# Low-Latency Trading

Joel Hasbrouck and Gideon Saar

This version: July 2012

Joel Hasbrouck is from the Stern School of Business, 44 West 4th Street, New York, NY 10012 (Tel: 212-998-0310, jhasbrou@stern.nyu.edu). Gideon Saar is from the Johnson Graduate School of Management, Cornell University, 455 Sage Hall, Ithaca, NY 14853 (Tel: 607-255-7484, gs25@cornell.edu). We are grateful for comments from Charles Jones, Andrew Karolyi, Albert Menkveld, Ciamac Moallemi, Maureen O'Hara, Harvey Westbrook, and seminar (or conference) participants at the Chicago Quantitative Alliance / Society of Quantitative Analysts, the Conference on Current Issues in Financial Regulation (University of Notre Dame), Cornell's Johnson School, Cornell Financial Engineering Manhattan, the CREATES Market Microstructure Symposium (Aarhus), ESSEC Business School, Humbolt University, the Investment Industry Regulatory Organization of Canada / DeGroote School, the National Bureau of Economic Research Market Microstructure Group meeting, New York University, Rutgers Business School, SAC Capital, University of Toronto, the Western Finance Association meetings, and the World Federation of Exchanges Statistics Advisory Group.

# Low-Latency Trading

## Abstract

We define low-latency activity as strategies that respond to market events in the millisecond environment, the hallmark of proprietary trading by high-frequency trading firms. We propose a new measure of low-latency activity that can be constructed from publicly-available NASDAQ data to investigate the impact of high-frequency trading on the market environment. Our measure is highly correlated with NASDAQ-constructed estimates of high-frequency trading, but it can be computed from data that are more widely-available. We use this measure to study how low-latency activity affects market quality both during normal market conditions and during a period of declining prices and heightened economic uncertainty. Our conclusion is that increased low-latency activity improves traditional market quality measures—lowering short-term volatility, decreasing spreads, and increasing displayed depth in the limit order book. Of particular importance is that our findings suggest that increased low-latency activity need not work to the detriment of long-term investors in the current market structure for U.S. equities.

# I. Introduction

Our financial environment is characterized by an ever increasing pace of both information gathering and the actions prompted by this information. Speed in absolute terms is important to traders due to the inherent fundamental volatility of financial securities. Relative speed, in the sense of being faster than other traders, is also very important because it can create profit opportunities by enabling a prompt response to news or market activity. This latter consideration appears to drive an arms race where traders employ cutting-edge technology and locate computers in close proximity to the trading venue in order to reduce the latency of their orders and gain an advantage. As a result, today's markets experience intense activity in the "millisecond environment," where computer algorithms respond to each other at a pace 100 times faster than it would take for a human trader to blink.

While there are many definitions for the term "latency," we view it as the time it takes to learn about an event (e.g., a change in the bid), generate a response, and have the exchange act on the response. Exchanges have been investing heavily in upgrading their systems to reduce the time it takes to send information to customers as well as to accept and handle customers' orders. They have also begun to offer traders the ability to co-locate the traders' computer systems next to theirs, thereby reducing transmission times to under a millisecond (a thousandth of a second). As traders have also invested in the technology to process information faster, the entire event/analysis/action cycle has been reduced for some traders to a couple of milliseconds.

The beneficiaries from this massive investment in technology appear to be a new breed of high-frequency traders who implement low-latency strategies, which we define as strategies that respond to market events in the millisecond environment. These traders now generate most message activity in financial markets and according to some accounts also take part in the majority of the trades.[1] While it appears that intermediated trading is

[1] See, for example, the discussion of high-frequency traders in the SEC's Concept Release on Equity Market Structure (2010).

on the rise (with these low-latency traders serving as the intermediaries), it is unclear whether intense low-latency activity harms or helps the market.

Our goal in this paper is to examine the influence of these low-latency traders on market quality. In other words, we would like to know how their combined activity affects attributes such as the short-term volatility of stocks, the total price impact of trades, and the depth of the market. To investigate these questions, we utilize publicly-available NASDAQ order-level data that are identical to those supplied to subscribers and which provide real-time information about orders and executions on the NASDAQ system. Each entry (submission, cancellation, or execution of an order) is time-stamped to the millisecond, and hence these data provide a very detailed view of activity on the NASDAQ system.

We begin by providing a discussion of the players in this new millisecond environment: proprietary and agency algorithms. We document periodicities in the time-series of market activity, which we attribute to activity by agency algorithms. We also look at the speed at which some traders respond to market events—the hallmark of proprietary trading by high-frequency trading firms—and find that the fastest traders have effective latency of 2-3 millisecond during our sample period.

We propose a new measure of low-latency activity based on "strategic runs" of linked messages that describe dynamic order placement strategies. We view this measure as a proxy for the activity of high-frequency traders. An advantage of our measure is that it can be constructed from publicly-available data, and therefore does not rely on specialty datasets that may be limited in scale and scope. We show that our measure is highly correlated with aggregate trading by the high-frequency trading firms featured in a small NASDAQ dataset that was studied in Brogaard (2011a, b, c) and Hendershott and Riordan (2011).

We use our measure to examine how the intensity of low-latency activity affects various market quality measures. We find that an increase in low-latency activity lowers short-term volatility, reduces quoted spreads and the total price impact of trades, and increases depth in the limit order book. These results suggest that increased activity of low-latency traders in the current market environment is beneficial to the traditional benchmarks of market quality. We use a multitude of econometric specifications and robustness tests to substantiate our conclusions.

Furthermore, we employ two distinct sample periods to investigate whether the impact of low-latency trading on market quality differs between normal periods and those associated with declining prices and heightened uncertainty. Over October 2007, our first sample period, stock prices were relatively flat or slightly increasing. Over our second sample period, June 2008, stock prices declined (the NASDAQ index was down 8% in that month) and uncertainty was high following the fire sale of Bear Stearns. We find that higher low-latency activity enhances market quality in both periods, and is especially beneficial in reducing volatility for small stocks during stressful times.[2]

Our paper relates to small but growing strands in the empirical literature on speed in financial markets and algorithmic trading (especially high-frequency trading). With regard to speed, Hendershott and Moulton (2011) and Riordan and Storkenmaier (2012) examine market-wide changes in technology that reduce the latency of information transmission and execution, but reach conflicting conclusions as to the impact of such changes on market quality. There are several papers on algorithmic trading that characterize the trading environment on the Deutsche Boerse (Gsell (2008), Gsell and Gomber (2008), Groth (2009), Prix, Loistl, and Huetl (2007), Hendershott and Riordan (2009)), Euronext and Chi-X (Jovanovic and Menkveld (2010), Menkveld (2011)), the interdealer foreign exchange market (Chaboud, Chiquoine, Hjalmarsson, and Vega (2009)), the futures market (Kirilenko, Kyle, Samadi, and Tuzun (2011)), and the U.S. equity market (Hendershott, Jones, and Menkveld (2009)). In particular, Brogaard (2011a, b, c) and Hendershott and Riordan (2011) attempt to evaluate the impact of high-frequency trading on various aspects of the U.S. market, a goal we share as well.

The rest of this paper proceeds as follows. The next section describes our sample and data. Section III provides an introductory discussion of proprietary and agency

[2] We note that this does not imply that the activity of low-latency traders would help curb volatility during extremely brief episodes such as the “flash crash” of May 2010.

algorithms with some evidence on their activity in the millisecond environment. Section IV lays down the measures and methodology we use for studying the impact of low-latency activity on market quality, while our main results and various robustness tests are reported in Section V. In Section VI we discuss related papers and place our findings within the context of the literature, and Section VII concludes.

## II. Data and Sample

### *II.A. NASDAQ Order-Level Data*

The NASDAQ Stock Market operates an electronic limit order book that utilizes the INET architecture (which was purchased by NASDAQ in 2005).[3] All submitted orders must be price-contingent (i.e., limit orders), and traders who seek immediate execution need to price the limit orders to be marketable (e.g., a buy order priced at or above the prevailing ask price). Traders can designate their orders to display in the NASDAQ book or mark them as "non-displayed," in which case they reside in the book but are invisible to all traders. Execution priority follows price, visibility, and time. All displayed quantities at a price are executed before non-displayed quantities at that price can trade.

The publicly-available NASDAQ data we use, TotalView-ITCH, are identical to those supplied to subscribers, providing real-time information about orders and executions on the NASDAQ system. These data are comprised of time-sequenced messages that describe the history of trade and book activity. Each message is time-stamped to the millisecond, and hence these data provide a detailed picture of the trading process and the state of the NASDAQ book.

We are able to observe four different types of messages: (i) the addition of a displayed order to the book, (ii) the cancellation (or partial cancellation) of a displayed order, (iii) the execution (or partial execution) of a displayed order, and (iv) the execution (or partial execution) of a non-displayed order. In other words, we observe every displayed order that arrives to the NASDAQ market, including the NASDAQ portion of

[3] See Hasbrouck and Saar (2009) for a more detailed description of the INET market structure.

Reg NMS Intermarket Sweep Orders and odd-lot orders. We do not observe submission and cancellation of non-displayed non-marketable limit orders, which are unobservable to market participants in real-time and hence are not part of the TotalView-ITCH data feed. Since we observe all trades (including odd-lots), however, we know when a non-displayed limit order is executed.[4]

### *II.B. Sample*

Our sample is constructed to capture variation across firms and across market conditions. We begin by identifying all common, domestic stocks in CRSP that are NASDAQ-listed in the last quarter of 2007.[5] We then take the top 500 stocks, ranked by market capitalization as of September 30, 2007. Our first sample period is October of 2007 (23 trading days). The market was relatively flat during that time, with the S&P 500 Index starting the month at 1,547.04 and ending it at 1549.38. The NASDAQ Composite Index was relatively flat but ended the month up 4.34%. Our October 2007 sample is intended to reflect a "normal" market environment.

Our second sample period is June 2008 (21 trading days), which represents a period of heightened uncertainty in the market, falling between the fire sale of Bear Stearns in March of 2008 and the Chapter 11 filing of Lehman Brothers in September. During June, the S&P 500 Index lost 7.58%, and the NASDAQ Composite Index was down 7.99%. In this sample period, we continue to follow the firms used in the October 2007 sample, less 29 stocks that were acquired or switched primary listing. For brevity, we refer to the October 2007 and June 2008 samples as "2007" and "2008," respectively.

[4] With respect to executions, we believe that the meaningful economic event is the arrival of the marketable order. In the data, when an incoming order executes against multiple standing orders in the book, separate messages are generated for each standing order. We view these as a single marketable order arrival, so we group as one event multiple execution messages that have the same millisecond time stamp, are in the same direction, and occur in a sequence unbroken by any non-execution message. The component executions need not occur at the same price, and some (or all) of the executions may occur against non-displayed quantities.

[5] NASDASQ introduced the three-tier initiative for listed stocks in July of 2006. We use CRSP's NMSIND=5 and NMSIND=6 codes to identify eligible NASDAQ stocks for the sample (which is roughly equivalent to the former designation of "NASDAQ National Market" stocks).

In our dynamic analysis we use summary statistics constructed over 10-minute intervals. To ensure the accuracy of these statistics, we impose a minimum message count cutoff. A firm is excluded from a sample if more than ten percent of the 10-minute intervals have fewer than 250 messages. Net of these exclusions, the 2007 sample contains 351 stocks, and the 2008 sample contains 399 stocks. In Section V we show that the results concerning the impact of low-latency trading on market quality are robust to imposing a less stringent screen that leaves more than 90% of the stocks in the sample.

Table 1 provides summary statistics for the stocks in both sample periods using information from CRSP and the NASDAQ dataset. Panel A summarizes the measures obtained from CRSP. In the 2007 sample, market capitalization ranges from $789 million to $276 billion, with a median of slightly over $2 billion. The sample also spans a range of trading activity and price levels. The most active stock exhibits an average daily volume of 77 million shares; the median is about one million shares. Average closing prices range from $2 to $635 with a median of $29. Panel B summarizes data collected from NASDAQ. In 2007 the median firm had 26,862 limit order submissions (daily average), 24,015 limit order cancellations, and 2,482 marketable order executions.[6]

## III. The Millisecond Environment: Proprietary vs. Agency Algorithms

Much trading and message activity in U.S. equity markets is commonly attributed to trading algorithms.[7] However, not all algorithms serve the same purpose and therefore the patterns they induce in market data and the impact they have on market quality could depend on their specific objectives. Broadly speaking, however, we can categorize algorithmic activity as agency or proprietary.

[6] These counts reflect our execution grouping procedure. In 2007, for example, the mean number of order submissions less the mean number of order cancellations implies that the mean number of executed standing limit orders is 45,508–40,943=4,565. This is above the reported mean number of marketable orders executed (3,791) because a single marketable order may involve multiple standing limit orders. As we describe in footnote 4, we group executions of standing limit orders that were triggered by a single marketable order into one event.

[7] The SEC's Concept Release on Equity Market Structure cites media reports that attribute 50% or more of equity market volume to proprietary "high-frequency traders." A report by the Tabb Group (July 14, 2010) suggests that buy-side institutions use "low-touch" agency algorithms for about a third of their trading needs.

Agency algorithms are used by buy-side institutions (and the brokers who serve them) to minimize the cost of executing trades in the process of implementing changes in their investment portfolios. They have been in existence for about two decades, but the last ten years have witnessed a dramatic increase in their appeal due to decimalization (in 2001) and increased fragmentation in U.S. equity markets (following Reg ATS in 1998 and Reg NMS in 2005). These algorithms break up large orders into pieces that are then sent over time to multiple trading venues. The key characteristic of agency algorithms is that the choice of which stock to trade and how much to buy or sell is made by a portfolio manager who has an investing (rather than trading) horizon in mind. The algorithms are meant to minimize execution costs relative to a specific benchmark (e.g., volume-weighted average price or market price at the time the order arrives at the trading desk) and their ultimate goal is to execute a desired position change. Hence they essentially demand liquidity, even though their strategies might utilize nonmarketable limit orders.

In terms of technological requirements, agency algorithms are mostly based on historical estimates of price impact and execution probabilities across multiple trading venues and over time, and often do not require much real-time input except for tracking the pieces of the orders they execute. For example, volume-weighted average price algorithms attempt to distribute executions over time in proportion to the aggregate trading and achieve the average price for the stock. While some agency algorithms offer functionality such as pegging (e.g., tracking the bid or ask side of the market) or discretion (e.g., converting a nonmarketable limit buy order into a marketable order when the ask price decreases), typical agency algorithms do not require millisecond responses to changing market conditions.

We believe that agency algorithms drive one of the most curious patterns we observe in the millisecond environment: clock-time periodicity. For a given timestamp *t*, the quantity mod(*t*,1000) is the millisecond remainder, i.e., a millisecond time stamp within the second. Assuming that message arrival rates are constant or (if stochastic) well-mixed within a sample, we would expect the millisecond remainders to be uniformly distributed over the integers {0,1,…,999}. The data, however, tell a different story.

Figure 1 depicts the sample distribution of the millisecond remainders. The null hypothesis is indicated by the horizontal line at 0.001. The distributions in both sample periods exhibit marked departures from uniformity: large peaks occurring shortly after the one-second boundary at roughly 10-30 ms and around 150 ms as well as broad elevations around 600 ms. We believe that these peaks are indicative of agency algorithms that simply check market conditions and execution status every second (or minute), near the second (or the half-second) boundary, and respond to the changes they encounter. These periodic checks are subject to latency delays (i.e., if an algorithm is programmed to revisit an order exactly on the second boundary, any response would occur subsequently). The time elapsed from the one-second mark would depend on the latency of the algorithm: how fast the algorithm receives information from the market, analyzes it, and responds by sending messages to the market. The observed peaks at 10-30 ms or at 150 ms could be generated by clustering in transmission time (due to geographic clustering of algorithmic trading firms) or technology.[8]

The similarities between the 2007 and 2008 samples suggest phenomena that are pervasive and do not disappear over time or in different market conditions. One might conjecture that these patterns cannot be sustainable because sophisticated algorithms will take advantage of them and eliminate them. However, as long as someone is sending messages in a periodic manner, strategic responses by others who monitor the market continuously could serve to amplify rather than eliminate the periodicity. The clustering of agency algorithms means that the provision of liquidity by proprietary algorithms or by one investor to another is higher at these times, and hence conceivably helps agency algorithms execute their orders by increasing available liquidity. As such, agency algorithms would have little incentive to change, making these patterns we identify in the data persist over time.[9] It is also possible, however, that the major players in the industry

[8] We checked with NASDAQ whether their systems that provide traders with more complex order types (e.g., RASH) could be the source of these clock-time periodicities. NASDAQ officials contend that their systems do not create such periodicities.

[9] This intuition is similar in spirit to Admati and Pfleiderer (1988), where uninformed traders choose to concentrate their trading at certain times in order to gain from increased liquidity even in the presence of informed traders.

that designs and implements agency algorithms were unaware of the periodicity prior to our research. If this is indeed the case, and the predictability of buy-side order flow is considered undesirable for various reasons, our findings in this paper could lead to changes in the design of agency algorithms that would eliminate such periodicities in the future.

Relative to agency algorithms, proprietary algorithms are more diverse and more difficult to concisely characterize. Nonetheless, our primary focus in this paper is a new breed of proprietary algorithms that utilizes extremely rapid response to the market environment. Such algorithms, which are meant to profit from the trading environment itself (as opposed to investing in stocks), are employed by hedge funds, proprietary trading desk of large financial firms, and independent specialty firms. These algorithms can be used, for example, to provide liquidity or to identify a trading interest in the market and use that knowledge to generate profit. Brogaard (2011a) reports that NASDAQ identifies 26 firms as being involved in high-frequency trading, but these firms generate most of the order flow in the market and are involved in 68.5% of NASDAQ dollar volume traded over his sample period.[10]

The hallmark of high-frequency proprietary algorithms is speed: low-latency capabilities. These traders invest in co-location and advanced computing technology to create an edge in strategic interactions. Their need to respond to market events distinguishes them from agency algorithms, and therefore we define low-latency trading as "strategies that respond to market events in the millisecond environment." How fast are the low-latency traders? The definition above, which is formulated in terms of speed of response to market events, suggests that an answer to this question could be found by focusing on market events that seem especially likely to trigger rapid reactions. One such event is the improvement of a quote. An increase in the bid may lead to an immediate trade (against the new bid) as potential sellers race to hit it. Alternatively, competing

[10] The NASDAQ classification excludes proprietary trading desks of large sell-side firms as well as direct-access brokers that specialize in providing services to small high-frequency trading firms, and therefore the total number of traders utilizing such low-latency strategies may be somewhat larger.

buyers may race to cancel and resubmit their own bids to remain competitive and achieve or maintain time priority. Events on the sell side of the book, subsequent to a decrease in the ask price, can be defined in a similar fashion.

We therefore estimate the hazard rates (i.e., the message arrival intensities) of the above specific responses subsequent to order submissions that improve the quote. In Figure 2 we plot separately the conditional hazard rates for same-side submissions, same-side cancellations, and executions against the improved quotes (pooled over bid increases and ask decreases). We observe pronounced peaks at approximately 2-3 ms, particularly for executions. This suggests that the fastest responders—the low-latency traders—are subject to 2-3 ms latency. For comparison purposes, we note that human reaction times are generally thought to be on the order of 200 milliseconds (Kosinski (2010)). The figure suggests that the time it takes for some low-latency traders to observe a market event, process the information, and act on it is indeed very short.

Since humans cannot follow such low-latency activity on their trading screens, one might wonder what it actually looks like. It is instructive to present two particular message sets that we believe are typical. Panel A of Table 2 is an excerpt from the message file for ticker symbol ADCT on October 2, 2007 beginning at 09:51:57.849 and ending at 09:53:09.365 (roughly 72 seconds). Over this period, there were 35 submissions (and 35 cancellations) of orders to buy 100 shares, and 34 submissions (and 33 cancellations) of orders to buy 300 shares. The pricing of the orders caused the bid quote to rapidly oscillate between $20.04 and $20.05. The difference in order sizes and the brief intervals between cancellations and submissions suggest that the traffic is being generated by algorithms responding to each other. Panel B of Table 2 describes messages (for the same stock on the same day) between 09:57:18.839 and 09:58:36.268 (about 78 seconds). Over this period, orders to sell 100 shares were submitted (and quickly cancelled) 142 times. During much of this period there was no activity except for these messages. As a result of these orders, the ask quote rapidly oscillated between $20.13 and $20.14.

The underlying logic behind each algorithm that generates such "strategic runs" of messages is difficult to reverse engineer. The interaction in Panel A could be driven by each algorithm's attempt to position a limit order, given the strategy of the other algorithm, so that it would optimally execute against an incoming marketable order. The pattern of submissions and cancellations in Panel B, however, seems more consistent with an attempt to trigger an action on the part of other algorithms and then interact with them. After all, it is clear that an algorithm that repeatedly submits orders and cancels them within 10 ms does not intend to signal anything to human traders (who would not be able to discern such rapid changes in the limit order book). Such algorithms create their own space in the sense that some of what they do seems to be intended to trigger a response from (or respond to) other algorithms. Activity in the limit order book is dominated nowadays by the interaction among automated algorithms, in contrast to a decade ago when human traders still ruled.

While agency algorithms are used in the service of buy-side investing and hence can be justified by the social benefits often attributed to delegated portfolio management (e.g., diversification), the social benefits of high-frequency proprietary trading are more elusive. If high-frequency proprietary algorithms engage in electronic liquidity provision, then they provide a similar service to that of traditional market makers, bridging the intertemporal disaggregation of order flow in continuous markets. However, the social benefits of other types of low-latency trading are more difficult to ascertain. One could view them as aiding price discovery by eliminating transient price disturbances, but such an argument in a millisecond environment is tenuous: at such speeds and in such short intervals it is difficult to determine the price component that constitutes a real innovation to the true value of a security as opposed to a transitory influence. The social utility in algorithms that identify buy-side interest and trade ahead of it is even harder to defend. It therefore becomes an empirical question to determine whether these high-frequency trading algorithms in the aggregate harm or improve the market quality perceived by long-term investors. Our paper seeks to answer this question.

## IV. Low-Latency Trading and Market Quality: Measures and Methodology

Agents who engage in low-latency trading and interact with the market over millisecond horizons are at one extreme in the continuum of market participants. Most investors either cannot or choose not to engage the market at this speed.[11] If we believe that healthy markets need to attract longer-term investors whose beliefs and preferences are essential for the determination of market prices, then market quality should be measured using time intervals that are easily observed by these investors. How does low-latency activity with its algorithms that interact in milliseconds relate to depth in the market or the range of prices that can be observed over minutes or hours? In this section we seek to answer this question by characterizing the influence of low-latency trading on measures of liquidity and short-term volatility observed over 10-minute intervals throughout the day.

### *IV.A. Measures*

To construct a measure of low-latency activity, we begin by identifying "strategic runs," which are linked submissions, cancellations, and executions that are likely to be parts of a dynamic algorithmic strategy. Our goal is to isolate instances of market activity that look like the interactions presented in Table 2. Since our data do not identify individual traders, our methodology no doubt introduces some noise into the identification of low-latency activity. We nevertheless believe that other attributes of the messages can used to infer linked sequences.

In particular, our "strategic runs" (or simply, in this context, "runs") are constructed as follows. Reference numbers supplied with the data unambiguously link an individual limit order with its subsequent cancellation or execution. The point of inference comes in deciding whether a cancellation can be linked to either a subsequent submission of a nonmarketable limit order or a subsequent execution that occurs when

[11] The recent SEC Concept Release on Equity Market Structure refers in this context to "long-term investors … who provide capital investment and are willing to accept the risk of ownership in listed companies for an extended period of time" (p. 33).

the same order is resent to the market priced to be marketable. We impute such a link when the cancellation is followed within 100 ms by a limit order submission or by an execution in the same direction and for the same size. If a limit order is partially executed, and the remainder is cancelled, we look for a subsequent resubmission or execution of the cancelled quantity. In this manner we construct runs forward throughout the day.

Our procedure links roughly 60 percent of the cancellations in the 2007 sample, and 54 percent in the 2008 sample. Although we allow up to 100 ms to elapse from cancellation to resubmission, 49 percent of the imputed durations are one or zero ms, and less than ten percent exceed 40 ms. The length of a run can be measured by the number of linked messages. The simplest run would have three messages, a submission of a nonmarketable limit order, its cancellation, and its resubmission as a marketable limit order that executes immediately (i.e., an "active execution"). The shortest run that does not involve an execution is a limit order that was submitted, cancelled, resubmitted, and cancelled or expired at the end of the day. Our sample periods, however, feature many runs of 10 or more linked messages. We identify about 46.0 million runs in the 2007 sample period and 67.1 million runs in the 2008 sample period.

Table 3 presents summary statistics for the runs. We observe that around 75% of the runs have 3 to 9 messages, but the longer runs (10 or more messages) constitute over 60% of the messages that are associated with strategic runs. The proportion of runs that are (at least partially) executed is 38.1% in 2007 and 30.5% in 2008. About 8.1% (7.1%) of the runs in the 2007 (2008) sample period end with a switch to active execution. That is, a limit order is cancelled and replaced with a marketable order. These numbers attest to the importance of strategies that pursue execution in a gradual fashion.

To construct a measure of low-latency trading that is more robust to measurement error, we transform the raw strategic runs in two ways. The first transformation is to use only longer runs—runs of 10 or more messages—to construct the measure. While our methodology to impute links between cancellations and resubmissions of orders can result in misclassifications, for a run with many resubmissions to arise solely as an artifact of such errors there would have to be an unbroken chain of spurious linkages. This suggests that longer runs are likely to be more reliable depictions of the activity of actual algorithms than shorter runs. While the 10-message cutoff is somewhat arbitrary, these runs represent more than half of the total number of messages that are linked to runs in each sample period, and we also believe that such longer runs characterize much low-latency activity. In Section V we provide robustness analysis demonstrating that our conclusions are unchanged when we include all runs.

The second transformation we use to reduce measurement error is to utilize time-weighting of the number of runs rather than simply aggregating the runs or the messages in runs. We define our measure of low-latency activity, *RunsInProcess*, as the time-weighted average of the number of strategic runs of 10 messages or more the stock experiences in an interval.[12] Time-weighting helps us combat potential errors because it ensures that roughly equivalent patterns of activity contribute equally to our measure, which can be demonstrated using the strategic run shown in Panel B of Table 2. This run, which lasts 78.5 seconds, contributes 0.129 (78.5/600) to *RunsInProcess* of stock ADCT in the interval 9:50-10:00am on October 2nd, 2007. What if we were wrong and the inferred resubmission at time 9:57:20.761 actually came from a different algorithm, so that the activity described in Panel B of Table 2 was generated by one 48-message algorithm and another 94-message algorithm rather than a single 142-message algorithm? This should not alter our inference about the activity of low-latency traders from an economic standpoint, because the two shorter algorithms together constitute almost the same amount of low-latency activity as the single longer algorithm. The time-weighting of *RunsInProcess* ensures that the measure computed from the two algorithms is almost identical to the one originally computed from the single algorithm (the two will differ only by 0.005/600=0.000008 due to the 5 millisecond gap between the end of the first

[12] The time-weighting of this measure works as follows. Suppose we construct this variable for the interval 9:50:00am-10:00:00am. If a strategic run started at 9:45:00am and ended at 10:01:00am, it was active for the entire interval and hence it adds 1 to the *RunsInProcess* measure. A run that started at 9:45:00am and ended at 9:51:00am was active for one minute (out of ten) in this interval, and hence adds 0.1 to the measure. Similarly, a run that was active for 6 seconds within this interval adds 0.01.

algorithm and the beginning of the second algorithm), and hence this type of error would not affect our empirical analysis.

It is important to recognize that our measure of low-latency activity does not have a positive relationship with market quality by construction. In fact, if liquidity is provided by patient limit order traders (which is the case most often described in theoretical models), depth in the book is maximized when the cancellation rate is zero. In other words, liquidity is highest when limit orders stay in the book until they are executed, in which case our measure *RunsInProcess* is equal to zero. As traders begin cancelling orders, liquidity in the book worsens and our measure increases. This suggest that holding everything else equal, *RunsInProcess* should be negatively related to liquidity, though liquidity may decline only modestly if traders cancel but replace limit orders with other limit orders rather than switch to marketable orders. However, the relationship between *RunsInProcess* and liquidity is more complex because low-latency traders may be willing to submit more limit orders and provide more depth if they have the technology to cancel limit orders quickly enough to lower the pick-off risk of their orders. Hence, we do not know a-priori whether the relationship between our measure of low-latency activity and market quality is positive or negative in equilibrium, and this is what we test in this section.

Our measure has one important advantage over the measures of high-frequency activity used in Brogaard (2011a, 2011b, 2011c) and Hendershott and Riordan (2011): it can be estimated from publicly-available data (NASDAQ's ITCH data). In contrast, the characterization of high-frequency trading in the aforementioned papers uses a specific sample constructed by NASDAQ of 26 high-frequency trading firms in 120 stocks during 2008 and 2009 (henceforth, the HFT dataset). Our measure may include more errors of inclusion relative to the NASDAQ proprietary data (i.e., we may capture activity that is not originated by high-frequency trading firms), but it has fewer errors of exclusion (the NASDAQ classification excludes proprietary trading desks of large sell-side firms as well as direct access brokers that specialize in providing services to small high-frequency trading firms).

Since our second sample period (June 2008) overlaps with the trading data in the HFT dataset, we looked at the correlation between our measure of low-latency activity and several measures that can be constructed from the HFT dataset. Let "H" denote high-frequency trading firms and "N" denotes other traders. Each trade in the HFT dataset is categorized by one of the following combinations: NH, NN, HN, or HH, where the first letter represents the party that takes liquidity and the second letter represents the party that supplies liquidity. We use this information to construct the following measures of high-frequency trading in each 10-minute interval:

1. Total HFT executed orders=NH+(2*HH)+HN
2. HFT-participated trades=NH+HH+HN
3. Total HFT executed orders that supplied liquidity=NH+HH
4. Net HFT executed orders that supplies liquidity=NH

The first two measures represent overall trading by high-frequency trading firms. The last two measures denote liquidity supplied by high-frequency trading firms, and we look at them to see whether, because strategic runs are comprised mostly of limit orders rather than marketable orders, our measure happens to be biased towards liquidity supply by high-frequency trading firms (as opposed to their overall trading). For each measure, we calculate two versions: one in terms of share volume and the other in terms of number of orders.

Of the 120 firms in the HFT dataset, 60 are NASDAQ stocks for which we use ITCH order-level data to construct the *RunsInProcess* measure.[13] Table 4 shows Spearman and Pearson correlations between the HFT dataset measures and *RunsInProcess* over all 10-minute intervals for all stocks. Two things are immediately apparent. First, the correlation between *RunsInProcess* and total high-frequency trading in the HFT dataset is very high: the Spearman correlation is over 0.8 irrespective of whether the measures are expressed in terms of number of orders or share volume.

[13] Out of the 60 stocks, 33 were in our June 2008 sample. We created the measure *RunsInProcess* for the 27 additional stocks to be able to estimate the correlations in Table 4 using all 60 stocks that are available in the HFT dataset.

Second, the correlations of *RunsInProcess* with the total high-frequency trading measures are very similar to its correlations with the measures of liquidity supplied by HFT firms. In other words, our measure of low-latency activity captures total high-frequency trading and is not biased toward capturing just liquidity-supplying trades.

We want to stress that both our *RunsInProcess* measure and the trading measures from the HFT dataset are only proxies for the activity of high-frequency trading firms. In particular, most of the activity by high-frequency traders involves orders that do not execute. The measures computed from the HFT dataset use only executed orders, and therefore do not necessarily reflect overall activity.[14] Still, the fact that our *RunsInProcess* measure and the measures of executed orders from the HFT dataset are highly correlated should be reassuring to researchers who carry out empirical analysis using either the publicly-available ITCH data or the HFT dataset to discern the overall impact of high-frequency trading firms.

In addition to our measure of low-latency activity, we use the ITCH order-level data to compute several measures that represent different aspects of NASDAQ market quality: a measure of short-term volatility and three measures of liquidity. The first measure, *HighLow*, is defined as the highest midquote in an interval minus the lowest midquote in the same interval, divided by the midpoint between the high and the low (and multiplied by 10,000 to express it in basis points). The second measure, *Spread*, is the time-weighted average quoted spread (ask price minus the bid price) on the NASDAQ system in an interval. The third measure, *EffSprd*, is the average effective spread (or total price impact) of all trades on NASDAQ during the ten-minute interval, where the effective spread is defined as the transaction price (quote midpoint) minus the quote midpoint (transaction price) for buy (sell) marketable orders. The fourth measure, *NearDepth*, is the time-weighted average number of (visible) shares in the book up to 10 cents from the best posted prices.

[14] The HFT dataset contains additional information, depth snapshots and quotes, for several short periods, but none of them overlaps with our sample period. Hence, we use the available information on executed orders to construct the measures we correlate with *RunsInProcess*.

*IV.B. Methodology*

The correlations between our measures of market quality and low-latency activity generally suggest a positive relation. For example, the correlations between *RunsInProcess* and short-term volatility (*HighLow*, an inverse market quality measure) are -0.15 in 2007 and -0.24 in 2008.[15] Our goal, however, is to assess the causal effects. Using *MktQuality*$_{i,t}$ as a placeholder for any of the market quality measures, pooled panel regression specifications of the form $MktQuality_{i,t} = a_0 + a_1 RunsInProcess_{i,t} + ... + u_{i,t}$ (for firm *i* and period *t*) can be motivated in the usual way, as linearized conditional expectations, with the $a_1$ coefficient capturing the impact of an exogenous change in *RunsInProcess*.

Estimation is complicated, however, by a strong possibility of simultaneity. For example, an exogenous drop in short-term volatility (*HighLow)* might establish a more attractive environment for low-latency activity. This mechanism induces correlation between *RunsInProcess* and *u*, rendering OLS estimates inconsistent, and motivating the use of instrumental variables. Our criteria for constructing an instrument are that it should be correlated with the explanatory variable (*RunsInProcess*$_{it}$), but not be directly affected by the dependent variable (*HighLow*$_{it}$). To this end we seek to measure the number of low-latency traders broadly active in a given interval, but excluding firm *i* and all firms that are likely to be related to *i* via correlated trading strategies.

Specifically, our instrument for *RunsInProcess*$_{i,t}$ is *RunsNotIND*$_{i,t}$, which is the average number of runs of 10 messages or more in the same interval for the other stocks in our sample excluding: (1) the INDividual stock, stock *i*, (2) stocks in the same INDustry as stock *i* (as defined by the four-digit SIC code), and (3) stocks in the same INDex as stock *i*, if stock *i* belongs to either the NASDAQ 100 Index or the S&P 500 Index.

[15] *RunsInProcess* is also negatively correlated with the quoted spread (-0.32 in 2007 and -0.37 in 2008) and the total price impact of trades (-0.16 in 2007 and -0.11 in 2008), and positively correlated with depth in the book (0.29 in 2007 and 0.35 in 2008).

The intent here is to remove the possible influence of algorithms that implement strategies across multiple stocks by excluding the most likely candidates for such cross-stock strategies. For example, statistical arbitrage strategies like pairs trading often utilize stocks in the same industry, which is why *RunsNotIND* does not contain any stock in the same four-digit SIC code. Similarly, algorithms that implement index strategies would have no impact on *RunsNotIND* because if a stock is in one of the two main indexes (the NASDAQ 100 or the S&P 500), we exclude all other stocks in that index from the computation of *RunsNotIND* for that stock.

Beyond stocks in the same industry and the same index that are explicitly omitted from the instrument, our results should be robust to multi-stock algorithms that utilize concurrent trading in a small number of stocks. The average (minimum) number of stocks that are used in the constructions of *RunsNotIND* is 322.7 (250) in 2007 and 371.3 (290) in 2008, making it insensitive to concurrent trading in a handful of related stocks. For robustness, we repeated the analysis with an instrument computed as the median of $RunsInProcess_{i,t}$ (excluding stock *i*, stocks in the same industry, and stocks in the same index) in each interval because the median should be even less sensitive to a handful of outliers. Our results with the median instrument are similar to those with *RunsNotIND*, suggesting that multi-stock algorithms are not a significant problem with respect to the validity of this instrument.

In specifications that jointly model NASDAQ market quality and *RunsInProcess* we also need an instrument for market quality. Here, we use $EffSprdNotNAS_{i,t}$, which is the dollar effective spread (absolute value of the distance between the transaction price and the midquote) computed for the same stock and during the same time interval but only from trades executed on non-NASDAQ trading venues (using the TAQ database). This measure reflects the general liquidity of the stock in the interval, but it does not utilize information about NASDAQ activity and hence would not be directly determined by the number of strategic runs that are taking place on the NASDAQ system, rendering it an appropriate instrument.

It might be argued that $EffSprdNotNAS_{i,t}$ won't be exogenous if many low-latency algorithms pursue cross-market strategies in the same security (i.e., if the same algorithm executes trades on both NASDAQ and another market). A cross-market strategy, however, cannot operate at the lowest latencies because an algorithmic program cannot be co-located at more than one market. This necessarily puts cross-market strategies at a disadvantage relative to co-located single-market algorithms. At least at the lowest latencies, therefore, we believe that the single-market algorithms are dominant.[16] Considerations of liquidity in multiple markets are also common in agency algorithms that create a montage of the fragmented marketplace to guide their order routing logic to the different markets. These, however, most likely do not give rise to the long strategic runs that we use to measure the activity of proprietary low-latency traders ($RunsInProcess_{i,t}$) and hence would not introduce reverse causality. Nonetheless, while we are able to significantly improve on the quality of *RunsNotIND* by excluding stocks in the same industry and index (the most likely candidates for cross-stock algorithms), we cannot improve on the quality of *EffSprdNotNAS* in a similar fashion. We therefore employ additional specifications that do not rely on this variable.

We use several econometric models that allow us to estimate the impact of low-latency on market quality under different sets of assumptions. The results from the various specifications combine to present a consistent picture as to the robustness of our conclusions. In Model I, we pool observations across all stocks and all time intervals and estimate the following two-equation simultaneous equation model for each market quality measure:

$$\begin{aligned} MktQuality_{i,t} &= a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + e_{1,i,t} \\ RunsInProccess_{i,t} &= b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + e_{2,i,t} \end{aligned} \tag{I}$$

where $i = 1,...,N$ indexes firms, $t = 1,...,T$ indexes 10-minute time intervals, *MktQuality* represents one of the market quality measures (*HighLow*, *EffSprd*, *Spread*, or *NearDepth)*, and the instruments are *EffSprdNotNAS* and *RunsNotIND*. To remove stock-

[16] Conversations with a NASDAQ official provided support to this view.

specific fixed effects, we standardize each variable by subtracting from each observation the stock-specific time-series average over the sample period and dividing by the stock-specific time-series standard deviation. The standardization eliminates the intercepts in the specification.

In Model I, *EffSprdNotNAS* serves as an instrument for *MktQuality* in the *RunsInProcess* equation, and also appears an exogenous variable in the *MktQuality* equation. If cross-market algorithms render *EffSprdNotNAS* less desirable (both as an instrument and as an exogenous variable), it is useful to consider an alternative. In Model II, we estimate the following single-equation specification for each market quality measure:

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 TradingIntensity_{i,t} + e_{1,i,t} \qquad \text{(II)}$$

In this variation, *TradingIntensity* is used (instead of *EffSprdNotNAS*) as an exogenous variable to capture the impact of intraday informational events or liquidity shocks. $TradingIntensity_{i,t}$ is defined as stock *i*'s total trading volume in the entire market (not just NASDAQ) immediately prior to interval *t* (i.e., in the previous 10 minutes), and therefore it is not subject to the simultaneity problem. As before, we estimate this equation using an IV estimator with *RunsNotIND* as an instrument for *RunsInProcess*.

Our third and fourth specifications are motivated by the tradition in finance that emphasizes commonalities in returns and volatilities. Could our results be explained by a return or volatility factor that drives both low-latency activity and market liquidity? Model III is therefore the following two-equation model:

$$\begin{aligned} MktQuality_{i,t} &= a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} \\ &\quad + a_3 R_{QQQQ,t} + a_4 \left| R_{QQQQ,t} \right| + e_{1,i,t} \\ RunsInProccess_{i,t} &= b_1 MktQuality_{i,t} + b_2 RunsNotI_{i,t} \\ &\quad + a_3 R_{QQQQ,t} + a_4 \left| R_{QQQQ,t} \right| + e_{2,i,t} \end{aligned} \qquad \text{(III)}$$

and Model IV is the single-equation model:

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 TradingIntensity_{i,t} + a_3 R_{QQQQ,t} + a_4 \left| R_{QQQQ,t} \right| + e_{1,i,t} \quad \text{(IV)}$$

Both models seek to rule out the possibility that our results are driven by the potential influence of omitted variables related to common factors by adding two independent variables to each equation: (i) the market return (in each 10-minute interval), and (ii) the absolute value of the market return. Since our sample consists exclusively of NASDAQ stocks, we use the NASDAQ 100 Index as a proxy for the market portfolio, and compute the 10-minute returns using the QQQQ Exchange Traded Fund that tracks the index.

The fifth and sixth specifications are motivated by theoretical models that give rise to intraday patterns in liquidly (as well as various empirical findings of time-of-day effects in liquidity measures). For example, models of adverse selection (e.g., Glosten and Milgrom (1985)) generally predict higher spreads in the morning compared to the rest of the day. An afternoon increase in spreads is consistent with inelasticity of demand (e.g., Brock and Kleidon (1992)), while the analysis in Admati and Pfleiderer (1988) could be used to justify morning and afternoon patterns driven by implicit or explicit coordination of traders in the market.

To account for potentially omitted time-of-day effects that could drive both low-latency trading and market liquidity we add dummy variables for the morning and afternoon periods as independent variables in each equation. Model V is therefore the following two-equation model:

$$\begin{aligned} MktQuality_{i,t} &= a_0 + a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} \\ &\quad + a_3 DumAM_{i,t} + a_4 DumPM_{i,t} + e_{1,i,t} \\ RunsInProccess_{i,t} &= b_0 + b_1 MktQuality_{i,t} + b_2 RunsNotI_{i,t} \\ &\quad + b_3 DumAM_{i,t} + b_4 DumPM_{i,t} + e_{2,i,t} \end{aligned} \qquad \text{(V)}$$

and Model VI is the single-equation model:

$$\begin{aligned} MktQuality_{i,t} &= a_0 + a_1 RunsInProcess_{i,t} + a_2 TradingIntensity_{i,t} \\ &\quad + a_3 DumAM_{i,t} + a_4 DumPM_{i,t} + e_{1,i,t} \end{aligned} \qquad \text{(VI)}$$

Here, *DumAM* is equal to one for intervals between 9:30am and 11:00am, and zero otherwise, and *DumPM* is equal to one for intervals between 2:30pm and 4:00pm, and zero otherwise. A constant term is added to each equation to represent mid-day effects (11:00am-2:30pm).

We estimate all six models using Two-Stage GMM, and consider two types of robust standard errors. The first type (which we report as Clust. p-value) is robust to arbitrary heteroskedasticity and clustering on two dimensions: (i) stocks, and (ii) time-intervals. Hence, the standard errors are robust to serial correlations in the time dimension (for each stock) and contemporaneous correlation of the errors across stocks (see Thompson (2011)). The second type (denoted in the tables as DK p-value) implements the estimator proposed by Driscoll and Kraay (1998). The DK estimator is an extension of the Newey-West heteroskedasticity-and-autocorrelation-consistent estimator that is also robust to very general spatial dependence (i.e., contemporaneous correlation of the errors across stocks).

## V. Low-Latency Trading and Market Quality: Results

Panel A of Table 5 presents the estimated coefficients of Model I side-by-side for the 2007 and 2008 sample periods. The most interesting coefficient is $a_1$, which measures the impact of low-latency activity on the market quality measures. We observe that higher low-latency activity implies lower posted and effective spreads, greater depth, and lower short-term volatility. Moreover, the impact of low-latency activity on market quality is similar in the 2007 and 2008 sample periods. The coefficients on the two instruments have the expected signs and are highly significant. In all regressions, Cragg-Donald (1993) statistics reject the null of weak instruments using the Stock and Yogo (2005) critical values.

To gauge the economic magnitudes implied by the $a_1$ coefficients, one can look at how the market quality measure for a representative stock changes when we increase the amount of low-latency activity. One standard deviation increase in *RunsInProcess* implies a decrease of 29% in short-term volatility (down 12.3 basis points from a mean value of 42.1 basis points) in the 2007 sample period, and similarly a decrease of 34% in the 2008 sample period. Depth within 10 cents from the best prices increases by 20% when we increase low-latency activity by one standard deviation in the 2007 sample period (up 2,199 shares from a mean of 11,271 shares) and an even greater increase—

34%—is observed in the 2008 sample period when the market is under stress. A similar pattern where low-latency activity has a greater positive impact on market quality in 2008 is also observed for spreads, where one standard deviation increase in *RunsInProcess* implies a decrease of 26% in 2007 and 32% in 2008.

The fact that low-latency activity decreases short-term volatility, lowers spreads, and increases depth even to a greater extent in the 2008 sample period—when the market is relentlessly going down and there is heightened uncertainty in the economic environment—is particularly noteworthy. It seems to suggest that low-latency activity creates a positive externality in the market at the time that the market needs it the most.

To ensure that our results are not driven by outliers and therefore are not an artifact of pooling the data, Figure 3 presents histograms of the $a_1$ coefficients from stock-by-stock estimations of the model. The first two panels of Figure 3, for example, show that almost all of the $a_1$ coefficients are negative when the market quality measure is short-term volatility (*HighLow*). The histograms of all other market quality measures demonstrate that the pooled results are not driven by outliers but rather represent a reasonable summary of the manner in which low-latency activity affects market quality in the cross section of stocks.

The results of the two-equation model also suggest that low-latency activity is attracted to more liquid and less volatile stocks (the estimated $b_1$ coefficients). However, this finding is dependent on the quality of the *EffSprdNotNAS* instrument. To rule out that our main results are also affected by the quality of this instrument, Panel B of Table 5 presents the coefficient estimates from Model II, where we do not use this instrument but rather focus exclusively on the impact of low-latency trading on market quality. Here as well we observe that higher low-latency activity implies lower posted and effective spreads, greater depth, and lower short-term volatility. The results appear even stronger than those of the two-equation specification in the sense that the magnitude of some of the coefficients suggests a larger effect.

Panel A of Table 6 presents the results of Model III, where we add common factor information (return and volatility of the market) to the two-equation model. Market volatility appears to be an important determinant of the market quality measures in both sample periods (the $a_4$ coefficient). As a determinant of low-latency activity, market volatility is significant only in 2007 (the $b_4$ coefficient). Market return has an impact on some of the market quality measures (especially short-term volatility and depth), but is not significant in the *RunsInProcess* equation. The important takeaway from this panel is that the inclusion of market return and volatility as independent variables does not eliminate the significant showing of the low-latency activity as a determinant of the market quality measures: all estimated $a_1$ coefficients have the same signs as in Panel A of Table 5 and are highly statistically significant.

Similar results are observed when we look at the results of Model IV in which market return and volatility are added to the single-equation specification (in Panel B of Table 6). In the presence of the trading intensity variable, market return is not a significant determinant of either market quality or low-latency trading. However, higher low-latency activity implies lower short-term volatility, lower spreads, and more depth exactly as before.

Table 7 presents the results of Model V (in Panel A) and Model VI (in Panel B), where we use dummy variables to account for potential time-of-day effects. The first thing to note looking at the column of $a_1$ coefficients in both models is that our results that greater low-latency trading implies lower short-term volatility, lower spreads and effective spreads, and greater depth remain extremely robust. In addition, we observe that time-of-day variables exert their expected influence on market activity: the $a_3$ coefficient, for example, demonstrates that stocks are less liquid in the first hour and a half of trading. Also, it is interesting to note that the intensity of low-latency trading is lower in the last hour and a half of trading (the $b_4$ coefficient in Panel A). Still, it does not appear as if omitted variables in the form of time-of-day effects or market factors diminish the impact of low-latency activity on market quality, increasing our confidence in the robustness of our conclusions.

### *V.A. Additional Robustness Tests*

The conclusion we draw from our main analysis is that low-latency trading positively impacts several standard measures of market quality. In this section we examine the robustness of this conclusion within subsets of our sample as well as to choices we make concerning sample construction and the definition of variables.

We begin by looking at whether the impact of low-latency activity on market quality differs for stocks that are somehow fundamentally dissimilar, like small versus large market capitalization stocks. Table 8 contains the estimates from Model I in subsamples consisting of four quartiles ranked by the average market capitalization over the sample period. The results using the other specifications (Models II-VI) are similar with respect to how low-latency trading impacts market quality (the $a_1$ coefficients), and are therefore omitted to economize on the size of the table. We observe that the $a_1$ coefficients in the subsamples have the same sign as in the full sample and are all statistically significant. While there is not much pattern across the quartiles in the 2007 sample period, the picture in the 2008 sample is different: it appears that during more stressful times, low-latency activity helps reduce volatility in smaller stocks more than it does in larger stocks.

Another interesting pattern can be observed in the coefficient $b_1$, which tells us how market quality affects low-latency trading. While better market quality implies more low-latency activity in larger stocks in the 2007 sample period, no such relationship is found for smaller stocks. During the stressful period of June 2008, however, the $b_1$ coefficients suggest a different behavior: higher liquidity encourages low-latency trading in smaller stocks but not in the top quartile of stocks by market capitalization.

We also performed the estimations separately on subsamples formed as quartiles of NASDAQ's market share of traded volume. Trading in the U.S. occurs on multiple venues, including competing exchanges, crossing networks, and Electronic Communications Networks. This fragmentation might jointly affect market quality and low-latency activity. Our results (not reported here), however, show no significant pattern across market-share quartiles. In other words, the beneficial impact of low-latency trading on the market quality measures is similar for stocks that have varying degrees of trading concentration on the NASDAQ system.

The second issue we address in this robustness section is that our sample selection procedure (described in Section II) screens for stocks with sufficient amount of message activity to reduce the noise in the measures. Specifically, we exclude stocks if more than ten percent of the 10-minute intervals have fewer than 250 messages. This reduces the number of firms we analyze by approximately 30% (20%) in the 2007 (2008) sample period. Panel A of Table 9 presents the results of estimating Model I on an alternative sample where we only exclude stocks if more than ten percent of the 10-minute intervals have fewer than 100 messages. This screen significantly increases the number of stocks in both sample periods (471 in 2007 and 456 in 2008), but the results are very similar to those presented in Table 5. In particular, all $a_1$ coefficients are highly statistically significant and have similar magnitudes. We obtain the same results when we estimate Models II-VI (from Panel B of Table 5 as well as Tables 6 and 7) on this modified sample of stocks.

Panel B of Table 9 presents a test that alters the definition of *RunsInProcess*, our measure of low-latency activity. The discussion in Section IV.A provides the rationale for focusing on longer runs (those with ten or more messages) as a way to mitigate the potential influence of errors in constructing the strategic runs. To ensure that omitting shorter runs does not materially affect our conclusions, however, we use all strategic runs to construct an alternative measure of low-latency activity: *AllRunsInProcess*, and carry out exactly the same analysis as before. The results in Panel B of Table 9 are similar to those in Table 5, and we reach the same conclusions when using Models II-VI.

Lastly, we computed an alternative market quality measure that attempted to isolate the cost of trading of “regular” investors who are not low-latency traders. The measure we created was an average effective spread that uses only trades that were not initiated by strategic runs. The results using this measure were very similar, in terms of sign and magnitude of the coefficients as well as their statistical significance, to those we obtained with the regular effective spread measure that includes all NASDAQ trades (*EffSprd*).

## VI. Related Literature

Our paper can be viewed from two related angles: (i) speed of information dissemination and activity in financial markets, and (ii) high-frequency trading (or algorithmic trading in general) and its impact on the market environment.

Regarding speed, Hendershott and Moulton (2011) look at the introduction of the NYSE's Hybrid Market in 2006, which expanded automatic execution and reduced the execution time for NYSE market orders from ten seconds to under a second. They find that this reduction in latency resulted in worsened liquidity (e.g., spreads increased) but improved informational efficiency. However, Riordan and Storkenmaier (2012) find that a reduction in latency (from 50 to 10 ms) on the Deutsche Boerse' Xetra system is associated with improved liquidity. It could be that the impact of a change in latency on market quality depends on how exactly it affects competition among liquidity suppliers (e.g., the entrance of electronic market makers who can add liquidity but also crowed out traditional liquidity providers) and the level of sophistication of liquidity demanders (e.g., their adoption of algorithms to implement dynamic limit order strategies that can both supply and demand liquidity).[17]

The literature on algorithmic trading seeks both to establish stylized facts related to algorithmic activity and to evaluate their impact on the market. Gsell (2008) shows that the majority of orders generated by algorithms on the German Xetra system demand rather than supply liquidity and are smaller than those sent by human traders, while Groth

[17] Cespa and Foucault (2008) and Easley, O'Hara, and Yang (2010) provide theoretical models in which some traders observe market information with a delay. The two papers employ rather different modeling approaches resulting in somewhat conflicting implications on the impact of differential information latency on the cost of capital, liquidity, and the efficiency of prices. Boulatov and Dierker (2007) investigate information latency from the exchange's perspective: how can the exchange maximize data revenue? Their theoretical model suggests that selling real-time data can be detrimental to liquidity but at the same time enhances the informational efficiency of prices. Pagnotta and Philippon (2012) model speed as a differentiating attribute of competing exchanges. Moallemi and Sağlam (2010) discuss optimal order placement strategy for a seller facing random exogenous buyer arrivals. In their model, the seller pursues a pegging strategy, and the delayed monitoring caused by latency leads to costly tracking errors.

(2009) finds that algorithmic orders have a higher execution rate than non-algorithmic orders. Gsell and Gomber (2008) show evidence consistent with pegging strategies on Xetra, while Prix, Loistl, and Huetl (2007) note that there are certain regularities in the activity of these algorithms. Hendershott and Riordan (2009) look at the 30 DAX stocks and find that algorithmic trades have a larger price impact than non-algorithmic trades and seem to contribute more to price discovery. Chaboud, Chiquoine, Hjalmarsson, and Vega (2009) look at algorithmic trading in the interdealer foreign exchange market and find no evidence of a causal relationship between algorithmic trading and increased exchange rate volatility. Boehmer, Fong, and Wu (2012a, 2012b) look at the impact of algorithmic trading across 39 exchanges. They conclude that greater intensity of algorithmic trading increases short-term volatility, but improves liquidity and informational efficiency. They also find that more algorithmic trading is associated with a decline in equity capital in the following year, mainly driven by an increase in repurchase activity.

Hendershott, Jones, and Menkveld (2011) use the arrival rate of electronic messages on the NYSE as a measure of combined agency and proprietary algorithmic activity. Using an event study approach around the introduction of autoquoting by the NYSE in 2003, the authors find that increase in normalized message count (their proxy for algorithmic trading) impacts liquidity only for large stocks. For these stocks, quoted and effective spreads decline, while quoted depth decreases. The largest stocks also experience improved price discovery. We, on the other hand, find an improvement in market quality using all measures, including depth and short-term volatility, and for all stocks rather than just the largest stocks.[18] Two considerations could account for the difference in findings. Firstly, our measure of low-latency trading is designed to capture the activity of high-frequency proprietary algorithms rather than that of agency algorithms. Secondly, prior to the NYSE's introduction of Hybrid Market in 2006,

[18] The average market capitalization (in billion dollars) of sample quintiles reported in Table 1 of Hendershott, Jones, and Menkveld (2009) is 28.99, 4.09, 1.71, 0.90, and 0.41. This corresponds rather well to our sample where the average market capitalization of quintiles is 21.4, 3.8, 2.1, 1.4, and 1.0, though we may have fewer very large and very small stocks compared to their sample.

specialists may have faced less competition from high-frequency proprietary algorithms. The 2003 autoquoting change, therefore, may have mostly affected the activity of agency algorithms.

In a set of contemporaneous papers, Brogaard (2011a, 2011b, 2011c) investigates the impact of high-frequency trading on market quality using two special datasets of 120 stocks: one from NASDAQ containing the activity of 26 high-frequency traders and the other from BATS with 25 high-frequency traders. He reports that high-frequency traders contribute to liquidity provision in the market, that their trades help price discovery more than trades of other market participants, and that their activity appears to lower volatility. Brogaard's results, therefore, complement our findings on market quality measures in Section V, which is especially important given two differences in the design of our study compared to his. First, Brogaard's data covers only a subset of firms that utilize low-latency algorithms.[19] Since our measure of low-latency trading relies on imputed strategic runs, we are more likely to capture a broader picture of high-frequency activity. Second, Brogaard's analysis does not focus on periods of market stress. His most detailed data is available for only one week in February 2010 when the NASDAQ Composite Index was basically flat, while our 2008 sample provides insights on what happens at times of declining prices and heightened uncertainty. The ability to study low-latency activity during a stressful period for the market is especially important when the conclusion from the analysis of "normal times" is that these traders improve, rather than harm, market quality.

We note, though, that traders engaged in low-latency activity could impact the market in a negative fashion at times of extreme market stress. The joint CFTC/SEC report regarding the "flash crash" of May 6, 2010, presents a detailed picture of such an event. The report notes that several high-frequency traders in the equity markets scaled down, stopped, or significantly curtailed their trading at some point during this episode.

[19] His data do not include two types of proprietary traders that utilize low-latency algorithms. First, they lack the proprietary trading desks of larger, integrated firms like Goldman Sachs or JP Morgan. Second, they ignore small firms that use direct access brokers (such as Lime Brokerage or Swift Trade) that specialize in providing services to high-frequency traders.

Furthermore, some of the high-frequency traders escalated their aggressive selling during the rapid price decline, removing significant liquidity from the market and hence contributing to the decline. Similarly, Kirilenko, Kyle, Samadi, and Tuzun (2011) investigate the behavior of high-frequency trading firms in the futures market during the flash crash. They define "high-frequency traders" in the S&P 500 E-mini futures contract as those traders that execute a large number of daily transactions and fit a certain profile of intraday and end-of-day net positions. The authors identify 16 high-frequency traders using this definition, and conclude that while these traders did not trigger the flash crash, their responses exacerbated market volatility during the event. Our study suggests that such behavior is not representative of the manner in which low-latency activity impacts market conditions outside of such extreme episodes.

Lastly, Hendershott and Riordan (2011) use the NASDAQ HFT dataset to investigate the role high-frequency trading plays in price discovery. They estimate a model of price formation and report that when high-frequency trading firms trade by demanding liquidity, they do so in the direction of the permanent price changes and in the opposite direction to transitory price changes. Hence, they conclude that high-frequency traders help price efficiency.

Several recent theoretical papers attempt to shed light on the potential impact of high-frequency trading in financial markets (Citanic and Kirilenko (2010), Gerig and Michayluk (2010), Hoffmann (2010), Jovanovic and Menkveld (2010), Biais, Foucault, and Moinas (2011), Cartea and Penalva (2011), Cohen and Szpruch (2011), Jarrow and Protter (2011), and Martinez and Rosu (2011)). Some of these papers have specific implications as to the relationships between high-frequency trading and liquidity or volatility, which we investigate empirically.

For example, Gerig and Michalyuk (2010) assume that automated liquidity providers are more efficient than other market participants in extracting pricing-relevant information from multiple securities. By using information from one security to price another security, these high-frequency traders are able to offer better prices, lowering the transactions costs of investors in the market. Hoffman (2010) introduces fast traders into the limit order book model of Foucault (1999). Their presence can (in some cases) lower transactions costs due to increased competition in liquidity supply. Cartea and Penalva (2011) construct a model in the spirit of Grossman and Miller (1988) except that they add high-frequency traders who interject themselves between the liquidity traders and the market makers. In equilibrium, liquidity traders are worse off in the presence of high-frequency traders and the volatility of market prices increases.

In general, the theoretical models demonstrate that high-frequency traders can impact the market environment (and other investors) positively or negatively depending on the specific assumptions regarding their strategies and the assumed structure of the economy (see, for example, the predictions in Jovanovic and Menkveld (2010) and Biais, Foucault, and Moinas (2011)). Since different types of proprietary algorithms may employ different strategies, a theoretical model that focuses on one type of strategy may shed light on the specific impact of such a strategy, but may not predict the overall effect that empirical studies find because the mixture of strategies in actual markets may overwhelm the effect of one strategy or the other. As such, while our results are more consistent with some models than others, we do not view them as necessarily suggesting that certain models are wrong. Rather, our results could point to the relative dominance of a subset of high-frequency traders who peruse certain strategies that improve market quality.

## VII. Conclusions

Our paper makes two significant contributions. First, we develop a measure of low-latency activity using publicly-available data that can be used to investigate the impact of high-frequency trading on the market environment. Second, we study the impact that low-latency activity has on market quality both during normal market conditions and during a period of declining prices and heightened economic uncertainty. Our conclusion is that in the current market structure for equities, increased low-latency activity improves traditional yardsticks of market quality such as liquidity and short-term volatility. Of particular importance is our finding that at times of falling prices and anxiety in the market, the nature of the millisecond environment and the positive influence of low-latency activity on market quality remains. However, we cannot rule out the possibility of a sudden and severe market condition in which high-frequency traders contribute to a market failure. The experience of the "flash crash" in May of 2010 demonstrates that such fragility is certainly possible when a few big players step aside and nobody remains to post limit orders. While our results suggest that market quality has improved, we believe it is as yet an unresolved question whether low-latency trading increases the episodic fragility of markets, and we hope that future research will shed light on this issue.

The millisecond environment we describe—with its clock-time periodicities, trading that responds to market events over millisecond horizons, and algorithms that "play" with one another—constitutes a fundamental change from the manner in which stock markets operated even a few years ago. Still, the economic issues associated with latency in financial markets are not new, and the private advantage of relative speed as well as concerns over the impact of fast traders on prices were noted well before the advent of our current millisecond environment.[20] The early advocates of electronic markets generally envisioned arrangements wherein all traders would enjoy equal access (see Mendelson and Peake (1979), for example). We believe that it is important to recognize that guaranteeing equal access to market data when the market is both continuous and fragmented (as presently in the U.S.) may be physically impossible.

The first impediment to equal access is the geographical dispersion of traders (see Gode and Sunder (2000)). Our evidence on the speed of execution against improved quotes suggests that some players are responding within 2-3 ms, which is faster than it would take for information to travel from New York to Chicago and back (1440 miles) even at the speed of light (about 8 ms). While co-location could be viewed as the ultimate equalizer of dispersed traders, it inevitably leads to the impossibility of achieving equal

[20] Barnes (1911) describes stock brokers who, in the pre-telegraph era, established stations on high points across New Jersey and used semaphore and light flashes to transmit valuable information between New York and Philadelphia. He notes that some of the mysterious movements in the stock markets of Philadelphia and New York were popularly ascribed to these brokers.

access in fragmented markets. Since the same stock is traded on multiple trading venues, a co-located computer near the servers of exchange A would be at a disadvantage in responding to market events in the same securities on exchange B compared to computers co-located with exchange B. Unless markets change from continuous to periodic, some traders will always have lower latency than others. It is of special significance, therefore, that our findings in this paper suggest that increased low-latency activity need not invariably work to the detriment of long-term investors in the post-Reg NMS market structure for U.S. equities.

## References

Admati, R. Anat, and Paul Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1(1), 3-40.

Barnes, A. W., 1911, *History of the Philadelphia Stock Exchange*, Philadelphia, Cornelius Baker.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2011, Equilibrium high frequency trading, Working paper, Toulouse School of Economics.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu, 2012a, International evidence on algorithmic trading, Working paper, EDHEC Business School.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu, 2012b, Algorithmic trading and changes in firms' equity capital, Working paper, EDHEC Business School.

Boulatov, Alex, and Martin Dierker, 2007, Pricing prices, Working paper, University of Houston.

Brock, William A., and Allan W. Kleidon, 1992, Periodic market closure and trading volume, *Journal of Economic Dynamics and Control* 16, 451-489.

Brogaard, Jonathan A., 2011a, The activity of high frequency traders, Working paper, University of Washington.

Brogaard, Jonathan A., 2011b, High frequency trading and market quality, Working paper, University of Washington.

Brogaard, Jonathan A., 2011c, High frequency trading and volatility, Working paper, University of Washington.

Cartea, Alvaro, and Jose Penalva, 2011, Where is the value in high frequency trading? Working paper, Universidad Carlos III de Madrid.

Cespa, Giovanni, and Thierry Foucault, 2008, Insiders-outsiders, transparency, and the value of the ticker, Working paper, Queen Mary University of London and HEC.

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2009, Rise of the machines: Algorithmic trading in the foreign exchange markets, Working paper, Board of Governors of the Federal Reserve System.

Citanic and Kirilenko, 2010, High frequency traders and asset prices, Working paper, Caltech.

Cohen, Samuel N., and Lukasz Szpruch, 2011, A limit order book model for latency arbitrage, Working paper, University of Oxford.

Cragg, John G., and Stephen G. Donald, 1993, Testing identifiability and specification in instrumental variable models, *Econometric Theory* 9, 222-240.

Driscoll, John C., and Aart C. Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, *Review of Economics and Statistics* 80, 549-560.

Easley, David, Maureen O'Hara, and Liyan Yang, 2010, Differential access to price information in financial markets, Working paper, Cornell University.

Gerig, Austin, and David Michayluk, 2010, Automated liquidity provision and the demise of traditional market making, Working paper, University of Technology, Sydney.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Gode, Dhananjay K. and Shyam Sunder, 2000, On the impossibility of equitable continuously-clearing markets with geographically distributed traders, Working paper, New York University.

Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure, *Journal of Finance* 43, 617–633.

Groth, Sven S., 2009, Further evidence on “Technology and liquidity provision: The blurring of Tradition Definitions,” Working paper, Goethe University, Frankfurt am Main.

Gsell, Markus, 2009, Algorithmic activity on Xetra, *Journal of Trading* 4, 74-86

Gsell, Markus, and Peter Gomber, 2008, Algorithmic trading versus human traders—Do they behave different in securities markets? Working paper, Goethe University, Frankfurt am Main.

Hasbrouck, Joel, and Gideon Saar, 2009, Technology and liquidity provision: The blurring of traditional definitions, *Journal of Financial Markets* 12, 143-172.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1-33.

Hendershott, Terrence, and Pamela C. Moulton, 2011, Automation, speed, and stock market quality: The NYSE's Hybrid, *Journal of Financial Markets* 14, 568-604.

Hendershott, Terrence, and Ryan Riordan, 2009, Algorithmic trading and information, Working paper, University of California at Berkeley.

Hendershott, Terrence, and Ryan Riordan, 2011, High-frequency trading and price discovery, Working paper, University of California at Berkeley.

Hoffmann, Peter, 2010, Algorithmic trading in a dynamic limit order market, Working paper, Universitat Pompeu Fabra.

Jarrow, Robert A., and Philip Protter, 2011, A dysfunctional role of high frequency trading in electronic markets, Working paper, Cornell University.

Jovanovic, Boyan, and Albert J. Menkveld, 2010, Middlemen in limit-order markets, Working paper, New York University.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2011, The flash crash: The impact of high frequency trading on an electronic market, Working paper, CFTC and the University of Maryland.

Kosinski, R. J., 2010, A literature review on reaction time, Working paper, Clemson University.

Martinez, Victor, and Ioanid Rosu, 2011, High-frequency traders, news and volatility, Working paper, Baruch College.

Mendelson, Morris, and Junius W. Peake, 1979, The ABCs of trading on a national market system, *Financial Analysts Journal* 35, 31-34+27-42.

Menkveld, Albert J., 2011, High frequency trading and the *new-market* makers, Working paper, VU University Amsterdam.

Moallemi, Ciamac C., and Mehmet Sağlam, 2010, The cost of latency, Working paper, Columbia University.

Pagnotta, Emiliano, and Thomas Philippon, 2012, Competing on speed, Working paper, New York University.

Prix, Johannes, Otto Loistl, and Michael Huetl, 2007, Algorithmic trading patterns in Xetra orders, *European Journal of Finance* 13, 717-739.

Riordan, Ryan, and Andreas Storkenmaier, 2012, Latency, liquidity, and price discovery, *Journal of Financial Markets,* forthcoming.

Securities and Exchange Commission, 2010, *Concept Release on Equity Market Structure* (Release No. 34-61358).

Stock, James H., and Motohiro Yogo, 2005, Testing for weak instruments in linear IV regression, in D.W.K. Andrews, and J.H. Stock, Eds, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rotenberg* (Cambridge: Cambridge University Press).

Thompson, Samuel B., 2011, Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* 99, 1-10.

U.S. Commodities Futures Trading Commission, and U.S. Securities and Exchange Commission, 2010, Findings regarding the market events of May 6, 2010, (Washington D.C.).

# Table 1

## Summary Statistics

This table presents summary statistics for the stocks in our sample. The universe of stocks used in the study is comprised of the 500 largest stocks by market capitalization on September 28, 2007. We investigate trading in these stocks in two sample periods: (i) October 2007 (23 trading days), and (ii) June 2008 (21 trading days). Since the main econometric analysis in the paper requires sufficient level of activity in the stocks, we apply the following screen to the stocks in each sample period: A firm is rejected if the proportion of 10-minute intervals with fewer than 250 messages is above 10%. A "message" for the purpose of this screen could be a submission, a cancellation, or an execution of a limit order. After applying the screen, our sample consists of 351 stocks in the October 2007 sample period and 399 stocks in the June 2008 sample period. In Panel A we report summary statistics from the CRSP database. *MktCap* is the market capitalization of the firms computed using closing prices on the last trading day prior to the start of the sample period. *ClsPrice* is the average closing price, *AvgVol* is the average daily share volume, and *AvgRet* is the average daily return. These variables are averaged across time for each firm, and the table entries refer to the sample distribution of these firm-averages. Panel B presents summary statistics from the NASDAQ market computed using TotalView-ITCH data. We report the average daily number of limit orders submitted and cancelled (or partially cancelled), marketable orders executions, and the average daily number of shares executed. The summary measures for the limit order book include the time-weighted average depth in the book, the time-weighted average depth near current market prices (i.e., within 10 cents of the best bid or ask prices), and the time-weighted average dollar quoted spread (the distance between the bid and ask prices). We also report the effective (half) spread, defined as transaction price (quote midpoint) minus the quote midpoint (transaction price) for a buy (sell) marketable order, averaged across all transactions.

Panel A: CRSP Summary Statistics

| | 2007 | | | | 2008 | | | |
|---|---|---|---|---|---|---|---|---|
| | *MktCap* *($Million)* | *ClsPrice* *($)* | *AvgVol* *(1,000s)* | *AvgRet* *(%)* | *MktCap* *($Million)* | *ClsPrice* *($)* | *AvgVol* *(1,000s)* | *AvgRet* *(%)* |
| Mean | 6,609 | 37.09 | 3,172 | 0.109 | 5,622 | 31.88 | 2,931 | -0.565 |
| Median | 2,054 | 29.08 | 1,074 | 0.130 | 1,641 | 24.96 | 1,111 | -0.516 |
| Std | 20,609 | 41.54 | 8,083 | 0.570 | 19,348 | 38.93 | 6,410 | 0.615 |
| Min | 789 | 2.22 | 202 | -2.675 | 286 | 2.32 | 112 | -3.449 |
| Max | 275,598 | 635.39 | 77,151 | 1.933 | 263,752 | 556.32 | 74,514 | 0.817 |

Panel B. NASDAQ (TotalView-ITCH) Summary Statistics

| | | Number of Limit Order Submissions | Number of Limit Order Cancellations | Number of Marketable Order Executions | Shares Executed (1,000s) | Depth (1,000s) | Near Depth (1,000s) | Quoted Spread ($) | Effective Spread ($) |
|---|---|---|---|---|---|---|---|---|---|
| 2007 | Mean | 45,508 | 40,943 | 3,791 | 1,400 | 486 | 57 | 0.034 | 0.025 |
| | Median | 26,862 | 24,015 | 2,482 | 548 | 147 | 11 | 0.025 | 0.019 |
| | Std | 73,705 | 68,204 | 4,630 | 3,231 | 1,616 | 257 | 0.032 | 0.021 |
| | Min | 9,658 | 8,013 | 695 | 130 | 26 | 1 | 0.010 | 0.009 |
| | Max | 985,779 | 905,629 | 62,216 | 32,305 | 15,958 | 3,110 | 0.313 | 0.214 |
| 2008 | Mean | 54,287 | 50,040 | 3,694 | 1,203 | 511 | 43 | 0.035 | 0.023 |
| | Median | 34,658 | 31,426 | 2,325 | 483 | 154 | 10 | 0.023 | 0.016 |
| | Std | 61,810 | 56,728 | 4,676 | 2,618 | 1,767 | 152 | 0.041 | 0.024 |
| | Min | 8,889 | 7,983 | 291 | 42 | 20 | 1 | 0.010 | 0.008 |
| | Max | 593,143 | 525,346 | 61,013 | 32,406 | 25,004 | 2,482 | 0.462 | 0.257 |

# Table 2

## Examples of Strategic Runs for Ticker Symbol ADCT on October 2, 2007

This table presents examples of “strategic runs,” which are linked submissions, cancellations, and executions that are likely to be parts of a dynamic strategy of a trading algorithm. The examples are taken from activity in one stock (ATC Telecommunications, ticker symbol ADCT) on October 2, 2007. We identify the existence of these strategic runs by imputing links between different submissions, cancellations, and executions based on direction, size, and timing. In the two cases presented below, the activity in the table constitutes all messages in this stock (i.e., there are no intervening messages that are unrelated to these strategic runs). In Panel A, we present order activity starting around 9:51:57am where two algorithms “play” with each other (i.e., they submit and cancel messages in response to one another). The messages sent by the second algorithm are highlighted in the table. The algorithms are active for one minute and 12 seconds, sending 137 messages (submissions and cancellations) to the market. In Panel B we present order activity starting around 9:57:18am where one algorithm submits and cancels orders. The algorithm is active for one minute and eighteen seconds, sending 142 messages (submissions and cancellations) to the market.

Panel A: ADCT Order Activity Starting 09:51:57.849

| Time | Message | B/S | Shares | Price | Bid | Offer |
|---|---|---|---|---|---|---|
| 09:51:57.849 | Submission | Buy | 100 | 20.00 | 20.03 | 20.05 |
| 09:52:13.860 | Submission | Buy | 300 | 20.03 | 20.03 | 20.04 |
| 09:52:16.580 | Cancellation | Buy | 300 | 20.03 | 20.03 | 20.04 |
| 09:52:16.581 | Submission | Buy | 300 | 20.03 | 20.03 | 20.04 |
| 09:52:23.245 | Cancellation | Buy | 100 | 20.00 | 20.04 | 20.05 |
| 09:52:23.245 | Submission | Buy | 100 | 20.04 | 20.04 | 20.05 |
| 09:52:23.356 | Cancellation | Buy | 300 | 20.03 | 20.04 | 20.05 |
| 09:52:23.357 | Submission | Buy | 300 | 20.04 | 20.04 | 20.05 |
| 09:52:26.307 | Cancellation | Buy | 300 | 20.04 | 20.05 | 20.07 |
| 09:52:26.308 | Submission | Buy | 300 | 20.05 | 20.05 | 20.07 |
| 09:52:29.401 | Cancellation | Buy | 300 | 20.05 | 20.04 | 20.07 |
| 09:52:29.402 | Submission | Buy | 300 | 20.04 | 20.04 | 20.07 |
| 09:52:29.402 | Cancellation | Buy | 100 | 20.04 | 20.04 | 20.07 |
| 09:52:29.403 | Submission | Buy | 100 | 20.00 | 20.04 | 20.07 |
| 09:52:32.665 | Cancellation | Buy | 100 | 20.00 | 20.04 | 20.07 |
| 09:52:32.665 | Submission | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.672 | Cancellation | Buy | 100 | 20.05 | 20.04 | 20.07 |
| 09:52:32.678 | Submission | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.707 | Cancellation | Buy | 100 | 20.05 | 20.04 | 20.07 |
| 09:52:32.708 | Submission | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.717 | Cancellation | Buy | 100 | 20.05 | 20.04 | 20.07 |
| 09:52:32.745 | Cancellation | Buy | 300 | 20.04 | 20.04 | 20.07 |
| 09:52:32.745 | Submission | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.746 | Submission | Buy | 300 | 20.05 | 20.05 | 20.07 |
| 09:52:32.747 | Cancellation | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.772 | Submission | Buy | 100 | 20.02 | 20.05 | 20.07 |
| 09:52:32.776 | Cancellation | Buy | 300 | 20.05 | 20.04 | 20.07 |
| 09:52:32.777 | Cancellation | Buy | 100 | 20.02 | 20.04 | 20.07 |
| 09:52:32.777 | Submission | Buy | 300 | 20.04 | 20.04 | 20.07 |
| 09:52:32.778 | Submission | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.778 | Cancellation | Buy | 300 | 20.04 | 20.05 | 20.07 |
| 09:52:32.779 | Submission | Buy | 300 | 20.05 | 20.05 | 20.07 |
| 09:52:32.779 | Cancellation | Buy | 100 | 20.05 | 20.05 | 20.07 |
| 09:52:32.807 | Cancellation | Buy | 300 | 20.05 | 20.04 | 20.07 |
| 09:52:32.808 | Submission | Buy | 100 | 20.02 | 20.04 | 20.07 |
| 09:52:32.808 | Submission | Buy | 300 | 20.04 | 20.04 | 20.07 |
| 09:52:32.809 | Cancellation | Buy | 100 | 20.02 | 20.04 | 20.07 |

… the interaction between the two strategic runs continues for 95 additional messages until a limit order of the 300-share run is executed by an incoming marketable order at 09:53:09.365.

Panel B: ADCT Order Activity Starting 09:57:18.839

| Time | Message | B/S | Shares | Price | Bid | Ask |
|---|---|---|---|---|---|---|
| 09:57:18.839 | Submission | Sell | 100 | 20.18 | 20.11 | 20.14 |
| 09:57:18.869 | Cancellation | Sell | 100 | 20.18 | 20.11 | 20.14 |
| 09:57:18.871 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:18.881 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:18.892 | Submission | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:18.899 | Cancellation | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:18.902 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:18.911 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:18.922 | Submission | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:18.925 | Cancellation | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:18.942 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:18.954 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:18.958 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:18.961 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:18.973 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:18.984 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:18.985 | Submission | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:18.995 | Cancellation | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:18.996 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:19.002 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:19.004 | Submission | Sell | 100 | 20.16 | 20.11 | 20.14 |
| 09:57:19.807 | Cancellation | Sell | 100 | 20.16 | 20.11 | 20.13 |
| 09:57:19.807 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.451 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.461 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.471 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.480 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.481 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.484 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.499 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.513 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.521 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.532 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.533 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.542 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.554 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.562 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.571 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.581 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.592 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.601 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.611 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.622 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.667 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.671 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.681 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.742 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |
| 09:57:20.756 | Cancellation | Sell | 100 | 20.13 | 20.11 | 20.14 |
| 09:57:20.761 | Submission | Sell | 100 | 20.13 | 20.11 | 20.13 |

… the strategic run continues for 89 additional messages until it stops at 09:58:36.268.

# Table 3

## Strategic Runs

This table presents summary statistics for "strategic runs," which are linked submissions, cancellations, and executions that are likely to be parts of a dynamic strategy. The imputed links between different submissions, cancellations, and executions are based on direction, size, and timing. Specifically, when a cancellation is followed within 100 ms by a submission of a limit order in the same direction and for the same quantity, or by an execution in the same direction and for the same quantity, we impute a link between the messages. The methodology that tracks the strategic runs also takes note of partial executions and partial cancellations of orders. We sort runs into categories by length (i.e., the number of linked messages), and report information about the number of runs, messages, and executions (separately active and passive) within each category. An active execution is when the run ends with a marketable limit order that executes immediately. A passive execution is when a standing limit order that is part of a run is executed by an incoming marketable order. One run could potentially result in both a passive execution and an active execution if the passive execution did not exhaust the order, and the reminder was cancelled and resubmitted to generate an immediate active execution

| | Length | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Of | Runs | Runs | Messages | Messages | Active | Active | Passive | Passive | Total | Exec. |
| | Runs | (#) | (%) | (#) | (%) | Exec. (#) | Exec. Rate | Exec. (#) | Exec. Rate | Exec. (#) | Rate |
| 2007 | 3-4 | 20,294,968 | 44.11% | 79,695,563 | 15.67% | 1,954,468 | 9.63% | 4,981,521 | 24.55% | 6,922,605 | 34.11% |
| | 5-9 | 13,540,437 | 29.43% | 89,204,570 | 17.54% | 1,012,573 | 7.48% | 4,715,922 | 34.83% | 5,706,905 | 42.15% |
| | 10-14 | 5,650,415 | 12.28% | 65,294,103 | 12.84% | 267,517 | 4.73% | 1,808,138 | 32.00% | 2,069,393 | 36.62% |
| | 15-19 | 1,854,002 | 4.03% | 31,229,102 | 6.14% | 153,839 | 8.30% | 654,241 | 35.29% | 805,414 | 43.44% |
| | 20-99 | 4,337,029 | 9.43% | 153,384,374 | 30.16% | 301,266 | 6.95% | 1,575,876 | 36.34% | 1,871,244 | 43.15% |
| | 100+ | 333,308 | 0.72% | 89,735,209 | 17.65% | 26,039 | 7.81% | 116,465 | 34.94% | 141,962 | 42.59% |
| | All | 46,010,159 | 100.00% | 508,542,921 | 100.00% | 3,715,702 | 8.08% | 13,852,163 | 30.11% | 17,517,523 | 38.07% |
| 2008 | 3-4 | 31,012,203 | 46.24% | 122,325,313 | 19.53% | 2,427,326 | 7.83% | 5,552,338 | 17.90% | 7,970,158 | 25.70% |
| | 5-9 | 19,758,076 | 29.46% | 130,370,772 | 20.82% | 1,287,276 | 6.52% | 5,436,189 | 27.51% | 6,705,727 | 33.94% |
| | 10-14 | 7,941,089 | 11.84% | 91,486,978 | 14.61% | 385,902 | 4.86% | 2,186,628 | 27.54% | 2,566,974 | 32.33% |
| | 15-19 | 2,533,217 | 3.78% | 42,663,802 | 6.81% | 219,403 | 8.66% | 795,483 | 31.40% | 1,012,340 | 39.96% |
| | 20-99 | 5,583,768 | 8.33% | 191,395,420 | 30.56% | 398,771 | 7.14% | 1,712,015 | 30.66% | 2,105,346 | 37.70% |
| | 100+ | 239,751 | 0.36% | 48,084,901 | 7.68% | 15,541 | 6.48% | 62,838 | 26.21% | 78,171 | 32.61% |
| | All | 67,068,104 | 100.00% | 626,327,186 | 100.00% | 4,734,219 | 7.06% | 15,745,491 | 23.48% | 20,438,716 | 30.47% |

# Table 4

## Correlation of *RunsInProcess* with High-Frequency Trading from the NASDAQ HFT Dataset

This table presents the correlations between our measure of low-latency activity and measures of trading by 26 high-frequency traders from a special NASDAQ HFT dataset. To measure the intensity of low-latency activity in a stock in each ten-minute interval, we use the time-weighted average of the number of strategic runs of 10 messages or more the stock experiences in the interval (*RunsInProcess*). Let "H" denote high-frequency trading firms and "N" denotes other traders. Each trade in the NASDAQ HFT dataset is categorized by one of the following combinations: NH, NN, HN, or HH, where the first letter in each pair identifies the liquidity taker and the second identifies the liquidity supplier. We construct four measures: (i) Total HFT Executed Orders=NH+2*HH+HN, (ii) Total HFT Trades=NH+HH+HN, (iii) HFT Liquidity Supplied in Executed Orders=NH+HH, and (iv) Net HFT Liquidity Supplied in Executed Orders=NH. The first two are measures of the overall trading activity of high-frequency trading firms. The last two measures denote liquidity supplied by high-frequency trading firms. For each measure, we calculate two variants: one in terms of share volume and the other in terms of number of orders. Of the 120 firms in the HFT dataset, 60 are NASDAQ stocks for which we use our order-level data to construct the *RunsInProcess* measure. The correlations are computed for our second sample period, June 2008, over all 10-minute intervals for all stocks (60*819=49,410 observations). P-values are computed against the null hypothesis of zero correlation.

| | | *RunsInProcess*<br>Spearman Corr | p-value | *RunsInProcess*<br>Pearson Corr | p-value |
|---|---|---|---|---|---|
| Total HFT Executed Orders | Shares | 0.812 | (<.001) | 0.654 | (<.001) |
| | Orders | 0.809 | (<.001) | 0.658 | (<.001) |
| Total HFT Trades | Shares | 0.818 | (<.001) | 0.666 | (<.001) |
| | Orders | 0.814 | (<.001) | 0.644 | (<.001) |
| HFT Liquidity Supplied in Executed Orders | Shares | 0.817 | (<.001) | 0.682 | (<.001) |
| | Orders | 0.810 | (<.001) | 0.634 | (<.001) |
| Net HFT Liquidity Supplied in Executed Orders | Shares | 0.816 | (<.001) | 0.685 | (<.001) |
| | Orders | 0.809 | (<.001) | 0.643 | (<.001) |

# Table 5

## Low-Latency Trading and Market Quality

This table presents analysis of the manner in which low-latency trading affects market quality. To measure the intensity of low-latency activity in a stock in each ten-minute interval, we use the time-weighted average of the number of strategic runs of 10 messages or more the stock experiences in the interval (*RunsInProcess*). We use ITCH order-level data to compute several measures that represent different aspects of NASDAQ market quality: (i) *HighLow* is the highest midquote minus the lowest midquote in the same interval, (ii) *EffSprd* is the average effective spread (or total price impact) of a trade, defined as the average across transactions of the transaction price (quote midpoint) minus the quote midpoint (transaction price) for buy (sell) marketable orders, (iii) *Spread* is the time-weighted average quoted spread (ask price minus the bid price), and (iv) *NearDepth* is the time-weighted average number of (visible) shares in the book up to 10 cents from the best posted prices. Panel A presents Model I, which is a two-equation model for *RunsInProcess* and each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + e_{1,i,t}$$
$$RunsInProccess_{i,t} = b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + e_{2,i,t}$$

As an instrument for $RunsInProcess_{i,t}$ we use $RunsNotIND_{i,t}$, which is the average number of runs of 10 messages or more in the same interval for other stocks in our sample excluding: (1) the INDividual stock *i*, (2) stocks in the same INDustry as stock *i* (as defined by its four-digit SIC code), and (3) stocks in the same INDex as stock *i* (if stock *i* belongs to either the NASDAQ 100 Index or the S&P 500 Index). As an instrument for the market quality measures we use $EffSprdNotNas_{i,t}$, which is the average dollar effective spread computed using the TAQ database from trades executed in the same stock and during the same time interval on other (non-NASDAQ) trading venues. Panel B presents Model II, a single-equation model for each of the market quality measures:

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 TradingIntensity_{i,t} + e_{1,i,t}$$

where $TradingIntensity_{i,t}$ is defined as stock *i*'s total trading volume in the entire market (not just NASDAQ) immediately prior to interval *t* (i.e., in the previous 10 minutes). We estimate both specifications by pooling observations across all stocks and all time intervals. We standardize each variable by subtracting from each observation the stock-specific time-series average and dividing by the stock-specific time-series standard deviation. Hence, this formulation essentially implements a fixed-effects specification. We estimate the system using Two-Stage GMM, and consider two types of robust standard errors. The first type (Clust. p-value) is robust to arbitrary heteroskedasticity and clustering on two dimensions: (i) stocks, and (ii) time-intervals. Hence, the standard errors are robust to serial correlations in the time dimension (for each stock) and contemporaneous correlation of the errors across stocks. The second type (DK p-value) implements the estimator from Driscoll and Kraay (1998). The DK estimator is an extension of the Newey-West HAC estimator that is also robust to very general spatial dependence (i.e., contemporaneous correlation of the errors across stocks). We report the coefficients and the p-values for both estimators (against a two-sided alternative) side-by-side for the 2007 and 2008 sample periods.

Panel A: Estimates of Model I (with Instruments *RunsNotIND* and *EffSprdNotNAS*)

| | | 2007 | | | | 2008 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **$a_1$** | **$a_2$** | **$b_1$** | **$b_2$** | **$a_1$** | **$a_2$** | **$b_1$** | **$b_2$** |
| *HighLow* | Coef. | -0.350 | 0.476 | -0.063 | 0.497 | -0.475 | 0.452 | -0.125 | 0.464 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *Spread* | Coef. | -0.534 | 0.567 | -0.052 | 0.494 | -0.615 | 0.526 | -0.107 | 0.461 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *EffSprd* | Coef. | -0.203 | 0.382 | -0.079 | 0.500 | -0.143 | 0.219 | -0.264 | 0.475 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *NearDepth* | Coef. | 0.380 | -0.236 | 0.123 | 0.484 | 0.716 | -0.116 | 0.378 | 0.360 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |

Panel B: Estimates of Model II (with Instrument *RunsNotIND*)

| | | **2007** | | **2008** | |
|---|---|---|---|---|---|
| | | $\mathbf{a_1}$ | $\mathbf{a_2}$ | $\mathbf{a_1}$ | $\mathbf{a_2}$ |
| *HighLow* | Coef. | -0.396 | 0.285 | -0.568 | 0.195 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) |
| *Spread* | Coef. | -0.493 | 0.075 | -0.598 | 0.030 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) |
| *EffSprd* | Coef. | -0.231 | 0.048 | -0.199 | 0.013 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) |
| *NearDepth* | Coef. | 0.467 | -0.106 | 0.813 | -0.007 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (0.702) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (0.715) |

# Table 6

## Low-Latency Trading and Market Quality: Market Factors

This table looks at how low-latency trading affects market quality when we add to the specifications market return and volatility factors to rule out the possibility that our results are driven by the potential influence of omitted variables related to common factors. We add to each equation (i) the market return (in each 10-minute interval), and (ii) the absolute value of the market return. We use the NASDAQ 100 Index as a proxy for the market portfolio, and compute the 10-minute returns using the QQQQ Exchange Traded Fund that tracks the index. Panel A presents Model III, which is a two-equation model for *RunsInProcess* and each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + a_3 R_{QQQQ,t} + a_4 \left| R_{QQQQ,t} \right| + e_{1,i,t}$$

$$RunsInProccess_{i,t} = b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + a_3 R_{QQQQ,t} + a_4 \left| R_{QQQQ,t} \right| + e_{2,i,t}$$

where all variables are standardized as in Table 5 to implement a fixed effect specification. As an instrument for *RunsInProcess$_{i,t}$* we use *RunsNotIND$_{i,t}$*, which is the average number of runs of 10 messages or more in the same interval for other stocks in our sample excluding: (1) the INDividual stock *i*, (2) stocks in the same INDustry as stock *i* (as defined by its four-digit SIC code), and (3) stocks in the same INDex as stock *i* (if stock *i* belongs to either the NASDAQ 100 Index or the S&P 500 Index). As an instrument for the market quality measures we use *EffSprdNotNas$_{i,t}$*, which is the average dollar effective spread computed using the TAQ database from trades executed in the same stock and during the same time interval on other (non-NASDAQ) trading venues. Panel B presents Model IV, a single-equation model for each of the market quality measures:

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 TradingIntensity_{i,t} + a_3 R_{QQQQ,t} + a_4 \left| R_{QQQQ,t} \right| + e_{1,i,t}$$

where only *RunsNotIND* is used as an instrument. We estimate the models using Two-Stage GMM, and report robust p-values using both two-dimensional-clustering and the Driscoll-Kraay methodology. We report the coefficients and the p-values (against a two-sided alternative) for the 2007 and 2008 sample periods.

Panel A: Estimates of Model III (with Market Factors and Instruments *RunsNotIND* and *EffSprdNotNAS*)

| | | | **a$_1$** | **a$_2$** | **a$_3$** | **a$_4$** | **b$_1$** | **b$_2$** | **b$_3$** | **b$_4$** |
|---|---|---|---|---|---|---|---|---|---|---|
| **2007** | *HighLow* | Coef. | -0.240 | 0.428 | -0.032 | 0.251 | -0.075 | 0.501 | -0.002 | 0.028 |
| | | Clust. p-value | (<.001) | (<.001) | (0.024) | (<.001) | (<.001) | (<.001) | (0.592) | (<.001) |
| | | DK p-value | (<.001) | (<.001) | (0.027) | (<.001) | (<.001) | (<.001) | (0.632) | (<.001) |
| | *Spread* | Coef. | -0.471 | 0.539 | -0.001 | 0.139 | -0.059 | 0.496 | 0.001 | 0.017 |
| | | Clust. p-value | (<.001) | (<.001) | (0.962) | (<.001) | (<.002) | (<.001) | (0.871) | (0.003) |
| | | DK p-value | (<.001) | (<.001) | (0.966) | (<.001) | (<.001) | (<.001) | (0.888) | (0.002) |
| | *EffSprd* | Coef. | -0.167 | 0.365 | 0.051 | 0.074 | -0.089 | 0.502 | 0.005 | 0.015 |
| | | Clust. p-value | (<.001) | (<.001) | (0.012) | (0.002) | (<.001) | (<.001) | (0.163) | (0.008) |
| | | DK p-value | (<.001) | (<.001) | (0.014) | (0.002) | (<.001) | (<.001) | (0.192) | (0.003) |
| | *NearDepth* | Coef. | 0.342 | -0.221 | 0.044 | -0.095 | 0.141 | 0.485 | -0.006 | 0.022 |
| | | Clust. p-value | (<.001) | (<.001) | (0.004) | (<.001) | (<.001) | (<.001) | (0.179) | (<.001) |
| | | DK p-value | (<.001) | (<.001) | (0.002) | (<.001) | (0.002) | (<.001) | (0.199) | (<.001) |
| **2008** | *HighLow* | Coef. | -0.448 | 0.430 | 0.176 | 0.243 | -0.125 | 0.464 | 0.007 | 0.005 |
| | | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.442) | (0.625) |
| | | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.420) | (0.617) |
| | *Spread* | Coef. | -0.595 | 0.509 | 0.040 | 0.124 | -0.105 | 0.461 | -0.010 | -0.012 |
| | | Clust. p-value | (<.001) | (<.001) | (0.250) | (<.001) | (<.001) | (<.001) | (0.202) | (0.137) |
| | | DK p-value | (<.001) | (<.001) | (0.256) | (<.001) | (<.001) | (<.001) | (0.223) | (0.149) |
| | *EffSprd* | Coef. | -0.132 | 0.210 | 0.022 | 0.067 | -0.261 | 0.475 | -0.009 | -0.009 |
| | | Clust. p-value | (<.001) | (<.001) | (0.291) | (<.001) | (<.001) | (<.001) | (0.348) | (0.417) |
| | | DK p-value | (<.001) | (<.001) | (0.260) | (<.001) | (<.001) | (<.001) | (0.355) | (0.414) |
| | *NearDepth* | Coef. | 0.714 | -0.114 | -0.056 | -0.047 | 0.368 | 0.363 | 0.009 | -0.003 |
| | | Clust. p-value | (<.001) | (<.001) | (0.043) | (0.068) | (<.001) | (<.001) | (0.542) | (0.835) |
| | | DK p-value | (<.001) | (<.001) | (0.029) | (0.057) | (<.001) | (<.001) | (0.508) | (0.839) |

Panel B: Estimates of Model IV (with Market Factors and Instrument *RunsNotIND*)

| | | **2007** | | | | **2008** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **$a_1$** | **$a_2$** | **$a_3$** | **$a_4$** | **$a_1$** | **$a_2$** | **$a_3$** | **$a_4$** |
| *HighLow* | Coef. | -0.324 | 0.261 | -0.004 | 0.297 | -0.512 | 0.184 | 0.005 | 0.702 |
| | Clust. p-value | (<.001) | (<.001) | (0.762) | (<.001) | (<.001) | (<.001) | (0.860) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (0.782) | (<.001) | (<.001) | (<.001) | (0.846) | (<.001) |
| *Spread* | Coef. | -0.461 | 0.064 | -0.004 | 0.132 | -0.580 | 0.027 | -0.026 | 0.212 |
| | Clust. p-value | (<.001) | (<.001) | (0.796) | (<.001) | (<.002) | (<.001) | (0.198) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (0.825) | (<.001) | (<.001) | (<.001) | (0.163) | (<.001) |
| *EffSprd* | Coef. | -0.220 | 0.044 | -0.001 | 0.046 | -0.193 | 0.012 | -0.005 | 0.056 |
| | Clust. p-value | (<.001) | (<.001) | (0.829) | (0.002) | (<.001) | (<.001) | (0.403) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (0.856) | (<.001) | (<.001) | (<.001) | (0.353) | (<.001) |
| *NearDepth* | Coef. | 0.431 | -0.094 | 0.034 | -0.146 | 0.801 | -0.004 | -0.033 | -0.149 |
| | Clust. p-value | (<.001) | (<.001) | (0.085) | (0.068) | (<.001) | (0.802) | (0.410) | (0.004) |
| | DK p-value | (<.001) | (<.001) | (0.094) | (<.001) | (<.001) | (0.811) | (0.394) | (0.004) |

# Table 7

## Low-Latency Trading and Market Quality: Time-of-Day Effects

This table looks at how low-latency trading affects market quality when we add to the specifications time-of-day dummy variables to rule out the possibility that our results are driven by the potential influence of omitted variables related to intraday time patterns. Panel A presents Model V, which is a two-equation model for *RunsInProcess* and each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_0 + a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + a_3 DumAM_{i,t} + a_4 DumPM_{i,t} + e_{1,i,t}$$

$$RunsInProccess_{i,t} = b_0 + b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + b_3 DumAM_{i,t} + b_4 DumPM_{i,t} + e_{2,i,t}$$

where $DumAM_{i,t}$ and $DumPM_{i,t}$ are dummy variables for (respectively) 9:30am to 11:00am and 2:30pm to 4:00pm. As an instrument for $RunsInProcess_{i,t}$ we use $RunsNotIND_{i,t}$, which is the average number of runs of 10 messages or more in the same interval for other stocks in our sample excluding: (1) the INDividual stock *i*, (2) stocks in the same INDustry as stock *i* (as defined by its four-digit SIC code), and (3) stocks in the same INDex as stock *i* (if stock *i* belongs to either the NASDAQ 100 Index or the S&P 500 Index). As an instrument for the market quality measures we use $EffSprdNotNas_{i,t}$, which is the average dollar effective spread computed using the TAQ database from trades executed in the same stock and during the same time interval on other (non-NASDAQ) trading venues. Panel B presents Model VI, a single-equation model for each of the market quality measures:

$$MktQuality_{i,t} = a_0 + a_1 RunsInProcess_{i,t} + a_2 TradingIntensity_{i,t} + a_3 DumAM_{i,t} + a_4 DumPM_{i,t} + e_{1,i,t}$$

where only *RunsNotIND* is used as an instrument. All non-dummy variables are standardized as in Table 5, but a constant term has been added to reflect the midday period. We estimate the models using Two-Stage GMM, and report robust p-values using both two-dimensional-clustering and the Driscoll-Kraay methodology. We report the coefficients and the p-values (against a two-sided alternative) for the 2007 and 2008 sample periods.

Panel A: Estimates of Model V (with Time Dummies and Instruments *RunsNotIND* and *EffSprdNotNAS*)

| | | | $a_0$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $b_0$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2007** | *HighLow* | Coef. | -0.128 | -0.220 | 0.466 | 0.372 | 0.184 | -0.005 | -0.067 | 0.503 | 0.029 | -0.008 |
| | | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.418) | (<.001) | (<.001) | (0.069) | (0.595) |
| | | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.496) | (<.001) | (<.001) | (<.001) | (0.074) |
| | *Spread* | Coef. | -0.022 | -0.437 | 0.555 | 0.215 | -0.121 | 0.002 | -0.055 | 0.498 | 0.016 | -0.026 |
| | | Clust. p-value | (0.099) | (<.001) | (<.001) | (<.001) | (<.001) | (0.680) | (<.001) | (<.001) | (0.324) | (0.056) |
| | | DK p-value | (0.361) | (<.001) | (<.001) | (<.001) | (0.002) | (0.826) | (<.001) | (<.001) | (0.325) | (0.125) |
| | *EffSprd* | Coef. | -0.002 | -0.184 | 0.380 | 0.039 | -0.032 | 0.004 | -0.082 | 0.502 | 0.008 | -0.023 |
| | | Clust. p-value | (0.830) | (<.001) | (<.001) | (<.001) | (<.001) | (0.534) | (<.001) | (<.001) | (0.660) | (0.097) |
| | | DK p-value | (0.876) | (<.001) | (<.001) | (0.027) | (0.009) | (0.732) | (<.001) | (<.001) | (0.636) | (0.170) |
| | *NearDepth* | Coef. | 0.033 | 0.150 | -0.207 | -0.494 | 0.350 | -0.001 | 0.149 | 0.499 | 0.078 | -0.072 |
| | | Clust. p-value | (0.040) | (<.001) | (<.001) | (<.001) | (<.001) | (0.848) | (<.001) | (<.001) | (<.001) | (<.001) |
| | | DK p-value | (0.307) | (0.002) | (<.001) | (<.001) | (<.001) | (0.914) | (<.001) | (<.001) | (<.001) | (0.002) |
| **2008** | *HighLow* | Coef. | -0.194 | -0.383 | 0.417 | 0.508 | 0.335 | 0.015 | -0.120 | 0.479 | -0.004 | -0.061 |
| | | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.071) | (<.001) | (<.001) | (0.835) | (0.031) |
| | | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.499) | (<.001) |
| | *Spread* | Coef. | -0.052 | -0.566 | 0.518 | 0.160 | 0.063 | 0.033 | -0.095 | 0.475 | -0.049 | -0.094 |
| | | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (0.064) | (<.001) | (<.001) | (<.001) | (0.006) | (0.001) |
| | | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | *EffSprd* | Coef. | -0.006 | -0.144 | 0.218 | 0.011 | 0.013 | 0.037 | -0.232 | 0.485 | -0.063 | -0.099 |
| | | Clust. p-value | (0.308) | (<.001) | (<.001) | (0.349) | (0.299) | (<.001) | (<.001) | (<.001) | (0.001) | (0.001) |
| | | DK p-value | (0.017) | (<.001) | (<.001) | (0.038) | (0.005) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | *NearDepth* | Coef. | 0.037 | 0.160 | -0.137 | -0.646 | 0.486 | 0.025 | 0.361 | 0.473 | 0.169 | -0.275 |
| | | Clust. p-value | (0.001) | (0.015) | (<.001) | (<.001) | (<.001) | (0.003) | (<.001) | (<.001) | (<.001) | (<.001) |
| | | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |

Panel B: Estimates of Model VI (with Time Dummies and Instrument *RunsNotIND*)

| | | **2007** | | | | | **2008** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathbf{a}_0$ | $\mathbf{a}_1$ | $\mathbf{a}_2$ | $\mathbf{a}_3$ | $\mathbf{a}_4$ | $\mathbf{a}_0$ | $\mathbf{a}_1$ | $\mathbf{a}_2$ | $\mathbf{a}_3$ | $\mathbf{a}_4$ |
| *HighLow* | Coef. | -0.159 | -0.224 | 0.257 | 0.411 | 0.070 | -0.249 | -0.318 | 0.148 | 0.643 | 0.214 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (0.024) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (0.189) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *Spread* | Coef. | -0.120 | -0.309 | 0.060 | 0.351 | -0.171 | -0.168 | -0.324 | 0.012 | 0.401 | -0.064 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.130) | (<.001) | (0.003) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.094) | (<.001) | (0.011) |
| *EffSprd* | Coef. | -0.059 | -0.157 | 0.042 | 0.134 | -0.069 | -0.046 | -0.130 | 0.008 | 0.102 | -0.013 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.034) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.089) |
| *NearDepth* | Coef. | 0.043 | 0.200 | -0.089 | -0.524 | 0.394 | 0.064 | 0.152 | 0.004 | -0.708 | 0.498 |
| | Clust. p-value | (0.020) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.048) | (0.761) | (<.001) | (<.001) |
| | DK p-value | (0.228) | (0.006) | (<.001) | (<.001) | (<.001) | (0.006) | (0.067) | (0.775) | (<.001) | (<.001) |

# Table 8

## Low-Latency Trading and Market Quality by Size Quartiles

This table presents the results of a two-equation model of low-latency trading and market quality estimated separately for stocks in each firm-size quartile. As in Table 5, we estimate Model I for each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + e_{1,i,t}$$

$$RunsInProccess_{i,t} = b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + e_{2,i,t}$$

with *RunsNotIND* and *EffSprdNotNas* as the instruments. We standardize each variable by subtracting from each observation the stock-specific time-series average and dividing by the stock-specific time-series standard deviation. We estimate the model using Two-Stage GMM, and report in the table p-values using standard errors that are robust to arbitrary heteroskedasticity and clustering on two dimensions: (i) stocks, and (ii) time-intervals. Hence, the standard errors are robust to serial correlations in the time dimension (for each stock) and contemporaneous correlation of the errors across stocks. We report the coefficients and the p-values (against a two-sided alternative) side-by-side for the 2007 and 2008 sample periods.

| | | | **2007** | | | | **2008** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Dep. Var.** | | | $a_1$ | $a_2$ | $b_1$ | $b_2$ | $a_1$ | $a_2$ | $b_1$ | $b_2$ |
| *HighLow* | Q1 | Coef. | -0.413 | 0.427 | 0.012 | 0.505 | -0.658 | 0.409 | -0.173 | 0.342 |
| | (small) | Clust. p-value | (<.001) | (<.001) | (0.688) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q2 | Coef. | -0.352 | 0.470 | 0.054 | 0.560 | -0.568 | 0.421 | -0.182 | 0.377 |
| | | Clust. p-value | (<.001) | (<.001) | (0.011) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q3 | Coef. | -0.361 | 0.472 | -0.050 | 0.511 | -0.432 | 0.473 | -0.099 | 0.514 |
| | | Clust. p-value | (<.001) | (<.001) | (0.035) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q4 | Coef. | -0.297 | 0.520 | -0.204 | 0.434 | -0.292 | 0.514 | 0.017 | 0.691 |
| | (large) | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.466) | (<.001) |
| *Spread* | Q1 | Coef. | -0.641 | 0.539 | 0.009 | 0.506 | -0.819 | 0.471 | -0.149 | 0.339 |
| | (small) | Clust. p-value | (<.001) | (<.001) | (0.688) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q2 | Coef. | -0.543 | 0.579 | 0.044 | 0.563 | -0.731 | 0.505 | -0.150 | 0.374 |
| | | Clust. p-value | (<.001) | (<.001) | (0.013) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q3 | Coef. | -0.525 | 0.598 | -0.039 | 0.509 | -0.557 | 0.544 | -0.085 | 0.511 |
| | | Clust. p-value | (<.001) | (<.001) | (0.035) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q4 | Coef. | -0.456 | 0.553 | -0.187 | 0.422 | -0.411 | 0.597 | 0.015 | 0.691 |
| | (large) | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.456) | (<.001) |
| *EffSprd* | Q1 | Coef. | -0.204 | 0.331 | 0.014 | 0.504 | -0.158 | 0.161 | -0.459 | 0.358 |
| | (small) | Clust. p-value | (<.001) | (<.001) | (0.700) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q2 | Coef. | -0.171 | 0.413 | 0.061 | 0.556 | -0.150 | 0.197 | -0.407 | 0.394 |
| | | Clust. p-value | (<.001) | (<.001) | (0.013) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q3 | Coef. | -0.196 | 0.408 | -0.058 | 0.514 | -0.133 | 0.262 | -0.182 | 0.524 |
| | | Clust. p-value | (<.001) | (<.001) | (0.036) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q4 | Coef. | -0.242 | 0.355 | -0.296 | 0.429 | -0.121 | 0.281 | 0.031 | 0.690 |
| | (large) | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.464) | (<.001) |
| *NearDepth* | Q1 | Coef. | 0.437 | -0.181 | -0.028 | 0.509 | 0.875 | -0.072 | 0.562 | 0.196 |
| | (small) | Clust. p-value | (<.001) | (<.001) | (0.689) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q2 | Coef. | 0.453 | -0.210 | -0.125 | 0.581 | 0.712 | -0.087 | 0.577 | 0.247 |
| | | Clust. p-value | (<.001) | (<.001) | (0.016) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q3 | Coef. | 0.402 | -0.229 | 0.100 | 0.499 | 0.686 | -0.116 | 0.326 | 0.417 |
| | | Clust. p-value | (<.001) | (<.001) | (0.028) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Q4 | Coef. | 0.238 | -0.319 | 0.326 | 0.426 | 0.598 | -0.203 | -0.044 | 0.705 |
| | (large) | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.480) | (<.001) |

# Table 9

## Robustness Analysis: Sample Screen and Definition of Strategic Runs

This table presents robustness analysis for our main results in Table 5 on the manner in which low-latency trading affects market quality. In Panel A, we estimate Model I (as in Table 5) on a modified sample created by rejecting firms if the proportion of 10-minute intervals with fewer than 100 messages is above 10% (which is less stringent than the 250-message cutoff we use to generate our main sample). After applying the screen, the modified sample consists of 471 stocks in the October 2007 sample period and 456 stocks in the June 2008 sample period. We estimate the following two-equation simultaneous equation model for *RunsInProcess* and each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + e_{1,i,t}$$

$$RunsInProccess_{i,t} = b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + e_{2,i,t}$$

with instruments *RunsNotIND* and *EffSprdNotNas*. All variables are defined as in Table 5. In Panel B, we estimate Model I on our regular sample but use an alternative measure of low-latency activity: the time-weighted average of the number of strategic runs the stock experiences in the interval (*AllRunsInProcess*). Unlike our main measure (*RunsInProcess*), this alternative definition includes runs shorter than 10 messages. We estimate the following two-equation simultaneous equation model for *AllRunsInProcess* and each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_1 AllRunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + e_{1,i,t}$$

$$AllRunsInProccess_{i,t} = b_1 MktQuality_{i,t} + b_2 AllRunsNotIND_{i,t} + e_{2,i,t}$$

with instruments *AllRunsNotIND* and *EffSprdNotNas*. We estimate the models by pooling observations across all stocks and all time intervals. We standardize each variable by subtracting from each observation the stock-specific time-series average and dividing by the stock-specific time-series standard deviation. We estimate the system using Two-Stage GMM, and consider two types of standard errors that are robust to heteroskedasticity, serial correlation, and contemporaneous correlation of the errors across stocks: the first implements two-dimensional clustering and the second implements the estimator from Driscoll and Kraay (1998). We report the coefficients and the p-values for both estimators (against a two-sided alternative) side-by-side for the 2007 and 2008 sample periods.

Panel A: Estimates of Model I on Modified Sample Screen

| | | **2007** | | | | **2008** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **$a_1$** | **$a_2$** | **$b_1$** | **$b_2$** | **$a_1$** | **$a_2$** | **$b_1$** | **$b_2$** |
| | Coef. | -0.326 | 0.474 | -0.029 | 0.518 | -0.516 | 0.440 | -0.131 | 0.439 |
| *HighLow* | Clust. p-value | (<.001) | (<.001) | (0.029) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Coef. | -0.500 | 0.584 | -0.024 | 0.517 | -0.658 | 0.514 | -0.111 | 0.436 |
| *Spread* | Clust. p-value | (<.001) | (<.001) | (0.031) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Coef. | -0.192 | 0.405 | -0.035 | 0.520 | -0.153 | 0.217 | -0.273 | 0.451 |
| *EffSprd* | Clust. p-value | (<.001) | (<.001) | (0.031) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | Coef. | 0.323 | -0.231 | 0.060 | 0.513 | 0.708 | -0.107 | 0.410 | 0.334 |
| *NearDepth* | Clust. p-value | (<.001) | (<.001) | (0.025) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |

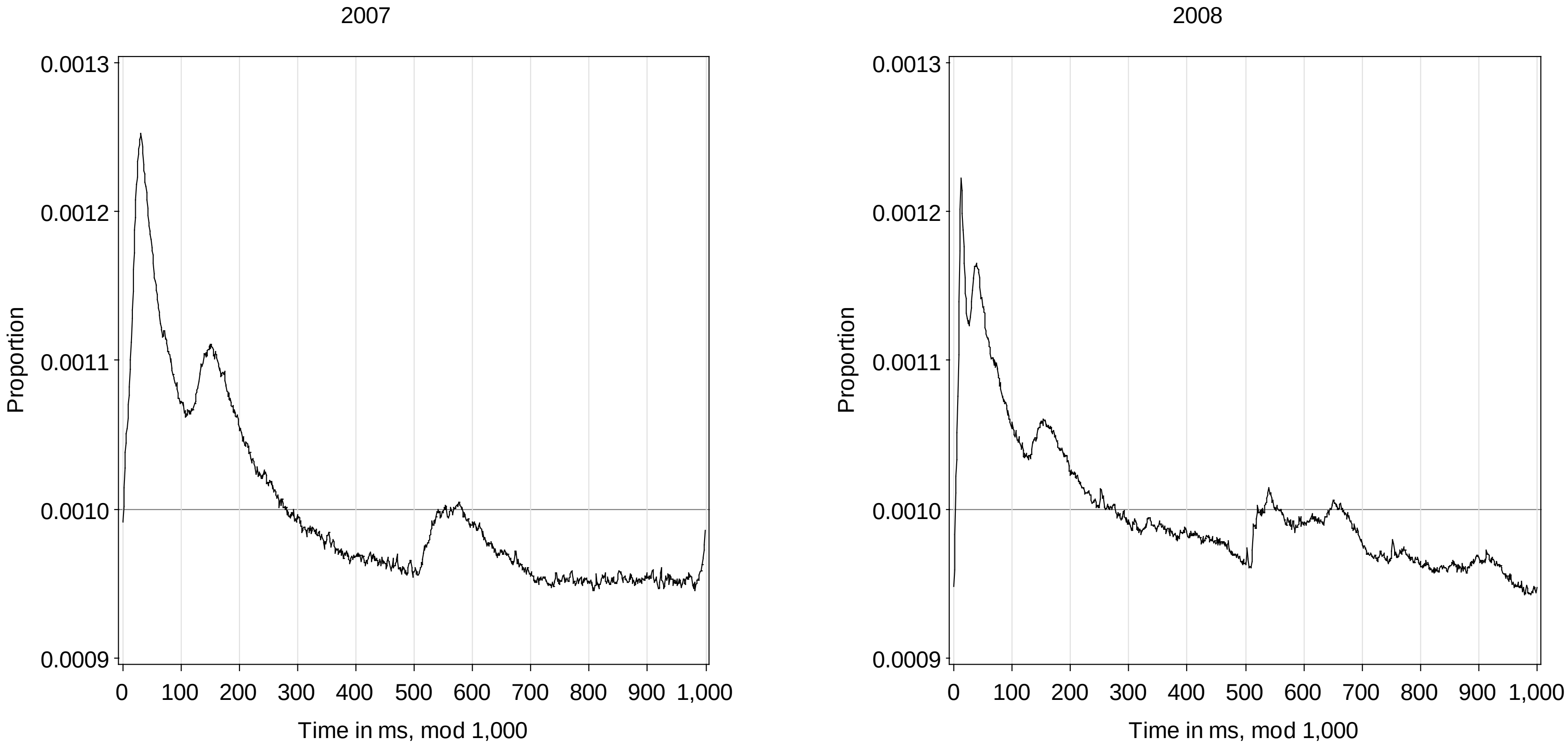
Panel B: Estimates of Model I using All Runs (with Instruments *EffSprdNotNAS* and *AllRunsNotIND*)

| | | **2007** | | | | **2008** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathbf{a_1}$ | $\mathbf{a_2}$ | $\mathbf{b_1}$ | $\mathbf{b_2}$ | $\mathbf{a_1}$ | $\mathbf{a_2}$ | $\mathbf{b_1}$ | $\mathbf{b_2}$ |
| *HighLow* | Coef. | -0.421 | 0.436 | -0.053 | 0.550 | -0.384 | 0.458 | -0.204 | 0.497 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *Spread* | Coef. | -0.614 | 0.517 | -0.044 | 0.547 | -0.544 | 0.518 | -0.177 | 0.487 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *EffSprd* | Coef. | -0.215 | 0.367 | -0.063 | 0.554 | -0.131 | 0.214 | -0.445 | 0.507 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| *NearDepth* | Coef. | 0.545 | -0.169 | 0.129 | 0.522 | 0.688 | -0.088 | 0.641 | 0.301 |
| | Clust. p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) |
| | DK p-value | (<.001) | (<.001) | (<.001) | (<.001) | (<.001) | (0.008) | (<.001) | (<.001) |

# Figure 1

## Clock-time Periodicities of Market Activity

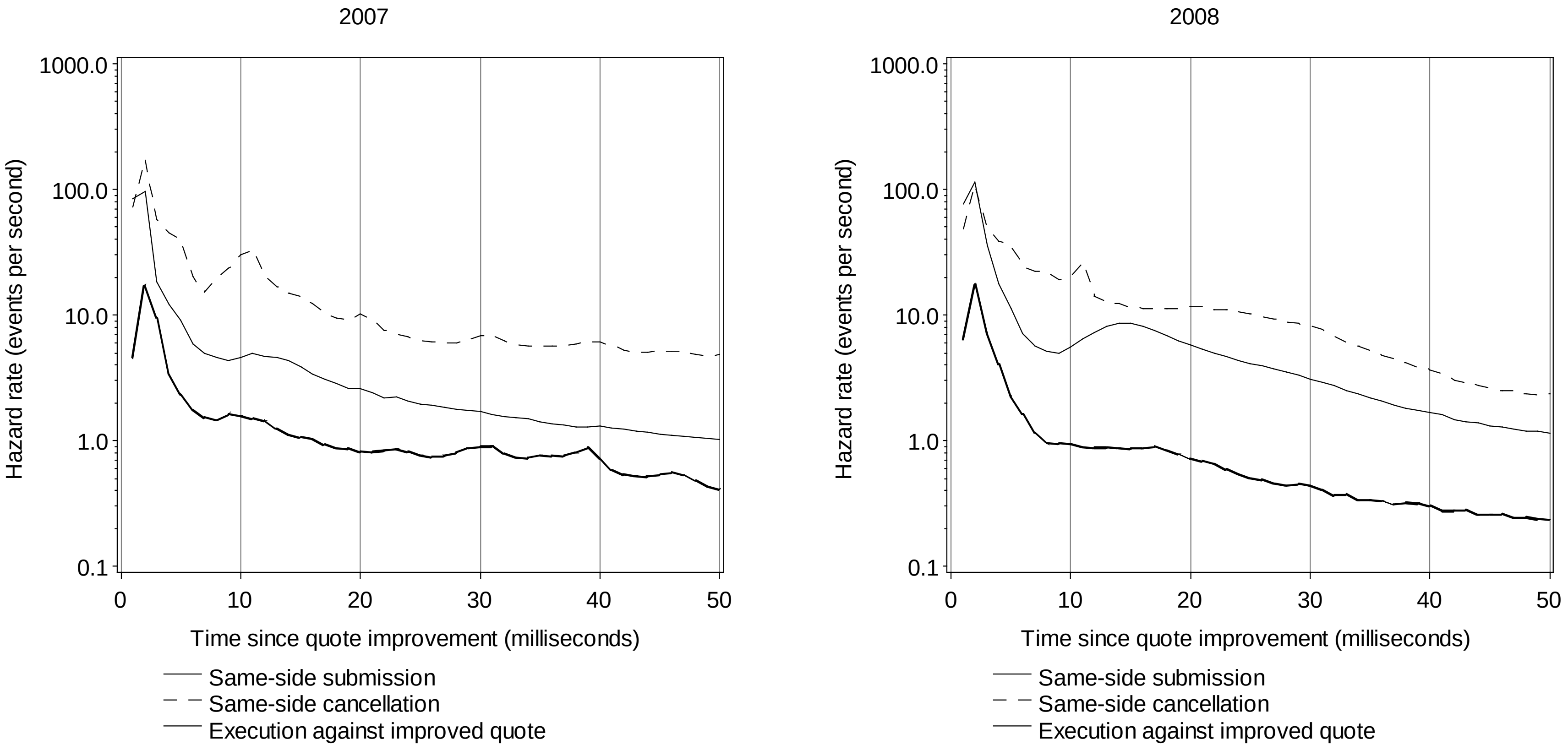
This figure presents clock-time periodicities in message arrival to the market. The original time stamps are milliseconds past midnight. The one-second remainder is the time stamp mod 1,000, i.e., the number of milliseconds past the one-second mark. We plot the sample distribution of one-second remainders side-by-side for the 2007 and 2008 sample periods. The horizontal lines in the graphs indicate the position of the uniform distribution (the null hypothesis).



# Figure 2

## Speed of Response to Market Events

This figure depicts response speeds subsequent to a specific market event. The market event is an improved quote via the submission of a new limit order—either an increase in the best bid price or a decrease in the best ask price. Subsequent to this market event, we estimate (separately) the hazard rates for three types of responses: (i) a limit order submission on the same side as the improvement (e.g., buy order submitted following an improvement in the bid price), (ii) a cancellation of a standing limit order on the same side, and (iii) an execution against the improved quote (e.g., the best bid price is executed by an incoming sell order). In all estimations, any event other than the one whose hazard rate is being estimated is taken as an exogenous censoring event. The estimated hazard rate plotted at time *t* is the estimated average over the interval [*t*–1 ms, *t*). The hazard rate for a response can be interpreted as the intensity of the response conditional on the elapsed time since the conditioning market event.



2007

2008

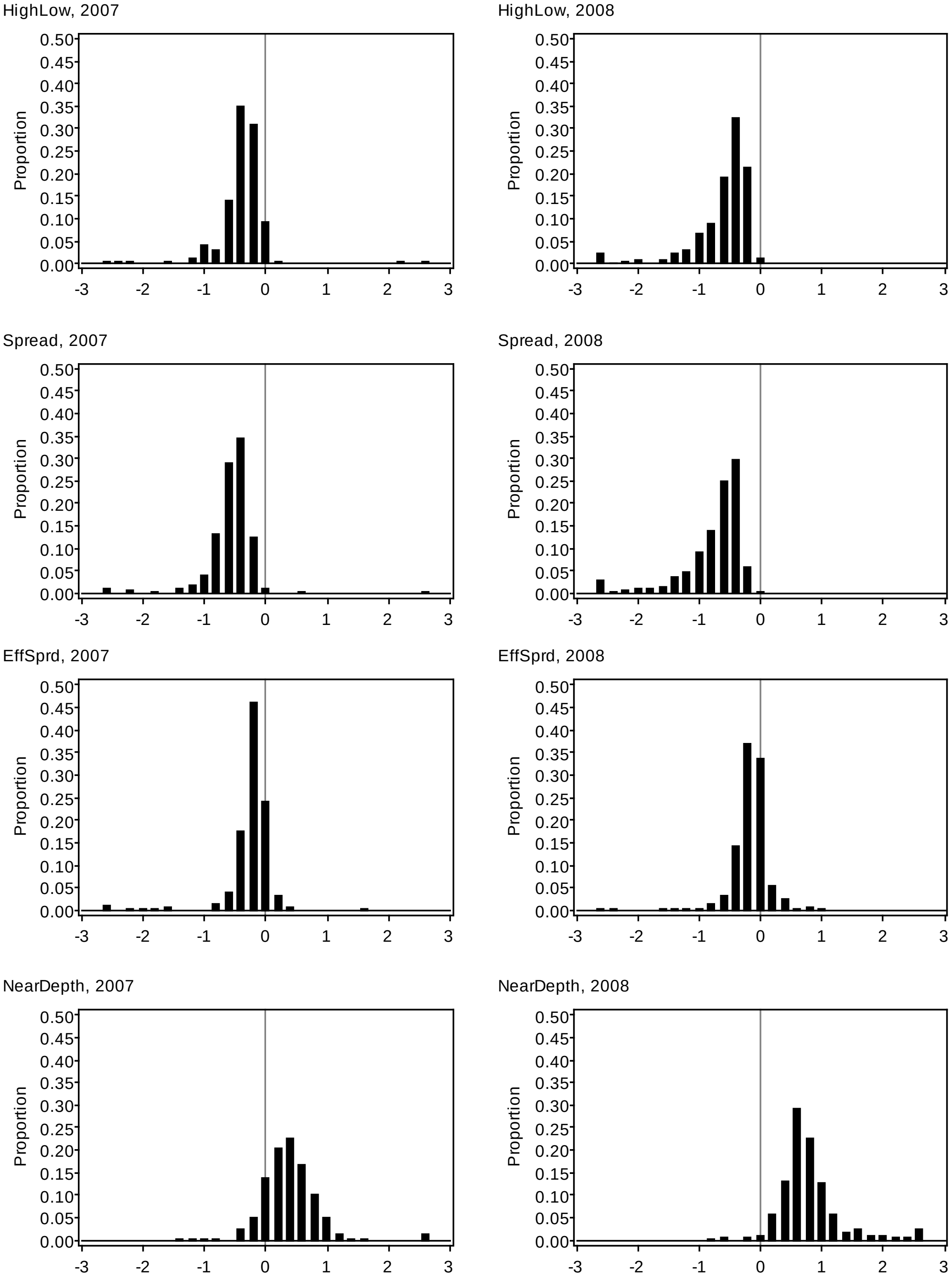
# Figure 3

## Histogram of Stock-by-Stock $a_1$ Coefficients

This figure presents further evidence on how low-latency activity affects market quality by providing histograms of the $a_1$ coefficients from stock-by-stock estimations of Model I for each of the market quality measures (*HighLow, EffSprd, Spread, and NearDepth*):

$$MktQuality_{i,t} = a_1 RunsInProcess_{i,t} + a_2 EffSprdNotNAS_{i,t} + e_{1,i,t}$$
$$RunsInProccess_{i,t} = b_1 MktQuality_{i,t} + b_2 RunsNotIND_{i,t} + e_{2,i,t}$$

The model is estimated using Two-Stage GMM with instruments *RunsNotI* and *EffSprdNotNas*. Each histogram shows the distribution of the $a_1$ coefficients that result from 351 (399) separate stock-specific estimations in 2007 (2008) for one of the four market quality measures.


HighLow, 2007
HighLow, 2008
Spread, 2007
Spread, 2008
EffSprd, 2007
EffSprd, 2008
NearDepth, 2007
NearDepth, 2008
Proportion