or sell-side. Traders with potentially losing quotes react with cancellations. In the Dahlström, et al. (2024) model, cancellations arise when a shock has undone the equilibrium of the limit order book (as modeled by Sandås (2001)). Their model suggests that cancellations are affected by changes in market depth and the relative position of an order on the book. Their empirical findings show that cancellations are primarily related to market-making decisions, which are usually conducted by algorithms.

Some studies view cancels as detrimental because they affect market quality—increase spreads and reduce depth—and the initial fleeting order may create a false sense of liquidity (Sornette and Van der Becke (2011); Van Ness, Van Ness, & Watson (2015)). Others have found that cancels are used by HFTs

in linked strategies—strategic runs—that may improve market quality (Hasbrouck & Saar, 2013). The standard view that cancels avoid adverse selection is also offered by Liu (2009) and Fong and Liu (2010).

We find that a large majority of cancelled orders by algorithms are re-submitted with the same terms, so depth returns quickly for fleeting orders. For in-sequence data, 72.4% of the cases have the same terms and for out-of-sequence data it is 60.2%.[10] Of these less than 2% are executable immediately, so the vast majority are using cancellations to change their execution priority by moving to the back of the book with a new order. This evidence suggests that many algorithms value the information gained when an order does not execute before being cancelled. That is, they know how long the order was on the book and the volume of trade at that limit price and other prices. This result parallels McInish, et al. (2020), who report that limit orders are more likely to be cancelled if they are closer to the top of the book and Fong and Liu (2010) who report that cancels are more likely if a limit order is near the best quotes.

III. METHODS

A. Cash Flows from Executions and Cancelled Orders

The approach to cash flows outlined here includes both executions and cancellations that would have executed if left on the book during the trading session. Consider the activity of a given trader using a strategy that involves limit orders, cancellations and possibly market orders. Denote by the pair (x_j, p_j) as the j^{th} order submission to trade quantity x_j at the limit price p_j. A buy is denoted by $x_j \in \mathbb{Z}^+$ and a sell by $x_j \in \mathbb{Z}^-$. A binary indicator variable z_j denotes whether the order executes.[11] If the j^{th} order is cancelled, a pseudo-order $(\hat{x}_j, \hat{p}_j)$ remains on the order book to determine whether subsequent trades would have executed against the cancelled order as it moves through the book. If the j^{th} order executes before cancellation, the pseudo-order quantity is set to zero ($\hat{x}_j = 0$).

[10] Baruch and Glosten ("Flickering Quotes," 2013) suggest that participants may cancel quotes quickly as part of a game where they reduce "undercutting" risks by resubmitting quotes at random price levels. There is almost no evidence of such randomness when the majority of cancelled quotes are re-submitted at exactly the same terms.

[11] The present discussion excludes the use of modification messages except that the price in a cancelled order is the last prevailing price for that order, so this price may have been generated by a modification message.

The binary indicator variable $\hat{z}_t$ denotes whether the cancelled order executes. To define pseudo executions, we examine if subsequent execution prices for limit orders on the same side as the pseudo-order are executed. We avoid rebuilding the book after each execution by creating vectors of future minimum buy prices (p_t^L) and future maximum sell prices (p_t^H), both of which are based on execution data. If at the point of a buy cancellation, the future minimum buy price is strictly less than the cancel price, then $\hat{z}_t = 1$ because the book depth at the cancelled buy price must be exhausted to avoid a trade-through rule violation. Similarly, for a sell cancellation, if the future maximum buy price is strictly greater than the cancel price, then $\hat{z}_t = 1$.[12] We have a potential problem only when the cancel price and the future comparison price are equal. Thus, we are not certain that the depth is exhausted at the cancel price. To address this issue, we report cash flow results for only the case where the cancel price is strictly greater or less than the future comparison price. The sample data show that only 2.8% of cancelled orders result in an ambiguous future execution, compared to the 95.8% of cancelled orders that would have executed if left on the book.

We set $\hat{z}_j = 0$ for any cancelled order that fails to execute before the end of the trading session. The scenarios above also rule out unfilled portions of a pseudo-order, but it is possible for a participant to have a pseudo inventory at the end of the day if there is an imbalance between their cancel buy and sell executions. The cash flow effect of this inventory is settled at the end-of-session settlement price (p^e), which is how any actual inventory is settled.

Figure 1 illustrates the identification of future minimum buy prices. The price path of buyside executions is shown as the main curve. Two prices are identified in the figure, $p^e$is the end-of-session settlement price, which corresponds to the last price on the curve, and p_1, the lowest traded price on the curve. The vector of minimum future prices will be defined by p_1 for all trades between the start of the session (t_0) and the absolute minimum price observed at t_1. Then as the trade price increases between t_1 and t_2, the minimum future price will change sequentially from p_1 to p^e, with the latter price observed at

[12] Note that if the cancelled order was for immediate execution, a variation of "fill or kill", and the pseudo-order quantity is greater than the comparison quantity, then $\hat{z}_j = 0$ as the order cannot be filled.

t_2. This is the shaded area in the figure. After the trade at t_2, the minimum future price remains constant at p^e. If new buyside limit orders followed the path of traded prices in the figure, then we would expect those cancelled between t_0 and t_1 to have executed because the future minimum buy price, p_1, would be less than the limit price. Those submitted along the path in the shaded area with limit price corresponding to the price path and cancelled during that interval would not be expected to execute. Lastly, those submitted at limits close to the path in the final interval and cancelled in that interval are expected to have executed as the last price, p^e, is expected to be the lowest buy price in that interval.

By including these pseudo executions, the end-of-session position of the participant equals the total size *submitted* on the buy side net of the total size *submitted* on the sell side.[13] This position is not equal to the net position determined solely from actual executions. The difference arises because all cancellations are followed to a pseudo execution when the pseudo-cancel reaches the top of the order book or if they do not execute then they are netted out when the end of the session is reached. The end-of-session net cash flow for a participant with regular and pseudo executions is as follows:

$$c = -\left[\sum_{x_j \in \mathbb{Z}^+} x_j p_j z_j + \sum_{x_j \in \mathbb{Z}^-} x_j p_j z_j + \sum_{\hat{x}_j \in \mathbb{Z}^+} \hat{x}_j \hat{p}_j \hat{z}_j + \sum_{\hat{x}_j \in \mathbb{Z}^-} \hat{x}_j \hat{p}_j \hat{z}_j\right] \equiv c^a + c^s \qquad (1)$$

where c^a is the cash flow of the actual executions given by the negative of the sum of the first two terms and c^s is the cash flow of the pseudo executions given by the negative of the sum of the last two terms. The initial negative sign corrects cash flows for the convention to use positive and negative trade sizes to denote buy and sell orders, respectively.

For a final calculation that accounts for end-of-session inventory, we mark-to-market the end-of-session net position at the settlement price:

$$c_e = -p^e \left[\sum_{x_j \in \mathbb{Z}^+} x_j z_j + \sum_{x_j \in \mathbb{Z}^-} x_j z_j + \sum_{\hat{x}_j \in \mathbb{Z}^+} \hat{x}_j \hat{z}_j + \sum_{\hat{x}_j \in \mathbb{Z}^-} \hat{x}_j \hat{z}_j\right] \equiv c_e^a + c_e^s \qquad (2)$$

[13] Note that $z_j = 0$ for outstanding orders at the end of the session, so the last order submitted in the session is not included in the submitted total unless it is cancelled or executes before the session ends.

where analogous to the above definitions, c_e^a is the cash flow of the actual net executions marked to market and c_e^s is the cash flow from any pseudo net executions marked to market. The difference, $c^a - c_e^a$, measures net trading revenues from actual executions plus the gain or loss on any outstanding net executions or rather end-of-session inventory. If $c^a - c_e^a > 0$ then the actual executions plus end-of-session net gains/losses increased the participant's cash flows. Equivalently, combining equations (1) and (2), if $c - c_e > 0$ then the combination of actual and pseudo executions creates a gain in net position value when marked to market. However, if $c^s - c_e^s > 0$ then the cash flow outcome from pseudo executions added to these gains (if any), so the decision to cancel was *not optimal* relative to retaining these orders on the order book. The empirical analysis summarizes of these calculations by participants across trading sessions. The regression analysis focuses on explaining $c^a - c_e^a - (c^s - c_e^s)$, when cancellations and executions are paired together. Specifically, do actions taken in formulating a new order that executes after a cancellation help to improve the net execution cash flows relative to the net cancel cash flows.

B. Cash Flows from Order Flow Buckets

The previous discussion focuses on all executions and cancellations for every participant over a daily session. However, each participant may react to market conditions as they perceive them during a trading session. As noted, they may engage in numerous possible strategies, such as market making, searching for the best price, chasing the market, reacting to adverse selection fears, or simply seeking a quick execution. Our goal is to identify order patterns that indicate actions consistent with each strategy—a bucket approach.

To implement the bucket approach, we isolate new orders that were cancelled and then followed by a new order that had an execution. This helps to balance the quantities involved in a participant's execution and pseudo execution order flow. Specifically, on average, participants will cancel more orders than those that execute. So, when adding up actual executions and pseudo executions to compute cash flows, there is a large quantity imbalance for new messages that execute versus those that are cancelled. This creates a relatively larger inventory imbalance at the end of the session on the cancel side.

The pairing of a prior cancelled order to an execution gives a more symmetric picture of cash flows and end-of-day inventory imbalances. To create the cash flow results by the various buckets, we collect the cash flow results for each of these cancel-new order pairs and mark any execution inventory or cancel inventory to market at the end of session price. This approach follows the formulae in equations (1) and (2). The cost of this simplification is that we are missing executions that are not preceded by cancellations and cancellations that are not followed by executions.

IV. SAMPLE DATA

A. Order Entry Dataset

As noted, the E-mini futures sample consists of 24 days in 2018 from October 22nd to November 23rd, excluding Thanksgiving on November 22nd. Each day contains between 8 to 14 million messages specifying instructions from participants (e.g., new orders, modifications of existing orders, or cancellations) or responses to participants from the order entry matching engine (e.g., executions, or system messages about an order). In total there are over 218 million messages in the sample.

We organized the sample data by first eliminating messages received prior midnight or after 4:00 p.m. (EST). This timeframe covers several hours of European trading and the hours where pre-market trading accumulates volume before the NYSE opens, plus the entire period of normal trading on the NYSE. [14] Although there is variation on each day, the time filter drops less than 2% of the message stream, which mainly occurs before midnight (EST). This filter left the sample with approximately 85.5 million unique new order identifiers. The next step removed orders that expired at the end of the trading session, such as day orders that were not cancelled, orders that contained only one message unless that was an execution, and orders that did not have a proper new order message, which implies that they were entered before our time window. This step removed approximately 4.3% of these messages. For the 24 days in the sample there remain over 81.8 million unique order identifiers with proper first messages (i.e., a new order).

[14] Bondarenko and Muravyev (2023) find that substantial price discovery in the e-mini futures market occurs during the earlier hours of European trading.

Table 1 offers selected details for the sample data. Of the 218 million plus messages, the most common is a new order (39.2%), followed by a cancel message (26.5%) and then an execution message, either a partial fill (9.3%) or a full fill (10.9%). Most of these orders are for proprietary accounts although approximately 10% are orders for customer accounts. Limit orders that rest on the book are the dominant order type at almost 98%, with market and stop orders comprising the rest. These limit orders are cancelled approximately 68% of the time, while executions—partial and/or full—are found in just over 30% of orders. Algorithmic-entry firms comprise slightly over 93% of order flow with the remaining being manual entry firms. Finally, the sample shows that 87% of orders contain only two messages, generally a new order and either a cancel or execution message, and that 97.1% of all orders contain five or fewer messages.

B. Sample Organization

Because these regulatory data provide a unique identifier for each participant, we organized the data by participants using two strategies: (1) sort messages by their timestamp for each participant identifier and (2) sort the time sequence of messages for a given order identifier, which were then blocked in time order for a given participant identifier. In most cases these strategies resulted in the same sequence of messages for a participant. That is, as shown in Table 1, most participants submitted only two messages: a new order message, which was followed by either a cancel or execution message. At which point the participant entered another new order and the process repeated. For the entire sample, there were 421,777 gaps between daily block sequences—sorted by participant ID, order ID, and time—for one participant's block group compared to a different participant's block group.

Figure 2 illustrates how message flows may differ based on participant/time sorting and participant/order identifier/block-time sorting. The message fields illustrated are new order (ORDR), cancel (CXL), modification (MOD), and execution (EXE). The top part of the figure shows three orders—Order ID #1, Order ID #2, and Order ID #3—for Acct#1. These are in in block-time order but would appear the same way in time order because this participant allows each order to resolve before placing a new order.

These are in-sequence messages because a new order message is entered after the end of the previous order sequence.

The bottom part of the Figure 2 illustrates order entry for Acct#2. This account has two orders: Order ID #4 and Order ID #5. The messages for these orders are sorted in participant-time order. This sorting shows that both new messages were entered contiguously so that the second new order message occurred before the ending message of the previous order: a cancel of Order ID #4. This result places the second order as out-of-sequence compared to the first order.

Note that in relation to our effort to create buckets that describe participant motives during the session, both Order ID #3 and Order ID #5 would be candidates for analyses. This is because in block time order they both follow cancel messages, and both have execution messages. However, we do make allowance for the different sequencing as Order ID #3 is in-sequence and Order ID #5 is out-of-sequence. Thus, if a participant submits a new order before a previous order has been cancelled (or executed) the motive for the participant's actions is less clear. Our regression analysis isolates these cases to determine if they provide similar results to the cases that follow an in-sequence time order.

The effects of participant/time and participant/order/block-time sorting in our sample are shown in Table 2. The sample data here are filtered to capture only new order records. Panels A-D report distributions of (lag) messages prior to a new order based on both time and block-time sorting. Panel A is limited to new orders that follow the timestamp of the last message in the previous order block when the new order is in-sequence relative to the end of the last order. Panel B is limited to new orders that do not follow the time sequence of the last message in the previous order block when block-time sorted. That is, the timestamp is earlier than the timestamp on the lag message. Panels A and B show lag messages from an overall time sorting. Panels C and D follow the restrictions in Panels A and B, respectively, but show lag messages for a block-time sorting of these orders.

When the data are found to be in-sequence (Panels A and C), most of the lag messages are either a cancel (approximately 64%) or an execution (approximately 35%). This is also true when they are out-of-sequence but sorted by participant-order identifier-block time as in Panel D. The exception arises when new

orders are out-of-sequence and the lag message is found after sorting by participant/time as in Panel B. Panel B shows that the majority of lag messages (62.8%) are new orders, so that when orders are considered out-of-sequence, it is most often that the new order follows another new order. In effect, the participants are stacking the submission of new order messages.

C. Price and Quantity Changes after Cancellations

Our focus is on whether cancellations help to improve participant cash flows. There are approximately 15 million more cancellations than executions in the sample, so we narrow our focus to the executions that are closely connected to a cancelled order. To do this, we first filter the sample data by identifying all new order messages that follow a cancelled order by the same account. We then identify all cases in which the new order for the same account had an execution. A breakdown of these results in reported in Table 3.

Table 3 provides information on four groups that are formed based on prior cancelled orders, new order sequencing, and whether the new order has an execution. New orders are either in sequence or out-of-sequence as previously discussed. The table reports how price and quantity changed relative to the cancelled order, which is the disposition of the previous order for that account. These statistics are shown for both algorithmic and manual accounts. The chi-squared test for independence between order entry type and either price or quantity changes is rejected in all cases with p-values less than 0.001.

Consider the price- and quantity-change results in Table 3. The in-sequence results for price changes show that manual-entry participants (98.7%) are more willing to alter prices in the next order compared to algorithmic-entry firms (38.2%). Similar results are found when the new order that follows a cancel has an execution. As a result, manual-entry participants show a greater likelihood of having an execution in the next order—49.2% versus 19.6% for algorithmic entry. The change in quantity results show a strong degree of stickiness. On average, 90.2% of manual-entry and 72.4% of algorithmic-entry participants retain the quantity posted on the canceled order in the new order submission. These percentages decrease slightly when the next order executes.

The out-of-sequence results offer similar findings for manual-entry participants but there are noticeable differences for algorithmic-entry participants. When new orders are out-of-sequence, algorithmic-entry participants are more likely to change price in the new order: 61.3% for all new orders and 73% for those new orders that have executions. However, this does not increase the chance of an execution compared to in-sequence orders for algorithmic entry. The chance of an execution decreases from 19.6% for in-sequence to 13.2% for out-of-sequence orders.

The timing of out-of-sequence orders and the nature of algorithmic coding may be a consideration. Manual traders may anticipate the decision to cancel the current order when submitting a new order out-of-sequence. The likelihood of execution for manual traders is still fairly high at 41.6%, but algorithmic traders must introduce such expectations in the coding, and often cancel decision require a trigger, such as time on the book or execution price changes, which is information that is not yet available. Thus, the lower chance of execution (13.2%) may reflect the algorithm's use of more limited information.

Table 4 offers additional support for the view that price changes may have improved the chance of executions for manual-entry traders. This table shows a cross-tabulation between price and quantity changes for the in-sequence data in Table 3. The use of the in-sequence data ensures that when the new order is submitted by an algorithm it is after the algorithm has cancelled the previous order. Even so, in over 60% of the cases with or without an execution, the algorithm does not change the price. In contrast, manual-entry participants change price approximately 94% of the time. In is reasonable then that these price changes result in the much higher chance of executions for manual-entry trades reported in Table 3.

C. Do Cancellations Execute?

This analysis requires knowledge of what would have happened to orders that were cancelled if they had remained on the order book. There may be no general equilibrium answer to this question because leaving the order on the book may have affected other participant's decisions to cancel their orders or submit other orders. Thus, our effort is necessarily qualified with the atomistic assumption that leaving any one order on the book has no effect on other order decisions.

Our approach is further simplified by not rebuilding the order book after each execution, order submission, cancellation, or modification. Rebuilding the order book is a problematic exercise but would provide two pieces of information: if the order would have executed *and* the time of such an execution. We give up the information on the time of execution so that we can use computing resources efficiently to determine if a cancelled order would have executed at *any* time during the remaining trading session. Our approach follows the description of Figure 1 (noted above).

The results for this approach showed that nearly all cancellations found a future execution. On the buy side, 98.2% of cancels were found to have a future execution. On the sell side, the fraction was somewhat lower at 95.8%. The reason for these results is that most new orders are placed near the current best quotes, so if an order is cancelled it may only require minimal future price volatility to pass the above test and confirm an execution. As a check, our method allowed us to compute the median time for cancels to equal the future minimum-bid or maximum-ask timestamps depending on the side submitted. On the buy side, the median was 35.8 sec and on the sell side it was 139.1 sec. As Figure 1 shows, the future minimum bid or maximum ask may be quite far away in the current trade stream. However, these median times while long in a high frequency environment, suggest that prices may be expected to move sufficiently within a few minutes to cause an execution.

D). Identifying Participant Motivation—The Buckets

The motive behind participant actions cannot be truly known, but we chose to infer likely motives from the actions taken after a cancelled order—a revealed preference approach. Table 5 reports how price and quantity changed relative to a cancelled order that preceded a new order message that had an execution. These changes are parsed by buy- and sell-side combinations, and by manual and algorithmic entry methods. All new orders are in sequence in this table.

The buy-buy and sell-sell pairs are the focus here for determining two obvious participant motives and associated buckets. If a participant seeks to chase a price (Bucket#1), then we expect that they will increase the price in the new order if they are buying and decrease the price if they are selling. Similarly, to avoid

adverse selection from toxic order flow (Bucket#2), we expect participants to decrease the price if they are buying and increase the price if they are selling.[15] Not changing price in these situations is more neutral and suggests other motives not directed at chasing the market or adverse selection concerns (Bucket#3).

Focusing on buy-then-buy and sell-then-sell orders in Table 5, we observe significant differences between algorithm-entry and manual-entry participants. For same-side decisions, manual-entry participants are more willing to either raise or lower prices in new orders compared to algorithmic-entry firms. For algorithms they appear to be chasing a price in about one-third of the cases, while manual traders seek a price in over 60% of these cases. Importantly, manual traders show a concern for possible adverse selection—lower price when buying and raise price when selling—in about 30% of cases, while algorithms only respond similarly in about 5.5% of cases. Arguably, algorithms have the potential to respond more quickly than manual-entry traders, so they may exhibit less fear of adverse selection problems. Alternatively, their speed advantage may allow them the first access to the market when new information arises, so they may create the adverse selection risk for the other participants.

Note that the stickiness of quantity changes here confirms the previous results in Table 3 that participants—both algorithmic and manual—stay with the same levels of quantity when submitting a new order after a cancellation. This is also true when conditioning on a price change as shown in Table 4. The most frequent order sizes are one or two contracts (median is two), so this size consistency may be chosen to neutralize any perceived information contained in the quantity submitted as suggested by Chakrabarty, Hendershott, Nawn, and Pascual (2017).

V. RESULTS

A. Overall Cash Flows

Table 6 provides a summary of the overall cash flows for the sample data. Both means and medians are shown for actual executions and for cancellations that would have executed if left on the book for the

[15] We also considered the actions of buy-then-sell or sell-then-buy as possible signals of adverse selection concerns. However, these signals are confounded by the behavior of market makers who routinely change sides to clear inventories.

trading session. These data are also filtered into cancel-execution pairs based on the previous discussion of in-sequence and out-of-sequence new orders. The sample sizes for these paired data are large, but much smaller than the full sample. In effect, we have excluded many cancellations that were followed by new orders that were also cancelled and many executions that were followed by new orders that may have been cancelled or had executions.

The average cash flows are shown for both intraday flows, end-of-session inventory settlement, and the net position after inventory settlement. These averages (and medians) are based on the calculations in equations (1) and (2) above. Two observations are apparent: the net average cash flows for both executions and cancellations tend to be small for both total cash flows and cash flow per contract, and the net average (or median) cash flows tend to be smaller for the in-sequence than the out-of-sequence sample. The first observation is likely due to the zero-sum nature of futures contracts, but the second is more of a puzzle. One might reason that placing a new order that eventually executes before cancelling an existing order (out-of-sequence) uses less information than placing the new order after the previous order is cancelled (in-sequence). These results suggest that while the information set differs, the timing factor may be more important for cash flow results, particularly if it is a fast-moving market.

Table 6 also shows that on average the net cash flows from cancellations that would have executed if left on the book are greater than the net execution cash flows. In terms of equations (1) and (2), the averages per contract show that $(c^s - c_e^s)/Q > 0$ and while $(c^a - c_e^a)/Q > 0$ the overall average indicates that $(c^s - c_e^s)/Q > (c^a - c_e^a)/Q$, so leaving the cancel on the book would have been a net benefit to the average participant. However, we do not know the strategy of the average participant and particularly how long it would take to execute the cancelled order, so the weight to place on this result is unclear.

The regression results that follow offer an approach to help sort out the different behavior of these participants, which allows us to measure whether actions to avoid adverse selection risk result in net benefits to these participants. In this approach, we model the relative difference between net execution and net cancellation cash flows. The intent is to isolate which participants are likely to gain from a cancellation and which are likely to lose, and test whether our expectations are correct.

B. Regression Analysis

The bucket approach described above provides guidance on what variables will affect net cash flows. Specifically, in the first bucket a participant seeks to avoid adverse selection risk, so we expect the participant to cancel an existing order and if they submit another order, then it will be on the books at a different price. Given the in-sequence and out-of-sequence pairing we have for cancelled orders and new orders that execute, a participant that cancels a buy order and then submits another buy order will reveal that adverse selection risk may be a motive if the new buy price is reduced relative to the cancelled price. Alternatively, if the cancel is a sell order and the new order is a sell order then we expect an increased sell price if adverse selection risk is a concern. We measure these two effects with regression variables labelled "AS_Effect_Buy" and "AS_Effect_Sell" respectively, and we expect that they will have a positive effect on net cash flows if cancels help avoid adverse selection risk.

For the second bucket, a similar argument can be made for momentum traders chasing a market price. These momentum variables are labelled "MO_Effect_Buy" and "MO_Effect_Sell" and are expected to have a negative effect on net cash flows. The third bucket collects the remaining participant behaviors and is used as a control group in our regressions. A complete list of the variables used in our regressions is shown in Table 7.

The dependent variable in these regressions is net cash flow from the execution of the new order in the cancel-new order pair minus the net cash flow for execution of the (pseudo) cancelled order had it remained on the book. This is computed on a per contract basis. As noted above, this represents $\frac{c^s - c_e^s}{Q} - \frac{c^a - c_e^a}{Q}$. The goal with this variable is to assess whether the bucket variables above have the expected effects. Specifically, do the adverse selection variables (AS_Effect_Buy and AS_Effect_Sell) tend to increase the net cash flow gains from an execution relative to what would have occurred had the cancel order executed. Similarly, we expect the momentum variables (MO_Effect_Buy and MO_Effect_Sell) to have the opposite effect; that is, to decrease the net execution gains relative to the net gains from a cancelled order.

Most of the variables in Table 7 are binary measuring the existence of a specific characteristic in the cancel-new order pairing. In addition to those noted above, we measure whether the cancel-new order pair changed sides (CN_Buy_Sell and CN_Sell_Buy), quantity increased or decreased (Qty_up, Qty_dn), price increased or decreased (Price_up, Price_dn), the trader used manual-entry methods for their orders (Manual), if the new order was for a customer versus proprietary (CTI_Equal_Four), if the new order that executes was the aggressor (Aggressor), and whether the new order was a stop limit order (New_is_Stop). Note that for price and quantity variables, the omitted category is for cases in which price and/or quantity did not change between the cancelled order and new order, which Tables 4 and 5 indicate to be fairly common. Other omitted categories include the same side variables (Buy and Buy or Sell and Sell), which interact with price changes in the first two buckets above.

The non-binary variables in these regressions include multiple execution counts (Mult_Exec_Cnt), total count of messages in the new order (Total_Cnt_of_Order), and the maximum number of records submitted for any order by the participant (Max_Record_Cnt). We expect these variables to act as size and activity controls. Specifically, participants who use iceberg orders are likely to have multiple executions in an order and participants who actively manage their orders have more messages in each order and may have a large total message count when all orders are compared.

Finally, we include two variables that measure time delay effects. For the in-sequence regressions we measure the time delay from the entry of the new order to the cancellation of the previous order (In_Seq_Time_Gap). The longer the delay, the more likely that market conditions have changed and the connection between cancellation and execution cash flows is less clear. This is also true for the out-of-sequence time variable (Out_Seq_Time_Gap). However, in this case the new order that executes was submitted before the existing order was cancelled.

Table 8 shows regression results with heteroscedasticity-adjusted standard errors for in-sequence data and Table 9 shows similar results using out-of-sequence data. Our focus is on the adverse selection variables. The results for AS_Effect_Buy and AS_Effect_Sell in Tables 8 and 9 support the thesis that cancels can improve cash flows when participant's reveal concerns about adverse selection risks. Both

coefficients are positive and significant in these regressions and suggest gains of $1.05 - $1.23 per contract on the buyside and $4.19 - $5.94 per contract on the sell side when participants react to adverse selection risks. These coefficients are relative to the omitted group in which the cancel-execution pair does not change price. Curiously, the momentum results are mixed with the buy side variable showing a positive effect in both tables and the sell side variable showing the expected negative result. We had expected losses to arise no matter which side of the market had the new order. Evidently, chasing the market on the buyside may end up profitable compared to the omitted group's no price change when inventory is settled at the end of a session.

For other results, we note that manual traders (Manual) show net cash flow losses relative to algorithms in both tables. Trading for customers (CTI_Equal_Four) tends to offer positive benefits when the new order timing is before the cancel order but is negative when the timing follows the cancel order. This may be correlated with being aggressive with some customer orders, although an interaction term (not shown) for this effect was not significant. Lastly, the time gap variables (In_seq_time_gap and Out_seq_time_gap) show that order entry delays will lower cash flows. So, waiting to place a new order after a cancel or waiting to cancel a previous order after placing a new order tends to lower net cash flows, although the effect is small.

C. Robustness Regressions

The above results on adverse selection hold up well when the data are re-estimated using daily observations. In fact, the F-test for whether the daily data should be combined into one sample indicate that daily regressions are more appropriate. Table 10 summarizes these results for adverse selection and momentum variables in twenty-four daily regressions. This table separates results for in-sequence and out-of-sequence data, and reports whether the variable coefficients are positive (pos), negative (neg), or not significant at the 5% level (ins). All the estimates that are significant are marked with an asterisk. The table includes the sample size and the adjusted R-squared for each regression. At the bottom of each variable

column is a summary of the number of significant positive coefficients and negative coefficients (#pos/#neg).

The summary information offers a clear directional story for the adverse selection variables. The adverse selection buy variable (AS_Buy) is significantly positive on 23 days and insignificant on one day. The sell variable (AS_Sell) is significant on 16 days, insignificant on four days, and negative and significant on four days. Out of 48 estimates, only four are significant with the wrong sign compared to 39 significant with the expected positive sign. For the out-of-sequence data, the results are not as strong but still indicate support for our thesis. Out of the 48 estimates, there are 27 with significant positive coefficients and 12 with significant negative coefficients.[16] As a general observation, it appears that our filter into the adverse selection bucket has more noise when participants react before rather than after a cancellation.

The momentum variables (MO_Buy and MO_Sell) offer parallel, somewhat weaker results except that we expect negative coefficients here. For the in-sequence daily data, 34 estimates show a significant negative coefficient compared to 7 significant positive coefficients. For the out-of-sequence daily data, 25 estimates show a significant negative coefficient compared to 22 significant positive coefficients, which is more of a mixed result.

An important consideration for the results in Table 10 is the R-squared for these daily regressions. With regressions that include a large number of observations, we are not surprised when the R-squared are low as in Tables 8 and 9, but the several coefficients show high levels of significance. In the daily regressions we have both high levels of significance and high R-squared values. For the in-sequence regressions, 20 have R-squared values greater than 25% and 16 are greater than 50%. The explanatory power is less in the out-of-sequence regressions where only 11 are greater than the 25% threshold. We interpret these results to indicate that our daily model is well=specified to explain the differential cash flows results between executions and cancelled orders.

[16] The multinomial probability distributions implies that the in-sequence counts are very unlikely ($pr < 0.0001$) for all four columns when the probability of positive and negative coefficients are each 0.475 and insignificant is 0.05 (type 1 error). The out-of-sequence counts are also unlikely ($pr < 0.05$).

We also estimated the regressions in Table 8 and 9 using fixed effects for both days and for individual accounts. For the individual account estimation, we excluded accounts that had fewer than 15 orders over the 24-day sample. The results for in sequence data were similar to those found in Table 8. For the adverse selection variables, AS_Effect_Buy and AS_Effect_Sell had significant coefficients of 0.812 and 3.881, respectively, for the case of only day effects, and 0.285 and 3.481 for the case of both day and account effects. The estimated R-squared increased marginally to 5.75%. The out-of-sequence results were slightly less in magnitude that those in Table 9 when controlling for day effects, but the sign of the coefficient on AS_Effect_Buy became significantly negative (-0.985) when controlling for both day and account effects. Although a negative result is not predicted, it suggests that individual effects may subsume the reaction to adverse selection risks when the new order is placed out of sequence to the cancelled order.

VI. CONCLUSIONS

Our goal in this analysis was to investigate whether cancellations served as a useful tool to avoid adverse selection risks. We started with over 280 million messages for the E-mini futures contract and developed an algorithm to determine whether cancelled orders would have executed if left on the order book. We also conjectured that the paired actions of cancellations and new orders would reveal a motive for the participant's actions. We focused on the adverse selection motive but also investigated the momentum case of "chasing the market." Our analysis found that in the presence of adverse selection risks, execution cash flows per contract were improved relative to the cash flows from leaving a cancel order on the book. Robustness tests also confirmed our findings, but we also found that the timing of new orders—in sequence or out-of-sequence compared to the cancel message—influenced the magnitude and sign of our cash flow results.

REFERENCES

Andersen, T.G., and Bondarenko, O. (2014). VPIN and the flash crash. *Journal of Financial Markets* 17 (2014), 1-46.

Bondarenko, O., and Muravyev, D. (2023). Market return around the clock, *Journal of Financial and Quantitative Analysis* 58(3), 939-967.

Chakrabarty, B., Hendershott, T., Nawn, S., and Pascual, R. (2017). Order exposure in high frequency markets, Working paper, Haas School of Business, UC-Berkeley.

Copeland, T. E., and Galai, D. (1983). Information effects on the bid-ask spread. *The Journal of Finance* 38(5), 1457-1469.

Dahlström, P., Hagströmer, P. and Nordén, L. (2024), The determinants of limit order cancellations, *The Financial Review* 59, 181-201.

Easley, D., and O'Hara, Maureen (1987). Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Easley, D., Lopez de Prado, M., and O'Hara, Maureen (2012). Flow toxicity and liquidity in a high-frequency world, *Review of Financial Studies* 25, 1457–1493.

Fishe, R.P.H. and Robe, M. (2004). The impact of illegal insider trading in dealer and specialist markets: Evidence from a natural experiment. *Journal of Financial Economics* 71, 461-488.

Fong, K. Y., and Liu, W-M. (2010). Limit order revisions. *Journal of Banking & Finance* 34, 1873–1885.

Griffin, J., Oberoi, J, and S. D. Oduro (2021). Estimating the probability of informed trading: A Bayesian approach. *Journal of Banking & Finance* 125, 2-22.

Harris, L.E. (1998). Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions and Instruments* 7, 1–76.

Hasbrouck, J. and Saar, G. (2009). Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets* 12 (2), 143-172.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets* 16 (4), 646-679.

Li, S., Wang, X., & Ye, M. (2021). Who provides liquidity, and when? *Journal of Financial Economics*, *141*(3), 968–980.

Lai, S., Ng, Lilian, and Zhang, Bohui (2014). Does *PIN* affect equity prices around the world? *Journal of Financial Economics* 114(1), 178-195.

Liu, W-M. (2009). Monitoring and limit order submission risks. *Journal of Financial Markets* (12) 104-141.

McInish, T. H., Nikolsko-Rzhevska, O., Nikolsko-Rzhevskyy, A., and Panovska, I. (2020). Fast and slow cancellations and trader behavior. *Financial Management* 49, 973–996.

Parlour and Seppi (2008). Limit-order markets: A Survey. *Handbook of Financial Intermediation & Banking*, edited by A.W.A. Boot and A.V. Thakor. New York., NY: Elsevier, 63-96.

Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies* 22, 4601-4641.

Sandås, P. (2001). Adverse selection and competitive market making: Empirical evidence from a limit order market. *Review of Financial Studies*, 14(3), 705–734.

Sornette, D. and Von der Becke, S. (2011). Crashes and high frequency trading, Working paper, Swiss Finance Institute Research Paper, 11-63.

Subrahmanyam, A. and Zheng, H., (2016). Limit order placement by high-frequency traders. *Borsa Istanbul Review*, 16(4), 185-209.

Van Ness, B. Van Ness, R. and Warr, R. (2001). How well do adverse selection components measure adverse selection?. *Financial Management* (Autumn), 5-30.

Van Ness, B., Van Ness, R., Watson, E.D. (2015). Cancelling liquidity. *Journal of Financial Research* 38 (1), 3-33.

Yang, H., and Xue, F., (2021). Analysis of stock market volatility: Adjusted VPIN with high-frequency data. *International Review of Economics & Finance* 75, 210-222.

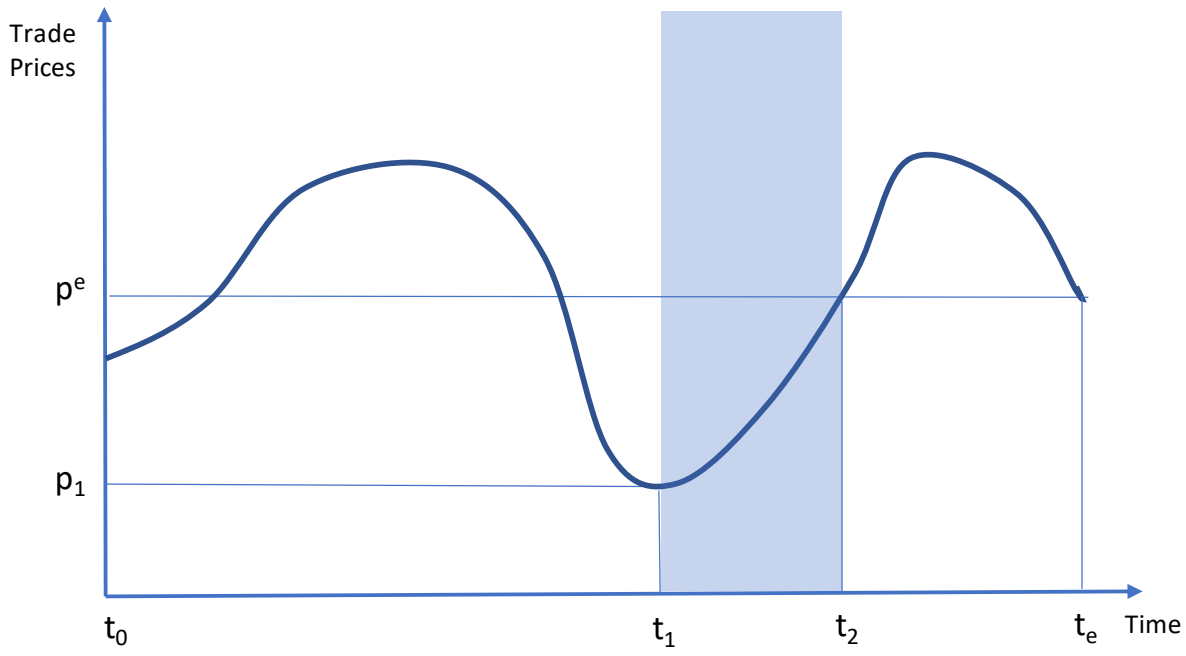


Figure 1
Price Path of Buyside Executions and Minimum Future Prices
In the figure, p^e is the end-of-session settlement price and p_1 is the lowest traded price on the execution price path. Minimum future prices equal p_1 for all trades between the start of the session (t_0) and the minimum price observed at t_1. Between t_1 and t_2, the minimum future price will change along the curve from p_1 to p^e, with the latter price observed at t_2 (shaded area). After t_2, the minimum future price remains constant at p^e.

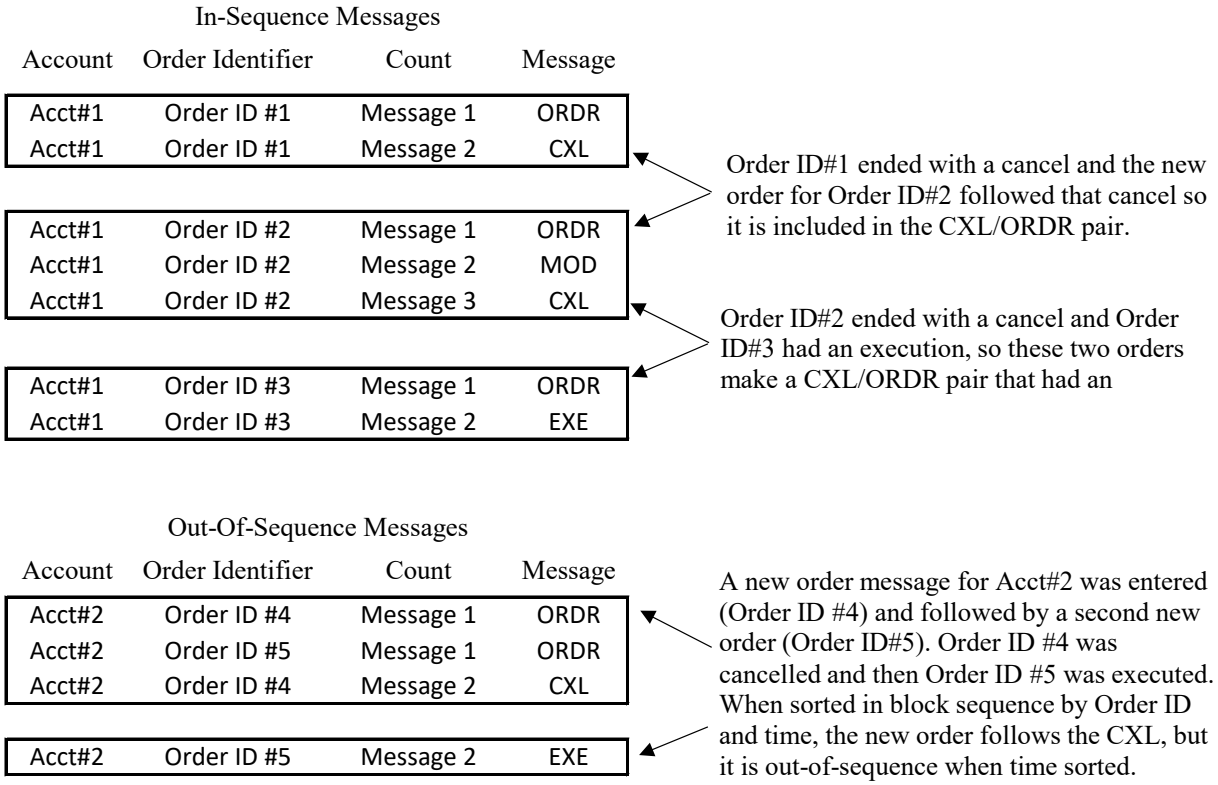


Figure 2
Order Flow for In-Sequence and Out-of-Sequence Messages
Message flow to the CME matching engine is illustrated to show how order flow for different participants may arise in sequential block-time sequence when sorted by participant, order identifier and then time, but be out of sequence when sorted by participant and time alone. The message fields illustrated are new order (ORDR), cancel (CXL), modification (MOD), and execution (EXE).

Table 1

Characteristics of E-Mini Futures Sample Data

The table shows how sample messages are distributed by message type, customer trade indicator, and order type. The percent of new orders that have executions or cancellations are given as well as the percent of messages that are submitted by manual-entry and algorithmic entry methods. The last group shows the cumulative message count for orders. There are 218,337,113 messages in the sample.

Description	Sample Percentage
Message Type	
New Order	39.2%
Partial Fill	9.3%
Full Fill	10.9%
Cancelled	26.5%
Modified	12.4%
Expired	1.7%
Customer Trade Indicator	
Member's Own Account	2.0%
Proprietary Account	86.8%
Trades for Another Member	0.0%
Customer Account	11.2%
Order Type	
Market	1.4%
Limit	97.7%
Stop Limit	1.0%
Other Characteristics	
Execution in Order	30.2%
Cancel in Order	67.7%
Algorithmic-Entry Order	93.2%
Manual-Entry Order	6.8%
Order Message Counts	
2 messages	87.0%
3 messages	93.2%
4 messages	96.0%
5 messages	97.1%

Table 2
Distribution of Message Type Based on Time and Block Sorting

The sample data are filtered to capture only new order records. The four panels report distributions of (lag) messages prior to a new order based on both time and block-time sorting. Time sorting arranges same participant messages by time order. Block-time sorting organizes data by same participant, order identification code in time sequence (first order messages for account, second order messages for account, etc.) and then sorts by time within the message block. Panel A is limited to new orders that follow timestamp of the last message in the previous order block when block sorted. Panel B is limited to new orders that do not follow the timestamp of the last message in the previous order block when block sorted. Panels A and B show lag messages from an overall time sorting. Panels C and D follow the restrictions in Panels A and B respectively, but show lag messages for a block sorting.

	Message Type				
Sample	Cancel	Execution	Modification	New Order	Totals
Panel A: Lag messages from a time sorting when new orders are in sequence by block time					
Lag Messages	27,663,874	15,072,709	467,307	186,611	43,390,501
	63.8%	34.7%	1.1%	0.4%	100.0%
Panel B: Lag messages from a time sorting when new orders are out-of-sequence in block time					
Lag Messages	9,675,335	2,451,988	1,865,898	23,646,587	37,639,808
	25.7%	6.5%	5.0%	62.8%	100.0%
Panel C: Lag messages from a block sorting when new orders are in sequence by block time					
Lag Messages	27,736,370	15,273,363	81,275	299,493	43,390,501
	63.9%	35.2%	0.2%	0.7%	100.0%
Panel D: Lag messages from a block sorting when new orders are out-of-sequence in block time					
Lag Messages	29,476,306	8,035,667	59,941	67,894	37,639,808
	78.3%	21.3%	0.2%	0.2%	100.0%

Table 3
New Order Changes after a Cancellation

The sample data are filtered by identifying all new order messages that follow a cancelled order by the same account. Four groups are then formed based on new order sequencing and whether the new order has an execution (EXE). New orders are considered as "In Sequence" if their time stamp follows the time stamp on the ending message of that account's previous order. It is "Out of Sequence" if the time stamp precedes the time stamp on the ending message of that account's previous order. The table reports how price and quantity changed relative to the cancelled order, which is the disposition of the previous order for that account. These statistics are shown for both algorithmic and manual accounts. The chi-squared test for independence between order entry type and either price changes or quantity changes is rejected in all cases with p-values less than 0.001.

New Order Characteristics		In Sequence All New Orders		In Sequence New Orders w/ EXE		Out of Sequence All New Orders		Out of Sequence New Orders w/ EXE	
		Manual	Algo	Manual	Algo	Manual	Algo	Manual	Algo
Price Change	Up	46.0%	18.9%	46.5%	19.7%	42.6%	30.9%	48.0%	36.2%
	Same	6.3%	61.8%	5.0%	60.4%	13.4%	38.7%	2.5%	27.0%
	Down	47.7%	19.3%	48.5%	19.9%	44.0%	30.3%	49.4%	36.8%
Quantity Change	Up	5.8%	14.4%	7.8%	22.9%	6.0%	22.8%	6.6%	26.1%
	Same	90.2%	72.4%	87.3%	58.8%	88.0%	54.2%	86.3%	51.4%
	Down	4.0%	13.2%	4.9%	18.3%	6.0%	23.1%	7.1%	22.5%
Message Pairs		1,471,080	26,188,187	724,475	5,140,366	931,270	28,542,064	387,506	3,765,826

Table 4
Comparing Price and Quantity Changes for New Orders in Sequence

The sample data from Table 3 are re-framed to show how price and quantity changes interact for in-sequence new orders that follow a cancellation. Manual- and algorithmic-entry results are shown separately in the two panels. The message pair counts are the same as in Table 3 for in-sequence data. The table reports how price and quantity changed relative to the specifications in the cancelled order.The chi-squared test for independence between price changes and quantity changes is rejected in all cases with p-values less than 0.001.

New Order Characteristics		All New Orders Quantity Change			New Orders w/ Executions Quantity Change		
		Up	Same	Down	Up	Same	Down
Panel A: Manual Entry							
Price Change	Up	1.7%	43.2%	2.7%	2.2%	42.5%	3.8%
	Same	0.7%	5.2%	0.5%	0.6%	4.0%	0.4%
	Down	1.6%	41.8%	2.6%	2.1%	40.8%	3.6%
Panel B: Algorithmic Entry							
Price Change	Up	2.8%	13.2%	3.3%	3.0%	12.1%	4.8%
	Same	7.7%	46.3%	7.8%	12.1%	34.8%	13.4%
	Down	2.7%	12.9%	3.3%	3.1%	11.9%	4.8%

Table 5

Inferring Motives from Buy- and Sell-Side Changes

The table reports how price and quantity changed relative to a cancelled order that preceded a new order message. The changes are parsed by buy- and sell-side combinations, and by manual and algorithmic entry methods. All new orders are in sequence for these results and have an execution in the new order. The buy-buy and sell-sell pairs are the focus for determining participant motives. Sample sizes for cancel-new order pairs are shown at the bottom of the table. The chi-squared test for independence between order side pairings and price changes or quantity changes for either manual- or algorithmic-entry types is rejected in all cases with p-values less than 0.001.

		Order Sides for Cancel Message then New Order Message							
New Order		Buy then Buy		Buy then Sell		Sell then Sell		Sell then Buy	
Characteristics		Manual	Algo	Manual	Algo	Manual	Algo	Manual	Algo
Price Change	Up	63.0%	33.3%	59.2%	35.8%	29.3%	5.5%	35.5%	4.5%
	Same	6.2%	60.9%	1.9%	59.6%	6.0%	61.0%	2.0%	59.3%
	Down	30.9%	5.8%	38.8%	4.6%	64.7%	33.5%	62.5%	36.2%
Quantity Change	Up	6.3%	19.6%	12.3%	28.9%	6.2%	19.5%	11.6%	28.8%
	Same	89.4%	64.3%	81.3%	48.9%	89.5%	64.5%	81.8%	49.1%
	Down	4.3%	16.1%	6.3%	22.2%	4.2%	16.0%	6.6%	22.1%
Message Pairs		265,620	1,642,635	98,370	933,211	262,746	1,630,969	97,739	933,551

Table 6

Cash Flows for Executions and Cancellations

The table shows average and median cash flows for the full sample of executions and cancellations, where the latter would have executed if they remained on the book. The regressions are estimated for cases in which a cancel is followed by a new order that executes in block-time sequence. The in-sequence sample is for new orders that actually follow the previous cancel. The out-of-sequence sample is for new orders that were submitted before the previous cancel, but in block-time sequence the cancel message was for the immediately prior order. Total cash flows and cash flow per contract are shown with entries following the designations in equations (1) and (2). The full sample after filtering equals 24,593,642 for executions and 54,089,576 for cancels. The insequence sample size is 5,141,405 and the out-of-sequqnce is 3,833,044. There is no distinction for the regression sample becasue cancels and executions are paired together.

	Full Sample		In-Sequence Sample		Out-Sequence Sample	
Description	Average	Median	Average	Median	Average	Median
Total Cash Flows:						
Execution Cash Flow (c^a)	$ (5.45)	$ (7.38)	$ (25.71)	$ (2,630)	$ (29.51)	$ (2,633)
Execution Cash Flow at End-of-Session Inventory (c^a_e)	$ (5.64)	$ 5.20	$ (25.74)	$ (2,630)	$ (29.71)	$ (2,630)
Net Execution Cash Flow ($c^a - c^a_e$)	$ 0.19	$ 0.10	$ 0.04	$ -	$ 0.20	$ 0.20
Cancel Cash Flow (c^s)	$ (5.36)	$ (2,634)	$ (29.10)	$ (2,629)	$ 53.46	$ 2,637
Cancel Cash Flow at End-of-Session Inventory (c^s_e)	$ (6.72)	$ (2,667)	$ (29.59)	$ (2,630)	$ 52.22	$ 2,630
Net Cancel Cash Flow ($c^s - c^s_e$)	$ 1.35	$ 0.68	$ 0.49	$ 0.15	$ 1.24	$ 0.90
Cash Flows per Contract:						
Execution Cash Flow (c^a)/Q	$ (0.56)	$ (9.13)	$ (7.53)	$ (2,622)	$ (11.10)	$ (2,626)
Execution Cash Flow at End-of-Session Inventory (c^a_e)/Q	$ (0.62)	$ 5.20	$ (7.57)	$ (2,630)	$ (11.27)	$ (2,630)
Net Execution Cash Flow ($c^a - c^a_e$)/Q	$ 0.06	$ 0.08	$ 0.03	$ -	$ 0.17	$ 0.10
Cancel Cash Flow (c^s)/Q	$ (1.81)	$ (2,633)	$ (6.33)	$ (2,621)	$ 20.63	$ 2,630
Cancel Cash Flow at End-of-Session Inventory (c^s_e)/Q	$ (2.26)	$ (2,667)	$ (6.49)	$ (2,630)	$ 20.21	$ 2,630
Net Cancel Cash Flow ($c^s - c^s_e$)/Q	$ 0.45	$ 0.40	$ 0.15	$ 0.10	$ 0.42	$ 0.50

Table 7

Explanation of Regression Variables

This table provides a description of the variables used in the regression analyses. The time gap measurements are in seconds and accurate to nine decimal places.

Variable	Description
Dependent Variable	Net cash flow from the execution of the new order in the cancel-new order pair minus the net cash flow from the cancel order, both on a per contract basis
Mult_Exec_Cnt	Count of Executions (partial and full fills) in execution order
New_is_Stop	Binary variable that indicates if the new order that executes is a stop limit order
Price_up	Binary variable is one is price increased relative to cancelled order
Price_dn	Binary variable is one is price decreased relative to cancelled order
AS_effect_buy	Binary variable is one if new and cancelled order are buys and price decreased
AS_effect_sell	Binary variable is one if new and cancelled order are sells and price increased
MO_effect_buy	Binary variable is one if new and cancelled order are buys and price increased
MO_Effect_Sell	Binary variable is one if new and cancelled order are sells and price decreased
CN_buy_sell	Binary variable is one if cancelled order is a buy and new order is a sell
CN_sell_buy	Binary variable is one if cancelled order is a sell and new order is a buy
Qty_up	Binary variable is one if quantity decreased from cancelled order to new order
Qty_dn	Binary variable is one if quantity increased from cancelled order to new order
Total_Cnt_of_Order	Total number of messages in new order
Max_Record_Cnt	Maximum number of messages submitted in any order for this participant
Manual	Binary variable equals one if submitted by manual entry
CTI_Equal_Four	Binary variable equals one if new order is a trade for a customer
Aggressor	Binary variable equals one if new order is the aggressor in the execution
In_seq_time_gap	Time between New Order message and Cancel message when data are in sequence
Out_seq_time_gap	Time between Cancel message and New Order message when data are out of sequence

Table 8

Effects of Adverse Selection on In-Sequence Data

Using E-mini futures data, the table provides regression estimates for net execution cash flows minus net cancel cash flows using paired cancel and new order data. Each new order has an execution and the cancelled order was found to execute at a later time during the trading session. The pairing uses in-sequence data so the cancellation arose for the participant's previous order before the new order was submitted. Heteroscedasticity-consistent standard errors are used to compute the probability of each t-statistic. The regression adjusted R-squared and the F-test for regression significance are shown at the bottom of the table. The sample contains 5,141,407 observations.

Variable	Coefficient	pr > | t |	Coefficient	pr > | t |	Coefficient	pr > | t |
Intercept	0.025	0.297	-0.050	0.049	0.206	<.0001
Mult_Exec_Cnt	0.003	0.591	0.002	0.654	0.003	0.568
New_is_Stop	3.014	<.0001	3.031	<.0001	3.426	<.0001
Price_up	-1.872	<.0001	-0.856	<.0001	-0.455	<.0001
Price_dn	1.681	<.0001	-0.147	0.0005	2.107	<.0001
AS_effect_buy	1.229	<.0001	3.178	<.0001		
AS_effect_sell	4.193	<.0001	3.296	<.0001		
MO_effect_buy	1.413	<.0001			-0.321	<.0001
MO_Effect_Sell	-2.758	<.0001			-3.502	<.0001
CN_buy_sell	-12.346	<.0001	-12.478	<.0001	-13.092	<.0001
CN_sell_buy	12.513	<.0001	13.184	<.0001	12.044	<.0001
Qty_up	-0.005	0.896	-0.031	0.439	0.034	0.408
Qty_dn	0.006	0.891	0.003	0.949	0.006	0.892
Total_Cnt_of_Order	0.000	0.669	0.000	0.688	-0.000	0.597
Max_Record_Cnt	-0.000	0.687	-0.000	0.722	0.000	0.631
Manual	-0.790	<.0001	-0.762	<.0001	-0.634	<.0001
CTI_Equal_Four	-0.554	<.0001	-0.638	<.0001	-0.418	<.0001
Aggressor	-0.066	0.055	-0.128	0.000	0.049	0.153
In_seq_time_gap	-0.000	0.074	-0.000	0.197	-0.000	0.041
Adjusted R-Squared	5.50%		5.50%		5.50%	
F-Test of Regression	<.0001		<.0001		<.0001	

Table 9

Effects of Adverse Selection on Out-of-Sequence Data

Following the explanation in Table 8, each new order has an execution and the cancelled order was found to execute at a later time during the trading session. The difference here is that the new order was placed on the book before the cancellation of the previous order, so these regressions use out-of-sequence data. Heteroscedasticity-consistent standard errors are used to compute the probability of each t-statistic. The regression adjusted R-squared and the F-test for regression significance are shown at the bottom of the table. The sample contains 3,833,044 observations.

Variable	Coefficient	pr > | t |	Coefficient	pr > | t |	Coefficient	pr > | t |
Intercept	-0.664	<.0001	-1.178	<.0001	0.042	0.710
Mult_Exec_Cnt	0.013	0.601	0.013	0.601	0.013	0.617
New_is_Stop	-3.216	<.0001	-3.353	<.0001	-3.255	<.0001
Price_up	-4.835	<.0001	-0.495	<.0001	-2.573	<.0001
Price_dn	6.615	<.0001	0.316	<.0001	6.842	<.0001
AS_effect_buy	1.053	0.018	7.890	<.0001		
AS_effect_sell	5.938	<.0001	2.135	<.0001		
MO_effect_buy	4.866	<.0001			1.797	<.0001
MO_Effect_Sell	-8.380	<.0001			-9.416	<.0001
CN_buy_sell	-7.236	<.0001	-9.479	<.0001	-9.548	<.0001
CN_sell_buy	5.241	<.0001	10.066	<.0001	4.273	<.0001
Qty_up	0.148	0.001	0.105	0.021	0.188	<.0001
Qty_dn	0.009	0.852	0.050	0.301	-0.043	0.371
Total_Cnt_of_Order	-0.000	0.020	0.000	0.006	0.000	<.0001
Max_Record_Cnt	0.000	<.0001	0.000	<.0001	0.000	<.0001
Manual	-0.469	0.063	-0.454	0.073	-0.554	0.028
CTI_Equal_Four	1.568	<.0001	1.549	<.0001	1.962	<.0001
Aggressor	0.728	<.0001	0.508	<.0001	0.864	<.0001
Out_seq_time_gap	-0.000	<.0001	-0.000	<.0001	-0.000	<.0001
Adjusted R-Squared	1.59%		1.46%		1.57%	
F-Test of Regression	<.0001		<.0001		<.0001	

Table 10
Effects of Adverse Selection and Momentum on Daily Data

Following Tables 8 and 9, this table summarizes results for daily regressions showing the effects of adverse selection (AS_buy and AS_Sell) and momentum (MO_buy and MO_Sell) variables. Coefficient estimates are signed as negative (neg) or positive (pos), and statistical significance at the 1% level using heteroscedasticity-consistent standard errors is indicated by an asterisk. Insignificance is labelled "ins". The regression adjusted R-squared and sample sizes are shown for each date. All F-tests of the regressions were found to be significant. Counts of significant positive and negative coefficients in a column are shown at the bottom of the table.

	In Sequence Data						Out-of-Sequence Data					
Date	AS_buy	AS_sell	MO_buy	MO_Sell	R-sqd	N Obs	AS_buy	AS_sell	MO_buy	MO_Sell	R-sqd	N Obs
Oct 22	pos*	pos*	neg*	neg*	49.2%	171,253	pos*	neg*	neg*	pos*	24.4%	129,795
Oct 23	pos*	ins	neg*	pos*	76.0%	279,424	neg*	pos*	pos*	neg*	8.3%	230,663
Oct 24	pos*	neg*	neg*	neg*	75.5%	286,062	neg*	pos*	pos*	neg*	39.2%	235,665
Oct 25	pos*	neg*	neg*	pos*	88.1%	250,788	neg*	pos*	pos*	neg*	45.8%	192,906
Oct 26	pos*	neg*	neg*	pos*	89.8%	338,541	neg*	pos*	pos*	neg*	47.3%	318,859
Oct 29	pos*	pos*	neg*	neg*	46.4%	287,039	pos*	neg*	neg*	pos*	20.4%	213,530
Oct 30	pos*	ins	neg*	neg*	79.1%	281,353	neg*	pos*	pos*	neg*	3.6%	221,290
Oct 31	pos*	pos*	ins	neg*	38.8%	225,167	pos*	pos*	neg*	pos*	1.0%	156,858
Nov 01	pos*	pos*	neg*	neg*	63.9%	193,082	neg*	pos*	pos*	neg*	26.1%	132,520
Nov 02	pos*	ins	neg*	ins	11.7%	254,770	pos*	ins	neg*	ins	0.4%	191,563
Nov 05	pos*	pos*	neg*	ins	64.7%	154,875	ins	pos*	pos*	neg*	30.6%	89,954
Nov 06	pos*	pos*	neg*	neg*	80.0%	138,403	neg*	pos*	pos*	neg*	36.8%	86,837
Nov 07	pos*	ins	neg*	neg*	75.9%	164,503	neg*	pos*	pos*	neg*	34.5%	118,482
Nov 08	pos*	pos*	neg*	neg*	12.4%	142,104	pos*	pos*	pos*	neg*	9.7%	103,099
Nov 09	pos*	pos*	neg*	neg*	20.4%	173,600	pos*	pos*	neg*	pos*	10.5%	132,992
Nov 12	pos*	pos*	pos*	neg*	62.2%	192,569	pos*	neg*	neg*	pos*	27.7%	152,668
Nov 13	pos*	pos*	ins	neg*	33.3%	217,506	pos*	pos*	neg*	pos*	8.3%	169,858
Nov 14	ins	pos*	pos*	neg*	58.2%	250,361	pos*	ins	neg*	pos*	4.6%	197,053
Nov 15	pos*	neg*	neg*	pos*	72.5%	268,898	neg*	pos*	pos*	neg*	35.8%	218,869
Nov 16	pos*	pos*	neg*	neg*	66.9%	207,391	ins	pos*	pos*	neg*	0.3%	155,532
Nov 19	pos*	pos*	pos*	neg*	24.2%	233,390	pos*	ins	neg*	neg*	9.1%	168,110
Nov 20	pos*	pos*	ins	neg*	58.6%	316,174	pos*	ins	neg*	pos*	27.5%	223,977
Nov 21	pos*	pos*	ins	neg*	74.0%	168,787	pos*	ins	neg*	pos*	37.8%	117,851
Nov 24	pos*	pos*	ins	neg*	71.9%	90,826	ins	ins	neg*	pos*	0.7%	60,452
Pos/Neg	23/0	16/4	3/16	4/18			12/9	15/3	12/12	10/13		