

Technology and liquidity provision: The blurring of traditional definitions[☆]

Joel Hasbrouck^{a,*}, Gideon Saar^{b,1}

^a*Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, USA*

^b*Johnson Graduate School of Management, Cornell University, 455 Sage Hall, Ithaca, NY 14853, USA*

Available online 13 June 2008

Abstract

Limit orders are usually viewed as patiently supplying liquidity. We investigate the trading of one hundred Nasdaq-listed stocks on INET, a limit order book. In contrast to the usual view, we find that over one-third of nonmarketable limit orders are cancelled within two seconds. We investigate the role these “fleeting orders” play in the market and test specific hypotheses about their uses. We find evidence consistent with dynamic trading strategies whereby traders chase market prices or search for latent liquidity. We show that fleeting orders are a relatively recent phenomenon, and suggest that they have arisen from a combination of factors that includes improved technology, an active trading culture, market fragmentation, and an increasing utilization of latent liquidity.

© 2008 Elsevier B.V. All rights reserved.

JEL classification: D40; G10; G20

Keywords: Limit order; Order strategy; Liquidity; Nasdaq

The usual economic perspective on a limit order emphasizes its role in supplying liquidity. In this capacity, it is often viewed as extending to the market a visible, ongoing, and persistent option to trade. Unlike a market order, it is passive and patient.

[☆]We thank INET for providing us with the data for this study, and are especially grateful for the help of Cameron Smith, Josh Levine, and Rob Newhouse. We benefited from the comments of the editor, an anonymous referee and seminar participants at Rice University.

*Corresponding author. Tel.: +1 212 998 0310.

E-mail addresses: jhasbrou@stern.nyu.edu (J. Hasbrouck), gs25@cornell.edu (G. Saar).

¹Tel.: +1 607 255 7484.

This characterization of a limit order arises most naturally from the viewpoint that a customer limit order is functionally equivalent to a dealer quote. Dealers are often modeled as risk-neutral liquidity suppliers, who are indifferent as to whether their bids and offers are hit, and who let their bids and offers persist until there is a trade. In other models, traders make strategic choices about whether to supply or demand liquidity. Under certain circumstances (e.g., the presence of a wide spread), even a trader with a strong desire to trade might submit a limit order. As in the dealer models, however, such limit orders are submitted in order to rest in the book, and interact with incoming market orders.

The evidence presented in this paper calls into question the traditional view of limit orders as patient suppliers of liquidity. We investigate the trading of 100 Nasdaq-listed stocks on INET, an electronic communication network organized as a limit order book; INET absorbed the SuperMontage and Brut systems in 2006 to become Nasdaq's primary trading platform. We observe that 36.69% of limit orders are cancelled within two seconds of submission. We term these "fleeting orders" and explore the role they play in trading strategies. Their sheer numbers and apparent defiance of easy classification within the usual framework of patient limit orders and impatient market orders poses a puzzle and a challenge to academic theories in this area.

We posit three hypotheses to explain why we observe fleeting orders. The *chasing hypothesis* is that they arise when a trader cancels and resubmits a limit order as the market moves away from the original limit price. The *cost-of-immediacy hypothesis* is that the cancelation occurs when the limit order is switched to a market order, in response to a drop in the cost of immediate execution. The *search hypothesis* posits that fleeting orders are outcomes of strategies that seek latent liquidity. "Latent" in this context comprises opposing hidden limit orders that are available for execution but are not displayed. Our usage of the term also extends, however, to counterparties who are actively monitoring the market and will immediately hit an aggressively priced limit order, but who are nevertheless unwilling to pre-commit to a price (with a displayed or undisplayed limit order). Thus, under the search hypothesis, a limit order is submitted within the spread in the hope of either immediately achieving execution against a standing hidden order, or by quickly attracting a new marketable order. If neither occurs within a brief interval, the limit order is cancelled, and we observe a fleeting order.

We carry out a variety of tests to characterize the use of fleeting orders and examine evidence directly related to the three hypotheses. We find that fleeting orders are priced somewhat more aggressively than limit orders with longer lives. Furthermore, the prevalence of fleeting orders at each price point exhibits a pattern very similar to that of executions against hidden orders: It progressively increases with the aggressiveness of the price. Both of these stylized facts are consistent with the search hypothesis, and suggest a strategy whereby traders start by searching at the most advantageous price points (e.g., one cent higher than the bid for a limit buy order) and sequentially move to worse prices.

We then use multinomial logit analysis to examine the market conditions that give rise to fleeting and "regular" (patient) limit orders. We find that fleeting orders are very different from more patient limit orders in their relation to variables such as lagged volume, lagged volatility, and prevailing spread. This means that a partition of the set of nonmarketable limit orders based on the rapidity of cancellation is in fact meaningful. The incidences of fleeting and market orders are affected in similar ways by market condition variables, suggesting that traders could be using them as components of strategies that are aimed at demanding liquidity.

The most comprehensive empirical tool we employ is a proportional hazards duration model of order cancellation. Our specification is an enhancement over prior studies in that it analyzes how market conditions that are determined *after* the order is submitted affect the cancellation decision. This models the strategic behavior of a trader who monitors the market after submission.

Consistent with the chasing hypothesis, we find that the cancellation probability of a limit order increases when the best price on the INET book on the same side as the order becomes more aggressive after the order is submitted, increasing the likelihood that the market is “running away” from the order. There is also weaker evidence that the cancellation intensity of a limit order increases when the best price on the opposing side of the INET book becomes more aggressive (say the ask price comes down after a limit buy order is submitted), reducing the cost of an immediate execution using a market order. The model also demonstrates that the probability of rapid cancellation increases the more aggressively the limit order is placed, in line with the search hypothesis.

Fleeting orders appear to be a relatively recent development in U.S. markets. We compare the distributions of times-to-execution and times-to-cancellation for three limit order samples: NYSE’s TORQ circa 1990, INET in 1999, and INET in 2004. While execution durations are similar across the three samples, cancellation intensities in the later periods are dramatically higher. We discuss key changes in the market environment that we believe are at the heart of this phenomenon: improved trading technology, the emergence of an active trading culture, fragmentation of the market structure, and increased utilization of latent liquidity.

We believe that, while improvement in trading technology is a necessary precondition of many of the dynamic strategies we discuss, it is the interplay of fragmentation and latent liquidity that holds the key to understanding the most curious feature of fleeting orders—their visibility. If the sole purpose of a fleeting limit order were to achieve execution against hidden orders already in the book, there would be no need to make the fleeting order itself visible. A trader could submit and quickly cancel a hidden limit order, or use an immediate-or-cancel order that does not enter the book. Since visibility can potentially reveal trading intentions, it entails obvious costs. Since it is also voluntary, there must be some offsetting benefits. We believe that the fragmentation of today’s trading environment among multiple trading venues creates a coordination problem: Patient traders need to decide where to post their hidden orders while impatient traders need to choose where to conduct their searches. The visibility of fleeting orders could be serving as a signal to potential latent liquidity providers that impatient traders are searching INET, encouraging the patient traders to submit hidden orders to INET and enabling both patient and impatient traders to better fulfill their trading needs.²

The evidence we document on the extensive use of fleeting limit orders in dynamic trading strategies that are aimed at demanding liquidity has important implications. First, it calls into question results from theoretical models that characterize limit orders as persistent and their traders as patient. The new trading environment we observe requires a different framework for thinking about optimal order choices in markets. Second, our results challenge the manner in which the execution quality of trading venues is evaluated. The Security and Exchange Commission’s rule 605 (formerly 11Ac1-5) requires market

²The visible limit orders could also trigger programmed trading algorithms that would quickly send a marketable order to effect an execution.

centers to report the fill rate of limit orders. A higher fill rate presumably indicates a greater likelihood of finding a counterparty and therefore a better market. We document a low fill rate for INET, yet the venue was highly successful and was ultimately chosen by Nasdaq as its primary platform. We argue that when many orders are quickly cancelled, the fill rate is a misleading and inappropriate metric of quality. We believe that recognizing the new ways in which trading and order choices have changed due to technology, active trading, fragmentation, and latent liquidity is important to academics, regulators, and investors.

We have organized the remainder of this paper as follows. The next section contains a literature review, and Section 2 discusses our sample and the INET data. Section 3 provides an initial characterization of fleeting orders and presents three hypotheses to explain why we observe them. In Section 4 we present a more structured empirical analysis of fleeting orders and test the specific hypotheses about their origins. Section 5 discusses the developments that have led to broader use of fleeting orders, and Section 6 provides the conclusion pertaining to our results.

1. Literature review

The notion that limit orders supply liquidity to the market suggests that they are similar in nature to dealers' quotes, and that the economic forces affecting limit order strategies should be similar to those considered in models of dealer markets. Dealers in the sequential trade models of asymmetric information are risk-neutral. They are subject to adverse selection, and the pricing of their bids and offers is ultimately determined by zero-expected profit conditions induced by competition (e.g., [Copeland and Galai, 1983](#); [Glosten and Milgrom, 1985](#); [Easley and O'Hara, 1987](#)). Agents who populate limit order books may be modeled from a similar perspective, subject to the important qualification that they cannot price their bids and offers conditional on the size of the incoming order ([Glosten, 1994](#); [Sandas, 2001](#); [Seppi, 1997](#)).

Risk neutrality on the part of dealers or limit order traders in the aforementioned models makes them indifferent as to whether or not their quotes or orders are hit (although they may have preferences concerning the total size of the order that triggers the execution).³ The sequential trade models also feature the rational expectations notion of "regret-free" prices, and therefore limit orders or quotes are changed only in response to new trades (or, as in [Easley and O'Hara, 1992](#), a period without trading). Since the interaction in a dealer market clearly distinguishes between liquidity suppliers and demanders, models of limit order books in this tradition also specify one class of traders who supply liquidity using limit orders and another class of liquidity demanders.

Models of traders' choices between market and limit orders offer a different perspective ([Cohen, Maier, Schwartz, and Whitcomb, 1981](#); [Angel, 1994](#); [Harris, 1998](#)). A trader's choice is usually heavily influenced by the probability that a limit order will execute. Since this probability is determined by the order choices of other traders, it is desirable to

³In many of these models the indifference to execution is mostly a consequence of assumed risk-neutrality. Risk-neutrality, however, is not essential to this result. The dealer in [Stoll \(1978\)](#) sets his bid to reflect the loss in expected utility in the suboptimal portfolio that will result if his bid is hit. At the optimum, however, expected utility conditional on the bid being hit is equal to that conditional on no trade. This implies an indifference to the execution.

analyze the equilibrium, as in Chakravarty and Holden (1995), Parlour (1998), Foucault (1999), Foucault, Kadan, and Kandel (2005), Goettler, Parlour and Rajan (2005, 2007), Kaniel and Liu (2006), and Rosu (2006).

The equilibrium order choice models do not for the most part attach importance to a limit order's duration. Typically, a randomly-drawn trader arrives at each instant and makes a choice between market and limit orders without the possibility of a subsequent trading opportunity. Order strategies are defined by type (market or limit), and (for a limit order) price. Individual strategies balance the lower trading costs of a limit order execution against the costs of delay or nonexecution. The duration of an order's exposure is not a key facet of these models because cancellation is nonstrategic.⁴ A noteworthy feature of these models is, however, that a trader's order choice influences the choices of subsequently arriving traders. For example, the probability that the next trader will use a market order increases if the current trader enters a limit order.

We should note here three theoretical papers that model dynamic strategies and enable the trader to cancel an order and resubmit a different one to actively seek an execution. Harris (1998) considers a trader who is trying to minimize the purchase price of a predetermined quantity, subject to a deadline. The optimal strategy is initially to place a limit order, then to re-price the order more aggressively as the deadline nears and, finally, if necessary, to use a market order. That is, limit orders are entered and revised pre-deadline even by agents who are ultimately constrained to trade. Bloomfield, O'Hara, and Saar (2005) provide evidence confirming the utilization of these strategies by constrained liquidity traders in experimental settings. They also show that traders with private information about the (common) value of a given security would begin trading with market orders but shift to limit orders as prices adjust to reflect their private information.

Large (2004) suggests that limit order cancellations arise from the refinement (over time) of a limit order trader's beliefs about the arrival rate of market orders, which is directly related to the expected time until the order's execution. Rosu (2006) proposes a model in which traders can update (cancel and resubmit) existing limit orders instantaneously. In Rosu's model, a limit order may be "fleeting," but only because it is immediately executed. Our fleeting orders, in contrast, are those that vanish because they are quickly cancelled.

The strategies discussed to this point are set in the context of a single execution venue. Fragmentation may increase the cost of exposing a limit order. Competing traders can use other venues to price-match the order, reducing its probability of execution (since there is no time priority across venues). Shortening the order's exposure time may be a way of controlling these costs. Also, sequential strategies involving fleeting orders may be used across venues in a fragmented market. There has been to our knowledge no theoretical work on this problem that is specific to limit order markets (although this could be considered more broadly as a search problem).

A number of empirical studies have sought to characterize limit order markets (e.g., Biais, Hillion, and Spatt, 1995; Hamao and Hasbrouck, 1995; Ahn, Bae, and Chan, 2001; Biais,

⁴Orders expire in one period in Foucault (1999), never expire in Parlour (1998) and Foucault, Kadan and Kandel (2005), and face random cancellation (with probabilities depending on the price path) in Goettler, Parlour, and Rajan (2005). Goettler, Parlour, and Rajan (2007) allow their traders to reenter the market (at a random time) and upon entering choose whether or not to cancel their limit order and resubmit at a different price. Reentry time is not, however, under the trader's control (it is exogenous), and therefore time to cancellation, for example, is not a strategic choice variable within their framework. For a recent survey of the theoretical models in this literature see Parlour and Seppi (2008).

Bisiere, and Spatt, 2003; and Hollifield, Miller, and Sandås, 2004). Only a few studies model timing of executions and cancellations. Cho and Nelling (2000) and Lo, MacKinlay and Zhang (2002) estimate duration models, but their focus is on execution, with cancellation being taken as an exogenous censoring process. Boehmer, Saar, and Yu (2005) estimate a duration model of limit order cancellation to characterize trader behavior around changes in pre-trade transparency (the introduction of NYSE's OpenBook service), and Chakrabarty, Han, Tyurin, and Zheng (2006) estimate a competing risk model of cancellation and execution times. We estimate a duration model in this paper to investigate fleeting orders, with the novelty (relative to the aforementioned papers) that we utilize time-varying covariates to look at how the probability of cancellation of a limit order is affected by changes in market condition after the order is submitted. We also use a multinomial logit specification to characterize order strategies. This approach is similar to that of Smith (2000), Ellul, Holden, Jain, and Jennings (2007), and Renaldo (2004). Our event classification, however, will involve outcomes as well as submission decisions.

Other studies have focused specifically on INET.⁵ Hasbrouck and Saar (2002) characterize the cross-sectional relation between volatility and different characteristics of INET's order flow and limit order book. Biais, Bisiere, and Spatt (2003) examine competitive behavior in order placement strategies on INET and Nasdaq. Hansch (2003) documents the extent of and variation in INET book depth, and Nguyen, Van Ness, and Van Ness (2004) describe changes related to INET's decision to shift trade reporting from Nasdaq to the Cincinnati (now National) Stock Exchange. Hendershott and Jones (2005) consider changes in INET's market quality in exchange-traded funds when, for regulatory reasons, it did not disseminate quotes or book data.

2. Sample and data

2.1. Sample construction

We rank all Nasdaq National Market domestic common stocks according to their equity market capitalization (from the CRSP database) on September 30, 2004. We then obtain a 100-stock sample by taking every fifth stock, hence creating a size-stratified subsample of 100 stocks from among the 500 largest Nasdaq stocks. The sample period is the month of October 2004 (21 trading days).

Table 1 presents summary statistics for the sample using information from the CRSP and Nastraq databases. The market capitalization of the smallest firm in the sample is 612 million dollars, while that of the median firm is about 1.5 billion dollars and that of the largest firm is about 76 billion dollars. The sample also spans a range of trading activity and price levels. The most active stock has a daily average of 43,805 trades over the sample period, while the median stock has about 2,759 trades on an average day, and the least actively-traded stock in the sample averages only 52 trades per day. Average daily CRSP closing prices range from \$2.52 to \$171.00, with a median of \$30.32. To provide a sense of the cross-sectional characteristics of the variables, we report medians for three groups constructed by ranking on market capitalization, average number of daily trades over the sample period (as a measure of trading activity), and standard deviation of daily returns (as a measure of volatility).

⁵Prior to the Island/Instinet merger, these studies used data provided by the Island ECN. The Island platform became INET after the merger with Instinet.

Table 1

Sample Summary Statistics

Our sample consists of 100 Nasdaq National Market stocks. To construct the sample we rank all Nasdaq National Market common domestic stocks using equity market capitalization as of September 30, 2004. We then take every fifth stock, thereby obtaining a size-stratified subsample of 100 stocks from among the 500 stocks with the largest market capitalization. The sample period is October 2004 (21 trading days). The following variables are calculated for each stock over the sample period using data from CRSP: AvgTrd is the average number of daily trades, AvgVol is the average daily share volume (in thousands), MedTurn is the median daily turnover (the number of shares traded divided by the number of shares outstanding), AvgPrc is the average daily closing price, and StdRet is the standard deviation of daily returns. MktCap is the market capitalization of the stocks on September 30, 2004, from CRSP. The following variables are calculated from the intraday Nastraq database for each stock: AvgSprd (in \$) is the average dollar spread (using all NBBO quotes in the sample period), and AvgSprd (in %) is the average relative spread (dollar spread divided by the quote midpoint). The table presents cross-sectional summary statistics for the entire sample and separately the medians for three groups (low, medium, and high) sorted by market capitalization, number of daily trades, and return standard deviation.

		MktCap (million \$)	AvgTrd	AvgVol (1,000 shrs)	MedTurn (in %)	AvgPrc (in \$)	StdRet (in %)	AvgSprd (in \$)	AvgSprd (in %)
Entire Sample	Mean	3,882	4,657	1,801	1.300	33.50	2.295	0.076	0.241
	Median	1,497	2,759	716	0.916	30.32	2.140	0.045	0.178
	Std.	8,873	6,433	3,486	2.152	22.20	0.881	0.134	0.232
	Min	612	52	10	0.022	2.52	0.695	0.010	0.040
	Max	75,759	43,805	27,399	15.680	171.00	4.699	1.257	1.684
MktCap Groups	Low	898	1,840	553	1.002	28.37	1.814	0.095	0.347
	Medium	1,580	3,857	1,585	1.524	30.13	1.769	0.059	0.221
	High	9,237	8,299	3,274	1.368	42.10	1.450	0.074	0.154
AvgTrd Groups	Low	1,567	812	221	0.380	35.94	1.271	0.144	0.395
	Medium	1,785	2,712	750	1.154	30.03	1.717	0.051	0.192
	High	8,357	10,506	4,465	2.371	34.63	2.046	0.032	0.136
StdRet Groups	Low	6,388	3,039	946	0.474	37.05	1.410	0.099	0.246
	Medium	2,977	4,017	1,810	0.991	34.39	2.121	0.075	0.231
	High	2,307	6,935	2,648	2.444	29.03	3.358	0.053	0.245

While all the empirical work we present in the paper uses the abovementioned sample, a previous version of the paper utilized an older sample of Nasdaq National Market stocks from the last quarter of 1999. We updated the sample in the course of the review process, but the phenomenon we identified (the “fleeting orders”) existed as early as 1999 and became even more pronounced with time. Where relevant, we mention in the text results from the 1999 sample to note any changes that occurred over time in the variables we investigate. While the two samples are comparable with respect to market capitalization and trading volume, they differ with respect to the magnitude of bid-ask spreads. The 1999 sample is from the period prior to decimalization, and therefore the average National Best Bid or Offer (NBBO) spread of stocks in the sample is 25.6 cents, and the average relative spread is 0.46% of the stocks’ NBBO midpoint. In the 2004 sample, the average NBBO spread is 7.6 cents, and the average relative spread is 0.24%.

2.2. *INET data*

INET was formed in 2002 by the merger of Island and Instinet, both electronic communications networks (ECNs). Nasdaq subsequently purchased INET, which then absorbed the SuperMontage and Brut systems, and became Nasdaq's primary trading platform. Although presently named SingleBook, the market was called "INET" during our sample period, and we use that name here.

INET operates a pure agency market. All orders must be priced. A trader who seeks immediate execution must price the limit order to be marketable, for example a buy order priced at or above the current ask price. For all intents and purposes, a marketable limit order in a pure limit order book is equivalent to a market order in floor or dealer markets. Such an order is never displayed in the book; rather, it is immediately executed upon arrival at the system. We use the terms 'market orders' and 'marketable limit orders' interchangeably in this paper, and reserve the term 'limit orders' for nonmarketable orders that enter the INET book.⁶ Orders may be visible or hidden, with the difference being that hidden orders reside in the INET book but are invisible to all traders.⁷ Execution priority follows price, visibility, and time. All visible quantities at a price are executed before any hidden quantities at that price can trade.

The INET data we use are identical to those supplied in real time to INET subscribers. These data are comprised of time-sequenced messages that completely describe the history of trade and book activity. The process may be summarized as follows. When an arriving order is marketable, i.e., it can be matched (in whole or part) against existing orders in the book, the system sends an Order Execution message. If the order can't be matched, i.e., it is nonmarketable, the system sends an Add Order message, which means that the order is being added to the limit order book. An Add Order message contains the direction (buy or sell), number of shares, limit price, and a unique identification number. If and when the order is executed, this number is reported in the Order Execution message. When an existing order is canceled or modified (in size), the system generates a Cancel Order message. The book, excepting the hidden orders, may be constructed by cumulating these messages from the start of the day onwards. Although the arrival time and quantity of a hidden order is never made available, the execution of such an order is signaled by a special trade message.

In presenting statistics based on the INET data, we take the individual stock as the unit of observation. That is, we first compute estimates for each stock, and then report summary statistics across stocks. Table 2 presents summary statistics on the number and size of orders that come into INET.⁸ In our sample, the stock with the most activity on INET has a daily average of 244,136 limit orders submitted, while the median stock has about 9,509 limit orders on an average day, and the least active stock in the sample averages only 280 limit orders per day. The average number of daily limit orders increases with market capitalization, trading activity, and volatility. There is also a wide range in intensity of market order executions across stocks: The most heavily traded stock has on

⁶An order may be submitted with an expiration time, using the time in force (TIF) attribute. If the TIF is set to zero, the order is not added to the book. It may, however, be matched on arrival, and is therefore equivalent to an immediate-or-cancel order.

⁷The option of complete invisibility differentiates INET's limit orders from the reserve ("iceberg") orders found in Euronext, where at least a portion of the limit order must be visible at all times.

⁸We consider only data from the regular trading session of the Nasdaq Stock Market (from 9:30 a.m. to 4:00 p.m.).

Table 2

Summary Statistics for INET Orders

This table presents summary statistics on the orders submitted to INET. We analyze a size-stratified sample of 100 stocks from among the 500 Nasdaq National Market common domestic stocks with the largest equity market capitalization on September 30, 2004. The sample period is October 2004 (21 trading days). All variables are calculated for each stock over the sample period using order-level data from INET: NumLMT is the average daily number of (visible) nonmarketable limit orders that enter the book, SizeLMT is the average size of a nonmarketable limit order in shares, NumCanc is the average daily number of order cancellations, SizeCanc is the average size of a cancelled order in shares, NumMKT is the average daily number of market orders (where a market order is defined as an order that is matched upon arrival and so never appears in the book), and SizeMKT is the average size of a market order in shares. The table presents cross-sectional summary statistics for the entire sample and separately the medians for three groups (low, medium, and high) sorted by market capitalization, number of daily trades, and return standard deviation.

		NumLMT	SizeLMT (in shares)	NumCanc	SizeCanc (in shares)	NumMKT	SizeMKT (in shares)
Entire Sample	Mean	16,660	297	15,531	296	1,584	191
	Median	9,509	239	8,880	238	846	153
	Std. Dev.	29,331	224	27,996	215	2,220	162
	Min	280	134	272	134	9	85
	Max	244,136	1,619	233,501	1,550	15,833	1,251
AvgCap Groups	Low	6,487	227	6,055	227	570	150
	Medium	11,346	292	10,528	287	1,108	208
	High	32,308	373	30,160	375	3,087	215
AvgTrd Groups	Low	3,429	199	3,221	200	273	126
	Medium	11,607	250	10,942	250	899	164
	High	35,097	444	32,568	440	3,600	284
StdRet Groups	Low	14,561	270	13,756	272	1,154	146
	Medium	12,799	290	11,826	289	1,376	185
	High	22,736	333	21,122	328	2,227	241

average 15,833 daily market orders (or executions) on INET, the median stock has on average 1,584 executions, and the least active stock has on average 9 executions a day. The average size of limit orders on INET is 297 shares, while market orders tend to be smaller than limit orders, with a mean of only 191 shares.⁹

3. Fleeting orders

3.1. The fill rate of limit orders

The view expressed by the literature on limit order markets is that the basic forces that determine market outcomes are those of supplying and demanding liquidity. Suppliers of liquidity use limit orders that are posted to the book, and these limit orders await execution by traders who demand liquidity using market orders. If the limit order fill rate

⁹For comparison, the average size of a limit (market) order in our 1999 sample was 572 (340) shares, testifying to a decrease in the size of submitted orders.

(probability of execution) is low, traders will submit fewer limit orders, increasing the costs of market orders. All else remaining equal, a trader would prefer to post a limit order in a venue with a high fill rate, and this motivates the disclosure of fill rates mandated by SEC Rule 605.

Table 3 describes the incoming order mix (i.e., the submission proportion of limit orders relative to all limit and market orders) and fill rates of limit orders. On average, limit orders account for 90.05% of the incoming orders. Of these (nonmarketable) limit orders, only 7.99% achieve even partial execution. Only 6.37% of the shares submitted in limit orders are executed. This fill rate is very low. By way of comparison, the average fill rate for the NYSE limit orders in the TORQ database (October 1989–January 1990) is 56%. Lo, MacKinlay, and Zhang (2002) report a fill rate of 53% for their sample of NYSE limit orders from 1994–1995. The fill rate in our 1999 INET sample is 18.4%.

Yet although the 2004 INET fill rate represents a sharp drop from the historical NYSE figure, INET is widely regarded as a successful trading venue. We now turn to an explanation of this apparent paradox, which calls into question the usefulness of the fill rate as a measure of market quality.

Table 3
Limit Order Submission Frequency and Fill Rates

This table presents summary statistics on the order mix and fill rate of limit orders on INET. The submission frequency of nonmarketable limit orders is defined as the number of nonmarketable limit orders (i.e., those that are stored in the book upon arrival to INET) divided by the sum of nonmarketable limit orders and market orders (where market orders are those that get executed upon arrival and never enter the book). The Fill Rate is the number of nonmarketable limit orders that were subsequently executed divided by the total number of nonmarketable limit orders submitted. We present three different fill rates: (i) counting in the numerator the limit orders that were at least partially executed, (ii) counting in the numerator only the limit orders that were fully executed, and (iii) computing the fill rate in terms of shares (i.e., shares in nonmarketable limit orders that were subsequently executed divided by the number of shares in all nonmarketable limit orders). All variables are calculated for each stock over the sample period using order-level data from INET. The table presents summary statistics for the entire sample and separately the medians for three groups (low, medium, and high) sorted by market capitalization, number of daily trades, and return standard deviation.

		Frequency of Limit Orders	Fill Rate (partially executed)	Fill Rate (fully executed)	Fill Rate (in terms of shares)
Entire Sample	Mean	90.05%	7.99%	6.96%	6.37%
	Median	90.47%	7.28%	6.24%	5.69%
	Std. Dev.	3.74%	3.45%	3.23%	3.54%
	Min	78.56%	1.99%	1.56%	1.05%
	Max	97.11%	19.78%	18.59%	22.21%
MktCap Groups	Low	91.23%	6.72%	5.90%	5.11%
	Medium	91.27%	6.82%	5.75%	5.39%
	High	89.51%	8.47%	7.30%	6.36%
AvgTrd Groups	Low	91.51%	5.96%	4.97%	4.53%
	Medium	91.27%	6.52%	5.58%	5.22%
	High	88.93%	9.06%	8.01%	6.99%
StdRet Groups	Low	91.81%	6.56%	5.65%	4.53%
	Medium	90.80%	6.95%	5.84%	5.48%
	High	90.27%	7.76%	6.82%	6.83%

Table 4

Cancellation and Execution Rates of Limit Orders

This table presents estimated cumulative probabilities of cancellation and execution within different time intervals (measured from the submission of the nonmarketable limit order). The probabilities are computed as $1-S(t)$, where $S(t)$ is the survival function (of cancellation or execution). The survival function is estimated using the life-table method using all nonmarketable limit orders for each of the 100 stocks in the sample. In the estimation for the cancellation process, execution is taken to be the censoring process (and vice versa).

Time	Cumulative Probability of	
	Cancellation	Execution
0.1 seconds	0.115	0.009
1	0.296	0.026
2	0.369	0.034
10	0.601	0.071
1 minutes	0.827	0.158
2	0.910	0.211
10	0.984	0.383
1 hours	0.997	0.568

3.2. Survival analysis

The first tool we use in analyzing what happens to limit orders after they are submitted to INET is survival analysis. Let τ denote the time between order submission and cancellation. The probability of cancellation in the interval $(0, t]$ is the distribution function $P_{\text{Cancel}}(t) = \Pr(\tau \leq t)$. This function is estimated separately for each stock using the life-table method, and taking execution as the censoring event (assumed to be exogenous). Table 4 presents the cross-sectional average of the cancellation probability. What is most striking in the table is that a large number of limit orders are canceled within a very brief interval of time after submission. $P_{\text{Cancel}}(2)$, the probability of cancellation within two seconds, is 0.369.¹⁰ By the time ten seconds have elapsed, this probability reaches 0.601. For completeness, the table also contains probabilities of execution (estimated by taking cancellation as the censoring event). Execution is clearly the less probable event, particularly in the first few seconds after submission.

The results in Table 4 directly explain INET's low fill rate. If a limit order is cancelled very quickly, there is little chance of an execution. The puzzle then shifts from “why is the fill rate so low?” to “why do so many limit orders get cancelled very quickly?” To reconcile the supposed contradiction between the popularity of INET as a trading system and the very low fill rate of limit orders that are submitted to INET, we must understand the driving forces behind this phenomenon.

3.3. Why do we observe fleeting orders?

In this section we discuss three different hypotheses to explain why so many “fleeting orders” (those limit orders that are cancelled very quickly) are observed on INET. We then

¹⁰The estimated probability of cancellation within two seconds (0.369, reported in Table 4 and discussed above) is corrected for censorship arising from execution. It is therefore slightly higher than the raw proportion of limit orders cancelled within two seconds (36.69%, reported in Table 5). The latter, more conservative, value is cited in the introduction and conclusion.

carry out empirical work to see if we find evidence consistent with any (or all) of these hypotheses.

One hypothesis, which we call the “chasing” hypothesis, is that traders pursue a dynamic strategy of cancelling a limit order and resubmitting another limit order (at a different price) as the market moves away from the price of the original limit order. In other words, if the trader wants to actively influence the likelihood of an execution, he would cancel an order if someone has placed a limit order ahead of it and resubmit the order at a more aggressive price. This strategy reflects some urgency in the trader’s desire to effect an execution, and therefore could be viewed as falling somewhere between a traditional patient limit order, which waits in the limit order book until the market price has reached it, and a market order that demands immediate execution at a greater cost. Chasing after the market price could also be a characteristic of certain trading strategies employed by high-frequency statistical arbitrage firms who attempt to earn market-making profits by placing limit orders near the prevailing market price (and therefore cancelling and resubmitting when it seems that the market price is changing).

Since our data do not identify the person submitting the order, we cannot unambiguously identify such strategies. We can investigate indirect evidence, however, by examining the relation between cancellation probability and movements in the same-side best bid or offer (BBO). It would be consistent with the chasing hypothesis, for example, if the cancellation probability of a limit buy order increases when a subsequently arriving bid betters the price of the original order.

A second hypothesis that might explain why we observe fleeting orders is that they are part of a dynamic strategy in which traders cancel a limit order and switch to a market order when the cost of immediate execution in the market decreases. In a sense, the “cost-of-immediacy” hypothesis reflects the basic tradeoff modeled already in [Cohen, Maier, Schwartz, and Whitcomb \(1981\)](#), whereby the “gravitational pull” of immediate execution using a market order increases when the spread decreases.

As with the chasing hypothesis, this cost-of-immediacy hypothesis implies that the original limit order was not meant to be a patient provider of liquidity, because the trader cancels the limit order to effect an immediate execution (either on INET or on a different market that trades the same stock). Therefore, this strategy combines elements of both supplying and demanding liquidity. Evidence consistent with this hypothesis would be observing an increase in the probability of rapid cancellation of a limit order when the other side of the BBO on INET approaches the limit price after the order is placed in the book.

The third hypothesis is that fleeting orders are a byproduct of a strategy meant to “search” for latent liquidity inside the INET spread. The search hypothesis implies that fleeting orders are intended to demand, rather than supply, liquidity. The order seeks immediate execution against a hidden order, and failing this it is cancelled rapidly. It is less aggressive than a market order strategy (which guarantees an immediate execution), but it is not a patient limit order, since the trader has no intention of keeping it in the book to benefit from execution against incoming order flow. Since the trader would presumably search for the latent liquidity only at prices inside the spread, one piece of evidence that would be consistent with this hypothesis is an increase in the probability of cancellation for very aggressive limit orders (that are inside the spread).

4. Analysis of order cancellation

In this section we use a variety of empirical tools to further investigate the behavior of fleeting orders, attempting to test whether they are driven by one or more of the hypotheses identified in Section 3. First, we look at the placement of limit orders to see whether fleeting orders are used in the same manner as “regular” limit orders (those that persist beyond two seconds) are, or whether they are more often used inside the spread as the search hypothesis implies. Second, we estimate a multinomial logit specification to ascertain whether fleeting orders have a relationship to the market environment that is similar to that of regular limit orders. Third, we use a duration model with time-varying covariates to test specific implications of the three hypotheses discussed in the previous section.

4.1. Placement of fleeting orders

To understand the rapid cancellation of limit orders, we start by asking whether limit orders that are cancelled very quickly are somehow different from “regular” limit orders, those that presumably sit in the book waiting to be filled. In other words, the large proportion of cancellations at short durations motivates consideration of these orders as a separate category from those limit orders that are traditionally characterized as patient providers of liquidity. We use, somewhat arbitrarily, two seconds as the break point: An order that is cancelled in two seconds or less is defined as a “fleeting order.”¹¹

Table 5 compares the pricing of fleeting orders (those that are cancelled within two seconds) with that of non-fleeting cancelled limit orders, i.e., orders that are cancelled after two seconds. The main difference that emerges from the table is that 35.38% of fleeting orders are priced ahead of the same-side INET BBO at submission, while only 20.73% of non-fleeting cancelled limit orders are priced ahead of the BBO. In other words, limit orders that are cancelled very quickly (fleeting orders) are priced more aggressively than limit orders that are cancelled less quickly.¹²

The information in Table 5 suggests that many fleeting orders are submitted at prices that better the same-side BBO by a small amount (a cent). Submitting a limit order at a slightly better price could be motivated by the desire to obtain price priority (i.e., jump to the head of the queue), or it could indicate a search for hidden orders whereby the searcher first tries the most favorable price (e.g., buying at a price that is just a cent above the bid in the market)

¹¹Since fleeting orders are characterized by the speed with which they are cancelled, it is useful to describe how such cancellations might occur. On INET (and in most limit order markets) a limit order can be cancelled by prearranged conditions that are set when the order is submitted. Alternatively, a trader or a computerized trading algorithm can continuously monitor the market and enter a cancellation request in response to changing market conditions. Ideally we would like to know the intended time in force (TIF) of the order, i.e., the value actually submitted with the order or the value that has been programmed into the trader's order-management system. Our data do not indicate this, however, and our inferences must therefore be based on the time the order was actually in the book.

¹²In our 1999 sample, 83.6% of fleeting orders were placed ahead of the same-side BBO, compared with 72.3% for all other limit orders. As we note in Section 2, the bid-ask spread in 1999 was much larger than in 2004, which could explain the higher frequency of both fleeting and regular limit orders submitted ahead of the BBO in the 1999 sample. Still, we documented the tendency of fleeting orders to be more aggressive than other limit orders already in the 1999 sample, and this tendency is even more pronounced in the 2004 sample.

Table 5
Pricing of Cancelled Limit Orders

This table presents information about the pricing of cancelled nonmarketable limit orders relative to the same side INET best bid or offer (BBO) by time between submission and cancellation. More specifically, we separate limit orders into those that had a cancellation within two seconds of submission (“fleeting orders”) and those that had a cancellation after more than two seconds have elapsed. The first line of the table reports the percentage of orders in each category relative to all limit orders submitted to INET. The next portion of the table reports sub-classifications within each category by the location of the limit order price relative to the same-side BBO prices in the INET book at the time the order was submitted: (i) behind BBO, (ii) at BBO, and (iii) ahead of BBO. For example, “Ahead of BBO” for a buy (sell) order means that it was submitted at a price that is higher (lower) than the best bid (ask). Percentages sum vertically to 100% within each category. The lower portion of the table breaks down the “Ahead of BBO” classification into the exact distance of the limit order price from the same-size BBO at time of submission. For example, the +0.01 line is for buy (sell) orders submitted at a price that is one cent higher (lower) than the best bid (ask). The percentages are computed separately for each stock, and the values in each cell in the table are cross-stock averages.

	Cancelled Limit Orders	
	Elapsed Time ≤ 2 s	Elapsed Time > 2 s
Percentage of Orders in Category (relative to all limit orders)	36.69%	56.24%
Relative to same-side BBO at submission		
Behind BBO	29.90%	46.47%
At BBO	34.72%	32.80%
Ahead of BBO	35.38%	20.73%
All	100.00%	100.00%
Breakdown of “Ahead of BBO” (by cents from same-side BBO)		
+0.01	17.28%	10.78%
+0.02	5.23%	3.30%
+0.03	3.08%	1.84%
+0.04	1.90%	1.01%
+0.05	1.11%	0.57%
> +0.05	6.77%	3.23%

and then sequentially searching for hidden orders at worse prices.¹³ If such a search process is taking place, we expect to find more buy (sell) executions against hidden orders at prices just one cent above (below) the bid (ask) than executions at higher (lower) prices. This is exactly what we observe in the data: The frequency of executions against hidden orders at one cent better than the same-side BBO is 58.51%, at two cents better it is 16.53%, at three cents it is 7.88%, and at four cents it is 4.05%. Hence, the placement pattern of fleeting orders in Table 5 seems to correspond to the pattern of executions against hidden orders. This observation is consistent with the search hypothesis: If the trader submits a limit order inside the BBO and finds hidden orders, we observe an execution, and if he does not obtain an execution, he cancels the limit order rapidly and we observe a fleeting order.

4.2. Multinomial logit analysis

As with the analysis of limit order placement, our goal with the multinomial logit analysis is to look for evidence that helps us determine whether we should be thinking

¹³About 15% of all executions on INET (representing 20% of executed shares) occur against hidden orders.

about fleeting orders as a separate category from that of limit orders that are traditionally characterized as patient providers of liquidity. The econometric model enables us to ask this question in a more structured framework in which we can also utilize a finer partition of the possible space of order outcomes. We begin this analysis by partitioning orders (or market events) into five classes. Limit orders (displayed nonmarketable orders that are added to the book) are partitioned according to their eventual outcome: limit orders that are subsequently cancelled within two seconds (“fleeting orders”), those that are subsequently executed within two seconds, and the remaining (“regular”) limit orders that subsequently persist on the book for more than two seconds. Executions are partitioned into an execution against a displayed quantity (“regular execution”) and an execution against a nondisplayed quantity (“hidden execution”).

This scheme does not, strictly speaking, describe a model of order choice, since classification relies in part on an outcome not known at the time of order submission. In particular, this differs from the practice in Smith (2000), Ellul, Holden, Jain, and Jennings (2007), and Renaldo (2004), wherein events are defined solely by reference to order characteristics and market conditions at the time the order arrives at the market. However, this classification could provide useful information if the eventual outcome (e.g., rapid cancellation) relates to the specification of a pre-determined dynamic strategy of the kind we discuss in the previous section (i.e., the three hypotheses about fleeting orders).

The set of five outcomes (indexed by $j = 0, \dots, 4$) is as follows: {regular limit order, fleeting limit order (cancelled within two seconds), limit order executed within two seconds, market order (execution against a visible quantity), hidden execution (execution against a hidden quantity)}. The sample comprises $i = 1, \dots, 100$ stocks, and $t = 1, \dots, 1000$ randomly-chosen events (or orders) for each stock. While within each stock a lower value of the index t represents an earlier event than one with a higher value of t , it should be noted that the events are essentially asynchronous across stocks even though the data for all stocks are taken from the same overall time period (October, 2004). For example, the event marked as $t = 10$ for one stock generally does not take place at the same instant as the $t = 10$ event for a different stock.

Let $\pi_{i,t,j}$ be the probability that event t for firm i has outcome j . The reference event is $j = 0$ (a regular limit order). The specification is a multinomial logit model with stock-specific fixed effects:

$$\log\left(\frac{\pi_{i,t,j}}{\pi_{i,t,0}}\right) = a_{i,j,0}d_{i,j} + a_{j,1}\left(\frac{\text{Proportional}}{\text{NBBO spread}_{i,t}}\right) + a_{j,2}\left(\frac{\text{lagged}}{\text{volume}_{i,t}}\right) + a_{j,3}\left|\frac{\text{lagged}}{\text{return}_{i,t}}\right| + a_{j,4}d_{i,j,t}^{\text{First hour}} + a_{j,5}d_{i,j,t}^{\text{Mid-day}} \quad (1)$$

where $d_{i,j}$ is a dummy variable set to one for firm i and outcome j , $d_{i,t,j}^{\text{First hour}}$ is a dummy variable set to one if the time is between 9:30 AM and 10:30 AM, and $d_{i,t,j}^{\text{Mid-day}}$ is a dummy variable set to one if the time is between 10:30 AM and 3:00 PM. Lagged absolute value of return and lagged volume are cumulated over the five-minutes preceding the event.

The explanatory variables comprise measures intended to capture dynamic variation in market conditions. The prevailing Nasdaq NBBO spread reflects the cost of obtaining immediacy in the market. Volume and volatility (absolute value of return) over the prior

five minutes are intended to capture variation in the general pace of market activity. Time-of-day dummy variables are included to capture deterministic intraday patterns. The spread, lagged volume and lagged absolute value of return are standardized within each firm to have zero mean and unit variance.

The standard errors usually computed for logit models assume independence of observations. We assess the independence assumption by examining the estimated residuals from a binomial model that contrasts regular limit orders with all other outcomes. Inclusion of the firm-dummies ensures that the average (within stock) residuals are essentially zero. The within-stock autocorrelations of these residuals are also small (the average first-order autocorrelation is less than 0.05). To assess independence across stocks, we construct for each stock a series of hourly mean estimated residuals, and compute the correlations between these residual series for all pairs of stocks. The average correlation is close to zero. These results provide little cause to question the independence assumption.

To facilitate interpretation, we compute the implied event probabilities (averaged across all stocks in the sample) when all explanatory variables are equal to their sample means. This is considered the “base” case. We then examine the implied probabilities when each of the variables, taken one at a time, increases by one standard deviation. Table 6 reports the mean probabilities and differences relative to the base case (with standard errors). The latter are computed using the delta method, and quantify estimation error arising from sample variation over time, holding constant the set of firms in the analysis.

It is immediately clear from these estimates that fleeting orders behave very differently from “regular” limit orders (those that stay in the book beyond two seconds). In fact, the effects of each of the variables on regular and fleeting limit orders run in the opposite direction. If the set of limit orders were homogeneous, a partition according to cancellation time would be meaningless and the effects of the economic environment on the probabilities of fleeting and regular limit orders would be identical. The differences between the effects on fleeting and regular limit orders suggest substantive economic differences as well.

The effect of increasing lagged volume or volatility is to decrease the probability of regular limit orders but to increase the probability of both fleeting orders and market orders (regular executions). This could be interpreted to suggest that fleeting orders behave as if they are meant to demand liquidity (like market orders) rather than to supply liquidity, in line with the search hypothesis. The response to an increase in the prevailing NBBO spread is different: The probability of a market order decreases while the probability of a fleeting order increases. The decreased use of market orders is consistent with the mechanism first noted by Cohen, Maier, Schwartz, and Whitcomb (1981) and found empirically in other studies (e.g., Biais, Hillion, and Spatt (1995)). A wider spread could lead to an increase in the use of fleeting orders if there are more price points at which to search for latent liquidity inside the spread. Under the search hypothesis, a limit order seeking latent liquidity achieves either a (hidden) execution, a rapid execution (under two seconds), or is quickly cancelled (and observed as a fleeting order). The estimates are consistent with this mechanism in that the coefficients of all variables for these three outcomes have the same signs.

We also examined the relation between the outcomes and two stock characteristics (market capitalization and turnover). In particular, we estimated a system of log odds ratios as seemingly unrelated regressions. This is a multinomial generalization of the

Table 6

Multinomial Logit Analysis

This table presents a multinomial logit analysis in which events (limit order submissions or executions) are partitioned into outcome classes. The set of five outcomes is: {*regular limit order*, *fleeting limit order (cancelled within two seconds)*, *limit order executed within two seconds*, *market order (execution against a visible quantity)*, *hidden execution (execution against a hidden quantity)*}, indexed by $j = 0, \dots, 4$. Regular limit orders are limit orders that persist on the book for more than two seconds. Fleeting orders are limit orders that are cancelled within two seconds. Limit orders that are executed within two seconds constitute a separate outcome. Regular executions are those market orders that execute against visible depth in the book, and hidden executions are executions against hidden orders. The sample consists of $i = 1, \dots, 100$ stocks, and $t = 1, \dots, 1000$ randomly-chosen events (for each firm). $\pi_{i,j,t}$ is the probability that event t for firm i has outcome j . The reference event is $j = 0$ (a regular limit order). The specification is a multinomial logit model with firm-specific fixed effects:

$$\log\left(\frac{\pi_{i,t,j}}{\pi_{i,t,0}}\right) = a_{i,j,0}d_{i,j} + a_{j,1}\left(\frac{\text{Relative}}{\text{NBBO spread}_{i,t}}\right) + a_{j,2}\left(\frac{\text{lagged}}{\text{volume}_{i,t}}\right) + a_{j,3}\left|\frac{\text{lagged}}{\text{return}_{i,t}}\right| \\ + a_{j,4}d_{i,j,t}^{\text{First hour}} + a_{j,5}d_{i,j,t}^{\text{Mid-day}}$$

where $d_{i,j}$ is a dummy variable set to one for firm i and outcome j (not reported for the sake of brevity), $d_{i,t,j}^{\text{First hour}}$ is a dummy variable set to one if the time is between 9:30 AM and 10:30 AM, and $d_{i,t,j}^{\text{Mid-day}}$ is a dummy variable set to one if the time is between 10:30 AM and 3:00 PM. Lagged volume and lagged absolute value of return are cumulated over the five-minutes preceding the event. The prevailing NBBO spread, lagged volume and lagged absolute value of return are standardized within each firm to have zero mean and unit variance. The top half of the table reports across-stock mean probabilities implied by the model estimates. In the Base Case column, all explanatory variables are set to their means. Each of the remaining columns reports the implied outcome probabilities when the indicated explanatory variable is set to one standard deviation above the mean. The bottom half of the table reports the probability shifts, or differences in probabilities, between the base and shifted cases. Standard errors computed using the delta method are reported in parentheses.

	Outcome	Base Case	NBBO Spread	Lagged Volume	Lagged Return
<i>Probability</i>	Regular limit order	0.5741 (0.0020)	0.5379 (0.0028)	0.5589 (0.0026)	0.5462 (0.0028)
	Fleeting limit order (cancelled ≤ 2 sec)	0.3169 (0.0019)	0.3637 (0.0026)	0.3273 (0.0025)	0.3381 (0.0026)
	Limit order executed (≤ 2 sec)	0.0195 (0.0006)	0.0199 (0.0007)	0.0220 (0.0008)	0.0216 (0.0008)
	Market order (regular execution)	0.0781 (0.0011)	0.0656 (0.0014)	0.0800 (0.0015)	0.0815 (0.0015)
	Hidden execution	0.0113 (0.0004)	0.0129 (0.0006)	0.0118 (0.0006)	0.0125 (0.0006)
Probability Shifts Relative to Base Case (implied by one standard deviation increase in the dependent variable)	Regular limit order		−0.0362 (0.0017)	−0.0152 (0.0017)	−0.0279 (0.0018)
	Fleeting limit order (cancelled ≤ 2 sec)		0.0468 (0.0016)	0.0103 (0.0016)	0.0212 (0.0016)
	Limit order executed (≤ 2 sec)		0.0003 (0.0004)	0.0024 (0.0006)	0.0021 (0.0004)
	Market order (regular execution)		−0.0125 (0.0008)	0.0020 (0.0010)	0.0034 (0.0009)
	Hidden execution		0.0015 (0.0003)	0.0004 (0.0003)	0.0012 (0.0003)

proportions model discussed in [Greene \(2002\)](#). As with the dynamic model, we examined the probabilities and probability shifts implied by the model. For the sake of brevity, the full results are not reported, but they may be summarized as follows. Larger stocks and more active stocks are characterized by a higher probability of fleeting orders and market orders, and a lower probability of “regular” limit orders that stay in the book for more than two seconds. This, again, is consistent with a meaningful partition of the limit order set, i.e., that fleeting orders are indeed different from the patient limit orders, and suggests that fleeting orders have more in common with market orders that demand liquidity.

4.3. Duration model for limit order cancellation

Neither the pricing investigation nor the logit analysis directly characterizes how market conditions *after* a limit order is submitted affect the cancellation decision. In fact, the drivers behind rapid cancellations of limit orders in two of the hypotheses (chasing and cost of immediacy) depend on what happens to the best prices in the book in the immediate aftermath of order submission. Both hypotheses specifically state that fleeting orders are a byproduct of dynamic strategies involving order revision in response to changing market conditions. As our data do not permit us to identify sequences of orders submitted by the same trader, the question cannot be answered definitively. We can, however, examine the dependence of cancellation on information that arrives subsequent to an order submission. In particular, we can assess the extent to which cancellations are driven by subsequent changes in the INET best bid or offer.

The statistical framework we use is a proportional hazards duration model with time-varying covariates ([Allison, 1995](#)). The event of interest is cancellation. Execution is viewed as a competing process. Let T denote the cancellation time for an order (relative to the submission time). The survival function is $S(t) \equiv \Pr(T > t)$. The hazard rate (the intensity of cancellation over the next instant) is $\lambda(t) = -d \log(S(t))/dt = S(t)^{-1} S'(t)$. For limit order i of firm j , the hazard rate is modeled with the semi-parametric form

$$\lambda_{ij}(t) = \lambda_{0j}(t) e^{X_{ij,t} \beta_j} \quad (2)$$

where $\lambda_{0j}(t)$ is an (unspecified) baseline hazard rate. The proportionality term allows for dependence on a vector of explanatory variables, $X_{ij,t}$. The components of $X_{ij,t}$ are known as of time t , but need not be known at the time the order is submitted. The coefficients are estimated in a partial-likelihood framework wherein the baseline hazard rate is left unspecified.

The specific model we estimate is

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp \left[\begin{aligned} & \beta_1 \left(\frac{\log \# \text{fleeting}}{\text{orders}_{ij}} \right) + \beta_2 \left| \frac{\text{lagged}}{\text{return}_{ij}} \right| + \beta_3 \left(\frac{\text{lagged}}{\text{volume}_{ij}} \right) \\ & + \beta_4 \left(\frac{\text{Proportional}}{\text{NBBO spread}_{ij}} \right) + \beta_5 p_{ij}^{\text{Relative}} + \beta_6 \Delta q_{ij,t}^{\text{Same}} + \beta_7 \Delta q_{ij,t}^{\text{Opposing}} \end{aligned} \right] \quad (3)$$

Volume and volatility (absolute value of return) over the prior five minutes and prevailing relative NBBO spread are used (as in the logit specification) to control for the

pace of trading and general market conditions. We include the log number of fleeting orders in the ten seconds prior to order submission in order to see if there is evidence of dynamic strategies that involve multiple rapid cancellations.¹⁴ These variables are standardized (within each stock) to have zero mean and unit variance.

The remaining three variables are meant to test the three hypotheses we discuss. For a limit buy order, the limit price aggressiveness variable is $p_{ij}^{\text{Relative}} \equiv [(\text{limit price}_{ij}) - (\text{INET bid}_{ij,t=0})] / (\text{INET bid}_{ij,t=0})$, where $t = 0$ denotes the time of order submission. This variable is more positive the higher the limit order price is, and is defined analogously for a limit sell order such that a more aggressive limit price (a lower price in this case) results in a more positive variable. The search hypothesis implies that the probability of cancellation of more aggressive orders should be higher.

Following limit order submission, the subsequent change to the best price in the limit order book on the same side as the order (henceforth, same-side BBO) is defined for a limit buy order as $\Delta q_{ij,t}^{\text{Same}} = [(\text{INET bid}_{ij,t}) - (\text{INET bid}_{ij,t=0^+})] / (\text{INET bid}_{ij,t=0^+})$, where $t = 0^+$ denotes the instant after submission. (This means simply that, if the arriving order sets a new bid, subsequent changes are measured relative to this bid.) The change in the same-side BBO is therefore positive if the bid price increases after the limit buy order is submitted, increasing the likelihood that the market is “running away” from the order and decreasing its probability of execution. The chasing hypothesis of fleeting orders implies that the probability of cancellation of a buy order should increase if the same-side BBO (the bid side) goes up because traders cancel their limit orders and resubmit at more aggressive prices. The change in the same-side BBO for a limit sell order is defined in analogous fashion so that a positive change would be associated with a decreased ask price.

We also include in the model changes in the opposing-side BBO subsequent to limit order submission, and for a limit buy order this variable is defined as $\Delta q_{ij,t}^{\text{Opposing}} = [(\text{INET ask}_{ij,t}) - (\text{INET ask}_{ij,t=0})] / (\text{INET ask}_{ij,t=0})$. The change in the opposing-side BBO is therefore negative if the ask side decreases after a limit buy order is submitted. The cost-of-immediacy hypothesis of fleeting orders implies that the probability of cancellation should increase if the opposing-side BBO (the ask) comes down because it becomes cheaper to demand liquidity by switching from a nonmarketable limit order to a market order. The change in the opposing-side BBO for a limit sell order is defined in an analogous fashion so that a negative change would be associated with an increased bid price.

Our approach is related to the models of limit order time to execution (or cancellation) suggested by Lo, MacKinlay, and Zhang (2002), Cho and Nelling (2000), and Boehmer, Saar, and Yu (2005), and the competing risk model of cancellation and execution times of Chakrabarty, Han, Tyurin, and Zheng (2006). The present specification differs from these in a key respect. The earlier studies restrict the conditioning set to variables known at the time the order is submitted. The present model allows the cancellation intensity to depend on price movements subsequent to submission, mimicking the strategic behavior of a trader who monitors the market, such as a trader who cancels the order when the price moves away. In other words, the key difference is our incorporation in the model of the time-varying covariates $\Delta q_{ij,t}^{\text{Same}}$ and $\Delta q_{ij,t}^{\text{Opposing}}$.

¹⁴The exact definition of the variable we employ is the log of the maximum of either one or the number of fleeting orders in the preceding ten seconds.

While the two time-varying covariates are introduced to investigate the chasing and cost-of-immediacy hypotheses, the incorporation of post-submission price movements improves the specification in another important respect. Most models of execution and cancellation times view the two processes as competing. One process is explicitly modeled, and the other is taken as a censoring process. The censoring process is usually assumed to be independent of the modeled event (conditional on the explanatory variables). In the limit order setting, when the conditioning variables are restricted to those known at the time of submission, this assumption is suspect. Price movements, for example, will change the joint probabilities of execution and cancellation. A price movement away from the limit price increase a trader's propensity to cancel, and also decreases the likelihood of execution. By incorporating post-submission conditioning information, these effects are brought into the model.¹⁵

The model is estimated separately for each stock using a random sample of 1,500 limit orders tracked through the first two seconds. Table 7 reports the mean and median of the coefficients' estimates across the stocks. We provide several different methods to help evaluate the strength of the results. First, we use *t*-tests and nonparametric tests for the sample's mean and median coefficients, taking into account cross-sectional variability in the estimated coefficient but not the model's estimation error. Second, we provide counts of both negative and positive coefficients, and separately the number of positive and negative coefficients that are significantly different from zero using the standard errors of the stock-specific coefficients.

These tests presume that the sample durations are independent over time and across firms. To investigate the validity of this assumption we examine the dynamic and cross-sectional properties of the estimated (martingale) residuals. The average within-stock first-order autocorrelation is approximately 0.005. To assess cross-sectional dependence, we construct for each stock an hourly mean residual, and compute the correlations of these series between all pairs of stocks. The average of these correlations is 0.004. These results support the independence assumption.

The mean coefficient on price aggressiveness (p_{ij}^{Relative}) is significantly positive (and so is the median) indicating that cancellation intensity increases the more aggressively the order is priced (yielding, e.g., a higher price for a limit buy order). This is consistent with the search hypothesis, as a search for hidden orders necessitates an aggressive order that betters the same-side quote. The mean coefficient on the change in same-side BBO is also significantly positive ($\Delta q_{ij,t}^{\text{Same}}$), indicating that cancellation intensity increases if the best bid (ask) price in the book increases (decreases) subsequently to the submission of a limit buy (sell) order. This is consistent with the chasing hypothesis, according to which a trader would cancel a limit order if the market price is moving away in order to resubmit a more aggressive limit order.

The effect of a change in the opposing-side BBO ($\Delta q_{ij,t}^{\text{Opposing}}$) is interesting in that it could represent the combined effects of more than one influence. For a buy order, the negative

¹⁵This improvement comes, however, at the cost of increased difficulty in forecasting. To predict the execution or cancellation durations for a hypothetical order at the time of submission, it is necessary to model the price evolution conditionally on order submission. It is not our present aim to develop a full model of the cancellation and execution processes. We frame the issue more narrowly as an attempt to investigate fleeting limit orders as strategic reactions to price movements.

Table 7

Duration Model of Limit Order Cancellation

This table presents analysis of limit order cancellations using a duration model. The cancellation hazard rate for limit order i of firm j at time t (relative to the time of order submission) is modeled as the proportional hazards duration model:

$$\lambda_{ij}(t) = \lambda_{0,j}(t) \exp \left[\beta_1 \left(\frac{\log \# \text{fleeting}}{\text{orders}_{ij}} \right) + \beta_2 \left| \frac{\text{lagged}}{\text{return}_{ij}} \right| + \beta_3 \left(\frac{\text{lagged}}{\text{volume}_{ij}} \right) + \beta_4 \left(\frac{\text{Relative}}{\text{NBBO spread}_{ij}} \right) + \beta_5 p_{ij}^{\text{Relative}} + \beta_6 \Delta q_{ij,t}^{\text{Same}} + \beta_7 \Delta q_{ij,t}^{\text{Opposing}} \right]$$

where $\lambda_{0,t}(t)$ is the (unspecified) baseline hazard rate. Lagged absolute value of return and lagged volume are cumulated over the preceding five minutes (relative to the time of order submission), and relative National Best Bid or Offer (NBBO) spread (dollar spread divided by the quote midpoint) is computed at the time of order submission. The “log # of fleeting orders” variable is the log of the maximum of either one or the number of fleeting orders in the preceding ten seconds. The preceding variables are standardized (within each stocks) to have zero mean and unit variance. The remaining variables are defined for a buy limit order as follows. Price aggressiveness $isp_{ij}^{\text{Relative}} \equiv [(\text{limit price}_{ij}) - (\text{INET bid}_{ij,t=0})]/(\text{INET bid}_{ij,t=0})$, where $t = 0$ denotes the time of order submission. Subsequent changes in same-side and opposing-side BBO are $\Delta q_{ij,t}^{\text{Same}} = [(\text{INET bid}_{ij,t}) - (\text{INET bid}_{ij,t=0^+})]/(\text{INET bid}_{ij,t=0^+})$, where $t = 0^+$ denotes the instant after submission, and $\Delta q_{ij,t}^{\text{Opposing}} = [(\text{INET ask}_{ij,t}) - (\text{INET ask}_{ij,t=0})]/(\text{INET ask}_{ij,t=0})$. For a sell order, p_{ij}^{Relative} , $\Delta q_{ij,t}^{\text{Same}}$, and $\Delta q_{ij,t}^{\text{Opposing}}$ are defined in an analogous fashion (with opposite signs). Unlike the remaining variables in the model, the time-varying covariates $\Delta q_{ij,t}^{\text{Same}}$ and $\Delta q_{ij,t}^{\text{Opposing}}$ evolve subsequent to order submission. The model is estimated separately for each stock using a random sample of 1,500 limit orders tracked through the first two seconds. The table reports the mean and median of the coefficients across our sample of 100 Nasdaq National Market stocks. We provide several different methods to help evaluate the strength of the results. First, we use t -tests and nonparametric tests to look at the hypotheses of zero sample mean and median coefficients, taking into account cross-sectional variability in the estimated coefficient but not the model’s estimation error. We report the two-tailed p -values associated with these tests. Second, we provide counts of both negative and positive coefficients, and separately the number of positive and negative coefficients that were significantly different from zero (using the standard errors of the stock-specific coefficients).

	Log # Fleeting Orders	Lagged Return	Lagged Volume	NBBO Spread	p_{ij}^{Relative}	$\Delta q_{ij,t}^{\text{Same}}$	$\Delta q_{ij,t}^{\text{Opposing}}$
Mean	0.357	0.028	0.003	0.079	87.255	341.562	−101.760
Median	0.347	0.040	0.016	0.067	24.303	163.802	−25.372
t -test (p -value)	<.001	0.005	0.803	<.001	<.001	<.001	<.001
Sign test (p -value)	<.001	<.001	0.133	<.001	<.001	<.001	<.001
Wilcoxon test (p -value)	<.001	<.001	0.184	<.001	<.001	<.001	<.001
# coefficients <0	1	26	42	17	18	5	76
# coefficients >0	99	74	58	83	82	95	24
# coef. with t -stat. <−1.96	0	2	4	1	4	1	21
# coef. with t -stat. > +1.96	96	8	10	28	54	83	3

mean coefficient implies that if the ask (the opposing best price in the book) moves down post-submission, cancellation intensity increases. It might be hypothesized that a trader who sees the opposing-side BBO moving closer believes that his order has an increased chance of execution, and therefore that the returns to waiting (not cancelling) are higher. On the other hand, a drop in the ask also decreases the cost of a market order, and

cancellation may intensify prior to the submission of a market buy order, which we termed the cost-of-immediacy hypothesis. The sign of the mean coefficient is consistent with the dominance of the second effect. It is also possible that the bidder interprets the drop in the ask as a signal of new negative private information, and the cancellation is a response to a perceived increase in information asymmetry. We observe that this coefficient is not statistically different from zero for many of the stocks (76 out of 100), while at the same time significant negative coefficients outnumber significant positive coefficients 21 to 3. Hence, there is some evidence consistent with the cost-of-immediacy hypothesis, but it is much weaker than the evidence that is consistent with the other two hypotheses.¹⁶

Turning to the non-price explanatory variables in the specification, the coefficients of lagged absolute return, volume, and NBBO spread are generally positive. This is consistent with their implied effects in the logit specification. The coefficients of (log) number of fleeting orders (in the preceding ten seconds) are positive in all but one case, and generally significant. This suggests that multiple fleeting orders may be used as part of a dynamic strategy or that fleeting orders arise as a response to earlier fleeting orders.

The duration model can also provide us with estimates of the probability of a fleeting order (i.e., the probability that a limit order is cancelled within two seconds) conditional on any given path of the time-varying dependent variables subsequent to submission. The model does not parameterize the baseline hazard rate, so this calculation relies on the estimated survivor function. Table 8 reports the cross-sectional averages of the implied two-second cancellation probabilities under various scenarios for a representative \$30 stock. The scenarios are created by intersecting several levels of price aggressiveness (from three cents behind the same-side BBO to three cents ahead of the same-side BBO) and different magnitudes of the change in same-side BBO after submission (from zero cents to five cents ahead of the order).

The chasing effect is clearly visible. An order submitted one cent ahead of the same-side BBO ($p_{i,j}^{\text{Relative}} = \0.01) has a 0.333 average probability of being cancelled over the next two seconds if the same-side BBO does not move ($\Delta q_{i,j,t}^{\text{Same}} = \0), but a 0.367 average probability of being cancelled if a new order was immediately placed one cent ahead of it (so that the same-side BBO moves one cent, $\Delta q_{i,j,t}^{\text{Same}} = \0.01). If the same-side BBO advances by five cents, the average probability of a fleeting order increases to 0.478. We estimate the standard error of each stock's survival function by the delta method, and present in the table the cross-sectional average of the standard errors for each average probability estimate. The average standard errors are much smaller than the average estimates themselves, providing us with some confidence that the pattern we observe across the columns in the table is indeed meaningful.

The table also demonstrates that a limit order that is more aggressively priced is more likely to be fleeting, consistent with the search hypothesis. In Section 4.1 we noted that one

¹⁶Note that the mean coefficient on the prevailing NBBO spread is positive, which suggests increased probability of cancellation at times of wider spreads. This result, similar in spirit to the one from the multinomial logit specification, is entirely consistent with the cost-of-immediacy hypothesis despite the fact that a narrower spread is associated with a lower cost of demanding liquidity. Since the prevailing spread is known when the order is submitted, a trader should submit a market order if the spread is narrow enough to warrant demanding liquidity (rather than submitting and then cancelling a limit order). The cost-of-immediacy hypothesis has to do with a change in the cost of immediacy *after* the order is submitted, which is best measured by the variable $\Delta q_{i,j,t}^{\text{Opposing}}$ (the change in opposing-side BBO subsequent to order submission) rather than the prevailing spread.

Table 8

Duration Model: Probabilities of Fleeting Orders

This table presents fleeting order probabilities (i.e., two-second cancellation probabilities) implied by a duration model of limit order cancellation. The cancellation hazard rate for limit order i of firm j at time t (relative to the time of order submission) is modeled as the proportional hazards duration model:

$$\lambda_{ij}(t) = \lambda_{0,j}(t) \exp \left[\beta_1 \left(\frac{\log \# \text{fleeting}}{\text{orders}_{ij}} \right) + \beta_2 \left| \frac{\text{lagged}}{\text{return}_{ij}} \right| + \beta_3 \left(\frac{\text{lagged}}{\text{volume}_{ij}} \right) + \beta_4 \left(\frac{\text{Relative}}{\text{NBBO spread}_{ij}} \right) + \beta_5 p_{ij}^{\text{Relative}} + \beta_6 \Delta q_{ij,t}^{\text{Same}} + \beta_7 \Delta q_{ij,t}^{\text{Opposing}} \right]$$

where $\lambda_{0,j}(t)$ is the (unspecified) baseline hazard rate. Lagged absolute value of return and lagged volume are cumulated over the preceding five minutes (relative to the time of order submission), and relative National Best Bid or Offer (NBBO) spread (dollar spread divided by the quote midpoint) is computed at the time of order submission. The “log # of fleeting orders” variable is the log of the maximum of either one or the number of fleeting orders in the preceding ten seconds. The preceding variables are standardized (within each stock) to have zero mean and unit variance. The remaining variables are defined for a buy limit order as follows. Price aggressiveness is $p_{ij}^{\text{Relative}} \equiv [(\text{limit price}_{ij}) - (\text{INET bid}_{ij,t=0})] / (\text{INET bid}_{ij,t=0})$, where $t = 0$ denotes the time of order submission. Subsequent changes in same-side and opposing-side BBO are $\Delta q_{ij,t}^{\text{Same}} = [(\text{INET bid}_{ij,t}) - (\text{INET bid}_{ij,t=0^+})] / (\text{INET bid}_{ij,t=0^+})$, where $t = 0^+$ denotes the instant after submission, and $\Delta q_{ij,t}^{\text{Opposing}} = [(\text{INET ask}_{ij,t}) - (\text{INET ask}_{ij,t=0})] / (\text{INET ask}_{ij,t=0})$. For a sell order, p_{ij}^{Relative} , $\Delta q_{ij,t}^{\text{Same}}$, and $\Delta q_{ij,t}^{\text{Opposing}}$ are defined in an analogous fashion (with opposite signs). Unlike the rest of the variables in the model, the time-varying covariates $\Delta q_{ij,t}^{\text{Same}}$ and $\Delta q_{ij,t}^{\text{Opposing}}$ evolve subsequently to order submission. The model is estimated separately for each stock using a random sample of 1,500 limit orders tracked through the first two seconds. We compute for each stock the estimated two-second cancellation probability for various values of p_{ij}^{Relative} and $\Delta q_{ij,t}^{\text{Same}}$, scaled to reflect \$0.01 increments for a \$30 stock price. In these calculations, all other variables are set to zero, and we use the product-limit estimate of the baseline survivor function. The table reports cross-stocks averages of the two-second cancellation probabilities. We estimate the standard error of each stock’s survival function by the delta method, and present in the table the cross-sectional average of the standard errors (in parentheses) below each average probability estimate.

p_{ij}^{Relative}	$\Delta q_{ij,t}^{\text{Same}}$ (change in same-side quote post submission)					
	\$0.00	\$0.01	\$0.02	\$0.03	\$0.04	\$0.05
\$0.03	0.351 (0.019)	0.386 (0.021)	0.418 (0.024)	0.445 (0.027)	0.469 (0.031)	0.492 (0.035)
\$0.02	0.342 (0.017)	0.377 (0.020)	0.409 (0.023)	0.437 (0.027)	0.462 (0.031)	0.485 (0.035)
\$0.01	0.333 (0.016)	0.367 (0.019)	0.400 (0.023)	0.428 (0.027)	0.454 (0.031)	0.478 (0.035)
\$0.00	0.325 (0.015)	0.358 (0.018)	0.391 (0.022)	0.420 (0.027)	0.446 (0.032)	0.470 (0.036)
−\$0.01	0.317 (0.015)	0.350 (0.018)	0.382 (0.022)	0.411 (0.027)	0.438 (0.032)	0.463 (0.036)
−\$0.02	0.309 (0.015)	0.341 (0.018)	0.374 (0.023)	0.403 (0.027)	0.430 (0.032)	0.455 (0.037)
−\$0.03	0.302 (0.016)	0.333 (0.018)	0.365 (0.023)	0.395 (0.028)	0.422 (0.033)	0.448 (0.038)

could claim that an aggressive limit order submitted one cent ahead of the same-side BBO is simply an attempt to gain price priority as opposed to an attempt to search for hidden orders. The price priority argument cannot, however, easily explain why a trader would submit a limit order two or three cents ahead of the same-side BBO. The evidence in Table 8 that such limit orders (submitted two or three cents ahead) are even more likely to be fleeting than a one-cent-ahead limit order is also consistent with the search hypothesis.¹⁷

5. Discussion

The use of fleeting orders in US markets is a relatively new phenomenon. The earliest widely available limit order dataset for a US market is TORQ, which covers a sample of NYSE stocks from October 1989 through January 1990 (see Hasbrouck, 1992). We formed a sample of all TORQ limit orders that were not tick-sensitive or otherwise qualified (roughly 300,000 orders). Fig. 1 depicts estimated distributions of time-to-cancellation (Panel A) and time-to-execution (Panel B) for three samples: (i) the TORQ sample, (ii) the 1999 INET sample used in an earlier version of this paper, and (iii) the 2004 INET sample.

The distributions of times-to-execution for the three samples (Panel B) are relatively similar (though the probability of an execution in the first few seconds after submission is somewhat higher in the INET samples). This implies that the likelihood of an execution for a “patient” limit order that remains in the book has not changed over the years. The distributions of time-to-cancellation (Panel A) are markedly different, however, and the dissimilarities explain the differences in fill rates discussed earlier. In other words, what has changed over the years is that patient limit orders are no longer the overwhelming majority of limit orders that are submitted to the market.

To what might the differences in cancellation patterns across these samples be attributed? One possible explanation involves the transparency of the trading process. During the TORQ sample period, information about the limit order book was unavailable away from the NYSE floor. Boehmer, Saar and Yu (2005) find that when the NYSE began distributing information on the book in real time (January, 2002), the cancellation rate of limit orders increased by 17% and the average time to cancellation decreased by 24%. This suggests that, as the NYSE became more similar to INET in terms of pre-trade transparency, NYSE trading practices moved closer to those observed on INET. Another consideration is tick size. The minimum tick is 1/8 in the TORQ sample, 1/16 in the 1999 INET sample, and one penny in 2004.¹⁸ A smaller tick size decreases the protection time priority offers to limit orders in the book, which may result in shorter durations.

Since fleeting orders appear to be a recent phenomenon, it is instructive to consider the relevant ways in which markets have changed. We propose that four factors are at the heart of the change in market environment: technology, active trading, fragmentation, and

¹⁷It is worth noting that even limit orders submitted three cents behind the same-side BBO have a 30% chance of being cancelled within two seconds. The chasing, cost-of-immediacy, and search hypotheses we propose are primarily applicable to orders submitted slightly behind, at, or within the BBO. The strategic rationale for deliberately pricing fleeting orders away from the market is not clear. We conjecture that some of these may arise due to latencies of order placement in rapidly moving markets. For example, a buy order priced to better the bid observed at the time of submission turns out to lie away from the bid prevailing when the order is actually received and processed. The submitter observes this, and quickly cancels the order, possibly resubmitting at a better price.

¹⁸While the minimum tick size on Nasdaq was 1/16 in 1999, the Island ECN (as the INET platform was called in 1999) permitted trading in finer increments.

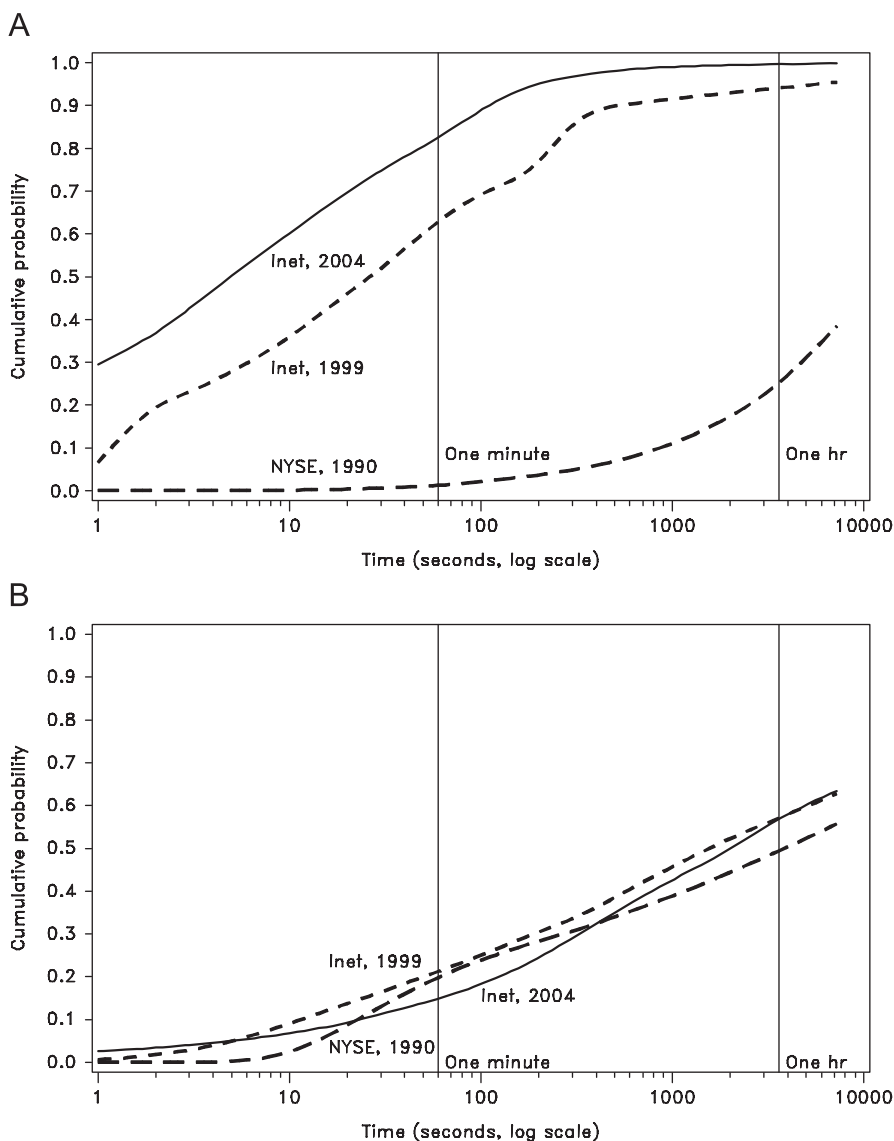


Fig. 1. Cancellation and Execution of Limit Orders: A Historical Perspective. This figure presents estimated distribution functions for time-to-execution and time-to-cancellation for three samples of limit orders covering different periods: NYSE in 1990, INET (or the Island ECN) in 1999, and INET in 2004. The functions are estimated using the life-table method. For time-to-execution estimates, cancellation is assumed to be an exogenous censoring event; for time-to-cancellation, execution is the censoring event. The INET 2004 sample is comprised of 100 stocks, identified as every fifth stock in the ranking of Nasdaq National Market stocks by equity market capitalization on September 30, 2004. The estimates are based on 1,000 randomly-chosen limit orders for each stock in October, 2004. The INET (Island ECN) 1999 sample is comprised of the 300 largest (by equity market capitalization) Nasdaq National Market stocks on September 30, 1999. The estimates are based on 1,000 randomly-chosen observations for each stock in the fourth quarter of 1999. The NYSE 1990 sample is from the TORQ database, and is comprised of 144 stocks. The estimates are based on all nonmarketable orders from November 1990 through January 1991. **Panel A:** Distribution of times to cancellation. **Panel B:** Distribution of times to execution.

the emerging importance of latent liquidity. The first of these, technology, particularly technology that is related to automated order management systems and algorithmic trading, appears to be a necessary precondition. It is difficult to imagine human traders managing fleeting orders efficiently, especially given their low probability of execution.

While the prevalence of automated order strategies has increased over time, indicative events include the 1995 establishment of the Financial Information (FIX) Protocol for electronic order transmission, and the development of sophisticated order routing algorithms at the end of the 1990s (like Tradescape.com's Smart Order Routing Technology). These systems enable traders to split orders across time and venues in an effort to obtain superior execution.

The second contributing factor, active trading, refers to a broad array of practices (day trading, breaking down of large orders by buy-side trading desks, "statistical arbitrage," etc.) that require frequent, repeated, and ongoing interaction with the market. The third contributing factor, market fragmentation, plays a role by increasing the use of search strategies, decreasing the execution probability of orders posted on any one book, and increasing the cost of limit order exposure due to the lack of time priority across market centers.

The risk of order exposure is further exacerbated by technological advancements that enable easy implementation of strategies that "front-run" existing limit orders in the book. This increase in the cost of order exposure is linked to the fourth factor—the emergence of latent liquidity. There are many pools of latent liquidity in today's markets, and some execution venues do not display any orders whatsoever (e.g., crossing networks like POSIT). In most limit order markets, traders have the ability to hide some or all of their orders. The existence of latent liquidity inside the spread on INET creates the opportunity for traders to search these prices in an attempt to obtain immediate execution for their orders at better prices.

This new form of interaction in the market seems prevalent because it satisfies traders' needs. Hidden orders help patient traders manage the risks associated with order exposure, but they may cause impatient traders to look for liquidity elsewhere if they are unaware of the hidden depth. The impatient traders need to balance their desire for immediacy with their willingness to incur a high price impact. It is their optimal choice that changes most noticeably with advanced technology and a commitment to active trading. With low-cost submission and cancellation of limit orders, impatient traders may be willing to search for latent liquidity. In the process, they are instigating a shift in the environment from one in which prices are posted (visible limit orders) into one in which searches (for latent liquidity) are needed in order to achieve better terms of trade.

We believe that this interplay between the providers of latent liquidity and those attempting to find it is behind the most curious feature of fleeting orders—the fact that they are visible orders that stay in the book for a short interval of time. An agent attempting to search for ("ping") hidden orders in the book could use hidden limit orders or immediate-or-cancel orders that would remain invisible unless executed. The visibility of fleeting orders may be serving as a signal to latent liquidity providers that there is increased likelihood of execution on INET, or may be used to trigger automated trading algorithms in the market.

This signaling device is important because market fragmentation creates a coordination problem: Patient traders need to decide where to post their hidden orders and impatient

traders need to decide where to search for the hidden depth. The visible fleeting orders on INET serve to solve the coordination problem by signaling to patient traders that impatient traders search this trading venue. This, in turn, encourages patient traders to submit hidden orders to the INET book, leading to the large number of executions against hidden orders that we observe.¹⁹

What are the overall welfare implications? If one considers technology, active trading, and fragmentation as exogenous factors, then both patient and impatient traders are better off in the new environment. Clearly patient traders can choose to submit visible limit orders, and their choice to use hidden orders must therefore be optimal. Similarly, the costly search of the impatient traders using fleeting orders makes sense only if they are better off carrying out the search than submitting marketable orders. Still, fragmentation and active trading could be viewed as responses to changing technological and regulatory environments, and hence endogenous. It is therefore unclear whether all traders fare better in the environment we observe than in a (hypothetical) alternative economy with a centralized trading venue that operates more as a posted price market than a search market.

Finally, while fleeting orders seem to figure prominently in Nasdaq stock trading, we should note that fleeting orders may not be so abundant in markets in which the trading rules do not allow for hidden orders or in which there are devices that discourage dynamic strategies involving cancellation and resubmission of orders. For example, the Euronext market (in Amsterdam, Brussels, Paris, and Lisbon) requires that a portion of each order be visible at all times. This, of course, prevents hidden orders from existing inside the spread and takes away some of the reasons to search for latent liquidity (although fleeting orders can still arise when traders “fish” for the existence of hidden depth behind visible orders at different price levels in the book). Euronext also presently charges for order submission if the ratio of orders submitted to executed trades of a member is more than five, which would discourage some dynamic strategies that involve extensive cancellation and resubmission of orders (Euronext, 2006).

6. Conclusion

A common economic perspective, and arguably the historical reality, is that a customer limit order closely resembles a dealer quote. In the present paper, however, we focus on a key attribute of current customer limit orders that strongly differentiates them from dealer quotes: the duration of the orders. Dealers seeking to establish and maintain reputations as dependable liquidity suppliers will tend to publish quotes that are highly persistent. We demonstrate that many customer limit orders, on the other hand, are cancelled after an extent that at first appears inexplicably brief.

¹⁹Another potential reason for the visibility of fleeting orders could be that it serves to induce competition among other quote setters, with the ultimate goal being execution of an order *in the opposite direction*. In its most blatant form, this has been considered market manipulation. In particular, wholesale market makers often guarantee execution of retail orders at the NBBO prices. A seller of, say, 2000 shares, might place an aggressively priced limit *buy* order for 100 shares, with the intent of establishing a favorable sale price, only for as long as necessary to effect the trade with the wholesaler. SEC litigation reports document several of these “spoofing” cases in recent years. The numbers are small (ten or fewer cases), but this may reflect the difficulty of detection. In light of the enormous number of fleeting orders on INET, however, we doubt that spoofing is at the core of this phenomenon.

We analyze a recent sample of limit order data from INET, and find that 36.69% of the (nonmarketable) limit orders are cancelled within two seconds. We term these “fleeting orders” and explore the puzzle they pose for traditional thinking about limit order markets. We provide three distinct hypotheses to explain why fleeting orders arise, and investigate the occurrence of these orders. We estimate a duration model with time-varying covariates that helps us analyze the strategic response of limit order traders to changing market conditions after the orders are submitted. The duration model uncovers evidence consistent with the use of dynamic limit order strategies by which traders attempt to accomplish their desired positions. These strategies seem to blur the traditional divide between supplying and demanding liquidity, as “component orders” of the strategy are placed and cancelled to affect the probability of achieving an execution.

We find that the incidence of limit order cancellation increases when market prices move away from the limit order price, which is consistent with our hypothesized chasing mechanism. We also find weaker evidence that limit order cancellations increase when the opposing quote moves closer, which is consistent with switching to a market order when the cost of immediacy goes down. Lastly, we document a number of results that are consistent with the hypothesis that fleeting orders are byproducts of searches for latent liquidity: (i) The relationship between the incidence of fleeting orders and market variables differs from that of regular limit orders, and bears some resemblance to that of market orders; (ii) limit orders that are priced more aggressively are more likely to be fleeting; and (iii) there are many executions that occur against hidden orders, especially at those price points where fleeting orders are found as well.

Since US samples of limit orders prior to the end of the 1990s do not exhibit such rapid cancellations, we believe that fleeting orders have arisen from improved trading technology, the emergence of an active trading culture, increased market fragmentation, and the ability to hide liquidity in limit order books. These factors induce changes in trader behavior and seem to have triggered a (partial) shift from a posted-price environment to a search environment. In particular, patient investors use hidden orders to supply liquidity to the book and impatient traders use limit orders priced inside the best visible prices in the book to search for the latent liquidity. While impatient traders bear the cost of the search, the larger supply of shares at better prices compensates them for it, and they are better off. The visibility of fleeting orders helps coordinate this interaction because it signals to patient traders that impatient traders are willing to search for liquidity on INET, which is one of several trading venues in a fragmented environment.

We view our documentation of fleeting limit orders and their role in the trading process as a contribution toward understanding the new realities of trading and as an impetus to changing the conventional dichotomy of market versus limit orders. Equilibrium in market microstructure models is usually characterized by a balance in the supply and demand for liquidity. Fleeting orders do not, however, readily fit into this perspective. They are more naturally viewed as intermediate steps in dynamic strategies that are actively pursuing executions. The “smart order” in modern markets is essentially a trading strategy that can use multiple orders sent at different prices to different venues in order to achieve a single execution.

Our results also suggest that limit order fill rates, computed and reported by market centers under SEC Rule 605, are driven more by the behavior of cancellations than by executions. Hence, these rates are not very informative about the likelihood of finding a counterparty.

Our analysis also relates to one of the conclusions from the double auction literature that high efficiency of the mechanism depends more on the institution itself—the double auction—than on the optimality of the traders' actions (e.g., Easley and Ledyard, 1993; Friedman, 1993; Gode and Sunder, 1993; Rust, Miller and Palmer, 1993). This is most vividly demonstrated by Gode and Sunder (1993), who show how markets populated by “zero-intelligence” traders demonstrate high efficiency and converge to the competitive equilibrium. Friedman (1993) notes that many different forms of double auctions exist in the world. He hypothesizes that efficiency could be very sensitive to environmental details, and that the most appropriate variant of double auction is different in each circumstance. It may be that some features of the market structure, like the ability to submit hidden orders, have been instituted in response to the changing technological environment or the fragmentation of the market. Our results suggest, however, that the behavior of traders also adjusts to the new environment and creates a whole new way of interacting in the double auction institution. It may therefore be that, in a complex environment, both changes in the rules of the double auction institution and the optimal responses of traders are important in achieving efficient outcomes.

References

- Ahn, H.-J., Bae, K.-H., Chan, K., 2001. Limit orders, depth and volatility: evidence from the Stock Exchange of Hong Kong. *Journal of Finance* 56, 767–788.
- Allison, P.D., 1995. *Survival Analysis Using the SAS System: a Practical Guide*. SAS Institute, Cary, NC.
- Angel, J.J., 1994. Limit versus market orders. Working paper. School of Business Administration, Georgetown University.
- Biais, B., Bisiere, C., Spatt, C., 2003. Imperfect competition in financial markets: Island vs. Nasdaq. Working paper. University of Toulouse.
- Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance* 50, 1655–1689.
- Bloomfield, R., O'Hara, M., Saar, G., 2005. The “make or take” decision in an electronic market: evidence on the evolution of liquidity. *Journal of Financial Economics* 75, 165–199.
- Boehmer, E., Saar, G., Yu, L., 2005. Lifting the veil: an analysis of pre-trade transparency at the NYSE. *Journal of Finance* 60, 783–815.
- Chakrabarty, B., Han, Z., Tyurin, K., Zheng, X., 2006. A competing risk analysis of executions and cancellations in a limit order market. Working paper. Center for Applied Economics and Policy Research, Indiana University.
- Chakravarty, S., Holden, C.W., 1995. An integrated model of market and limit orders. *Journal of Financial Intermediation* 4, 213–241.
- Cho, J.-W., Nelling, E., 2000. The probability of limit order execution. *Financial Analysts Journal* 56, 28–33.
- Cohen, K.J., Maier, S.F., Schwartz, R.A., Whitcomb, D.K., 1981. Transaction costs, order placement strategy, and existence of the bid-ask spread. *Journal of Political Economy* 89, 287–305.
- Copeland, T., Galai, D., 1983. Information effects and the bid-ask spread. *Journal of Finance* 38, 1457–1469.
- Easley, D., Ledyard, J.O., 1993. Theories of price formation and exchange in double oral auctions. In: Friedman, D., Rust, J. (Eds.), *The Double Auction Market Institutions, Theories and Evidence*, Proceedings vol. XIV. Santa Fe Institute, Addison-Wesley, Reading, MA, pp. 63–97.
- Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. *Journal of Financial Economics* 19, 69–90.
- Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. *Journal of Finance* 47, 576–605.
- Ellul, A., Holden, C.W., Jain, P., Jennings, R.H., 2007. Order dynamics: Recent evidence from the NYSE. *Journal of Empirical Finance* 14, 636–661.
- Euronext, 2006. Cash market equities and rights trading fee structure (effective 22/12/2006).
- Foucault, T., 1999. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets* 2, 99–134.

- Foucault, T., Kadan, O., Kandel, E., 2005. Limit order book as a market for liquidity. *Review of Financial Studies* 18, 1171–1217.
- Friedman, D., 1993. The double auction market institution: a survey. In: Friedman, D., Rust, J. (Eds.), *The Double Auction Market Institutions, Theories and Evidence*, Proceedings vol. XIV. Santa Fe Institute, Addison-Wesley, Reading, MA, pp. 3–26.
- Glosten, L.R., 1994. Is the electronic open limit order book inevitable? *Journal of Finance* 49, 1127–1161.
- Glosten, L.R., Milgrom, P.R., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Gode, D.K., Sunder, S., 1993. Allocative efficiency of markets with zero-Intelligence traders: markets as a partial substitute for individual rationality. *Journal of Political Economy* 101, 119–137.
- Goettler, R.L., Parlour, C.A., Rajan, U., 2005. Equilibrium in a dynamic limit order market. *Journal of Finance* 60, 2149–2192.
- Goettler, R.L., Parlour, C.A., Rajan, U., 2007. Microstructure effects and asset pricing. Working paper. Haas School of Business, U.C. Berkeley.
- Greene, W.H., 2002. *Econometric Analysis*. Macmillan, New York.
- Hamao, Y., Hasbrouck, J., 1995. Securities trading in the absence of dealers: trades and quotes on the Tokyo Stock Exchange. *Review of Financial Studies* 8, 849–878.
- Hansch, O., 2003. Island tides: exploring ECN liquidity. Working paper. Smeal College of Business, Penn State University.
- Harris, L.E., 1998. Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions and Instruments* 7, 1–76.
- Hasbrouck, J., 1992. Using the TORQ Database. Working paper. New York Stock Exchange.
- Hasbrouck, J., Saar, G., 2002. Limit orders and volatility in a hybrid market: the Island ECN. Working paper. Department of Finance, Stern School, NYU.
- Hendershott, T., Jones, C.M., 2005. Island goes dark: transparency, fragmentation and regulation. *Review of Financial Studies* 18, 743–793.
- Hollifield, B., Miller, R.A., Sandás, P., 2004. Empirical analysis of limit order markets. *Review of Economic Studies* 71, 1027–1063.
- Kaniel, R., Liu, H., 2006. So what orders do informed traders use? *Journal of Business* 79, 1867–1913.
- Large, J., 2004. Cancellation and uncertainty aversion on limit order books. Working paper. Nuffield College, Oxford University.
- Lo, A.W., MacKinlay, A.C., Zhang, J., 2002. Econometric models of limit order execution. *Journal of Financial Economics* 65, 31–71.
- Nguyen, V.T., Van Ness, B.F., Van Ness, R.A., 2004. The reporting of Island trades on the Cincinnati Stock Exchange. *Journal of Applied Finance* 14, 30–39.
- Parlour, C.A., 1998. Price dynamics in limit order markets. *Review of Financial Studies* 11, 789–816.
- Parlour, C.A., Seppi, D.J., 2008. Limit order markets: a survey. In: Boot, A.W.A., Thakor, A.V. (Eds.), *Handbook of Financial Intermediation and Banking*. Elsevier, Amsterdam.
- Renaldo, A., 2004. Order aggressiveness in limit order book markets. *Journal of Financial Markets* 7, 53–74.
- Rosu, I., 2006. A dynamic model of the limit order book. Working paper. University of Chicago.
- Rust, J., Miller, J.H., Palmer, R., 1993. Behavior of trading automata in a computerized double auction market. In: Friedman, D., Rust, J. (Eds.), *The Double Auction Market Institutions, Theories and Evidence*, Proceedings vol. XIV. Santa Fe Institute, Addison-Wesley, Reading, MA, pp. 155–198.
- Sandas, P., 2001. Adverse selection and competitive market making: evidence from a pure limit order book. *Review of Financial Studies* 14, 705–734.
- Seppi, D.J., 1997. Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies* 10, 103–150.
- Smith, J.W., 2000. Market vs. limit order submission behavior at a Nasdaq market maker. Working paper. National Association of Securities Dealers (NASD), NASD Economic Research.
- Stoll, H., 1978. The pricing of security dealer services: an empirical study of Nasdaq stocks. *Journal of Finance* 33, 1153–1172.