

LIMIT ORDER MARKETS: A SURVEY ¹

CHRISTINE A. PARLOUR
University of California Berkeley

DUANE J. SEPPI
Carnegie Mellon University

Current Draft: April 9, 2007

¹Forthcoming 2008 in *Handbook of Financial Intermediation & Banking*, edited by A.W.A. Boot and A.V. Thakor. We thank the authors of the research reviewed here for their insights into the microstructure of limit order markets and apologize for any errors or misrepresentations in our discussion of their work. We thank Arnoud Boot for asking us to undertake this survey and for his patience while we completed it.

When Walras sought inspiration in the nineteenth century for his eponymous model of markets, the Paris Bourse ran batch auctions. Periodically, an auctioneer aggregated orders and announced a market-clearing price. Later in the 1980s when Kyle [1985] and Glosten and Milgrom [1985] published their own eponymous theories of financial markets, the intermediation activities of NYSE specialists, the Tokyo Saitori, Nasdaq and London dealers, and floor traders in the Chicago futures pits were central to the trading process. Now in 2007 most equity and derivative exchanges around the world are either pure electronic limit order markets or at least allow for customer limit orders in addition to on-exchange market making.¹ This is specifically true of Euronext Paris, the successor to the Paris Bourse. The NYSE has progressively expanded the role of customer limit orders in its own trading process and, in addition, has recently acquired two limit order markets, Archipelago and Euronext. Nasdaq has had to adapt to the growing market share of ECN limit order markets while the electronic futures market ICE has taken market share away from floor-based futures exchanges. Given the prevalence of limit order trading, this article assays what we currently know and don't know about the economics of limit order markets.

A limit order is an ex ante pre-commitment (t, j, x, p) made on date t to trade up to a given amount x of a security j at a pre-specified limit price p . The order is in force until filled or cancelled. Unexecuted limit orders queue up in a limit order book. Limit orders are executed when other investors submit market orders or marketable limit orders. In particular, a market order is a request to trade immediately at the best price currently available in the market. Market clearing of limit orders is discriminatory: Each limit order executed in a transaction is filled at its respective limit price. It is this discriminatory execution property that distinguishes a limit order market from markets with a uniform market-clearing price (e.g., as in Walras or Kyle [1989]).

Markets typically impose price and time priority rules on limit order execution. Price priority means that limit orders offering better terms of trade — limit sells at lower prices and limit buys at higher prices — execute ahead of limit orders at worse prices. Time priority means that, at each price p , older limit orders are executed before more recent limit orders. The queuing discipline is thus “first in, first out” which rewards first-movers providing liquidity at a given price. Taken together, the price and time priority of a limit order translates directly into a probability distribution over execution timing.

Other market design issues also affect limit orders. Some exchanges restrict trading to limit orders and market orders exclusively. Others permit additional ex post liquidity provision by on-exchange market makers who decide how much to trade after a market order arrives. The specialist on the NYSE behaves in this way. Exchanges have also adopted a range of regimes for informational transparency. In an open book, all limit orders are observable to all investors while, in a closed book, traders cannot see the book. Some exchanges only disclose limit orders at a restricted set of prices. Others allow “iceberg” orders where part of a limit order is hidden from other traders. In addition, the information disclosed about investor identity varies across exchanges.

The basic economics of the trading process with limit orders all follows from limit orders being ex ante commitments to provide liquidity. Demsetz [1968] highlights the importance of inventory and waiting costs due to delays in limit order execution. Cohen, Maier, Schwartz and Whitcomb [1981] describes the trade-off between execution probability and price improvement in the optimal choice between limit orders and market orders. They also recognize that the sequential arrival of investor orders affects the entire trading process. As a result, they derive a non-infinitesimal aggregate bid-ask spread from the *gravitational pull* of existing quotes — that is to say, market

¹See Jain [2005] and Swan and Westerholm [2006].

orders to trade at existing quotes have guaranteed execution whereas even very aggressive limit orders do not. Copeland and Galai [1983] points out that *ex ante* commitments to trade, like limit orders and binding dealer quotes, give options to other traders to trade at the quoted prices. As such, limit orders are at an informational disadvantage since they can be *picked off* by later investors who receive updated public information or who have private information.

The ongoing research challenge is, theoretically, to model and analyze these basic intuitions in a rigorous equilibrium framework and, empirically, to quantify the importance of the various causal relations and, operationally, to develop optimized algorithms for practical use. This is no easy task. Despite the simplicity of limit orders themselves, the economic interactions in limit order markets are complex because the associated state and action spaces are extremely large and because trading with limit orders is dynamic and generates non-linear payoffs. A limit order executes against future market orders and competes against existing limit orders and against limit orders that may be submitted in the future. Thus, when choosing limit prices and quantities for (potentially multiple) limit orders and choosing quantities for market orders, a trader needs to condition on everything that can affect the future evolution of the trading process. This potentially includes a complete description of the existing limit order book — namely, all quantities for multiple orders at multiple prices from multiple past investors at multiple points in time — as well as the histories of all past trades and orders. The high dimensionality of limit order markets is a challenge for theoretical modelling and empirical estimation as well as, more practically, for trading. Dynamic trading strategies also involve decisions about how frequently to monitor changing market conditions and when and how to modify or cancel unexecuted limit orders. Lastly, limit orders have non-linear payoffs. In some future states they execute (and have linear payoffs in future cash flows) while in others they do not.

Research on limit orders is a area of intense activity. Over the last decade this effort has produced a number of significant new insights. Consequently, now is a good time to take stock of what has been accomplished and what is still left to be done. Our survey describes the main conceptual insights about limit orders and points out connections between theory and the empirical evidence. We also highlight modelling obstacles and the devices used to surmount them. Some of the broad thematic punch lines are as follows:

Price formation: The process of price formation in dynamic limit order markets differs fundamentally from sequential Walrasian markets and dynamic dealer markets. The Walrasian “market-clearing” price reflects an aggregation of supply and demand throughout the entire economy. In contrast, investors arrive and trade asynchronously in a limit order market so there is no unique market-wide “market-clearing” price. Rather, there is a sequence of bilateral transaction prices at which endogenously matched pairs of investors choose to trade over time. Similarly, the changing identity of limit order submitters is different from the Ptolemaic market makers in Kyle [1985] and Glosten and Milgrom [1985] who continuously set quotes at the informational and economic center of the market.

Liquidity: The distinction between liquidity supply and demand blurs in limit order markets. Investors with active trading motives may choose to post limit orders that are more aggressive than those a disinterested liquidity provider would use but less aggressive than market orders. Such limit orders are something in between pure liquidity supply and pure liquidity demand. In the extreme, limit buys (sells) can be posted above (below) the “efficient” price given public information.²

²The efficient price is a term of art used to describe a statistically derived component of asset prices that excludes high-frequency microstructure “noise” due to inventory effects and compensation for liquidity.

Thus, quotes in limit order markets cannot always be decomposed into an efficient price plus a non-negative compensation for liquidity provision.

Dynamics: Limit order books change over time in response to parametric changes in the environment and because of random ebbs and flows in the realized supply and demand for liquidity. Trades and prices in limit order markets can also exhibit path dependencies given the sequence in which buyers and sellers arrive in the market.

Information aggregation: Given the risk of being adversely picked off and of costly non-execution, limit order books should impound forward-looking information about future price volatility, the intensity of future adverse selection, and future order flow. This has been confirmed empirically. A richer picture is also emerging about the interaction between information and the supply and demand for liquidity. Limit orders are not just susceptible to being picked off by informed trading, they are also potentially a vehicle for informed trading themselves.

Inter-market competition: Glosten [1994] shows that competitive limit order markets can provide maximal liquidity in the face of adverse selection frictions. In such environments, limit order markets are “inevitable” in the sense that they can implement “competition-proof” price schedules. However, limit order markets are not inevitable with non-informational market frictions. In particular, hybrid markets combining dealers and limit orders can co-exist with, and even drive out, pure limit order markets when there are order submission costs.

Understanding the economics of trading processes generally, and of limit orders specifically, is important for at least three audiences. First and most practically, investors and trading desks want to reduce their trading costs. Limit orders are potentially executed at better prices than market orders, but they run the risk of non-execution and are exposed to a winner’s curse problem of being adversely picked off if the security’s value moves past the limit price before the limit order can be cancelled. The optimal choice depends on the dynamics of future order submission decisions of other investors.

Second, exchanges are businesses that face competitive pressures to make their product (the ability to trade) more attractive to their customers (investors). The fact that so many exchanges are organized as limit order markets suggests that this market design attracts investors and, thus, business for exchanges. The reasons why and conditions under which limit order trading is attractive are, however, rooted in the economics of the interactions between investors that limit orders facilitate. Exchanges also grapple with how best to implement limit order trading in terms of market transparency and whether to have solely limit orders or whether to have a hybrid structure with both investor limit orders and market makers.

Third, economists outside of market microstructure are recognizing a deeper connection between trading, liquidity and asset pricing. The fact that an asset can be traded makes asset valuation a social activity. Optimal risk sharing and consumption smoothing requires heterogeneous investors with higher valuations to buy securities from investors with lower valuations. In a market with low trading frictions, securities can be valued under the expectation that cash flows will be received over time by the investors who attach the highest valuations to them. As Harrison and Kreps [1978] show, the resale option associated with a tradable asset determines its value. Hence, trading is not just a mechanism for price discovery; trading also creates value by allowing investors to reshuffle security ownership over time as their personal valuations change.

Frictions that prevent investors from trading, and realizing gains-from-trade, actually lower the ex ante value of assets. The frictions of interest here are not, however, exogenous costs but rather coordination problems that arise when investors arrive to trade asynchronously with different

information about asset cash flows and about the availability of potential counter-parties. To the extent that the rules of trade affect which potential trades are actually consummated, the choice of the trading mechanism can affect allocations and, hence, social welfare. The growing literature on liquidity and asset pricing suggests, moreover, that the interaction between trading mechanisms and asset prices is significant.³ A natural question, therefore, is whether society is better off because of the global adoption of limit order markets.

Our survey is preceded by several excellent earlier reviews. O'Hara [1994] is the first comprehensive overview of the microstructure literature. Madhavan [2000] and Biais, Glosten and Spatt [2005] describe subsequent advances in microstructure theory and Hasbrouck [2007] explains tests and methods used in empirical microstructure. Harris [2003] reviews lessons and insights of microstructure research for practitioners and policy makers. By contrast, our survey is focused specifically on limit order markets. This more narrow focus is justified since today limit order markets are the dominant institution for trading equities and other exchange-traded securities today.

1 Modelling limit orders

Microstructure questions of optimal trading and price discovery are usually considered separately from questions of portfolio choice and asset pricing and vice versa. This is mathematically convenient but potentially misleading. Investor trading decisions should ultimately be understood in the context of investor portfolio choices.

In the canonical portfolio problem, an investor i chooses a portfolio strategy θ_i consisting of holdings $(\theta_{i1t}, \dots, \theta_{iNt})$ in N securities at each date t to maximize her lifetime expected utility from consumption

$$\max_{\theta_i} u_i(c_{it_0}) + E_{t_0} \left[\sum_{t=t_1}^T e^{-\rho_i(t-t_0)} u_i(c_{it}) \right] \quad (1)$$

subject to a budget constraint on consumption $c_{it} = \sum_{j=1}^N (\theta_{ijt-1} - \theta_{ijt}) P_{jt} + \theta_{ijt-1} D_{jt}$. Here, D_{jt} are cash distributions paid at date t by asset j . This standard formulation assumes a competitive Walrasian market. At each date t there is a market-clearing price P_{jt} for stock j at which the investor's trades $x_{ijt} = \theta_{ijt} - \theta_{ijt-1}$ are executed. Thus, the investor solves (1) taking market-clearing prices and the ability to trade at those prices as given. Indeed, the fact that the problem is formulated in terms of asset holdings θ_{ijt} rather than trades x_{ijt} implicitly presumes that trade execution is both certain and effortless. The corresponding asset pricing process is usually represented as a rational expectations equilibrium.

Definition 1 *A rational expectations equilibrium in a Walrasian market is a set of asset prices and portfolio holding strategies such that at each date: (i) the supply and demand for each security are equated, (ii) each investor's portfolio strategy is optimal given the market-clearing prices, and (iii) investor beliefs are rational given the available information.*

Market institutions have evolved since the batched call auctions of Walras' time to allow for continuous trading. The fact that investors trade asynchronously complicates both market-clearing

³See Amihud and Mendelson [1986]; Brennan and Subrahmanyam [1996]; Easley, Hvidkjaer and O'Hara [2002]; and Pastor and Stambaugh [2003].

(i.e., connecting buyers and sellers) and price discovery (i.e., aggregating information to value future cash flows). When the arrival asynchronicity is too severe, dealers intermediate trades between investors. In most high volume markets, however, early investors can use limit orders, effectively, to negotiate trades with later investors.

The individual investor portfolio optimization problem changes dramatically in limit order markets. Rather than submitting a single order x_{ijt} for an exact amount to be traded at a known market-clearing price at a precise date t , investors potentially submit vectors of market and limit orders so as to react to random fluctuations in buying and selling interest over time. Since limit order execution is uncertain, investors do not know with certainty how much they will actually trade at date t given their submitted orders. This leads to random slippage between the investor's ideal portfolio and her actual holdings depending on how many limit orders are executed. In other words, portfolio holdings are stochastic. Consequently, the order submission decision can be viewed as inducing an optimal probability distribution from which an investor's realized trades and trade prices will be drawn.

Given the priority rules of a limit order market, an investor i arrives at date t with current security holdings θ_{ijt-1} and possibly a set x_{ijt}^O of old orders still outstanding. She then submits instructions x_{ijt}^I consisting of new limit and market orders and any cancellations of old orders. Given her orders and the subsequent flow of orders M_{jt} from all other investors in the market, let $x_{ijt} = x(x_{ijt}^O, x_{ijt}^I, M_{jt})$ denote the realized number of shares traded by investor i between date t and the next time, $t + 1$, she enters the market. Let $\bar{P}_{ijt} = \bar{P}(x_{ijt}^O, x_{ijt}^I, M_{jt})$ denote the average price for these trades. Investor i does not know the flow of future orders from other investors when she submits her instructions x_{ijt}^I . Thus, the investor's problem in a limit order market is to use a dynamic order submission strategy that maximizes lifetime expected utility from consumption

$$\max_{x_i^I} E_{t_0} \left[u_i \left(\sum_{j=1}^N \theta_{ij0} D_{j0} - x_{ij0} \bar{P}_{ij0} \right) + \sum_{t=t_1}^{\infty} e^{-\rho_i(t-t_0)} u_i \left(\sum_{j=1}^N \left[\theta_{ij0} + \sum_{s=0}^{t-1} x_{ijs} \right] D_{jt} - x_{ijt} \bar{P}_{ijt} \right) \right] \quad (2)$$

given the uncertainty in consumption induced by randomness in the cash flow process D_{jt} and by randomness in the order flow process M_{jt} .

The optimization problem in (2) is more complex than the standard problem in (1) for three reasons. First, the action space at each decision date t is larger. Rather than just submitting a single order x_{ijt} , the investor in (2) makes multi-dimensional decisions about order type (i.e., whether to submit market orders, limit orders, or some combination of the two), limit order aggressiveness (i.e., at what prices to post limit orders), and order quantities (i.e., how many shares for each order). Second, the state space is larger. Rather than just conditioning on cash flow information and the corresponding market-clearing prices P_{jt} , the investor in (2) also conditions on everything that can affect the aggregate order flow process M_{jt} since M_{jt} affects the probability distribution over which orders will execute, $x(x_{ijt}^O, x_{ijt}^I, M_{jt})$, and over the prices, $\bar{P}(x_{ijt}^O, x_{ijt}^I, M_{jt})$, at which they will execute. This includes the composition of the current book and the history of past order submissions. Third, the decision dates t_0, t_1, \dots themselves in (2) are chosen by investor i rather than being predetermined dates for aggregate market clearing. Continuously monitoring the market is costly so investors do not trade continuously. Thus, the order submission dates for investor i can be modelled as Poisson events that occur with an intensity $\gamma(t, \mathcal{J}_{it})dt$ that depends on agent i 's information \mathcal{J}_{it} set at time t . The content and dynamics of \mathcal{J}_{it} is agent-specific and can

include private and public information about cash flows, common cross-investor trading motives, and investor-specific private value motives to trade.

The trading process in a limit order market is a continuous-time game in which a sequence of investors randomly enter (and re-enter) the market to solve portfolio/trading problems as in (2). In particular, each investor has her own individual Poisson order submission dates t . When aggregated together, the actions of all of the investors collectively determine the dynamics of the market-wide order flows M_{jt} . The economics of market clearing in such an environment is dramatically different from a Walrasian market. In particular, the notion of aggregate supply and demand being equated at a market-clearing price is replaced with the weaker notion of a Nash equilibrium in investor trading strategies where prices are simply the outcome of a series of bilateral transactions.

Definition 2 *A rational expectations equilibrium in a dynamic limit order market is a set of prices and order submission strategies such that at each date: (i) trades occur when arriving investors prefer trading with existing limit orders via market orders rather than submitting new limit orders of their own, (ii) these transaction prices satisfy the market’s priority rules, (iii) each investor’s order submission strategy is optimal given the order flows from the other investors, and (iv) investors’ beliefs are rational given their available information about future cash flows and about the endogenous dynamics of the market-wide flow of orders M_{jt} .*

No existing models, to our knowledge, formally imbed dynamic limit order submission decisions in a dynamic portfolio choice problem as in (2) or integrate aggregate limit order flow dynamics with consumption-based equilibrium asset pricing. Instead, issues of “how” investors trade are decoupled from issues of “why” they trade.

Once the order submission problem is detached from the portfolio problem, it is necessary to specify reduced-form trading preferences. Clearly, investors want to execute at the most favorable prices possible. More fundamentally, however, a trading benefit is needed to proxy for the consumption utility derived from trading. Otherwise, there would be no trading at all. One approach is to penalize traders if they fail to achieve a trading target. Another is to assume investors have private values, due to tax or hedging considerations, for particular portfolio positions. These potential private payoffs depreciate over time until trades are completed. Yet another approach penalizes execution waiting time directly. An important point in Engle and Ferstenberg [2006], however, is that reduced-form trading preferences ultimately should be compatible with investors’ consumption preferences. Extreme trading risk aversion, for example, is probably not consistent with low consumption risk aversion. Moreover, investors should be indifferent between trading strategies that achieve comparable consumption flows.

A variety of modelling assumptions reduce the dimensionality of the investor action and state spaces and simplify interactions between investors. Our taxonomy of models highlights assumptions about the order type decision, the timing of trades, the informational environment, and the extent of competition. Some models assume that the use of limit orders or market orders is exogenous while others explicitly model the choice between limit and market orders. The timing of trade can be static or dynamic. In static batch models, orders are aggregated across multiple investors and executed simultaneously in one round of trade. The trading uncertainty is about execution risk: Limit orders may or may not be executed. In sequential arrival models, traders arrive in the market and submit orders one at a time. Execution uncertainty is augmented with timing uncertainty about when limit orders will execute. The information environment in different models sometimes allows for adverse selection. When some investors have private information, limit orders are also

exposed to valuation risk since the value of the underlying asset may be correlated with the states in which limit orders execute. Models also differ in whether there is perfect or imperfect competition in liquidity provision and about the role of contemporaneous competition versus intertemporal competition via asynchronous limit order submissions at different dates.

Similar problems of dimensionality are encountered in empirical studies of limit order data. For tractability, empirical tests focus on a small set of economic actions — order type choices, order quantities, order aggressiveness, and order and transaction timing — and condition on a relatively small number of empirical summary statistics for the state of the market.

1.1 Static equilibrium models

The first equilibrium limit order models are static and have trading by investors with sharply differentiated demands for immediacy. Rock [1996] started this approach followed by Glosten [1994] and Seppi [1997]. At an initial date 1, passive *liquidity suppliers* submit limit orders into a limit order book. These investors have no intrinsic motive to trade. They only trade to be compensated for providing liquidity to other investors with a demand for immediacy. At a later date 2 an *active trader* arrives and demands immediacy via a market order for a random number of shares x which is then crossed against the limit order book from date 1. The goal is to describe the shape of the aggregate limit order book given perfect contemporaneous competition among liquidity providers.

Limit orders are exposed to a variety of costs and risks. For concreteness, we focus on limit sells and let Q_j denote the cumulative quantity of limit sells at or below a generic price p_j . First, there is the possibility of non-execution. In particular, the trading rules of an exchange determine the set Γ_j of market orders which cause the marginal (i.e., last) limit order submitted at p_j to execute. For example, time priority in a pure limit order market implies that $x \in \Gamma_j$ if the market order is large enough, $x \geq Q_j$, to fill the entire queue up through p_j .⁴ Second, there is valuation risk due to public and private information. The expected asset value conditional on the realized market order x is represented by a monotone function $v(x)$ which reflects “picking off” risk, as in Copeland and Galai [1983] when market orders are conditioned on future public information, and the possibility of active traders trading on private information. This leads to the *upper tail expectation* property whereby liquidity providers recognize that the value of the asset is conditional on the information content of the market orders that trigger execution of different limit orders. Third, there may be up-front order submission costs c_j as in Seppi [1997] and ex post order execution costs g_j as in Sandås [2001] and Foucault and Menkveld [2005].

The shape of the aggregate book is determined by the the ex ante profitability of the marginal limit order at each price

$$\pi_j = [p_j - E(v(x)|x \in \Gamma_j) - g_j]\text{Prob}(x \in \Gamma_j) - c_j. \quad (3)$$

Competition drives expected profits from limit orders to zero. As more limit orders are submitted, cumulative depths increase which causes execution probabilities $\text{Prob}(x \in \Gamma_j)$ to fall and causes

⁴Pro rata allocation and randomization are other possible order allocation rules for pure limit order markets. Seppi [1997] characterizes the executable sets Γ_j in a hybrid market with a specialist in terms of thresholds that depend on the specialist’s profit from undercutting or not undercutting limit orders at price p_j . A similar intuition is implicit in Ready [1999] when specialists have the option to “stop” execution of a market order and then condition their undercutting decision on subsequent information. See Kavajecz [1999] and Harris and Panchapagesan [2005] for empirical evidence of strategic specialist behavior vis-à-vis the limit order book.

expected gross profits conditional on execution, $p_j - E(v(x)|x \in \Gamma_j)$, to shrink while leaving submission costs the same. In equilibrium the book satisfies a break-even condition: The equilibrium cumulative depths Q_1, Q_2, \dots set $\pi_j = 0$ at each price p_j with positive depth.

These models proved to be useful for policy purposes. For example, they explain why decimalization reduced market liquidity due to the impact of “penny jumping” on the incentive to submit limit orders.⁵ However, these models are also unrealistic in several ways. Most importantly, there is no order type decision. Investors either have an inelastic motive to trade, and are willing to pay for immediate execution via market orders, or they are entirely disinterested liquidity providers with no reason to trade other than to be compensated for supplying liquidity via limit orders. The static nature of these models also limits their ability to speak to order flow dynamics. Changes in the book occur over time only if structural parameters of the underlying costs and distributions change. Lastly, there is no market power in limit order submission. There are always enough competitive liquidity providers to insure that the limit order book is break-even rather than having ebbs and flows in limit order profitability.

Empirical evidence: Sandås [2001] interprets intraday snapshots of the limit order book as observations of a repeated one-period model. He then conducts the first structural GMM estimation of a limit order model. Two moment conditions are used. The first is a break-even condition. Recognizing that the actual marginal expected profit π_{jt} at any given time t may deviate from zero — either because of delays in the arrival of sufficient limit orders (in which case $\pi_{jt} > 0$) or because of active liquidity demand in the limit order book (in which case $\pi_{jt} < 0$) — the break-even condition is relaxed to mean just that, on average, the expected marginal profit is zero: $E[\pi_{jt}] = 0$. The second moment condition is rational valuation. Assuming that the conditional value $v(x)$ is linear in the market order size x , Sandås tests an over-identifying restriction that the price impact $v(x)$ impounded in the cross-section of depths in the limit order book is consistent with the time-series price impact of market orders.

Unconditional and conditional versions of the break-even condition are rejected using Sandås’s test for actively traded stocks on the Stockholm Stock Exchange. The model’s main difficulty in fitting the data is that the estimated impact of order flow implicit in the limit order book is greater than the observed time series price impact. In other words, the limit order book is, on average, not deep enough to drive average expected profits to zero. One possible interpretation is that limit orders do not arrive fast enough. Supporting the idea of adjusted lags, the expected profits on limit orders are decreasing as the length of time between market orders, during which limit orders can accumulate, is longer. A conditional model, allowing for time variation in price impacts and other variables as functions of changing state variables (e.g., price volatility), fares better than the unconditional (constant parameter) version, but is still rejected. A second difficulty is that the estimated order execution costs γ_j are negative. This suggests that limit orders are submitted, not by disinterested investors with trading costs, but rather by investors with private trading motives.

1.2 Equilibrium models with static order choice and a terminal penalty

One step towards full multi-period optimization is to introduce a terminal penalty for non-execution into a static model. Investors presumably dislike trading costs and also dislike deviations from

⁵See Goldstein and Kavajecz [200] and Jones and Lipson [2001] for evidence consistent with predictions in Harris [1994] and Seppi [1997].

trading targets. This suggests a representation of the investor's problem in which a vector of market orders and limit orders x_i^I is submitted to solve

$$\min_{x_i^I} E(g[x(x_i^I, M) - \omega_i] + f[c(x_i^I, x(x_i^I, M))]) \quad (4)$$

where g is a penalty function given the realized deviation of investor i 's actual executed trades $x(x_i^I, M)$ from a personal trading target ω_i and f is a penalty function for order submission and execution costs $c(x_i^I, x(x_i^I, M))$. The expectation is taken over the random vector M of the aggregate order flow from all investors. The penalty function is a reduced-form for the continuation value in the Bellman equation. A shortcoming of this approach is that g is ad hoc rather than derived from an explicit dynamic programming problem.

Kumar and Seppi [1994] is an example of this approach. They model a market in which two different types of traders use limit orders. *Value traders* submit limit orders simply to exploit profit opportunities in the limit order book but do not need to trade per se. In contrast, *liquidity traders* have an active motive to trade in response to random individual liquidity shocks ω_i . Market clearing is a simultaneous move game in which buyers and sellers submit market and limit orders at the same time. Randomness in the trading demand of the liquidity traders leads to price risk for market orders and execution risk for limit orders.

Assuming a quadratic specification for (4) leads to optimal orders $x_{ij} = b_j \omega_i$ that are linear in the individual trading targets ω_i . The coefficient b_j is a function of the expected costs and probabilities of execution. After integrating over a continuum of small price-taking liquidity traders and then solving a fixed point problem for the equilibrium b_j coefficient, the model produces aggregate market and limit order flows which have an endogenous linear factor structure. This factor structure is qualitatively consistent with a block diagonal correlation matrix in which buy (sell) market orders are positively correlated with buy (sell) limit orders. In addition, if repeated over time, the target deviation in period t will induce autocorrelations in order submissions over time since unfilled orders at date t will roll over into additional trading demand at date $t + 1$. In sum, randomness in the limit order book should have a factor structure and that investors should submit vectors of limit and market orders rather than single orders.

Empirical evidence: Cao, Hansch and Wang [2004] finds that different points of the limit order book are co-integrated. Aitken, Almeida, Harris and McNish [2005] specifically confirms the use of vector order submissions by institutional investors on the Australian Stock Exchange.

1.3 Dynamic optimal control models for single agents

Static competition models focus primarily on the shape of the aggregate limit order book rather than on individual investor order submissions. However, order submission strategies themselves are of interest for at least two reasons. First, market-wide order flows M_{jt} are the aggregation of individual investors' order submissions. Thus, dynamic equilibrium models (discussed in section 1.4) focus on order submissions rather than on the shape of the book. Second, the growth of automated algorithmic trading has stimulated interest in order submissions purely as an optimal control problem. Theoretical and numerical analysis in Harris [1998], Angel [1994], and Obizhaeva and Wang [2005] solves for optimal trading strategies that minimize expected costs for a risk-neutral investor. In this work the dynamics of aggregate order flows M_{jt} are deemed exogenous.

Empirical evidence: The earliest empirical studies of limit orders focus on execution costs rather than on order submission decisions. Hasbrouck and Harris [1993] estimates expected trading costs for actual orders while Handa and Schwartz [1996] uses a back-testing approach in which hypothetical executions are simulated for fictitious small orders given actual price time series. Both studies find that, conditional on execution, limit orders have costs that are favorable to market orders, but that costs associated with non-execution can be significant. More recently, Nevmyvaka, Kerns, Papandreou and Sycara [2005] uses a mean/variance criterion to evaluate various back-tested limit order submission strategies where these strategies are contingent on market conditions.

1.4 Multi-period equilibrium models

Recent research represents limit order markets as sequential games rather than as static batch markets. Foucault [1999]; Parlour [1998]; Foucault, Kadan and Kandel [2005]; Goettler, Parlour and Rajan [2005a, 2005b, 2007]; and Rosu [2005] take this approach. All of these models imbed a discrete choice order submission problem in a variant of a dynamic multi-agent bargaining game. Risk neutral investors arrive sequentially and submit orders to maximize their expected gains from trade. In particular, the investor arriving at date t values the asset as the sum of an investor-specific random private value y_t plus possibly a random common component v_t . The order submission decision is formulated as a discrete choice problem with a penalty for non-immediate execution. Investors choose whether to use market orders or limit orders rather than being assumed to use a particular order type.

Investors in these models have local temporal market power in providing liquidity. This market power comes from two sources. First, quantity constraints restrict the number of shares any one investor can submit as limit orders. This prevents investors from individually driving the book to the break-even competitive depths. Second, only a small number of investors (often just one) are monitoring the market on any given date t and are able to act in real time. Investors who are not “present” — in that they are not actively monitoring the market or in that they have not yet arrived — cannot respond to the actions of investors who are present. This creates a window of time $[t, \tau]$, between t and the next time τ a competitor reacts, during which the only direct constraints on the market power in liquidity provision of investor t are the price and time priority of limit orders already in the book. As a result, there are too few liquidity providers — in contrast to the competitive batch models and their deep break-even books. Paradoxically, in equilibrium the shortage of liquidity leads not only to positive expected profits for some limit orders, but also, in other cases, to “desperate” limit order submissions that, while optimal, have *negative* expected profits.

The main goal of this line of research is to model endogenous order choice and the resulting patterns of order flow autocorrelation. An influential early impetus to this work was empirical evidence on intra-day order submissions on the Paris Bourse in Biais, Hillion and Spatt [1995]. In their study, orders are classified in terms of “aggressiveness,” ranging from market orders that “walk the book” and move prices (most aggressive) to limit orders placed behind the inside quotes (least aggressive). Using this schema, Biais, Hillion and Spatt [1995] document two important facts. First, order submissions are contingent on the “state” of the market. For example, a wide inside bid-ask spread increases the probability of price-improving limit orders and reduces the probability of market orders. Second, order submissions are autocorrelated. For example, there is a “diagonal effect” whereby orders with a particular level of aggressiveness tend to be followed by similar orders.

Subsequent research has confirmed these empirical regularities in many different markets.⁶

The order flow and trade dynamics in these multi-period models are derived from intertemporal bargaining by buyers and sellers on opposite sides of the market and intertemporal competition by traders on the same side of the market. Investors arrive and make trading decisions asynchronously which precludes Bertrand competition since future investors cannot respond contemporaneously to the actions of earlier investors. However, a type of imperfect intertemporal competition is still possible since the knowledge that more investors will arrive in the future affects the trades to which rational investors agree at earlier dates. Thus, an investor submitting a limit buy at date t competes indirectly with future potential buyers. If her bid is not sufficiently attractive, future sellers will submit limit sells in the hope of trading with future buyers rather than trading with the date t limit buy. Thus, intertemporal competition imposes dynamic incentive compatibility constraints on limit order submissions: Limit prices must be set such that at least some future traders will choose to trade with existing limit orders rather than submitting limit orders of their own on the other side of the market. In other words, bids must be set so that, for at least some future seller, the “bid in the hand is worth more than an ask in the bush” where the continuation value of the potential future ask itself depends endogenously on incentive compatibility constraints involving potential trading decisions of investors at even more distant future dates.

Modelling chains of incentive compatibility constraints is difficult. The first models to do this for limit order markets were Foucault [1999] and Parlour [1998]. A number of models followed which differ from each other in the progressive complexity and realism of the investor decisions and information sets and, specifically, in their assumptions about what happens after limit orders are submitted: How long do limit orders last before being cancelled? How frequently do investors return and modify their orders? These timing assumptions determine the bargaining power of the investor at date t relative to investors who arrived in the past and relative to investors who will arrive in the future.

Foucault [1999] identifies price quotation as an essential aspect of dynamic limit order trading. In particular, at what prices will investors post limit orders? To keep his analysis tractable, limit orders are assumed to survive for just one period. If unfilled after one period, they are exogenously cancelled. This timing assumption effectively turns limit orders into “take it or leave it” offers of liquidity to the next arriving investor. Foucault also assumes that the common value process v_t evolves on a binomial tree with equiprobable increments σ or $-\sigma$ and that the private value y_t takes one of two possible values, L or $-L$. Thus, there are four possible fundamental states for the arriving investor: $(+\sigma, +L)$, $(-\sigma, +L)$, $(+\sigma, -L)$, and $(-\sigma, -L)$. The resulting order submission and trade dynamics are intuitive. If the limit order book is empty, arriving investors with positive (negative) private values post limit buy (sell) orders in hopes of trading with a negative (positive) private value investor next period. The challenge is to determine the equilibrium bid and ask prices where limit orders will be posted when the book is empty. The fact that there are only four possible states next period and the fact that buyer and seller valuations can, given particular parameters, be ranked leads to two equations in two constant quote spreads, $a^* = A_t^* - v_t$ and $b^* = v_t - B_t^*$, above and below the (changing) common value v_t . The solution is the stationary equilibrium spreads.

The Foucault model does not make realistic empirical predictions about order flow dynamics. Indeed, given the one-period limit order shelf-life, there is at most one limit order in the book at any time. Rather, the main result is an analysis of the impact of Copeland and Galai [1983] “picking off”

⁶See Griffiths, Smith, Turnbull and While [2000] for the Toronto Stock Exchange; Rinaldo [2004] for Swiss Stock Exchange; Cao, Hansch and Wang [2004] for the Australian Stock Exchange; and related research discussed below.

risk on the equilibrium mix of limit and market orders. Increased common value volatility weakly increases the bid-ask spread which reduces the number of states in which investors submit market orders to trade with existing limit orders, thereby lowering the welfare gains from consummated trades. The intuition is that when value volatility is low, the required compensation for the risk of being picked off is sufficiently small that limit sells from a low private value $-L$ investor at t are executed in both the $(+\sigma, +L)$ and $(-\sigma, +L)$ states at $t + 1$. However, when volatility is high and the compensation for picking off risk must be large, then limit sells at the ask $v_t + a^*$ are only executed in the $(+\sigma, +L)$ state, not in the $(+L, -\sigma)$ state. In particular, an investor with a valuation $v_t - \sigma + L$ submits a limit buy at $v_t - \sigma - b^*$ despite the presence of a limit sell in the book. Thus, higher asset volatility increases the proportion of limit order submissions, reduces the welfare gains from consummated trades, and widens the bid-ask spread.

Empirical evidence: Ranaldo [2004] and others confirm that the inside limit order bid-ask spread is indeed increasing in price volatility. Furthermore, Ahn, Bae and Chan [2001] finds that the volume of limit order submissions is increasing in price volatility. These results are consistent with Foucault’s prediction. An alternative explanation, however, is that, rather than measuring potential picking off risk from fundamental valuation randomness, high lagged volatility may instead simply reflect the mechanical effect that prices are more volatile in thin markets. In this case, the observed positive volatility/limit order submission correlation could be spurious in that high volatility may indicate a thin book and a profitable trading opportunity which stimulates increased submission of limit orders.

Handa, Schwartz and Tiwari (2003) derives and tests another prediction of the Foucault model: The bid-ask spread should be greater in “balanced” markets than in unbalanced markets with unequal numbers of (high private value) buyers or (low private value) sellers.⁷ In unbalanced markets, the scarce type of traders have greater market power which lets them extract most of the gains-from-trade. Since they extract these gains-from-trade irrespective of whether they post limit orders directly or simply threaten to do so and thereby coerce more advantageous limit orders from their more numerous, desperate-to-trade counterparties, the result is that bid-ask spreads should be tighter in unbalanced markets. This prediction is confirmed empirically for the CAC40 stocks on the Paris Bourse.

Parlour [1998] models dynamic queue formation as another essential aspect of limit order trading. In particular, when will investors choose to join an existing queue of limit orders? Holding bid and ask quotes fixed for tractability, investors decide whether, given the current book, to submit a limit order of their own or to submit a market order. Limit orders in Parlour [1998] are long-lived and remain in the book indefinitely. This leads to a rich set of possible book dynamics as limit orders accumulate and are executed over time. This allows for more detailed predictions about state-contingent order flow autocorrelations than does the three-state book in Foucault [1999]. Limit orders are risky because they only execute if enough market orders arrive in the future to execute them plus all of the limit orders with priority ahead of them in the queue. In the model, investors trade to shift consumption between two dates given differences in their intertemporal rates of substitution. Investors with extreme time preferences have large gains to trade and endogenously demand liquidity from investors with less extreme time preferences. The critical time preference where the optimal order changes depends endogenously on the state of the limit order book.

⁷The Foucault bid-ask spread is, strictly speaking, a “shadow” spread between an actual limit order and a hypothetical order on the other side of the market rather than an actual spread between two concurrent limit orders.

The main result in Parlour [1998] is that the autocorrelations of transactions and order flow submissions reproduce a version of the diagonal effect: Market orders become more likely after market orders on the same side of the market. The intuition is that market buys, for example, reduce the available liquidity at the ask, thus making future liquidity provision at the ask more profitable, thereby shifting the critical time preference and causing more future sellers to choose to submit limit sells rather than market sells. More generally, serial correlation in order flow is shown to arise from liquidity dynamics as well as from informed trading.

Empirical evidence: The synergy between theory and empirics has been particularly fruitful in research into order submission dynamics. Taking advantage of the recent willingness of exchanges worldwide to provide order flow data, the empirical literature has disentangled and identified multiple factors driving order flows at different frequencies. Ellul, Holden, Jain and Jennings [2007] find strong positive serial correlation in orders at high frequencies (the diagonal effect) but negative autocorrelation at lower frequencies. They interpret this as waves of competing order flows arriving in quick succession in response to market events (e.g., due to mimicking, competition, and order splitting) within a stable cycle of random liquidity depletion and replenishment.

For the most part, reduced-form regressions have been used to test qualitative predictions about order submissions. An exception is Hollifield, Miller and Sandås [2004] which derives and tests structural restrictions on optimal order submissions in a model with sequentially arriving investors. Consider an investor who arrives at a date t with a high total common plus private valuation $v_t + y_t$ and who is restricted to submit at most a single limit or market order for q_t shares. Given the existing book and the parameters of the prevailing market environment, the investor's expected profit per share using a buy order at price p_j is

$$\pi_t(p_j, q_t) = \psi_t(p_j, q_t)(v_t + y_t - p_{jt}^{\text{trade}}) + \xi_t(p_j, q_t) - c \quad (5)$$

where $\psi_t(p_j, q_t)$ is the expected fraction of the order that will eventually be filled, p_{jt}^{trade} is the limit price p_j (for a limit order) or the volume-weighted execution price (for a market order), and $\xi_t(p_j, q_t)$ is the expected picking off risk due to future expected changes in the common value component given order execution, and c is an order submission cost.

The fact that the expected profit for each different order is linear in the private value y_t — with a slope equal to order j 's expected fill ratio $\psi_t(p_j, q_t)$ — means the optimal order submission strategy has a simple representation: There will be a set of intervals in the private values y for which different orders' profit lines are maximal. For each of these intervals, the order corresponding to the maximal profit line is, by construction, optimal. These optimal orders will be ordered as follows: Market buys are optimal given very high private valuations because they have the greatest slope/expected fill ratio $\psi(p_j, q_t)$. Limit buys with progressively lower bids and progressively lower expected fill ratios are optimal for realized private valuations in progressively lower intervals. A symmetric result holds for sell orders.

The key testable insight is that the thresholds delimiting these intervals — which are computed by equating the expected profit lines for the adjoining optimal orders — should be monotone decreasing as the expected fill ratios fall. The HMS statistic tests the monotonicity of estimated thresholds using empirical estimates of $\hat{\psi}(p_j, q_t)$ and $\hat{\xi}_t(p_j, q_t)$ in (5).⁸ Using a single Swedish stock

⁸This is just a test of order submission optimality since no market-clearing condition is imposed requiring the fitted individual investor optimal orders to aggregate up to market order flows that are consistent with the empirically estimated $\hat{\psi}(p_j, q_t)$ and $\hat{\xi}_t(p_j, q_t)$ functions.

to illustrate their methodology, the monotonicity restriction is rejected using buy and sell orders jointly. It is not known, unfortunately, how general this rejection is for other stocks. However, Hedvall, Niemeyer and Rosenqvist [1997] and Rinaldo [2004] also find reduced-form evidence of asymmetries in investor behavior on the two sides of the market.

Foucault, Kadan and Kandel [2005] combines endogenous quote determination on a multi-price grid, as in Foucault [1999], with queuing behavior given long-lived orders, as in Parlour [1998]. This allows for trade-offs between limit order price choices and execution waiting times. Limit orders are again infinitely lived and cannot be cancelled or changed. Investors' heterogeneous preferences for immediacy are captured by an explicit penalty on waiting time. Analytic expressions are obtained for the equilibrium trading strategies and the expected time until execution, but at the cost of several additional strong assumptions. Investors arrive sequentially and alternate deterministically between buyers and sellers. There is no quantity choice (all orders are for one share) and only quote-improving limit order submissions are allowed. In particular, limit order submissions deeper in the book are not allowed by assumption. The effect of these assumptions is that the inside spread becomes a sufficient statistic for the state of the limit order books. Price priority reduces to something we might call *spread priority*: The equilibrium execution priority of a limit order — irrespective of which side of the market it is on — increases the smaller the spread the order causes. On the same side of the market, this is automatic given price priority but, on the other side, it follows from alternating buyers and sellers and the quote-improving restriction.

The goal of their analysis is to develop predictions about the temporal properties of order submissions and trades rather than about order flow autocorrelation. A result that is directly relevant to the conditional autoregressive duration of transactions (see Engle and Russell [1998]) is that the frequency of transactions is weakly decreasing in the bid-ask spread. This is a consequence of the fact that both patient and impatient investors use market orders when the spread is at its minimum, but only impatient investors use market order when the spread is wider. Limit order books in the model also have “holes,” ranges of prices which investors jump over when submitting limit orders. Holes are a common feature of empirical limit order books (e.g., see Biais, Hillion and Spatt [1995]). This leads to the concept of *resiliency* which is measured as the probability that enough limit orders will arrive to return the book to the minimum bid-ask spread before the next transaction. Intuitively, the more potential holes there are, the fewer limit orders it takes to tighten the spread.⁹ The analysis also delivers comparative static results about “fast” and “slow” markets as measured by the frequency of order arrivals. For example, slower markets are shown here to have narrower spreads and to be more resilient.

Rosu [2005] models a continuous-time market similar to Foucault, Kadan and Kandel [2005] but with the innovation that investors can dynamically modify limit orders in real time. The result is the first fully dynamic but still analytic model of a limit order market. This is in contrast to previous models in which the market is dynamic but the individual investor decision problem is static. The ability to modify limit orders in real time is important because now the number of investors actively present in the market varies randomly and, in particular, can be more than just one. Consequently, liquidity providers are no longer local monopolists constrained solely by intertemporal competition. Now there is also contemporaneous competition as in the competitive batch models. Surprisingly, rather than complicating the model, the analysis is actually simplified

⁹There is some ambiguity about the notion of resiliency in the model. While holes lead to rapid spread recovery when limit orders arrive, they also cause rapid spread deterioration when market orders arrive.

due to a key insight: In equilibrium all investors with limit buys in the limit order book must have equal expected utilities (and analogously for investors with limit sells). Otherwise, with continuous prices, lower utility investors would revise their limit orders to undercut higher utility orders by an infinitesimal amount.

Rather than leading to Bertrand competition, the Rosu equilibrium has a set of M possible ask prices $\{a_1, a_2, \dots, a_M\}$ and N possible bid prices $\{b_1, b_2, \dots, b_N\}$ where limit orders are possible. Arriving patient investors fill in the book starting from the maximal ask a_1 and the minimal bid b_1 and then working in via quote-improving limit orders — which endogenizes what is an assumption in Foucault, Kadan and Kandel [2005]. In particular, the m th interior ask a_m has the property that it is the maximum price such that the investors who previously submitted the $m - 1$ other limit sells at a_1, \dots, a_{m-1} do not have an incentive to cancel their existing order and undercut the new limit sell. An analogous statement is true for interior bids. In this model, the number of buyers and sellers (n_t, m_t) with outstanding limit orders at each date t is a sufficient statistic for the state of the limit order book.

Rosu’s analysis leads to predictions about the shape of the limit order book, order flow autocorrelation, and the temporal properties of orders and trades. As in Rock [1996] and the other batch models, the shape of the limit order book depends on the probability distribution for arriving market orders. For example, sufficiently high probabilities of large (multi-unit) orders can lead to hump-shaped limit order books. Rosu also models patient and impatient investor arrival rates separately. This leads to a more intuitive result about the effect of fast markets: High impatient trader arrival rates lead to tighter spreads. Lastly, the limit order book is full when the “gravitational pull” of using a market order to trade with the best quote on the other side of the market outweighs the price improvement (net of expected waiting cost) from a limit order. A new phenomenon in Rosu is “fleeting” limit orders. Once the book is full, a patient investor on one side of the quotes may submit a short-lived trial limit order at an intermediate price proposing to “split the difference” with the patient investor on the other side of the market. This is one possible explanation for very short-lived limit orders documented in Hasbrouck and Saar [2002].

Empirical evidence: Causality in the relation between execution time and limit order submissions runs in both directions. On the one hand, Lo, MacKinlay and Zhang [2002] use survival analysis to show that limit order execution times are decreasing in the aggressiveness of limit prices. This is both a mechanical consequence of price priority rules and the potentially endogenous effect of aggressive order inducing latent demand for trade (i.e., aggressive limit orders reward investors on the other side of the market for submitting market orders rather than limit orders). On the other hand, the premise in Foucault, Kadan and Kandel [2006] and Rosu [2005] is that investors care about execution time and that expectations about execution time affect order submissions. Tkatch and Kandel [2006] use a simultaneous equations specification to test for a causal impact of expected execution time on the decision of which orders are submitted while controlling for the causal impact of aggressiveness on execution time. They find that investors do appear to care about the expected execution time when trading equities and bonds on the Tel Aviv Stock Exchange.

Goettler, Parlour and Rajan [2005a] models limit order trading dynamics with a large decision set. Investors can submit multiple limit orders at different prices and choose order quantities. This step forward in terms of realism comes at the cost of analytic tractability. The equilibrium must be computed numerically. The difficulty is that the many order submission possibilities causes the dimensionality of the information set to explode. For example, if there are L possible depths at N

possible prices, then the number of possible states of the limit order book is L^N . Even numerically the curse of dimensionality can be severe.

Investors arrive sequentially to trade a risky asset that has a random common value component v_t and an investor-specific private value component y_t . The total value an investor receives/gives up upon execution of an order at a date $\tau \geq t$ is $v_\tau + y_t$ per share traded — where the common value component changes over time but an investor’s private value is unchanged. The size of each investor’s feasible trade is also bounded by a random variable z_t . The sequence of investor types (y_t, z_t) is uncorrelated over time. Given the cumulative limit order book L_t at the time she arrives, investor t submits a vector of market and limit orders X . As in Hollifield, Miller and Sandås [2004], unexecuted limit orders are subject to stochastic cancellation over time which acts like a discount rate. Making the cancellation probability a function of limit order mispricing relative to the changing common value is a reduced-form for market monitoring by limit order submitters.

Investors submit orders to maximize their expected gain-from-trade. While this is a one-time decision for individual investors, their optimization takes into account random order cancellation and internalizes the impact of their orders on the dynamics of future investors’ trading decisions. An equilibrium is a fixed point in the execution probability function μ_t^e and the expected common value function condition on order execution Δ_t^v (i.e., the risk of being picked off). Since these are high dimensional functions, the model is solved based on a numerical algorithm which limits the updating of probabilities and strategies to the set of numerically recurrent states.

The model produces among the richest set of conditional order flow dynamics yet derived. Perhaps as importantly, the analysis illustrates the fundamental differences between quote-driven markets — such as intermediated dealer markets and limit order markets in which only disinterested value traders provide liquidity — and order-driven limit order markets. For example, the common value v_t is frequently outside of the inside bid and ask quotes in the numerical simulations. Moreover, this is not solely due to stale quotes. When the sell side of the book is thin and the buy side is deep, potential buyers with a large positive private value y_t optimally submit limit buys at prices $p_j > v_t$. Such orders encourage future investors to submit market sells rather than limit sells and yet are still profitable relative to investor t ’s private value so long as $v_t < p_j < v_t + y_t$.

Empirical evidence: Lo and Sapp [2005] extend the empirical methodology on order choice by considering the order size decision jointly with the order aggressiveness decision. Using a simultaneous equations probit model, they find that aggressiveness and size are negatively correlated. This study is also noteworthy for using data from the foreign exchange market. As in equity markets, FX limit order submitters appear to trade off execution probability against price concessions.

1.5 Limit orders and private information

Kyle [1985] and Glosten and Milgrom [1985] have been the workhorse frameworks for adverse selection in securities markets. However, both make strong assumptions about the interaction between information and liquidity. In particular, liquidity providers are taken to be uninformed while informed investors demand liquidity via market orders. Similarly, the intuition in Copeland and Galai [1983] is that infrequently monitored limit orders are susceptible to being picked off by later better informed investors. The first formal limit order models, Rock [1996] and Glosten [1994], also treat market orders as potentially informed and limit orders as uninformed.¹⁰

¹⁰Limit orders are equally vulnerable to being picked off by investors with private information and by investors who can condition on subsequent public news faster than limit orders can be cancelled. In either case, the information

The recent focus on endogenous order choice has led to interest in rational expectations equilibria in which informed investors use both limit and market orders. This is a hard problem, but there has been some progress. One early model with informed limit orders is Chakravarty and Holden [1995]. If there is uncertainty about where uninformed investors will supply liquidity on the other side of the market (or about the random market orders from noise traders with batched market clearing), then strategic informed investors may also employ limit orders as insurance to bound the (random) price at which their market orders will trade. Another early informational model is Kumar and Seppi [1994]. Given that (as discussed in section 1.2) uninformed investors trade using packages mixing market and limit orders, informed investors must trade using the same mix of market and limit orders to avoid detection. More recently, Kaniel and Liu [2006] investigates the choice between market and limit orders by informed investors and and patient uninformed investors. Their model extends to an equilibrium setting earlier optimal order submission strategy results in Angel [1994] and Harris [1998] showing that informed investors can optimally use limit orders when private information is sufficiently persistent. Indeed, Bloomfield, O’Hara and Saar [2005] argues that informed traders are actually natural liquidity suppliers. In an experimental market they find that informed traders initially demand liquidity via market orders but then switch to provide liquidity via limit orders. Because informed traders know the value of the asset, they are the first to know when prices have adjusted to a level such that limit orders cannot be “picked off.”

Goettler, Parlour and Rajan [2007] numerically solves the first dynamic model of limit orders with asymmetric information. Briefly, this is a continuous-time game in which agents arrive randomly and may trade one share in an open electronic limit order market. Investors value the asset for its cash flows (the common value) and for portfolio motives (private value). The structure of the game differs from the earlier Goettler, Parlour and Rajan [2005a] model in two significant respects. First, the individual investor trading problem is now dynamic: Agents revisit the market probabilistically at which time they may revise or cancel previous orders. Thus, the model accounts for the endogenous order cancellation option. Second, there is endogenous acquisition of asymmetric information. Before the start of trade, investors decide whether or not to pay a fixed fee to receive private information in the future. Investors with the lowest private motive for trade, dubbed *speculators*, have the highest willingness to pay for information. This is intuitive since their strategies are most affected by small changes in the value of the asset. On average, the speculators are liquidity suppliers; therefore limit orders are on average submitted by informed traders. The same “race to trade” by informed investors as in Holden and Subrahmanyam [1992] operates to mitigate adverse selection in the limit order book. Interestingly, there is an inverse relationship between the informativeness of the limit order book and the volatility of the cash flow common value. When the underlying common value is volatile, informed traders are less likely to supply liquidity and do so at more conservative prices. As a result, the limit order market acts as a volatility multiplier: Small changes in underlying asset volatility lead to larger changes in transaction price volatility. In addition, the correlation between fundamental value changes and changes in the transitory component of prices (i.e., the difference between the transaction price and the common value) — which can bias asset pricing variables such as estimated betas — can vary cross-sectionally with stocks’ common value volatility.

Empirical evidence: Research into the information content of limit order submissions has largely

set of the market order submitter is superior to the information on which uninformed limit orders are conditioned at the time they are submitted. Of course, the mechanism through which information is revealed (and the limit order book is updated over time) is very different if information is announced or if it must be inferred from trading.

concentrated on high frequency return predictability. The initial evidence was mixed. Biais, Hillion and Spatt [1995] finds that price revisions move in the direction of previous limit order flows. This suggests that later investors infer information from prior limit order submissions. However, Griffiths, Smith, Turnbull and White [2000] finds a significant price impact of non-marketable limit orders in the opposite direction. More recent evidence, however, supports the hypothesis that limit orders are used by informed investors and, thus, reveal information. Cao, Hansch and Wang [2004] finds that lagged limit order book imbalances are informative about future price changes. Kaniel and Liu [2006] actually find evidence that informed traders may use limit orders more frequently than market orders.

One weakness with high frequency return predictability evidence is that it is unclear *what* limit orders are informative about. For example, Kavajecz and Odders-White (2004) suggests that limit orders may, in part, be informative about pockets of future liquidity rather than about future fundamentals. However, Caglio and Berber [2005] avoids this critique by investigating order submissions around events prone to private information (e.g., earnings announcements) and finds that the direction of limit order flow is correlated with subsequent realized events. For example, more buy limit orders are placed before positive earnings announcements. Lastly, while most of the evidence relates to *directional* information about the mean of subsequent prices, Foucault, Moinas and Theissen [2005] finds that the depth of the limit order book on Euronext Paris can be used to forecast future price volatility.

2 Market design

No one, to date, has formulated the mechanism design problem to which a dynamic limit order market is the solution. Thus, it is difficult to evaluate whether the limit order market structure is optimal. A complete mechanism design analysis would need to address a number of questions. Given the similarity between limit order markets and multi-unit auctions, does the discriminatory execution of limit orders prevent potential manipulation of uniform price mechanisms as in Back and Zender [1993]? Does time priority discourage collusion by liquidity providers as in Dutta and Madhavan [1997]? However, some progress has been made on three market design issues: The robustness of limit order markets to competition, the welfare properties of limit order markets, and optimal limit order transparency.

2.1 Competition and limit order markets

Glosten [1994] comes closest to addressing the optimal market design question. He demonstrates that competitive limit order markets provide the maximal liquidity in the presence of adverse selection and monopsonistic liquidity demand. This leads to a striking result: Under certain conditions, limit order markets are competition-proof — the entry of a rival market cannot profitably improve the liquidity provided by a competitive limit order market — and inevitable — the entry of a limit order market can provide additional liquidity if existing markets earn non-negative profits on liquidity provision. The intuition is as follows: In the model there are competitive risk neutral liquidity providers and a single liquidity demander who maximizes quasi-convex preferences over shares and cash balances. Given her market power, the monopsonistic liquidity demander decides

how much to trade based on the marginal cost of liquidity.¹¹ Thus, when trading on multiple competing liquidity supply schedules, the liquidity demander splits up her total trade x to equate marginal costs in all markets in which she trades (up to any caps on how much a liquidity provider will trade). For liquidity providers, the expected profit from providing the Q th incremental share of aggregate liquidity is an upper-tail expectation given the asset value conditional on the information revealed by the total amount traded. Since competition among risk neutral liquidity providers in a stand-alone limit order market drives these expected profits to zero, the entry of a new rival market providing additional liquidity at any particular marginal cost level can only drive the expected profit negative. Similarly, a stand-alone liquidity supply schedule which earns non-negative expected profits, but which differs from the competitive limit order book, must have at least one price where the expected marginal profit from incremental liquidity is positive. A limit order can then enter and profitably provide liquidity at that point.

The world-wide ascendancy of limit order markets might appear to validate Glosten’s result, but inevitability is not assumption-free. Thus, the full reach of the inevitability result is still an open theoretical question. For example, although both market and limit order quantities are endogenously derived in Glosten [1994], the order type choice is exogenous. The optimal mechanism with endogenous order type choice is not known. There are, however, other caveats to limit order inevitability about which more is known.

One caveat is that non-informational trading costs are empirically significant. Huang and Stoll [1997] estimate that order processing costs account for over 80 percent of the bid-ask spread. Parlour and Seppi [2003] specifically considers heterogeneous non-informational submission costs and finds that the impact on inter-market competition is quite different from that of adverse selection. Unlike asymmetric information costs which depend on information revealed by the total trade of the active investor across all markets, order submission costs are independent of what happens on other markets. Parlour and Seppi [2003] extends Seppi [1997] specifically to model competition between a hybrid limit order market — with both a limit order book and a specialist who can provide ex post price improvement by undercutting the limit order book after the market order has arrived — and a pure limit order market with no specialist. They find multiple equilibria in which the outcome depends on the tie-breaking “preferencing” rules investors use to split orders between the two markets when indifferent. In particular, equilibria exist in which the hybrid market dominates the pure limit order market and in which the two markets co-exist. Foucault and Melkved [2005] uses a similar analysis to show that, with time priority (rather than pro rata rationing as in Glosten [1994]), the cumulative limit order depth with multiple competing pure limit order markets can exceed that of a single pure limit order market.

A second caveat is that equilibrium outcomes depend on who is trading as well as on institutional structure. Changing the characteristics of traders can lead to very different outcomes with the same market structure. The extent of the differences can be illustrated using the model of Seppi [1997]. His model gives the equilibrium in a stand-alone pure limit order market (PLM) with value traders and one active trader and also the equilibrium in a stand-alone hybrid market (HLM) with value traders, one active trader, and a specialist who offers ex post price improvement by undercutting the limit order book after the market order arrives but before limit orders are executed. Consider what happens when the person who would have been the specialist — who we call here the *wouldbe specialist* — joins the other traders in a pure limit order market. In particular, suppose the *wouldbe*

¹¹In contrast, competitive liquidity demanders in a batch market would trade based on the price of the marginal share.

specialist continues to monitor the market in real-time but is stripped of his ability to ex post undercut the limit order book by interposing his order in between market orders and limit orders.

Proposition 1 *If (i) the active trader is not limited to market orders but can submit a single limit order or market order and (ii) if the wouldbe specialist retains the specialist’s bilateral bargaining power vis-a-vis the active trader,¹² then the HLM equilibrium can be implemented on a pure limit order market.*

The difficulty in implementing the HLM equilibrium is that the wouldbe specialist cannot undercut limit orders unilaterally on a pure limit order market. To circumvent this difficulty requires the active trader’s cooperation. The active trader submits a marketable limit order (rather than a market order) which crosses with limit orders up through a quantity-appropriate stop-out price at which point the unexecuted residual is posted as a limit order. The wouldbe specialist monitors the market and, seeing the advantageously priced limit order at the stop-out price, submits a market order to clean up the residual limit order using the same liquidity supply schedule he would as the specialist in a hybrid market. The active trader knows this schedule and, given that the wouldbe specialist retains the bargaining power, submits the right residual order at the appropriate stop-out price given the total amount she wants to trade. The active trader is willing to enable the wouldbe specialist’s undercutting of the limit order book because this reduces her overall trading costs relative to trading with limit orders at even worse prices.¹³ The limit order traders, knowing that the wouldbe specialist and the active investor will cooperate this way, rationally submit the HLM (rather than the PLM) limit order quantities. Thus, although the wouldbe specialist cannot unilaterally implement the HLM equilibrium — he has no special status on a PLM entitling him to intervene in the mechanical crossing of a market order with limit book — the HLM equilibrium can be collectively implemented. Thus, market institutions are not uniquely associated with equilibrium outcomes. In particular, the allocation implemented on a pure limit order market depends critically on the sophistication of the active traders and the presence or absence of a wouldbe specialist.

A third caveat, mentioned in Glosten [1994], is the absence of direct communication between traders. Communication is clearly an important channel for information aggregation and contracting in dealer markets. The impact of ignoring communication on limit order market inevitability is, however, unclear. On the one hand, communication may intensify informational asymmetries by reducing the amount of anonymous trading noise in which informed traders can hide. On the other hand, communication may also reduce the incentive to acquire information. We also note that in a dynamic context, there is some limited scope for communication in limit order markets. Hasbrouck and Saar [2002] empirically document a large number of fleeting limit orders, which are placed and then immediately cancelled, which, they suggest, may be a communication device to negotiate and propose possible divisions of gains-from-trade.

¹²Assumptions (i) and (ii) simplify the bargaining problem between the wouldbe specialist and the active trader. The assumption that the active trader can only submit a single limit order keeps her from tricking the wouldbe specialist into providing liquidity and returning to trade again. Similarly, the fact that the active trader actively wants to trade is assumed to prevent her from submitting credible “take it or leave it” offers to extort better liquidity from the wouldbe specialist who only trades to earn a profit.

¹³Biais, Hillion and Spatt [1995] find evidence of investors posting marketable limit orders to draw out unposted (or hidden) liquidity on the Paris Bourse. Short-lived fleeting limit orders in Hasbrouck and Saar [2002] may also be “advertising” by wouldbe-specialists that they are present and monitoring the market.

2.2 Imperfect competition and limit orders

An institutional mechanism which performs well under perfect competition may perform less well under oligopolistic or monopolistic conditions. The asynchronous trading models discussed in section 1.4 analyze intertemporal imperfect competition in limit order markets. Biais, Martimort and Rochet [2000] is an elegant analysis of contemporaneous imperfect competition. In their model, a group of N risk neutral liquidity providers precommit to quotation schedules to provide liquidity. After the schedules are posted, a risk averse investor arrives with both private inventory motives and private cash flow information. She decides how much to trade and then splits up her market orders to trade optimally on the various quoted schedules. As previously shown in Bernhardt and Hughson [1997], competition in price schedules need not lead to zero profits for the liquidity providers. Given adverse selection, price schedules are quantity-sensitive and, given order splitting, the competition (unlike in Kyle [1985]) is not of the “all or nothing” type that leads to Bertrand competition. The Biais et al model establishes the existence and uniqueness¹⁴ of a symmetric equilibrium in convex schedules (i.e., the price paid for the marginal share is increasing in the order size) where the liquidity providers earn positive profits. As the number of competitors grows, the equilibrium converges in the limit to the competitive limit order market in Glosten (1994).

The model provides one of the first characterizations of limit order books in a static non-competitive environment. In particular, the fact that the liquidity schedules are convex means that they are equivalent to a collection of limit orders. To see this, note that the total payment $T_i(x_i)$ associated with a market order x_i to market maker i can be written as

$$T_i(x_i) = \int_0^{x_i} t_i(z) dz \quad (6)$$

where $t_i(z)$ is the marginal price of the z th unit. If the schedule is convex, then the marginal prices $t_i(z)$ are increasing the quantity z just as for a schedule of limit orders. The analysis also illustrates that inter-market competition in liquidity provision and cost-minimizing order splitting in the absence of priority rules can mimic intra-market competition between liquidity providers on a limit order market with priority rules. However, the model cannot be viewed as competing ECNs since the quotation schedule submitted by an individual strategic liquidity supplier will not be the same as the aggregate schedule submitted by multiple investors on an ECN.

Instead of modelling competition between markets, Viswanathan and Wang [2002] asks whether liquidity demanders would prefer trading in an oligopolistic dealer market or trading in an oligopolistic limit order market. In each market alternative, liquidity providers compete by quoting price schedules and then the liquidity demander splits up her total trade across the competing schedules. In the dealer market, customer market orders are executed at a uniform price while, in the limit order market, market orders are executed in a discriminatory fashion. The assumption of a finite number N of liquidity providers with inventory costs means liquidity providers have market power. This leads to “bid shading” — that is, paying less than their actual marginal valuations for shares they buy (and, analogously, over-charging for shares sold). The aggregate limit order book price schedule has a zero-quantity spread and bid-shading which decreases at larger quantities. In contrast, the dealership market schedule is steeper but has no zero-quantity spread. As a result, small

¹⁴While the equilibrium is unique within the class of convex schedules and the equilibrium in convex schedules is an equilibrium within the class of all schedules, it is not established that the equilibrium is unique within the larger class.

orders receive better execution in the dealer market while larger orders receive better execution in the limit order market.

The welfare analysis is conducted ex ante before the realized shares traded is known. For a large family of bounded market order probability distributions, the expected selling proceeds are always greater in the limit order market. Thus, risk neutral liquidity demanders prefer oligopolistic limit order markets over oligopolistic dealer markets. However, the greater concavity (convexity) of the dealer market total proceeds (cost) for sell (buy) orders, given the steeper price schedules, means there is some level of volatility aversion such that risk averse liquidity demanders will prefer trading in an oligopolistic dealer market.

2.3 Dealer markets

Limit order markets and dealer markets are the two dominant forms of financial markets today, so understanding the similarities and differences between them is important. Back and Baruch [2005] prove an equivalence result for dynamic limit order markets and a class of dynamic dealer markets when investors can split up their trades over time. Their model is continuous in time and prices and has a strategic long-lived informed trader. The analysis begins by noting that discriminatory pricing and ex ante liquidity provision in a competitive limit order market means that limit prices are upper-tail conditional expectations: A market order for a block of x shares is executed in a discriminatory fashion at a sequence of limit prices where the limit price for the q th share of the order is $E_t(v|x_t^L \geq q)$ given the informed trader's strategy x_t^L in a limit order market at date t . In contrast, uniform pricing and ex post liquidity provision in a competitive dealer market implies that market-clearing prices are simple conditional expectations: A market order for x is all executed at the break-even value $E_t(v|x_t^D = x)$ given the informed trader's strategy x_t^D in a dealer market at date t . Next, the possibility is introduced of a *worked block* which is a rapid sequence of one-share market orders submitted essentially instantaneously. In this case, when a dealer sees the q th one-share market order arrive within a given instant, she can only condition on the knowledge that the total worked order size is at least q . Thus, she executes the q th unit at the upper-tail expectation $E_t(v|x_t^{WO} \geq q)$ given the informed trader's strategy x_t^{WO} for submitting work blocks in a dealer market at date t . The main result is that whatever outcomes can be implemented on a limit order market can also be implemented in a dealer market if investors use worked blocks.

2.4 Welfare

Separate from whether limit order markets are immune to competition is the question of whether limit order markets are socially desirable. The question here is not which mechanism minimizes the cost of liquidity, but rather which is more efficient in allowing investors to realize gains-from-trade. Thus, market power and private information, which lead to transfers between agents, can be ignored unless they impede efficient trades.

Answers to the efficiency question require measures of the investor costs and benefits from trading. Hollifield, Miller, Sandås and Slive [2006] use the first-order condition for the optimal order choice from Hollifield, Miller and Sandås [2004] to recover a probability distribution over investors' private values implied by observed order submissions. The model is estimated using data from the Vancouver Stock Exchange. The model can then be used to compute and compare the realized gains-from-trade from actual trading and the maximum possible gains-from-trade in a

frictionless benchmark. The results suggest that there is substantial variation in private values¹⁵ and that the VSE limit order trading mechanism achieves 90 percent of the maximum possible gains-from-trade.

These results are dramatic, but they are also subject to some caveats: The structural estimation assumes that arriving investors only submit orders once and that there is no asymmetric information. The fact that VSE stocks are generally thinly traded has the advantage of emphasizing strategic interactions in the market but also the disadvantage that ignored informational asymmetries may be substantial. Treating differences in information as differences in private values may cause the estimated dispersion in private values to be overstated.

In numerical simulations in Goettler, Parlour, Rajan [2005b], an open limit order market achieves 92 percent of a theoretical benchmark with no frictions. The paper does not derive the optimal trading mechanism with private values and asymmetric information, but it does find that social welfare with limit orders is better than under several alternate incentive compatible mechanisms.

These welfare comparisons are conservative since both papers take the set of traders participating in the market as given. It is reasonable to suppose, however, that investor arrival frequencies would change if the costs of trading were to go down. In other words, the composition of investors who choose trade in a market may be determined, in part, by the market design. Lastly, there are other aspects of a market that may also be important for welfare. For example, markets provide a public externality in the form of price discovery.

2.5 Robustness

Market failure occurs when there is no market-clearing price for liquidity. It is well known from Glosten [1989] that adverse selection problems can cause competitive dealer markets to fail when uninformed traders are price sensitive. Glosten [1994] shows the same is true for limit order markets. Given asymmetric information, there may not be enough price-sensitive uninformed demand to support any price schedule with a non-infinite slope. Portniaguina, Bernhardt and Hughson [2006] shows that limit order markets can fail even in the absence of adverse selection problems. They extend Seppi [1997] by making market orders price sensitive. The intuition for market failure is that if the limit order book is too thin, then price elastic market order submitters will scale back their market order submissions. However, as the endogenous distribution of submitted market order quantities shifts towards zero, the probability of limit order execution falls which, given ex ante limit order submission costs, leads to fewer limit orders and, thus, a thinner book. If market order submissions are sufficiently elastic, the limit order book may fail. As an example, they show that, in a hybrid market, cutting the tick size can lead to market failure since a smaller tick makes it easier for the specialist to undercut the book which, in equilibrium, makes the book thinner.

2.6 Transparency

Optimal limit order transparency has recently begun to receive attention.¹⁶ In a limit order market, transparency is a continuum going from a *closed book* in which the public knows nothing about the book to intermediate cases in which investors can choose to hide part of their orders (e.g.,

¹⁵This is consistent with Handa, Schwartz and Tiwari (2003) which uses GMM to estimate the deep parameters of a Foucault model for the Paris Bourse. The implied gains-to-trade are large — up to four francs for their sample.

¹⁶Rindi [2002] considers transparency in a rational expectations framework.

via iceberg orders) to an *open book* with real-time order disclosure. In terms of the granularity of information disclosed, exchanges might reveal aggregated depths at all prices or at just subset of prices. They might even reveal individual orders themselves. Order information is sometimes accompanied by investor identity information (e.g., broker codes). This can be useful if traders are differentially informed so that reputation matters.

Baruch [2005] constructs a static model of a hybrid market with a specialist in which liquidity traders and possibly an informed trader submit market orders. Limit order traders submit price contingent orders and the specialist sets a *stop out price* at which the market clears. If the limit order book is open, then liquidity suppliers compete more fiercely and, *ceteris paribus*, submit more aggressive orders. A counterbalancing effect is that a deeper book encourages the informed trader to submit larger orders, increasing adverse selection. The competition effect outweighs the adverse selection effect and (under specified conditions) displaying the limit order book is good for market order traders. They benefit both from a smaller price impact of their orders and because prices reveal more information. In sum, limit order traders and specialists extract fewer informational rents when the book is open.

Empirical evidence: In 2002, OpenBook allowed off-exchange investors to see the whole NYSE limit order book instead of just the best bids and offers. Boehmer, Saar and Yu [2005] find that order submission strategies appear to change. In particular, there is a higher cancellation rate and a shorter time to cancellation for limit orders once the book is open. The volume executed by floor brokers and specialists declined, suggesting that investors substituted away from floor brokers to limit orders and crowded out the specialists consistent with Baruch [2005]’s predictions. Further, characteristics of overall market quality such as the price impact of orders and price efficiency improved. This result may not be true for a pure limit order book: Madhavan, Porter and Weaver [2005], found the move to transparency on Toronto led to a decrease in overall liquidity and an increase in transaction costs and volatility. Sinaan, Weaver and Whitcomb [2003] finds that market makers compete more aggressively when they can post anonymous limit orders on ECNs.

Foucault, Moinas and Thiessen [2005] uses a natural experiment on Euronext for an event study on identity information disclosure. In 2001 Euronext stopped displaying trader IDs publicly. An important intuition from Copeland and Galai [1983] is that limit order submitters give away free options for others to trade at their limit prices. The value of these trading options is increasing in the underlying price volatility. Thus, strategic liquidity suppliers will condition the spread between their limit buy and sell orders on any private information they have about future price volatility. Uninformed liquidity suppliers will then attempt to infer volatility information from the limit order book. Specifically, they undercut this spread if they believe the spread in the book is too large, and match it if they believe that the spread correctly reflects future price volatility. If the market is transparent, liquidity suppliers who potentially have information about future price movements will sometimes bluff and post wide spreads, even if they know that they are unwarranted, to increase their profits. However, if there is anonymity, then they will only post the wide spreads when the price is indeed, going to be more volatile (i.e., they cannot bluff about their information). Thus, the introduction of anonymity can lead to both improved liquidity (the informed liquidity traders do not bluff) in terms of on average lower spreads and less informative quotes. The idea that limit orders impound forward-looking information about future volatility is also tested in Foucault, Moinas and Theissen [2005]. As noted above, they find that the depth of the limit order book on Euronext Paris can be used to forecast future price volatility.

3 Questions for future research

There is still much we do not know about limit order markets. In terms of the basic modelling of optimal trading strategies and market equilibrium, only very stylized environments have been studied thus far. Joint decisions about order aggressiveness and quantity have not been fully modelled and the role of optimal monitoring strategies in limit order trading is unexplored territory. The interplay between the use of limit and market orders and information aggregation also still needs to be worked out more fully. For example, how can order flow correlations due to liquidity dynamics be distinguished from order splitting and correlated trading on private information? An indication that limit order modelling is still in its infancy is that empirical research has largely focused on testing qualitative predictions of theory but not structural functional forms. In the few exceptions, such as Sandås [2001], the structural model is usually rejected. Much like the “equity premium puzzle” stimulated a wave of asset pricing theory, microstructure theory and empirics might benefit from greater attention to the quantitative and structural predictions of theory. For example, what individual investor order submission strategies aggregate into the observed aggregate order flow process?

The integration of trading strategies and portfolio optimization is still to be done. Since order execution depends on the arrival of counterparties, anything that affects future investors’ willingness to trade can change the price/execution probability tradeoff including systematic market-wide events. Some questions here are: How do investors value the riskiness of particular trading strategies? How does the fact that investors trade groups of stocks affect their order submission decisions vis-à-vis an investor trading just one stock? If investors have a demand for certain generic stock characteristics (e.g., growth/value, industry) rather than for a specific stock, how does that affect their order submission choices across stocks?

A fundamental question of interest to financial economists is why investors trade. Limit order submissions are potentially a useful window through which to observe investor heterogeneity (e.g., private trading motives, urgency for trading, etc.). This suggests, for example, potential interaction between limit order book characteristics and liquidity-based asset pricing.

Optimal market design and competition between markets poses some very timely questions and issues. As competition between demutualized profit-seeking exchanges intensifies, market design will be one front in that competition. Theory can provide guidance to regulators, customers, and the exchanges themselves. Some important questions still outstanding are: To what social welfare problem is a limit order market the solution? What are the welfare and competitiveness properties of limit order markets with random liquidity provision (via customer limit orders) as well as random liquidity demand? What is the theoretical basis for the apparently good welfare performance of limit order markets with asynchronous dynamic trading? Does information get aggregated more quickly via trading in limit order markets or in dealer markets? How do different transparency regimes and other market design decisions affect the efficiency and competitiveness of limit order markets? The large number of natural experiments involving changes in market design in different exchanges means that these questions can be examined both theoretically and empirically. A challenging question for structural estimation would be to see if the deep structural parameters of the trading economy are unchanged given changes in market institutions.

Theoretical modelling may also help with some significant methodological challenges in empirical limit order research. One challenge is data summary and representation. To handle the enormous order flow data sets, observations are typically aggregated. However, absent clear theoretical guidance, the appropriate form of aggregation is not known. Another challenge is that many

observables from limit order markets are endogenous and are simultaneously determined. Attempts to deal with endogeneity, such as Tkatch and Kandel [2006] and Lo and Sapp [2005], would benefit from more realistic theory that could identify theoretically justifiable exogenous instruments.

References

- [1] Ahn, H., K. Bae, and K. Chan, 2001, Limit Orders, Depth and Volatility: Evidence from the Stock Exchange of Hong Kong, *Journal of Finance*, 56, 767–788.
- [2] Aitken, M., N. Almeida, F. Harris, and T. McInish, 2005, Order Splitting and Order Aggressiveness in Electronic Trading, working paper.
- [3] Amihud, Y. and H. Mendelson, 1986, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics* 17, 223–249.
- [4] Anand, A. , S. Chakravarty, and T. Martell, 2004, Empirical Evidence on the Evolution of Liquidity, Choice of Market versus Limit Orders by Informed Traders, working paper, Syracuse University.
- [5] Angel, J., 1992, Limit versus Market Orders, working paper, Georgetown University.
- [6] Back, K. and S. Baruch, 2005, Working Orders in Limit-Order Markets and Floor Exchanges, *Journal of Finance* forthcoming.
- [7] Back, K., and J. Zender, 1993, Auctions of Divisible Goods: On the Rationale for the Treasury Experiment, *Review of Financial Studies* 6, 733–764.
- [8] Baruch, S., 2005, Who Benefits from an Open Limit-Order Book? *Journal of Business* 78, 1267–1306.
- [9] Berber, A. and C. Caglio, 2004, Order Submission Strategies and Information: Empirical Evidence from the NYSE, working paper, University of Lausanne.
- [10] Bernhardt, D., and E. Hughson, 1997, Splitting Orders, *Review of Financial Studies* 10, 69–101.
- [11] Biais, B., L. Glosten and C. Spatt, Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications, *Journal of Financial Markets* 8, 217–264.
- [12] Biais, B., P. Hillon and C. Spatt, 1995, An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse, *Journal of Finance* 50, 1655–1689.
- [13] Biais, B., D. Martimort and J. Rochet, 2000, Competing Mechanisms in a Common Value Environment, *Econometrica* 68, 799–837.
- [14] Bloomfield, R., M. O’Hara and G. Saar, 2005, The “Make or Take” Decision in an Electronic Market: Evidence on the Evolution of Liquidity, *Journal of Financial Economics* 75, 165–199.
- [15] Boehmer, E., G. Saar and L. Yu, 2005, Lifting the Veil: an Analysis of Pre-Trade Transparency at the NYSE, *Journal of Finance* 60, 783–815.

- [16] Brennan, M., and A. Subrahmanyam, 1996, Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns, *Journal of Financial Economics* 41, 441–464.
- [17] Cao, C., O. Hansch and X. Wang, 2004, The Informational Content of an Open Limit Order Book, working paper, Pennsylvania State University.
- [18] Chakravarty, S., and C. Holden, 1995, An Integrated Model of Market and Limit Orders, *Journal of Financial Intermediation* 4, 213–241.
- [19] Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb, 1981, Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread, *Journal of Political Economy* 89, 287–305.
- [20] Copeland, T. and D. Galai, 1983, Information Effects on the Bid-Ask Spreads, *Journal of Finance* 38, 1457–1469.
- [21] Demsetz, H., 1968, The Cost of Transacting, *Quarterly Journal of Economics* 82, 33–53.
- [22] Dutta, P., and A. Madhavan, 1997, Competition and Colusion in Dealer Markets, *Journal of Finance* 52, 245–276.
- [23] Easley, D., S. Hvidkjaer, and M. O’Hara, 2002, Is Information Risk a Determinant of Asset Returns?, *Journal of Finance* 57, 2185–2221.
- [24] Ellul, A., C. Holden, P. Jain and R. Jennings, 2007, Determinants of Order Choice on the New York Stock Exchange, working paper, Indiana University.
- [25] Engle, R., and R. Ferstenberg, 2006, Execution Risk, working paper, New York University.
- [26] Engle, R. and J. Russell, 1998, Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data *Econometrica*, 66, 1127–1162.
- [27] Foucault, T., 1999, Order Flow Composition and Trading Costs in a Dynamic Limit Order Market, *Journal of Financial Markets* 2, 99–134.
- [28] Foucault, T., O. Kadan and E. Kandel, 2005, The Limit Order Book as a Market for Liquidity, *Review of Financial Studies* 18, 1171–1217.
- [29] Foucault, T., and A. Menkveld, 2005, Competition for Order Flow and Smart Order Routing Systems, working paper, HEC.
- [30] Foucault, T., S. Moinas, E. Thiessen, 2005, Does Anonymity Matter in Electronic Limit Orders Markets?, working paper, HEC.
- [31] Glosten, L., 1989, Insider Trading, Liquidity and the Role of the Monopoly Specialist, *Journal of Business*, 62, 211–235.
- [32] Glosten, L. (1994), Is the Electronic Open Limit Order Book Inevitable? *Journal of Finance* 49, 1127–1161.
- [33] Glosten, L., and P. Milgrom, 1985, Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics* 21, 123–144.

- [34] Goettler, R., C. Parlour and U. Rajan, 2005a, Equilibrium in a Dynamic Limit Order Market, *Journal of Finance* 60, 2149–2192.
- [35] Goettler, R., C. Parlour and U. Rajan, 2005b, Information Acquisition in a Limit Order Market, working paper, Carnegie Mellon.
- [36] Goettler, R., C. Parlour and U. Rajan, 2007, Microstructure effects and Asset pricing, working paper, UC Berkeley.
- [37] Goldstein, M., and K. Kavajecz, 2000, Eighths, Sixteenths, and Market Depth: Changes in Tick Size and Liquidity Provision on the NYSE, *Journal of Financial Economics* 56, 125–149.
- [38] Griffiths, M., B. Smith, A. Turnbull and R. White, 2000, The Costs and Determinants of Order Aggressiveness, *Journal of Financial Economics* 56, 65–88.
- [39] Handa, Puneet and Robert A. Schwartz, 1996, Limit Order Trading, *Journal of Finance* 51, 1835–1861.
- [40] Handa, P., R. Schwartz, and A. Tiwari, 2003, Quote Setting and Price Formation in an Order Driven Market, *Journal of Financial Markets* 6, 461–489.
- [41] Harris, L., 1994, Minimum Price Variations, Discrete Bid–Ask Spreads and Quotation Sizes, *Review of Financial Studies* 7, 149–178.
- [42] Harris, L., 1998, Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems, *Financial Markets, Institutions & Instruments* 7, 1–76.
- [43] Harris, L., 2003, *Trading & Exchanges: Market Microstructure for Practitioners*, Oxford University Press.
- [44] Harris, L., and J. Hasbrouck, 1996, Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy, *Journal of Financial and Quantitative Analysis* 31, 213–231
- [45] Harris, L., and V. Panchapagesan, 2005, The Information Content of the Limit Order Book: Evidence from NYSE Specialist Trading Decisions, *Journal of Financial Markets* 8, 25–67.
- [46] Harrison, J. Michael and David M. Kreps, 1978, Speculative Behavior in a Stock Market with Heterogeneous Expectations, *The Quarterly Journal of Economics* 92, 323–336.
- [47] Hasbrouck, J., 2007, *Empirical Market Microstructure*, Oxford University Press.
- [48] Hasbrouck, J., and G. Saar, 2002, Limit Orders and Volatility in a Hybrid Market: The Island ECN, working paper, New York University.
- [49] Hedvall, K., J. Niemeyer, and G. Rosenqvist, 1997, Do Buyers and Sellers behave similarly in a limit order book? A high-frequency data examination of the Finnish Stock Exchange, *Journal of Empirical Finance* 4, 279–293.
- [50] Holden, C., and A. Subrahmanyam, 1992, Long-Lived Private Information and Imperfect Competition, *Journal of Finance* 47, 247–270.

- [51] Hollifield, B., R. Miller, and P. Sandås, 2004, Empirical Analysis of Limit Order Markets, *Review of Economic Studies* 71, 1027–1063.
- [52] Hollifield, B., R. Miller, P. Sandås, and J. Slive, 2006, Estimating the gains from trade in limit order markets, *Journal of Finance* 16, 2753–2804.
- [53] Huang, R., and H. Stoll, 1997, The Components of the Bid-Ask Spread: A General Approach, *Review of Financial Studies*, 10, 1035–1064.
- [54] Irvine, P., G. Benston and E. Kandel, 2000, Liquidity beyond the inside spread: Measuring and using information in the limit order book, working paper, Emory University.
- [55] Jain, P., 2005, Financial Market Design and the Equity Premium: Electronic versus Floor Trading, *Journal of Finance* 60, 2955–2985.
- [56] Jones, C., and M. Lipson, 2001, Sixteenths: Direct Evidence on Institutional Execution Costs, *Journal of Financial Economics* 59, 253–278.
- [57] Kaniel, R., and H. Liu, 2006, What Orders do Informed Traders Used, *Journal of Business* 79, 1867–1913.
- [58] Kavajecz, K., 1999, A Specialist’s Quoted Depth and the Limit Order Book, *Journal of Finance* 54, 747–771.
- [59] Kavajecz, K. and E. Odders-White, 2004, Technical Analysis and Liquidity Provision, *Review of Financial Studies* 17, 1043–1071.
- [60] Kumar, P., and D. Seppi, 1994, Limit and Market Orders with Optimizing Traders, working paper, Carnegie Mellon University.
- [61] Kyle, A., 1985, Continuous Auctions and Insider Trading, *Econometrica* 53 p 1315–1335.
- [62] Kyle, A., 1989, Informed Speculation with Imperfect Competition, *Review of Economic Studies* 56, 317–355.
- [63] Lo, A., A. MacKinlay, and J. Zhang, 2002, Econometric Models of Limit-Order Executions, *Journal of Financial Economics* 65, 31–71.
- [64] Lo, I. and S. Sapp, 2005, Price Aggressiveness and Quantity: How Are They Determined in a Limit Order Market?, working paper.
- [65] Madhavan, A., 2000, Market Microstructure: A Survey, *Journal of Financial Markets* 3, 205–258.
- [66] Madhavan, A., D. Porter and D. Weaver, 2005, Should securities markets be transparent? *Journal of Financial Markets* 8, 265–287.
- [67] Nevmyvaka, Y., M. Kearns, M. Papandreou, K. Sycara, 2005, Electronic Trading in Order-Driven Markets: Efficient Execution, *E-Commerce Technology: Seventh IEEE International Conference*, 190–197.

- [68] Obizhaeva, A., and J. Wang, 2005, Optimal Trading Strategy and Supply/Demand Dynamics, working paper, MIT.
- [69] O'Hara, M., 1995, *Market Microstructure Theory*, Blackwell.
- [70] Parlour, C., 1998, Price Dynamics in Limit Order Markets, *Review of Financial Studies* 11, 789–816.
- [71] Parlour, C., and D. Seppi, 2003, Liquidity-Based Competition for Order Flow, *Review of Financial Studies* 16, 301–343.
- [72] Pastor, L., and R. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642–685.
- [73] Portniaguina, E., D. Bernhardt, and E. Hughson, 2006, Hybrid Markets, Tick Size and Investor Trading Costs, *Journal of Financial Markets* 9, 433–447.
- [74] Ranaldo, A., 2004, Order Aggressiveness in Limit Order Book Markets, *Journal of Financial Markets* 7, 53–74.
- [75] Ready, M., 1999, The Specialist's Discretion: Stopped Orders and Price Improvement, *Review of Financial Studies* 12, 1075–1112.
- [76] Rindi, B., 2002, Transparency, liquidity and price formation, Working paper, Bocconi University.
- [77] Rock, 1996, The Specialist's Order Book and Price Anomalies, working paper.
- [78] Rosu, I. , 2005, A Dynamic Model of the Limit Order Book, Working paper, University of Chicago.
- [79] Sandås, P, Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market, *Review of Financial Studies* 14, 705–734.
- [80] Seppi, D., 1997, Liquidity Provision with Limit Orders and a Strategic Specialist, *Review of Financial Studies* 10, 103–150.
- [81] Simaan, Y., D. Weaver, and D. Whitcomb, 2003, Market Maker Quotation Behaviour and Pre-Trade Transparency, *Journal of Finance* 58, 1247–1267.
- [82] Swan, P., and P. Westerholm, 2006, Market Architecture and Global Exchange Efficiency, working paper.
- [83] Tkatch, I., and E. Kandel, 2006, Demand for the Immediacy of Execution: Time *is* Money, working paper Georgia State University.
- [84] Viswanathan, S., and J. Wang, 2002, Market Architecture: limit-order books versus dealership markets, *Journal of Financial Markets* 5, 127–167.