Liquidity Provision with Limit Orders and a Strategic Specialist

Duane J. Seppi
Carnegie Mellon University

This article presents a microstructure model of liquidity provision in which a specialist with market power competes against a competitive limit order book. General solutions, comparative statics and examples are provided first with uninformative orders and then when order flows are informative. The model is also used to address two optimal market design issues. The first is the effect of "tick" size —for example, eighths versus decimal pricing — on market liquidity. Institutions trading large blocks have a larger optimal tick size than small retail investors, but both prefer a tick size strictly greater than zero. Second, a hybrid specialist/limit order market (like the NYSE) provides better liquidity to small retail and institutional trades, but a pure limit order market (like the Paris Bourse) may offer better liquidity on mid-size orders.

The provision of liquidity is the raison d'être for organized financial markets.[1] Investors value liquidity because it facilitates better risk sharing and encour-

I thank Praveen Kumar for many invaluable discussions and Franklin Allen (the editor), an anonymous referee, David Brown, Bob Dammon, Michael Fishman, Larry Glosten, Joel Hasbrouck, Craig Holden, Ronen Israel, Greg Kadlec, Bart Lipman, Ananth Madhavan, Susan Monaco, Uday Rajan, Patrik Sandas, Stephen Smith, Chester Spatt, Lara Wolfson and seminar participants at the Atlanta Finance Forum, Carnegie Mellon, the 1995 Indiana Symposium, University of Michigan, Washington University, Virginia Polytechnic Institute and the 1995 EFA, 1995 WFA, and 1996 AFA meetings for helpful comments. Address correspondence to Duane Seppi, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213.

[1] Amihud and Mendelson (1986) and more recently Brennan and Subrahmanyam (1994) show that the liquidity of a stock represents a significant part of its value.

The Review of Financial Studies Spring 1997 Vol. 10, No. 1, pp. 103–150

ages the collection of costly information and corporate control contests leading to improved resource allocation. The equilibrium liquidity in a market is the outcome of trading decisions taken by liquidity suppliers — investors who sell at prices above or buy at prices below their individual pretrade valuations of a security — and liquidity demanders — investors who trade at a premium or discount for the right to trade quickly. In a liquid market, the per-share impact of trades on prices is small. Trading affects prices because of adverse selection and/or market imperfections such as compensation for immediacy, taxes and order-processing costs, and any market power or inventory considerations of the marginal trader.

This article models the liquidity provision process in an institutional context motivated by features of the NYSE. In the model, a specialist trades strategically to maximize his profits from executing arriving market orders in the face of competition from limit orders from off-exchange liquidity suppliers. These value traders do not *need* to trade per se, but they do submit orders if their expected profit is positive.

Limit orders are price-contingent orders to sell (buy) if the price rises above (falls below) a prespecified limit price. As on the NYSE, price and public priority must be respected in that all public limit orders at prices at or below (above) the price at which the specialist proposes to sell (buy) must first be executed in full. The drawback of a limit order is that value traders only know the ex ante distribution of market orders when they submit their orders. In contrast, the specialist can wait to see the realized size of the arriving market order before deciding how to trade. Waiting is valuable if order submission is costly due to brokerage fees or to possible intervening revisions in the security's value (i.e., from public announcements or informative order flows).

The main results from this model are the following:

1. The value traders' and specialist's equilibrium strategies have a simple recursive structure. The general solution is analytic up to the evaluation of an inverse cumulative density (with uninformative order flows) or an inverse conditional expectation (with informative orders).

2. Analytic solutions are provided in the special cases of uninformative exponential (truncated or untruncated) and uniformly distributed market orders.

3. The relation between tick size and liquidity is non-monotone and discontinuous.

4. Retail and institutional investors have heterogeneous preferences about the optimal tick size with small investors favoring smaller ticks and large investors favoring larger ticks. Both, however, prefer discrete

over continuous pricing. There are also strong grounds to conjecture that mid-size investors prefer some discreteness in prices, too.

5. Small retail and large institutional investors prefer hybrid specialist/limit order markets (such as the NYSE), while some mid-size investors may prefer pure limit order markets (like the Paris Bourse).

This article is closely related to four lines of prior work. First, in a very important article, Rock (1996) solves for the equilibrium limit order book when off-exchange investors (with a risk-bearing advantage) compete with risk-averse competitive specialists (with an informational advantage). Byrne (1993) adapts the model to allow for price discreteness.[2] In this article, the tension between on- and off-exchange liquidity providers stems from the specialist's market power rather than from risk-related inventory effects. I initially abstract from issues of price discovery, but then later add in a simple price discovery problem due to informative market order flows, as in Rock (1996).

The second antecedent is Harris (1994a), which empirically estimates the effect of the NYSE moving from the current pricing on "eighths" to decimal pricing. It suggests that with decimal pricing, the bid/ask spread would shrink significantly, the depth at the inside bid/ask quotes would drop, and market order volume would increase, but it cannot say how depth further into the limit order book would change. Since limit orders beyond the inside bid/ask quotes are a source of liquidity for large blocks, this is an issue of some importance to institutional and other large traders. The theory in this article does address this issue.

A third line of work concerns the role of market power in market making, as in Glosten (1989) and Leach and Madhavan (1993). Specifically, how much market power does the advantaged position of specialists in the trading process confer? And how effectively do limit orders curb this market power?[3] Such questions — long of concern to academics and regulators — have come to the fore in the recent NASDAQ controversy. [See *Wall Street Journal* (1994).] In par-

[2] Other equilibrium models of limit orders include Glosten (1994), who studies competition between exchanges with limit orders but no specialist; Kumar and Seppi (1994), who investigate optimal order placement strategies when informed and liquidity traders can use limit orders too; Chakravarty and Holden (1995), who obtain analytic solutions to a limit order model with informed traders; Brown and Holden (1995), who consider forms of "smart" limit orders; and the dynamic models in Foucault (1993) and Parlour (1994). In addition, Angel (1992), Cohen, Maier, Schwartz, and Whitcomb (1981); and Harris (1994b) describe optimal limit order strategies in partial equilibrium settings. For empirical evidence on limit orders see Biais, Hillion, and Spatt (1995), Frino and McCorry (1995), Greene (1996), Harris and Hasbrouck (1996), Hollifield, Miller, and Sandas (1996), and Kavajecz (1996).

[3] The role of competition from an on-exchange trading crowd and price continuity rules are also considered. The issue of whether knowledge of the limit order book itself gives the specialist an unfair advantage in price discovery does not arise here since my limit order book is open and *limit* order flows are predictable and hence uninformative.

ticular, Christie and Schultz (1994) and others suggest that allegedly wide NASDAQ spreads may in part be due to inadequate limit order priority on the NASDAQ system.

The fourth work involves efforts to model the noninformational component of prices more explicitly. While Glosten and Milgrom (1985) and Kyle (1984, 1985) profoundly deepened our understanding of the role of adverse selection, theory on noninformational liquidity effects has only recently begun — most notably in Grossman and Miller (1988), Keim and Madhavan (1996), Leach and Madhavan (1993), and Spiegel and Subrahmanyam (1995) — to catch up with our empirical knowledge about the transitory component of prices [e.g., see Hasbrouck (1991, 1993)].

This article is organized as follows. Section 1 describes a market in which order flows are uninformative. Section 2 finds the unique equilibrium in this setting and presents some examples. Section 3 takes up optimal market design issues relating to tick size and the role of the specialist. Section 4 considers the interplay between price discovery and liquidity provision by extending the model to allow for market orders that are informative. Section 5 revisits the market design issue in this richer environment. Section 6 summarizes my findings and discusses possible extensions. All proofs are in Appendix A.

1. A Model without Price Discovery

Consider a liquidity provision game in a simple static market with four types of investors. An *active trader* demands liquidity inelastically by submitting a market order to buy/sell stock. Liquidity is supplied by (1) competitive risk neutral *value traders* who post limit orders from off of the exchange, (2) a single strategic *specialist* who clears the market subject to certain priority rules, and (3) a large *trading crowd* on the exchange, which (potentially) enters to provide unlimited liquidity if the profit from doing so exceeds a fixed hurdle rate.

The timing of events is depicted in Figure 1. Value traders first submit limit orders at time 1. Some time later an active trader arrives and submits a market order either to buy a random number of shares B or to sell a random S shares. In particular, α is the probability that he buys, and F is a continuous distribution over possible volumes (i.e., B or S). The specialist then maximizes his expected profit by choosing a *clean up* price p^* at which he clears the market after first executing any limit orders with priority (see below) and giving the crowd — which also sees B and p^* but has a higher reservation price — a chance to trade ahead of him.

Prices are restricted to a discrete grid $\mathcal{P} = \{\ldots, p_{-1}, p_1, p_2, \ldots\}$. I initially abstract from issues of price discovery by assuming that the

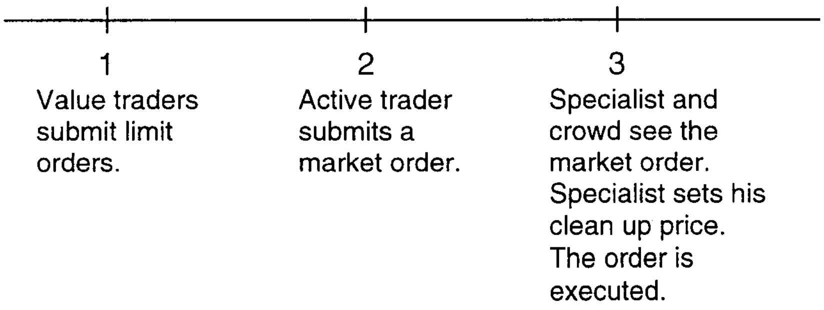
1
Value traders submit limit orders.
2
Active trader submits a market order.
3
Specialist and crowd see the market order. Specialist sets his clean up price. The order is executed.

Figure 1
Timing of events

potential liquidity providers — that is, the specialist, value traders, and the crowd — have a common valuation v, which is known to all investors, for the stock.[4] Prices are indexed on $\mathcal{P}$ by their ordinal position above or below v. If v itself is on the grid, then it is indexed as p_0. Later in Section 4, a price discovery complication is added.

The ability of limit orders to compete against the specialist hinges on certain priority rules. As on the NYSE, the specialist must respect the price, public, and time priority of other traders' orders. *Price priority* means that if the specialist wants to sell (buy) at a price p^*, then any limit sell (buy) orders at lower (higher) prices $p < p^* (p > p^*)$ must first be filled in full. *Public priority* means that any limit or trading crowd orders at p^* must also be filled first. *Time priority* stipulates that if the limit orders (and any from the trading crowd) exceed the available market order volume, then orders at p^* are executed sequentially in the order in which they arrived. When executed, limit orders trade at their posted limit price, which may be less (more) than p^*.

For time priority to make sense, the value traders are modeled as Bertrand competitors with stochastic arrival times. One of them randomly arrives first and — seeing an empty limit order book and knowing that other value traders will follow shortly — submits limit orders until the marginal expected profit on additional orders at each price is driven to zero.[5] Unlike on the NYSE, the limit order book here

[4] The active trader clearly has a different private valuation since he is willing to trade at a discount/premium relative to v to achieve immediacy.

[5] Alternatively, to avoid having a single value trader generate the entire limit order book, imagine liquidity constrained value traders arriving sequentially and each posting small limit orders until the expected profit from additional submissions is again driven to zero.

is open (i.e., publicly observable), so a value trader always knows the current depth at each price at the time she considers submitting an order.

In submitting limit orders, the value traders incur an up-front submission cost of c per share. This cost is important because it bounds depth in the limit order book, thereby giving the specialist "room" to exercise market power. Since brokerage fees, in practice, are only paid *after* execution, c is best interpreted as a reduced form for the idea that limit orders can be "picked off" based on information from public announcements, or, alternatively, as capturing any incremental opportunity or shoeleather costs investors bear when trading from off the exchange.[6] This cost and the fact that the probability of a market buy is $\alpha < 1$ together imply that limit orders are ex ante profitable only at prices above $v + c/\alpha$ (for limit sells) and below $v - c/(1-\alpha)$ (for buys).

The trading crowd has a slightly different constraint on their willingness to trade. They provide unlimited liquidity whenever the stock's price at time 3 differs from its value by r or more, but otherwise do not participate — that is, they buy if $p^* \leq v - r$ and sell if $p^* \geq v + r$ — thereby effectively bounding (when $r < \infty$) the scope of the specialist's market power. The premium r represents both compensation for any trading costs the crowd may incur plus its required net profit. The idea is that the crowd only recognizes or intervenes on gross mispricing (e.g., spreads of fifty cents) because it is otherwise busy carrying out a more profitable broker function.[7]

Since the buy and sell sides are separable and symmetric in this model, I focus expositionally on limit sells/market buys. I define $p_{max} = p_{j_s^{max}}$ as the lowest price above $v + r$ (i.e., the trading crowd's reservation price), where $j_s^{max} = \min(j \mid v + r \leq p_j)$. In addition, if $c/\alpha < r$ (the natural case), then $p_{min} = p_{j_s^{min}}$ is the lowest price above $v + c/\alpha$ (i.e., the starting point for limit sell orders), where $j_s^{min} = \min(j \mid v + c/\alpha < p_j)$. In addition, let $S_1, S_2, \ldots$ denote the number of shares posted as limit sells at prices $p_1, p_2, \ldots$, respectively,

[6] Neither the specialist nor the crowd gives away free trading options since they can update their pricing faster than off-exchange traders to reflect any news that may arrive [as in Brown and Holden (1995) and Kumar and Seppi (1994)] after limit orders are submitted but *before* market orders arrive. Allowing the ex ante costs $c_1, c_2, \ldots$ to depend on the price $p_1, p_2, \ldots$ where a limit order is posted is straightforward. While the idea that off-exchange investors are at a cost disadvantage seems reasonable, little changes qualitatively if the specialist *does* pay c except that intraspread trading is then confined to prices between $v + c$ and $v + c/\alpha$ (see below).

[7] The crowd can be reinterpreted as an "upstairs" market for large blocks where r is the premium/discount demanded by potential counterparties, as in Keim and Madhavan (1996). However, for the upstairs market to attract positive volume, the specialist's maximum position must be bounded since otherwise he will *always* undercut the upstairs price if $p_1 < v + r$.

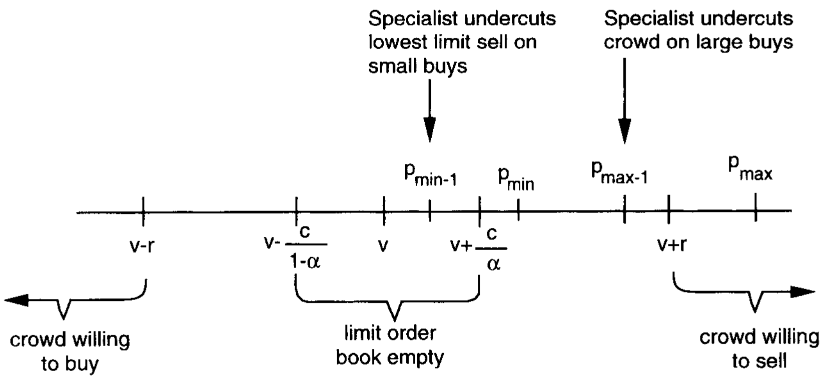

Figure 2
Restrictions on the specialist's clean up price imposed by the limit book and the crowd

and let $Q_j = \sum_{i=1}^{j} S_i$ be the cumulative depths at or below p_j. All order quantities B, S_j, and Q_j are unsigned (non-negative) volumes.

The specialist's two advantages — that is, not paying the incremental cost c, and trading after he sees the realized volume B rather than based on ex ante expectations — make him the low-cost liquidity provider in this model. However, he provides liquidity at favorable prices only if he has an incentive to do so, which is precisely what the price and public priority rules accomplish. They force the specialist to undercut competing liquidity providers.

Competition from the limit book and the threat of entry by the crowd constrain the specialist's market power in a number of ways. One immediate way is that he never sells at a p^* less than p_1 (if the book starts at $p_{min} = p_1$) or $p_{min-1} = p_{j_s^{min}-1}$ (if it starts at $p_{min} > p_1$). The p_{min-1} case follows from the fact that the specialist only needs to undercut the book by one tick to get small trades. Similarly, he cannot ask more than $p_{max-1} = p_{j_s^{max}-1}$ and still undercut the trading crowd on large trades.

This presupposes, of course, that the grid $\mathcal{P}$ includes (as in Figure 2) prices in the appropriate intervals. Otherwise, the specialist may be unable to trade given price and public priority. In particular, if he cannot undercut low limit sells (i.e., if $\mathcal{P} \cap (v, v + c/\alpha]$ is empty), then he only provides liquidity as a "counterparty of last resort" on larger trades $B > S_1$. If he cannot undercut the trading crowd [i.e., if $\mathcal{P} \cap (v, v + r)$ is empty], then he cannot trade at all.

Consider the specialist's formal order execution problem. Given a market order to buy B shares, he maximizes his profit by clearing the market at the price p^* that solves[8]

$$\max_{v<p<p_{max}} \pi(p) = \left[B - \min\left(B, \sum_{i \text{ s.t. } p_i \le p} S_i\right)\right](p - v). \qquad (1)$$

Restricting the choice set to prices between v and the trading crowd's reservation price p_{max} simply recognizes that (1) the specialist requires a positive profit to sell, and (2) failure to deter entry by the crowd drives his volume (and hence profits) to zero. The "min" term reflects the fact that the specialist never *buys* from the book or crowd at $p^* > v$. A perhaps less obvious point is that under his optimal strategy, the posted limit orders S_j at each price p_j (except p_1) are executed either in full or not at all.[9] Another useful result is a monotonicity property that p^* is weakly increasing in volume.

Proposition 1. *If $B > B'$, then $p^*(B) \ge p^*(B')$.*

The key to the specialist's problem is that, in raising p^*, the benefit of a greater profit per share $p^* - v$ must be balanced against the cost of a reduced number of shares $B - \sum_{i \text{ s.t. } p_i \le p^*} S_i$ left over for him after executing competing orders with priority.[10] For example, if the market buy is big enough to fill limit orders at prices p_1 through p_j (i.e., $B > S_1 + \cdots + S_j$), the specialist's profit from selling at $p^* = p_{j-1}$ is

$$\pi_{j-1} = \left(B - \sum_{i=1}^{j-1} S_i\right)(p_{j-1} - v), \qquad (2)$$

and from selling at p_j is

$$\pi_j = \left(B - \sum_{i=1}^{j} S_i\right)(p_j - v). \qquad (3)$$

Comparing Equation (3) with Equation (2) shows that he earns more

[8] Price continuity rules have virtually no effect on this problem since the specialist, after doing the bulk of his trades at p^*, can always use the last few shares he sells to "walk" back down the price schedule to recenter prices at v again.

[9] A partial fill at $p^* > p_1$ would mean that the specialist went too far into the book and exhausted the supply of market orders without trading himself. However, a partial fill of limit orders at p_1 is possible since it is not profitable to undercut the lowest price above v. Thus, how many of the S_1 limit orders are executed is beyond the specialist's control in that it depends entirely on the volume B.

[10] The model differs from the "dominant firm" model in that the competitive fringe's supply schedule is derived given that limit orders trade at their posted inframarginal prices $p_j \le p^*$ rather than all at the marginal price p^*.

trading at p_j than at p_{j-1} if and only if the market order is sufficiently large:

$$B > Q_{j-1} + S_j \frac{p_j - v}{p_j - p_{j-1}} \equiv S_{j-1,j}. \tag{4}$$

As is shown below, *in equilibrium*, comparisons with the adjacent prices alone suffice to determine the specialist's optimal trading rule. In principle, however, deciding whether p_j solves Equation (1) requires checking not only that trading at the *adjacent* prices p_{j-1} or p_{j+1} is less profitable, but also that trading at *any* other price is less profitable. For example, the specialist's profit at $p_k < p_{j-1}$,

$$\pi_k = \left(B - \sum_{i=1}^{k} S_i\right)(p_k - v), \tag{5}$$

is less than Equation (3) if and only if

$$B > Q_k + \left(\sum_{i=k+1}^{j} S_i\right)\left(\frac{p_j - v}{p_j - p_k}\right) \equiv S_{k,j}. \tag{6}$$

Monotonicity of p^* from Proposition 1 lets us represent the specialist's strategy in terms of minimum volume *thresholds*

$$\hat{S}_j = \inf\left(B \geq Q_j \mid p^* \geq p_j\right) \tag{7}$$

governing whether limit sells at $p_1, p_2, \ldots$ are executed in full. At prices p_j, where $p_1 < p_j < p_{max}$, each threshold $\hat{S}_j$ is a crossing point $\hat{S}_{k,i}$ for some pair of prices p_k and p_i, where $p_k < p_j \leq p_i$. More structure on the $S_1, S_2, \ldots$ is still needed, however, to say *which* two prices p_k and p_i. Since limit orders are either executed in toto or not at all at $p_j > p_1$, these execution thresholds are, in turn, equal to pricing thresholds

$$S_j^* = \inf(B \mid p^* \geq p_j), \tag{8}$$

so that $\hat{S}_j = S_j^*$ for $j > 1$. The "inf" notation used in Equations (7) and (8) simply reflects a convention that the specialist sets $p^*(B) \geq p_j > p_1$ only when $B > \hat{S}_j$.

Pricing and limit order execution at p_1 is somewhat different. Neither the specialist nor the crowd can profitably undercut p_1, so full execution of S_1 depends simply on whether $B \geq S_1$. Thus, $\hat{S}_1 = S_1$ and $S_1^* = 0$. In addition, in contrast to $p_j > p_1$, we have $p^*(\hat{S}_1) = p_1$ when $S_1 > 0$.

Now consider the value traders' problem. Since the book is open, each value trader knows the current depth when she arrives and can readily calculate the expected profit from additional orders. In equilibrium, their total order submissions $S_1, S_2, \ldots$ are guided by the marginal expected profit from an additional limit order at each p_j:

$$e_{sj} = \alpha F(B > \hat{S}_j)(p_j - v) - c, \tag{9}$$

where $\alpha F(B > \hat{S}_j)$ is the probability of a market order B arriving that is large enough to trigger execution of all limit sells at p_j.[11] Since value traders are risk-neutral, competitive, and do not *need* to trade per se, they simply submit limit orders opportunistically until any positive expected marginal profits from supplying liquidity are eliminated.

Putting the behavioral assumptions about the specialist, values traders, and trading crowd together gives the following definition.

Definition 1. *An equilibrium is a set of total value-trader limit orders $S_1, S_2, \ldots$ such that the marginal expected profit $e_{sj} = 0$ (if $S_j > 0$) or $e_{sj} \leq 0$ (if $S_j = 0$) at each price p_j where the specialist's execution thresholds $\hat{S}_1, \hat{S}_2, \hat{S}_3, \ldots$ satisfy Equation (7).*

Before continuing, a brief comparison of this model with the competitive limit order models of Rock (1996) and Byrne (1993) may be helpful. The first difference is that a noninformational spread arises here from the strategic execution of orders by a monopolistic specialist in competition with a competitive fringe rather than from risk-related inventory effects. The second is the absence of adverse selection. Limit orders are costly so far simply because of the submission fee c.

2. Results

This model is tractable because, *in equilibrium*, the specialist's execution strategy $\hat{S}_j$ takes a particularly simple form. To appreciate this, consider first his strategy against an *arbitrary* limit order book. Plotting the specialist's profits π_1, π_2, and π_3 from selling at p_1, p_2, and p_3, respectively, against volume B, one sees that if the depth S_3 is high (as in Figure 3a), then the specialist sells at p_1 if $B \leq S_{1,2}$, at price p_2 if $S_{1,2} < B \leq S_{2,3}$, and at p_3 only if $S_{2,3} < B$. In contrast, if S_3 is low (as in Figure 3b), then he clears the market at p_1 until $B > S_{1,3}$ at which point his profit-maximizing price jumps over p_2 directly to p_3. Thus, a priori, the relative ranking of the crossing points $S_{1,2}$, $S_{2,3}$, and $S_{1,3}$ is ambiguous. With only three prices, this is manageable, but having

[11] Since limit sells at p_1 are executed in full when $B \geq S_1 = \hat{S}_1$ (i.e., rather than only when $B > \hat{S}_1$), their probability of execution is actually $F(B \geq \hat{S}_1)$ rather than $F(B > \hat{S}_1)$. However, with a continuous distribution F, these two probabilities are the same.

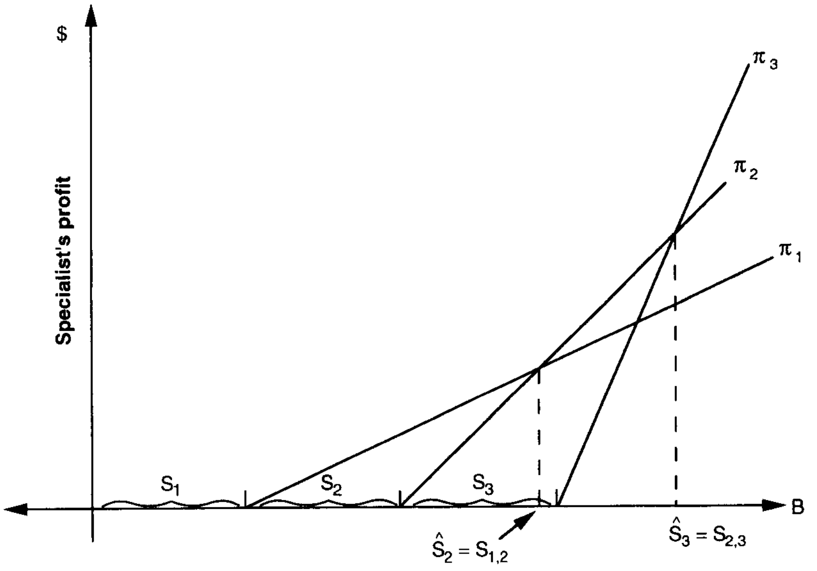

Figure 3a
Specialist's profit with big depth S_3
The lines π_1, π_2, and π_3 give the specialist's profit from selling at clean up prices p_1, p_2, and p_3, respectively, given possible buy volumes B in excess of the cumulative limit sells Q_1, Q_2, and Q_3 at each of these prices. His optimal strategy is the upper envelope of these lines. The execution thresholds $\hat{S}_1$, $\hat{S}_2$, and $\hat{S}_3$ are just the depth S_1 and the crossing points $S_{1,2}$ (for lines π_1 and π_2) and $S_{2,3}$ (for lines π_2 and π_3).

to check all pairwise comparisons with decimal pricing to determine the $\hat{S}_j$ would be a computational mess.

A related complication (again with arbitrary books) is that greater depth at higher prices may or may not affect the execution threshold for limit orders at lower prices. In Figure 3b, a small increase in S_3 moves the profit line π_3 slightly to the right, thereby raising the crossing point $S_{1,3}$ and thus thresholds $\hat{S}_3$ and $\hat{S}_2$; whereas in Figure 3a, increasing S_3 has no effect on $\hat{S}_2$ (since it is determined by $S_{1,2}$ rather than $S_{1,3}$).

This complexity is worrisome since solving for the equilibrium involves finding a fixed point. The value traders' total limit order submissions $S_1, S_2, \ldots$ depend through $e_{s1}, e_{s2}, \ldots$ on the specialist's execution thresholds $\hat{S}_1, \hat{S}_2, \ldots$, which depend, in turn, on the submitted limit orders $S_1, S_2, \ldots$. Fortunately, the next proposition cuts this Gordian knot.

Proposition 2. *In equilibrium, if $S_j > 0$ limit orders are posted at a price p_j, then $S_{j+1} > 0$ limit orders are also posted at p_{j+1}, provided $p_{j+1} < p_{max}$. In addition, given the equilibrium depths $S_1, S_2, \ldots$ in*

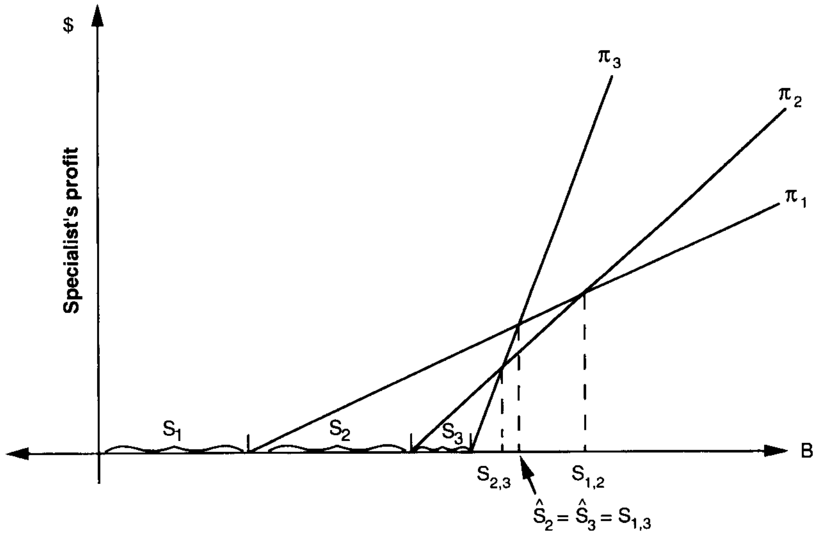

Figure 3b
Specialist's profit with small depth S_3
This figure is identical to Figure 3a except that depth S_3 is much less than in Figure 3a. Consequently, the execution thresholds are $\hat{S}_1 = S_1$ and $\hat{S}_2 = \hat{S}_3 = S_{1,3} < S_{1,2}$.

the limit order book, the specialist's execution thresholds are just the adjacent price crossing points

$$\hat{S}_j = S_{j-1,j} = Q_{j-1} + S_j \frac{p_j - v}{p_j - p_{j-1}}. \tag{10}$$

The intuition is the following. Take, for example, $j = 2$ and suppose $S_2 > 0$. If S_3 is too small, as in Figure 3b, then value traders can increase their expected profit by shifting $\epsilon > 0$ limit orders from p_2 up to p_3. This is because (1) the realized profit *conditional* on execution at p_3 is greater than at p_2, and (2) the execution probability is the same as long as $S_{1,2} \geq S_{1,3}$. In particular, since $\hat{S}_2 = \hat{S}_3 = S_{1,3}$, the common threshold from Equation (6) depends only on the *sum* $S_2 + S_3$ and not on the levels of S_2 and S_3 separately. Furthermore, $S_{1,3}$ stays to the left of $S_{1,2}$ (as in Figure 3b) with $S_2 - \epsilon$ at p_2 and $S_3 + \epsilon$ at p_3, as long as $\epsilon < S_2 - (S_2 + S_3)\frac{(p_3 - v)(p_2 - p_1)}{(p_3 - p_1)(p_2 - v)}$. Thus, a depth S_3 that is too small is inconsistent with equilibrium since it implies $e_{s3} > e_{s2} = 0$.[12] To zero out the marginal expected profit at p_3, we need sufficient depth S_3 so that [from Equations (9) and (10)] limit orders at p_2 are sometimes

[12] The equality $e_{s2} = 0$ follows because, in equilibrium, $S_j > 0$ implies $e_{sj} = 0$.

executed when limits at p_3 are not or $F(B > \hat{S}_2) > F(B > \hat{S}_3)$. Thus, in equilibrium, the book looks like Figure 3a rather than (the more complicated) Figure 3b.

This result is important because it lets us construct our equilibrium recursively. Starting at price p_1 and working our way up, the depth S_j at each price p_j is either zero [if $e_{sj}(0) \leq 0$] or can be calculated by (1) inverting the zero profit condition $e_{sj}(\hat{S}_j) = 0$ to obtain the break-even threshold $\hat{S}_j$ and then (2) inverting $\hat{S}_j = S_{j-1,j}$ from Equation (10) — together with the previously determined depths $S_1, \ldots, S_{j-1}$ — for the depth S_j that induces this threshold. In particular, Proposition 2 assures us that depths S_k at higher prices $p_k > p_j$ do not affect $\hat{S}_j$ and S_j in equilibrium.

At this point, some notation is needed. Let $H(F)$ be the inverse of the upper-tail cumulative probability $F(B > z)$ and define

$$H_j = \begin{cases} H\left(\frac{c/\alpha}{p_j - v}\right) & \text{if } \frac{c/\alpha}{p_j - v} \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

In words, H_j is the number that sets the marginal probability of execution $F(B > H_j) = \frac{c/\alpha}{p_j - v}$ (i.e., when this is a proper probability), which is what $e_{sj}(\hat{S}_j) = 0$ requires $\hat{S}_j$ to do.

Proposition 3. *The equilibrium limit order book has a recursive structure such that*

$$\begin{aligned} S_1 &= H_1, \\ S_j &= \gamma_j (H_j - Q_{j-1}), \end{aligned} \tag{12}$$

$$\begin{aligned} Q_1 &= H_1, \\ Q_{j-1} &= \gamma_{j-1} H_{j-1} + (1 - \gamma_{j-1}) Q_{j-2}, \end{aligned} \tag{13}$$

and

$$\gamma_j = \frac{p_j - p_{j-1}}{p_j - v} \tag{14}$$

at prices $p_j < p_{max}$ and $S_j = 0$ thereafter, provided that $p_1 < p_{max}$. The specialist's execution strategy is simply

$$\hat{S}_j = \begin{cases} H_j & \text{if } p_j < p_{max} \\ \infty & \text{if } p_j = p_{max}. \end{cases} \tag{15}$$

If instead $p_1 = p_{max}$, then the specialist never trades, and $S_1 = H_1$ and $S_j = 0$ for all $j > 1$.

There are several things to notice. First, the equilibrium is unique since it is constructed recursively. That is, S_1 is unique and, given S_1, then S_2 is unique, and so on. Second, $\gamma_j \leq 1$ is shorthand for the reciprocal of the $\frac{p_j - v}{p_j - p_{j-1}}$ term multiplying S_j in the expression for $\hat{S}_j$ in Equation (10). Intuitively, a small γ_j reflects reduced limit order depth due to more aggressive specialist undercutting of the book [i.e., $\hat{S}_j$ exceeds the cumulative depth Q_j by $S_j(\frac{1}{\gamma_j} - 1) > 0$]. Third, when $S_1 > 0$, the ***inframarginal*** limit sells at p_1 actually earn positive expected profits. This is because their probability of execution is higher than that of the (break-even) *marginal* sell at p_1, due to the possibility of partial execution based on time priority when $B < S_1$. In contrast, inframarginal orders at $p_j > p_1$ simply break even in expectation since they are executed either in toto along with the marginal order at p_j (i.e., when $B > \hat{S}_j$) or not at all. Fourth, the posted bid/ask spread reflects orders in the book (i.e., $p_{ask} = p_{min}$) since my specialist does not post quotes up-front. However, if the grid includes prices between v and $v + c/\alpha$, then he engages in intraspread trading at p_{min-1} to undercut the book whenever $B < H(\frac{c/\alpha}{p_{min} - v})$ — thereby tightening the effective spread for small trades relative to the posted spread. Intraspread trading is an attractive feature of the model, given its prevalence on the NYSE.[13] Fifth, a competitive market (i.e., $p^* = p_1$ for all B) is a special case when $c = 0$ and/or $r \leq p_1 - v$. Sixth, the cumulative depth Q_j is a weighted average of the intervening $H_j, H_{j-1}, \ldots$ with weights $\gamma_j, (1 - \gamma_j)\gamma_{j-1}, \ldots$.[14] This in turn leads to Proposition 4.

Proposition 4. *If a distribution F' first-order stochastic dominates distribution F and/or the adjusted cost $c'/\alpha' < c/\alpha$, then the market with F' and c'/α' is more liquid in that cumulative depth is greater ($Q_j' \geq Q_j$ for all j), execution thresholds are higher ($\hat{S}_j' \geq \hat{S}_j$ for all j), and the book is deeper at p_{min}' ($S_{min'}' \geq S_{min'}$).*[15]

[13] Shapiro (1993) reports that 66% of all trades occur between the bid and the ask when the spread is more than $\frac{1}{8}$. In Chakravarty and Holden (1995), intraspread trades arise from "liquidity discovery" by off-exchange investors. The two models can be distinguished empirically since intraspread trades there involve public orders meeting public orders, while here the specialist is on one of the sides.

[14] If the tick size $p_j - p_{j-1}$ is everywhere a constant Δ, and $p_1 = v + q\Delta$ (i.e., possibly a fraction q of a tick above v), then $\gamma_j = \frac{1}{j-1+q}$, and the cumulative depths at $p_j > p_1$ simplify to $Q_j = \frac{qH_1 + \sum_{i=2}^{j} H_i}{j-1+q}$.

[15] The effect, for example, of c/α on S_j at prices $p_j > p_{min}'$ is hard to sign since $\frac{\partial S_j}{\partial c/\alpha} = \gamma_j \left(\frac{\partial H_j}{\partial c/\alpha} - \frac{\partial Q_{j-1}}{\partial c/\alpha} \right)$ can go either way, depending on whether $\frac{\partial H_j}{\partial c/\alpha}$ or $\frac{\partial Q_{j-1}}{\partial c/\alpha}$ is larger.

Proposition 5. *Fix a value $\varphi > v$. The cumulative depth $Q(\varphi, \mathcal{P}) = \sum_{i \text{ s.t. } p_i \leq \varphi} S_i$ at prices below φ is bounded by $H(\frac{c/\alpha}{\varphi - v})$ for any price grid $\mathcal{P}$.*

The first result says markets are more liquid when submission costs are low and/or large market orders are frequent. Thus, while large trades use up liquidity when they arrive, a high *probability* of large trades actually attracts liquidity ex ante. The second result on the boundedness of the book is important for market design in Section 3.

The solution in Proposition 3 is "almost" analytic in that the functional form of the inverse distribution function H is still needed. In certain parametric cases, this is available. However, even without closed forms for H, equilibria are readily calculated numerically.

Examples. Closed-form solutions for H exist in three special cases, where the market buy B is (1) exponentially distributed with parameter θ such that

$$F(B > z) = e^{-z/\theta}, \tag{16}$$

in which case the inverse distribution function is

$$H(F) = -\theta \ln(F); \tag{17}$$

(2) truncated exponential with an upper bound K such that

$$F(B > z) = \frac{e^{-z/\theta} - e^{-K/\theta}}{1 - e^{-K/\theta}}, \tag{18}$$

and hence

$$H(F) = -\theta \ln[(1 - e^{-K/\theta})F + e^{-K/\theta}]; \tag{19}$$

or (3) uniformly distributed on $[0, K]$ so that $F(B > z) = 1 - \frac{z}{K}$ and

$$H(F) = (1 - F)K. \tag{20}$$

Substituting these expressions in Proposition 3 gives the desired strategies. For example, suppose volume is untruncated exponential with a mean (and standard deviation) of θ shares.[16] If the tick size $p_j - p_{j-1}$ is a constant Δ with $p_1 = v + q\Delta > v + c/\alpha$ (where $0 < q \leq 1$ allows for p_1's less than a full tick above v), then, in equilibrium, the

[16] The exponential density f starts at a peak of $1/\theta$ when $B = 0$ and declines monotonically — not an unreasonable pattern for volume.

limit order book is

$$S_j = \begin{cases} \theta \ln\left(\frac{q\Delta}{c/\alpha}\right) & \text{if } j = 1 \\ \theta \frac{\ln(j-1+q) - W_{j-1}}{j-1+q} & \text{if } j_s^{max} > j > 1, \end{cases} \tag{21}$$

where

$$W_{j-1} = \begin{cases} \ln(q) & \text{if } j-1 = 1 \\ \left(\frac{1}{j-2+q}\right) \ln(j-2+q) + \left(\frac{j-3+q}{j-2+q}\right) W_{j-2} & \text{if } j-1 > 1. \end{cases} \tag{22}$$

An interesting feature of this solution is that the book's depth at prices above p_1 just depends on the ratio q and the ordinal ranking of prices j (i.e., second, third, etc. above v), but surprisingly *not* on the tick size Δ (e.g., whether p_j is \$0.05 or \$0.50 above p_{j-1}, as long as $p_j < v + r$), the probability α, or even the submission cost c (i.e., beyond the stipulation $c/\alpha < q\Delta$).

Figure 4 shows the limit book (in the first column) and the execution thresholds (in the second) for the following parameters: stock value $v = \$20$, probability of a buy $\alpha = 0.5$, expected volume $\theta = 25$ round lots, a constant tick $\Delta = p_j - p_{j-1} = \$0.125$ with the grid centered at $p_0 = v$ (i.e., $q = 1$), and a crowd hurdle rate $r = \$0.50$. The submission cost $c = \$0.019$ per share was calibrated to match the 1994 NYSE specialist participation rate (i.e., specialist volume to total volume) of 8.6%. (See the *1994 NYSE Fact Book*.) Depth in the book is concentrated at prices $\frac{1}{8}$ (30 round lots) and $\frac{2}{8}$ (another 8 round lots) above v, which seems reasonable.[17] The specialist in this example is willing to trade up to 17 ($= \hat{S}_2 - S_1$) round lots at \$20$\frac{1}{8}$ but ups his price to \$20$\frac{1}{4}$ when $B > 47$ and to \$20$\frac{3}{8}$ when $B > 57$, where (given $p_{max} = \$20.5$) he provides infinite depth.

A general property of the clean up price (illustrated by this example) is that $p^*(B)$ — due to its monotonicity and boundedness — is concave in B close to p_{max-1}. This is consistent with empirical evidence in Keim and Madhavan (1996, Table 4) showing that the temporary price impact $p^* - v$ on block trades is concave in volume. However, a locally *convex* $p^*(B)$ is possible on smaller trades.

Empirical issues. The model is suitable for structural estimation on tractability grounds and since it distinguishes — as do NYSE (TORQ) or Paris Bourse audit data — between arriving orders and the ensu-

[17] The book depicted here is monotone (i.e., depth decreases at higher prices), but it is easy to construct non-monotone books. From Equation (21), simply increase c/α until $S_1 < S_2$.

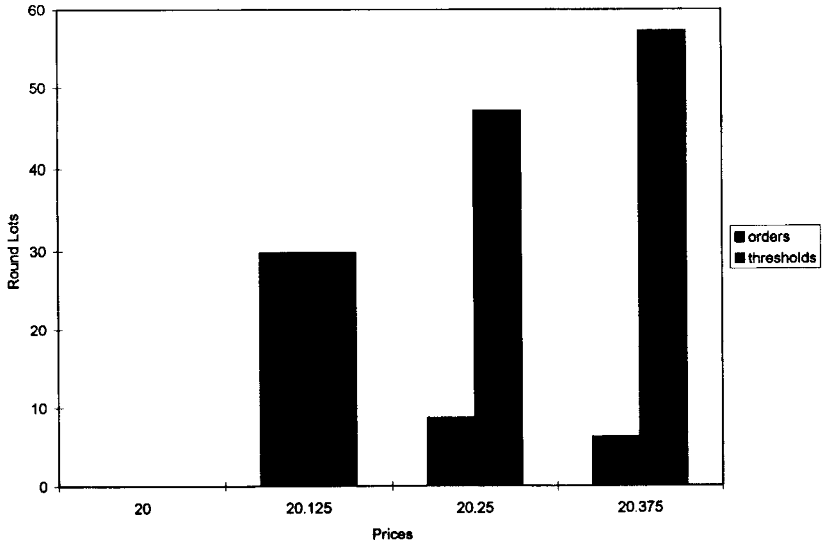

Figure 4
A numerical example of limit orders and thresholds with exponential market orders
This figure gives the equilibrium depths $S_1, \ldots, S_3$ and thresholds $\hat{S}_1, \ldots, \hat{S}_3$ (in round lots) for a market where the stock value $v = \$20$, the probability of a buy $\alpha = 0.5$, the expected volume $\theta = 25$ round lots, the price grid is centered at $p_0 = v$ with a constant tick size $\Delta = p_j - p_{j-1} = \$0.125$, and the crowd hurdle rate is $r = \$0.50$. The submission cost $c = \$0.019$ per share was calibrated to match the 1994 NYSE specialist participation rate of 8.6%.

ing sequence of *transactions* with the book and specialist. Beyond this, the model (even in this simple symmetric information form) is capable of explaining a number of cross-sectional and intertemporal patterns in specialist participation rates and profits. For example, the finding in Madhavan and Sofianos (1994, Table 5) that wider spreads are correlated with higher specialist participation rates is consistent with heterogeneity in c. In particular, if c proxies for the ex ante cost of being adversely "picked off" (see footnote 6), then stocks where $p_{min} > p_1$ (i.e., where c is big due to infrequent order arrivals or volatile announcements) will have more specialist undercutting (and hence higher participation rates) than stocks where $p_{min} = p_1$ (i.e., where c is low). In addition, the model may be able to replicate significant cross-sectional relationships between stocks' empirical volume distributions F and specialist participation [see Madhavan and Sofianos (1994, Table 3)]. In particular, more mid-size volume implies deeper books (reducing the specialist's share of volume), while more blocks $B > \hat{S}_{max-1}$ imply more trades in which the specialist's share $B - Q_{max-1}$ is large compared with the bounded liquidity Q_{max-1} provided by the book. Lastly, evidence in Sofianos (1995, Table 7) of a U-shaped pattern in specialist total revenue conditioned on trade size may reflect frequent specialist undercutting on small trades and a profitable "counterparty of last resort" role on blocks.

Robustness. The robustness of a static analysis is naturally a concern. For example, in a multi-period market, what prevents liquidity demanders from repeatedly hitting cheap limit sells at p_{min} and then — rather than paying up to hit orders further in the book — simply waiting for the book to replenish itself, and then buying more at p_{min}? Clearly the answer must be a strong demand for immediacy.[18] A maintained assumption is that only the specialist and the (otherwise occupied) crowd are *continuously* present in the market. In other words, waiting for new value traders to arrive and replenish the book must take longer than liquidity demanders are willing to wait. Such an assumption is not unreasonable for less active (e.g., non-DJIA) stocks, where the rate of both market and thus limit order arrival is slower.

A second robustness concern is the exogeneity of the volume distribution F. This concern is compounded by the fact that the average price paid on a market buy jumps *discontinuously* at each $B = H_j > 0$ for $j > 1$. In particular, when $B \leq H_j$, the specialist sells up to $B - Q_{j-1}$ at p_{j-1}, whereas once $B > H_j$, then any liquidity he provided at p_{j-1} "relocates" to p_j (i.e., leaving only the S_{j-1} in the book at p_{j-1}). It is hard to imagine that a trader — if he has any flexibility — would not prefer a slightly smaller buy $B = H_j$ with a clean up price p_{j-1} over market orders in an interval $(H_j, H_j + \epsilon)$ for at least some small ϵ. Such preferences can lead to discontinuities or mass points in F.[19] Although explicitly solving for a fixed point in F given a formal optimization problem for market orders is beyond the scope of this article, Appendix B does explain how to allow for *exogenous* mass points.

Having thus far taken the market structure as given, I now analyze the impact of various market design issues on liquidity.

3. Optimal Market Design

Competition for order flow among the NYSE, crossing networks and regional exchanges — along with the organization/deregulation of exchanges in eastern Europe and in emerging markets — has led to a growing interest in the way financial markets are structured. This

[18] The active trader may also try to post a *limit buy* order at p_1 as a "take it or leave it offer" to coerce better execution from the specialist. However, a strong demand for immediacy undermines the credibility of such a strategy.

[19] I thank Larry Glosten for pointing out this issue. Of course, *between* H_{j-1} and H_j the average price paid is continuous in B.

section considers two specific design issues. The first is the choice of tick size, and the second is the role of the specialist.

The main message is that the equilibrium liquidity of a market depends on the *interaction* between the specialist's strategy, the limit order book, and latent competition from the crowd. In some situations, the specialist is simply a "trader of last resort" on larger trades (i.e., when $p_1 = p_{min}$). In others, he is also the primary source of liquidity for very small orders (i.e., when $p_1 < p_{min}$). Moreover, while the specialist is the lowest cost provider of *marginal* liquidity [i.e., if $\mathcal{P} \cap (v, v + r)$ is non-empty], the demand for marginal liquidity at p^* is reduced by *inframarginal* liquidity supplied by the book (i.e., limit sells below p^*). Thus, investors' market design preferences — that is, how they evaluate trade-offs between deeper books $S_1, S_2, \ldots$ and lower clean up prices p^* — follow directly from their differing demands for inframarginal vs. marginal liquidity, as induced by their predispositions to use different sizes of market orders.[20]

My main focus is the welfare impact of different market designs on active investors. After all, they — unlike the value traders, specialist, and crowd — actively need to trade. The liquidity/illiquidity of a market, once the valuation v is fixed, refers to the temporary (or noninformational) price impact of different sizes of market orders. In particular, if j^* is the index for the clean up price (i.e., $p_{j^*} = p^*$), then I define the *average ask premium* above v on a market buy B, given a grid $\mathcal{P}$, as

$$A(B, \mathcal{P}) = \frac{S_1 p_1 + \cdots + (B - Q_{j^*-1})p^*}{B} - v. \tag{23}$$

3.1 Optimal tick size

Tick size refers to the difference between adjacent prices on a grid $\mathcal{P}$. The relation between tick size and liquidity is non-monotone, discontinuous, and differs across market order size. This is because the choice of tick size affects the location of prices relative to the entry points $v + c/\alpha$ (for value traders) and $v + r$ (for the trading crowd).

To see this, fix v and consider two possible tick sizes $\Delta' < \Delta$ with associated grids $v < p'_1 < v + c/\alpha < p_1 < p'_2 < p_2 < p'_3$. The picture of who wins and loses on the two grids is quite complicated. Small investors submitting orders $B \leq H(\frac{c/\alpha}{p'_2 - v}) \equiv \hat{S}'_2$ prefer buying at p'_1 rather than p_1, but investors submitting somewhat larger orders $\hat{S}'_2 < B \leq H(\frac{c/\alpha}{p_2 - v}) \equiv \hat{S}_2$ prefer trading at p_1 rather than p'_2.

[20] An important caveat here is the assumption of a monopolistic specialist. Bertrand competition among multiple risk-neutral dealers leads to infinite liquidity at p_1 (and p_{-1}). This article takes no stand on *why* the specialist has market power. It can arise innocently if market making represents a natural monopoly — that is, if only one trader is willing to stand ready continuously to commit his capital to trade at a profit less than the crowd's r — or less innocently from collusion.

The reason for this switch is that any liquidity from the specialist at p_1' "disappears" when p_1' becomes inframarginal. The picture is even more complicated for orders larger than $\hat{S}_2$ since they are partially executed at multiple prices.

The liquidity-maximizing tick size can be considered both from an ex post (i.e., given a particular v) as well as from an ex ante perspective. In the ex ante case, while market participants know the realized v when they submit their orders and trade, the exchange must choose its grid $\mathcal{P}$ *beforehand*, knowing only the probability distribution over possible values v might take in the future.

Consider first very large market orders when v is known.[21] If $B > \hat{S}_{max-1}$, then the specialist sells at $p^* = p_{max-1}$ — just low enough to deter entry by the trading crowd — after first filling any limit orders at lower prices. However, from Proposition 5, the relative contribution of depth in the (bounded) book below $v + r$ to the average ask premium $A(B, \Delta)$ shrinks as B gets *very* large. Thus, the key to creating liquidity for very large blocks is to choose the tick size Δ that minimizes the maximum entry-deterring price p_{max-1}. For example, given a price grid centered at $p_0 = v$ with a constant tick Δ,

$$p_{max-1} = \begin{cases} v + \Delta & \text{if } \frac{r}{2} \leq \Delta < r \\ v + j\Delta & \text{if } \frac{r}{j+1} \leq \Delta < \frac{r}{j}. \end{cases} \tag{24}$$

As can be seen in Figure 5, this relation is non-monotone and discontinuous. The discontinuities arise when shrinking the tick size Δ by a small δ causes p_{max-1} to jump up from the jth price above v on the $\mathcal{P}_\Delta$ grid [i.e., where $v + j\Delta < v + r \leq v + (j+1)\Delta$] to the $j+1$th price on the $\mathcal{P}_{\Delta-\delta}$ grid [i.e., where $v + j\Delta < v + (j+1)(\Delta - \delta) < v + r \leq v + (j+2)(\Delta - \delta)$]. From Equation (24), p_{max-1} is minimized when $\Delta_\infty = \frac{r}{2}$.

However, *all* market buys pay $v + \Delta_\infty$ with this tick choice. This leads to a clear conflict between large institutions (which are ex ante predisposed to submit large blocks) and retail investors (who trade using smaller orders) who prefer a smaller tick size if (as seems likely) $c/\alpha < r$.

Interestingly, although small investors may prefer a finer grid, even they dislike continuous pricing. Instead they want to maximize the amount the specialist undercuts the *book* (i.e., since the crowd is irrelevant for them as a source of liquidity). The argument is similar to the one above except that (1) $v + c/\alpha$ replaces $v + r$ as the price the specialist needs to beat, and (2) the tick size must be strictly

[21] Madhavan and Cheng (1997) report that the downstairs (floor) market on the NYSE accounts for over 70% of blocks even as large as 50,000 shares.

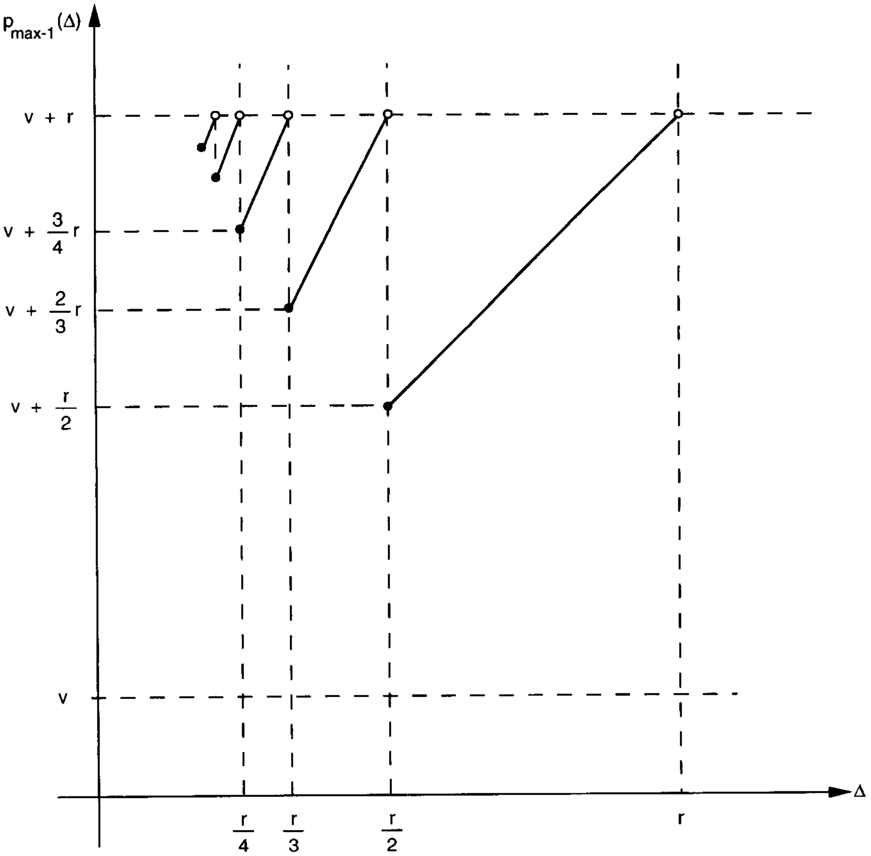

Figure 5
Maximum entry-deterring prices and price tick size
The highest clean up price p_{max-1} that undercuts the crowd (given their hurdle rate r) is plotted against the tick size Δ of different price grids (all of which are centered at $p_0 = v$).

greater than (rather than equal to) $\frac{c/\alpha}{2}$. This second point is because the potential depth in the limit book at $v + c/\alpha$ is zero, whereas the crowd still provides unlimited depth even if $p_{max} = v + r$.[22]

Although the specifics of this example clearly depend on anchoring the price grid at $p_0 = v$, the idea of maximizing the amount the specialist undercuts the book or the crowd carries over to price grids centered at other points.

[22] Strictly speaking, the liquidity-maximizing tick size for very small trades does not exist since it involves minimizing on an open set. Thus, we must restrict ourselves to relative comparisons.

Consider now an exchange when it chooses $\mathcal{P}$. Since price grids are not easily recustomized every time v changes, the exchange's ex ante choice of a (fixed) Δ affects the location of prices for many different v's in the future.[23] In choosing between two grids, it turns out that a very weak criterion still leads to interesting results.

Definition 2. *Liquidity, given a market design $\mathcal{M}$, ex ante dominates liquidity, given $\mathcal{M}'$, for a buy B if the average ask premiums are ranked $A(B, \mathcal{M}) \leq A(B, \mathcal{M}')$ for every v [and analogously for a sell S if the average bid discounts are $A(S, \mathcal{M}) \leq A(S, \mathcal{M}')$].*

Proposition 6. *For every Δ not in the interval $[\frac{r}{2}, 2r]$, there is a tick size Δ_∞ in $[\frac{r}{2}, 2r]$ that ex ante dominates Δ for both large market buys B and sells S in the limit as $B, S \to \infty$.*

Proposition 7. *For every Δ not in $[\frac{c/\alpha}{2}, 2c/\alpha]$, there is a tick size Δ_0 in $[\frac{c/\alpha}{2}, 2c/\alpha]$ that ex ante dominates Δ for small market buys B and sells S in the limit as $B, S \to 0$.*

While large and small investors may disagree about particular tick sizes between $\frac{c/\alpha}{2}$ and $2r$, they are unanimous in disliking ticks larger than $2r$ and less than $\frac{c/\alpha}{2}$. A tick size smaller than $\frac{c/\alpha}{2}$ simply allows the specialist to undercut $v + c/\alpha$ (and $v + r$, too) by less, while a tick larger than r means[24] that for some v's there may be no way to undercut the crowd or the book. In terms of current NYSE policy, where decimal pricing (i.e., $\Delta = .01$) and pricing on eighths (i.e., $\Delta = .125$) fall in these ranges is an interesting empirical question.

Turning from the extremes of very large and small orders, I conjecture that investors using intermediate-size market orders also prefer discrete over continuous pricing. First, Proposition 5 and the presence of the trading crowd insure that $p^*(B)$ is bounded by $\max(p_1, \min(p_{max}, v + \frac{c/\alpha}{F(B)}))$, where the third term [i.e., involving the upper tail probability $F(B)$] comes from inverting $B = H(\frac{c/\alpha}{p-v})$ for p. A discrete tick size Δ forces the specialist to undercut $\min(p_{max}, v + \frac{c/\alpha}{F(B)})$ by a nontrivial amount. Second, from Proposition 5, the cumulative depth $Q(p)$ at each inframarginal price p (i.e., $p^* > p > v + c/\alpha$) is maximized when prices are discrete and $p = p_1$ (i.e., so the specialist cannot undercut it).

[23] Although the exchange does not know v when it chooses its grid $\mathcal{P}$, investors do know the current v when they submit their orders.

[24] Of course, only when $\Delta > 2r$ can we guarantee that moving to a finer grid will not change pricing in intervals that already contain a price.

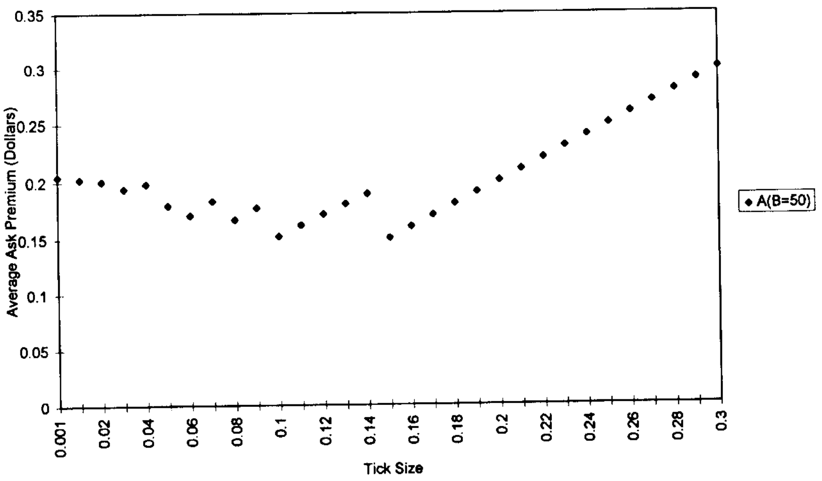

Figure 6
A numerical example of the relation between tick size and liquidity on mid-size trades
The average ask premium $A(B, \Delta)$ paid on a block of 50 round lots is minimized at a tick size Δ strictly greater than zero in a market where the stock value $v = \$20$, the probability of a buy $\alpha = 0.5$, the expected exponential volume $\theta = 25$ round lots, the price grid is centered at $p_0 = v$ with a constant tick size $\Delta = p_j - p_{j-1} = \frac{1}{8}$, the crowd hurdle rate $r = \frac{3}{8}$ (the only difference from Figure 4) and a submission cost $c = \$0.019$.

Figure 6 shows that my conjecture is true for at least some mid-size trades. In this figure, the average ask premium A on a market buy of $B = 50$ round lots is plotted as a function of Δ in a market with the same parameterization — except for a more realistic hurdle $r = \frac{3}{8}$ — as in Figure 4.

Although I have focused on the liquidity demand side's welfare, the preferences of the liquidity providers over continuous versus discrete pricing are easily understood. First, value traders earn positive expected profits on inframarginal limit orders at p_1 *only* if $p_1 > v + c/\alpha$. However, clearly $p_1 < v + c/\alpha$ with continuous pricing. In addition, p_1 with continuous pricing is also less than $v + r$, so the crowd never trades.

In contrast, the specialist's profits are actually *maximized* with continuous pricing. This is because as the tick size Δ shrinks, it becomes less costly for him to undercut the book, which leads, in turn, to an endogenously thinner book in equilibrium.

Proposition 8. *(1) The cumulative depth $Q(\varphi, \Delta)$ at or below any price $\varphi \geq v + c/\alpha$ on grids $\mathcal{P}$ such that $\varphi \in \mathcal{P}$ is minimized in the limit as $\Delta \to 0$. (2) The specialist's profit for every order size B is maximized in the limit as $\Delta \to 0$.*

3.2 Specialist vs. pure limit order markets

Another policy issue concerns the pros and cons of a hybrid specialist/limit order market (as on the NYSE) versus a pure limit order market (as on the Paris Bourse). The present model provides some insights into the role and viability of the specialist as a provider of liquidity.

The first step is to explain the mechanics of how the crowd can provide residual liquidity in a pure limit order market. Large buyers will sequentially hit limit orders at prices p_1 up through p_{max} and then — as Biais, Hillion, and Spatt (1995) find on the Paris Bourse — post the residual at p_{max}. The crowd sees this "distress call" and enters and submits offsetting orders to clean up.

Proposition 9. *In a pure limit order market, value traders place orders until, in equilibrium, the limit order book is*

$$S_1 = H_1, \tag{25}$$

and

$$S_j = H_j - H_{j-1}$$

at prices $p_j \le p_{max}$ and $S_j = 0$ thereafter.

The pure market depth S_1 is the same, given a common tick size Δ, as in the hybrid market. However, above p_1 the cumulative depth $\sum_{i=1}^{j} S_i = H_j$ on the Bourse is greater than the weighted average Q_j in the hybrid market. This does not mean, however, that a pure limit order market is necessarily preferable to the hybrid market. Indeed, investors who value marginal over inframarginal liquidity strictly prefer having a specialist despite the reduced depth in the book. To see this, consider two possible cases for the tick size.

First, suppose Δ is such that $p_1 > v + c/\alpha$. Retail investors submitting small market buys $B \le H_1$ are indifferent between the two market structures since they trade at p_1 in either case. Moreover, investors with orders $H_1 < B \le H_2$ strictly prefer the hybrid market since the specialist is willing to sell up to an additional $H_2 - H_1$ at p_1 to undercut limit orders at p_2. Mid-size investors' preferences are more complicated, however. Consider, for example, someone buying $B = H_3 + \epsilon < H_4$. In this case, the average ask premium in the pure limit order market,

$$A^{pure} = \frac{H_1 p_1 + (H_2 - H_1)p_2 + (H_3 - H_2)p_3 + \epsilon p_4}{B} - v, \tag{26}$$

is clearly (given $\gamma_2 < 1$) less than in the hybrid market,

$$A^{hybrid} = \frac{H_1 p_1 + \gamma_2 (H_2 - H_1)p_2 + (H_3 + \epsilon - Q_2)p_3}{B} - v \tag{27}$$

for ϵ small enough. Of course, for a larger ϵ, this might reverse again. Thus, mid-size investors like buying up to H_{j+1} at prices at or below p_j in the hybrid market (i.e., because the specialist undercuts orders at p_{j+1}) versus only H_j shares in the limit order market. However, the lower hybrid clean up price comes at the cost of reduced limit order depth at inframarginal prices $p_i < p^*$ (i.e., due to their lower execution probabilities again because of specialist undercutting). For some mid-size orders, the "lower depth" effect outweighs the "lower clean up price" effect. Investors trading these quantities will thus prefer a pure limit order market's deeper book. Lastly, institutions prefer a hybrid system if their orders are large enough (i.e., given the bounded liquidity in either book) since the specialist provides unlimited liquidity for large buys at p_{max-1}, which is less than the p_{max} charged by the trading crowd in the pure limit order market.

If instead Δ is such that $p_1 < v + c/\alpha$, then the large and mid-size investors' preferences are qualitatively unchanged. However, retail investors submitting *very* small orders $B \leq H_{j_s^{min}}$ will like the fact that the specialist undercuts the book at p_{min} by selling at p_{min-1} on their trades. Thus, under these conditions, small retail investors strictly prefer a hybrid market.

So far, these comparisons assume that the tick size Δ is the same in the two markets. Actually, a stronger result is possible.

Proposition 10. *(1) In a pure limit order market, the average ask premium is minimized for all B with continuous pricing. (2) A hybrid market with any discrete tick* $min(c/\alpha, r) > \Delta > 0$ *ex ante dominates a pure limit order market with continuous pricing for, simultaneously, very small orders* $B \to 0$ *and very large orders* $B \to \infty$.

Since neither market mechanism dominates the other across all order sizes, which one will prevail is an interesting topic for future research. It appears, though, that small retail and large institutional investors would both lobby for a hybrid structure — even if this involves some specialist market power — as the preferred "second best" when competition among multiple risk-neutral dealers is infeasible.

4. Price Discovery

This section generalizes the basic model to allow [as in Rock (1996)] for price discovery due to market order flows that are informative. In broad terms, the main insight is that the underlying logic and tractability of the analysis carry over in this richer setting. There are, however, at least a few empirically testable qualitative differences. First, the interplay between price discovery and liquidity provision leads to the possibility of both (1) partial execution (and hence inframarginal profitability) of limit sells *above* p_1 and (2) actual trading by the crowd. Second, in contrast to Proposition 4, increasing the probability of large *informative* orders does not necessarily induce a deeper book. Third, the book may now have "holes" in it — that is, prices with zero depth $S_j = 0$ even though higher and lower prices have positive depth.

Consider a market identical to the one above except that the conditional value of the stock is now a non-decreasing continuous function $v(B)$ of the size of the arriving market buy B. Let $\alpha F(x, y)$ be the cumulative probability of a buy $x < B \leq y$ and $\mu(x, y)$ the expected value of $v(B)$ conditional on $x < B \leq y$, and define $\alpha F(y)$ and $\mu(y)$ analogously for $B > y$. Prices $\ldots, p_{-2}, p_{-1}, p_1, p_2, \ldots$ on a grid $\mathcal{P}$ are indexed relative to the zero-volume valuation $v(0)$. If a price on $\mathcal{P}$ equals $v(0)$, it is indexed as p_0. Define the limit depth S_j, cumulative depth Q_j, and submission cost c as above.

One other change is a new priority rule protecting market orders. In particular, I assume that the specialist cannot "front run" arriving market orders in that, as public orders, they have priority in trading with the limit book. This rule prevents the specialist from cleaning out limit sells at prices below $v(B)$ ahead of a market order B.[25]

The specialist once again exercises market power subject to competition from the value traders, the crowd, and the obligations of price, public order, and time priority. In trading off price and quantity, his profit also depends on B through the conditional value $v(B)$. Given B and a limit book $S_1, S_2, \ldots$, he picks the clean up price p^* to solve

$$\max_{p>v(0)} \pi(p) = \tag{28}$$
$$\begin{cases} \left[B - T(B, p)\right]\left[p - v(B)\right] & \text{if } p \geq v(B) \\ \sum_{p_j \leq p} \max(0, S_j - \max(0, B - Q_{j-1}))\left[v(B) - p_j\right] & \text{otherwise,} \end{cases}$$

where the total selling by rival liquidity providers, given the trading crowd's reservation price $p_{max}(B)$, is

$$T(B, p) = \begin{cases} \min\left(B, \sum_{i \text{ s.t. } p_i \leq p} S_i\right) & \text{if } p < p_{max}(B) \\ B & \text{if } p \geq p_{max}(B). \end{cases} \tag{29}$$

The first branch in Equation (28) is the specialist's profit from liquidity provision (i.e., selling at p^*), while the second allows for the possibility of his *buying* from the book (e.g., cleaning out excess liquidity) at successively higher prices $p_j < v(B)$ if $B < Q_j$.[26]

[25] Kyle (1992) suggests that the desire to prevent front running is a key motive behind investors' demand for immediacy.

[26] The possibility of the specialist buying, given a buy B, is included simply for completeness.

The crowd's reservation price p_{max} is now a function of B, which they (along with the specialist) also see. In particular, given B, the crowd is willing to provide liquidity by selling $B - \sum_{i \text{ s.t. } p_i \leq p^*} S_i$ whenever $p^* \geq v(B) + r$.[27] As in Section 1, the potential entry by the crowd at $p_{max}(B)$ [i.e., the lowest price at or above $v(B) + r$] effectively constrains the specialist to set p^* to be no more than $p_{max-1}(B)$ [i.e., the highest price below $v(B) + r$] if he wants positive volume. Since it is sometimes more convenient to work with constraints expressed in terms of B, let κ_j be the volume that solves $v(\kappa_j) + r = p_j$. In other words, κ_j is the cutoff for B such that when $B > \kappa_j$, selling at p_j undercuts the crowd's $p_{max}(B)$.

Value traders are still Bertrand competitors with stochastic arrival times. They submit limit orders until the marginal expected profit at each price p_j is driven to zero (if $S_j > 0$) or is non-positive (if $S_j = 0$). In so doing, they recognize the interaction between limit order execution and stock values. While larger orders B lead to execution of limit sells at higher prices, they also imply higher valuations $v(B)$ for the stock sold. In particular, a limit sell at p_j is profitably executed if and only if — given the monotonicity of v — it is executed against an order $B < \beta_j$, where β_j solves $v(\beta_j) = p_j$. Thus, the marginal expected profit for a limit sell at p_j is

$$\begin{aligned} e_{sj} &= \alpha F(Z_j^-)[p_j - E(v \mid Z_j^-)] \\ &\quad + \alpha F(Z_j^+)[p_j - E(v \mid Z_j^+)] - c, \end{aligned} \tag{30}$$

where Z_j^- is the set of volumes $B \leq \beta_j$ such that limit sells at p_j are executed in full and $F(Z_j^-)$ is its associated probability, and where the set Z_j^+ and probability $F(Z_j^+)$ are defined analogously for $B > \beta_j$. Thus, the three terms in Equation (30) represent the ex ante expected gain from profitable executions ($B \in F_j^-$), the ex ante loss from unprofitable executions ($B \in F_j^+$), and any up-front submission costs, respectively.

Definition 3. *An equilibrium is a set of value trader limit orders $S_1, S_2, \ldots$ such that the marginal expected profit $e_{sj} = 0$ (if $S_j > 0$) or $e_{sj} \leq 0$ (if $S_j = 0$) at each price p_j, given that the specialist's strategy $p^*(B)$ satisfies Equation (28).*

Proposition 11 shows that, in equilibrium, this is never optimal. In addition, the specialist clearly never buys from the crowd, as they are both discretionary traders and thus cannot simultaneously profit on a trade.

[27] Liquidity provision from the crowd is now limited rather than infinite at a price p since larger orders B can move $v(B) + r$ above p.

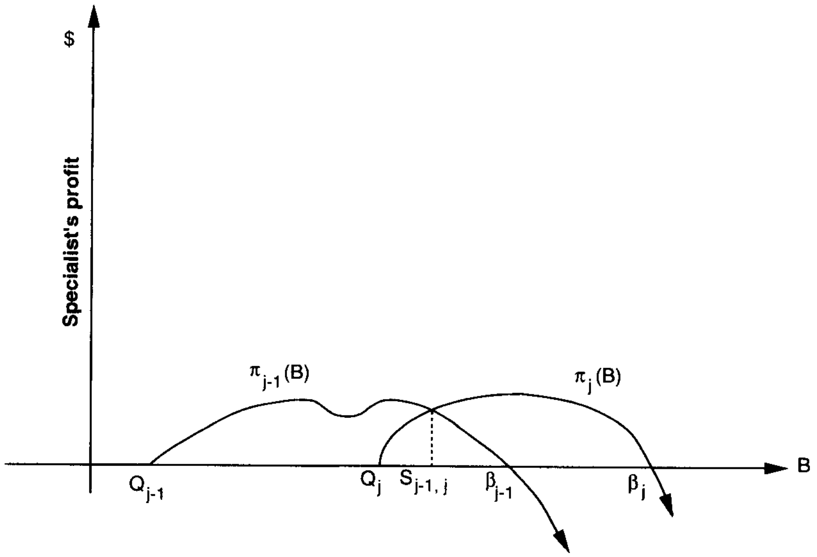

Figure 7
A simple example of the specialist's profit function
The curves $\pi_{j-1}(B)$ and $\pi_j(B)$ illustrate the specialist's profit for different possible volumes B in the case where $Q_{j-1} \geq \kappa_{j-1}$ and $Q_j \geq \kappa_j$ (i.e., the cumulative depths both exceed the respective cutoffs for undercutting the crowd at p_{j-1} and p_j). They are zero at volumes β_{j-1} and β_j, where the respective valuations equal the prices p_{j-1} and p_j. The crossing point $S_{j-1,j}$ is the volume where the two profit levels are equal.

The analysis can be simplified with two preliminary results. The first is an immediate implication of Equation (30).

Proposition 11. *In equilibrium, the cumulative depths satisfy $Q_j \leq \beta_j$, and thus the specialist never buys from the limit order book.*

This lets us ignore the bottom branch of Equation (28) because if $B < Q_j$, then $v(B) < p_j$, and the specialist is unwilling to *buy* at a loss at p_j to make up any shortfall $Q_j - B$. Thus, he either defers to the book/crowd (i.e., does not trade) by setting $p^* = p_j$ or he actively provides liquidity (i.e., sells) by undercutting the book/crowd at $p^* < p_j$.

The specialist's profit $\pi_j = (B - Q_j)[p_j - v(B)]$ when selling at p_j has two important properties, as illustrated in Figure 7. First, the function π_j is initially increasing in B,[28] but is eventually pulled to zero as $v(B) \rightarrow p_j$ and stays negative thereafter. Second, its slope $\frac{d\pi_j}{dB} = p_j - v(B) - \frac{dv(B)}{dB}(B - Q_j)$ is increasing in p_j. Thus, despite their

[28] Whether π_j is increasing at some $B > Q_j$ depends on how slowly $v(B)$ increases in B.

non-monotonicity, each pair of profit functions π_j and π_k (i.e., given prices $p_j \neq p_k$) has at most one positive crossing point $S_{j,k}$. This *single crossing property* leads to our next result.

Proposition 12. *In equilibrium, if $B > B'$, then $p^*(B) \geq p^*(B')$.*

Monotonicity of the clean up price $p^*(B)$ again implies the existence of execution thresholds $\hat{S}_1 \leq \hat{S}_2 \leq \ldots$, where $\hat{S}_j = \inf(B \geq Q_j \mid p^* \geq p_j)$ as in Equation (7). With informative orders, however, these execution thresholds no longer necessarily equal the price thresholds $S_j^* = \inf(B \mid p^* \geq p_j)$ since partial execution of limit sells *above* p_1 cannot be ruled out now. In particular, any time $B > \beta_{j-1}$ [i.e., so that $v(B) > p_{j-1}$], the specialist cannot profitably undercut limit sells at p_j and thus must defer to the book even if $\beta_{j-1} < B \leq Q_j$.

Accordingly, e_{sj} can be rewritten in terms of the thresholds $\hat{S}_1$, $\hat{S}_2, \ldots$. Since the full execution of S_j depends on whether $B > \hat{S}_j$, the marginal expected profit at p_j simplifies to

$$\begin{aligned} e_{sj}(\hat{S}_j) &= \alpha F(\hat{S}_j, \beta_j)[p_j - \mu(\hat{S}_j, \beta_j)] \\ &\quad + \alpha F(\beta_j)[p_j - \mu(\beta_j)] - c. \end{aligned} \tag{31}$$

Depth S_j influences the marginal profit e_{sj} only indirectly via the threshold $\hat{S}_j$. Thus, to determine the equilibrium depth of the limit book, we need the relation between the S_j's and $\hat{S}_j$'s. As in Section 3, the execution thresholds $\hat{S}_j$ are closely related to the adjacent price crossing points $S_{j-1,j}$. Once again, the crossing point $S_{j-1,j}$ is found by comparing — given an order $B > Q_j$ — the specialist's profit from selling at p_j,

$$\pi_j = (B - Q_j)[p_j - v(B)], \tag{32}$$

with the profit from selling at p_{j-1},

$$\pi_{j-1} = (B - Q_{j-1})[p_{j-1} - v(B)]. \tag{33}$$

Thus, $\pi_j > \pi_{j-1}$ if and only if

$$B > Q_{j-1} + S_j \frac{p_j - v(B)}{p_j - p_{j-1}}, \tag{34}$$

where now — unlike in the uninformative order flow case — the RHS also depends on B via $v(B)$. Combining Equation (34) and the uniqueness of the crossing point for π_j and π_{j-1} gives the following implicit definition:

$$S_{j-1,j} = Q_{j-1} + S_j \frac{p_j - v(S_{j-1,j})}{p_j - p_{j-1}}. \tag{35}$$

Two considerations may prevent the specialist from setting a threshold $\hat{S}_j$ equal to $S_{j-1,j}$. First, he must also undercut the trading crowd to win any volume. In particular, if $S_{j-1,j} \le \kappa_j$, then his share of buys $S_{j-1,j} \le B \le \kappa_j$ would be zero at $p^* = p_j$. To avoid this, the specialist strategically sets p^* — provided that $B < \beta_{j-1}$ — to undercut the crowd at p_j. Second, if $\beta_{j-1} < \max(Q_j, \kappa_j)$, then there is no price $p > v(B)$ for the specialist to use to undercut the book's/crowd's willingness to sell at p_j on buys $\beta_{j-1} < B < \max(Q_j, \kappa_j)$. Under these conditions, he simply steps aside and lets such B's cross with the book/crowd. [29] Thus, now the crowd may actually trade as an active (rather than merely latent) source of liquidity.

Proposition 13. *In equilibrium, each threshold $\hat{S}_j$ is determined by one of four cases. First, at price p_1, the threshold is $\hat{S}_1 = S_1$. Second, at prices $p_j > p_1$, where $\kappa_j \le 0$, the threshold is*

$$\hat{S}_j = \begin{cases} S_{j-1,j} & \text{if } e_{sj}(\beta_{j-1}) < 0 \\ Q_j & \text{if } e_{sj}(\beta_{j-1}) \ge 0. \end{cases} \tag{36}$$

Third, at prices $p_j > p_1$, where $0 < \kappa_j < \beta_{j-1}$, then

$$\hat{S}_j = \begin{cases} \kappa_j & \text{if } e_{sj}(\kappa_j) \le 0 \\ S_{j-1,j} & \text{if } e_{sj}(\kappa_j) > 0 > e_{sj}(\beta_{j-1}) \\ Q_j & \text{if } e_{sj}(\beta_{j-1}) \ge 0. \end{cases} \tag{37}$$

Fourth, at prices $p_j > p_1$, where $\beta_{j-1} < \kappa_j$, then

$$\hat{S}_j = \begin{cases} \beta_{j-1} & \text{if } e_{sj}(\beta_{j-1}) < 0 \\ Q_j & \text{if } e_{sj}(\beta_{j-1}) \ge 0. \end{cases} \tag{38}$$

The specialist profit function in Figure 8 illustrates the logic behind these cases. As in Figure 7, the profit from selling at different prices is plotted for various possible B's, given the hypothetical cumulative depths $Q_1, Q_2, \ldots$. For small buys $B \le Q_1$, the specialist cannot (in this example) undercut the book so he simply defers to it. He sells to larger B's at p_1 until the crossing point $S_{1,2}$, where he switches to p_2. Since limit sells at p_3 cannot be profitably undercut on still larger buys $\beta_2 < B \le Q_3$ [i.e., since $\beta_2 < B$ implies $p_2 < v(B)$], the possibility of partial execution of S_3 means $\hat{S}_3 = Q_3$. Buys $B \ge Q_4$ are examples of the third case in Proposition 13. Selling at p_5 (rather than p_4) would

[29] This just generalizes what happens at p_1. Geometrically, $\beta_{j-1} < Q_j$ corresponds to a *negative-profit* crossing point $S_{j-1,j}$ to the left of Q_j. In addition, it does not matter whether $F(\hat{S}_j, \beta_j)$ and $\mu(\hat{S}_j, \beta_j)$ in Equation (31) are defined with S_j executing only when $B > \hat{S}_j$ (i.e., as when the specialist trades) or when $B \ge \hat{S}_j = Q_j$ (i.e., when the specialist defers to the book) since, once again, with a continuous distribution F, they are the same.

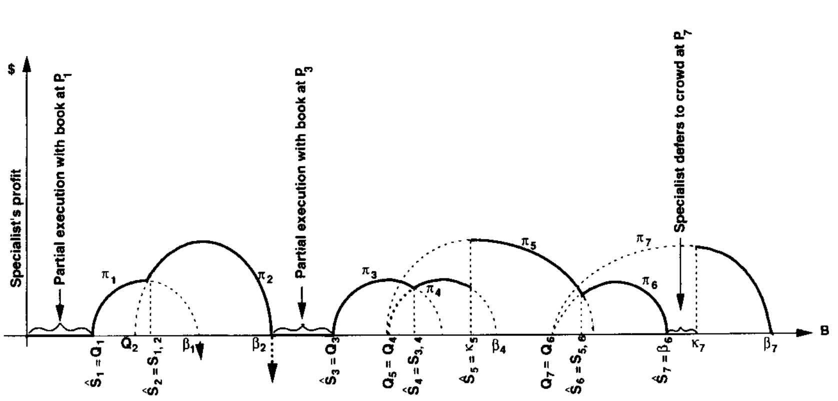

Figure 8
A more complicated example of the specialist's profit function
This figure illustrates each of the possible ways execution thresholds $\hat{S}_j$ can be determined when market orders B are informative. A generic π_j function gives the specialist's profit if he could sell at price p_j after first executing all Q_j limit sells at or below price p_j *without regard* to whether p_j undercuts the crowd's reservation price $p_{max}(B)$. The specialist's optimal strategy is the upper envelope of the π functions subject to a constraint that he has positive volume at p_j only if B is greater than κ_j [i.e., where $v(\kappa_j) + r = p_j$] so that selling at p_j undercuts the crowd. The π_j functions each equal zero at a volume β_j where $v(\beta_j) = p_j$. The crossing points $S_{j-1,j}$ are the volumes such that the profit from selling at p_{j-1} and p_j are equal. One idiosyncrasy in this example is that the tick sizes are *unequal* since $\kappa_5 < \beta_4$ but $\beta_6 < \kappa_7$ implies that $p_5 - p_4 < p_7 - p_6$. This is simply to illustrate two particular cases that depend on different relative ordering of β_{j-1} and κ_j.

seem to be the natural alternative to p_3 (since this book happens to be empty at p_5) except that the crowd would enter and supply all of the residual demand at p_5 whenever $B \leq \kappa_5$. Thus, the threshold for p_5 is $\hat{S}_5 = \kappa_5$ rather than $S_{3,5}$. The rest of the figure is self-explanatory until $\beta_6 < B < \kappa_7$. Here, the crowd is willing to sell at p_7, but the specialist cannot profitably undercut them by selling at p_6 [i.e., since $B > \beta_6$ implies $p_6 < v(B)$], so he has no choice but to defer to them. Thus, $\hat{S}_7 = \beta_6$, which illustrates the fourth case in Proposition 13.

Value traders again submit orders in equilibrium until the execution thresholds $\hat{S}_j$ are high enough to eliminate any marginal expected profit in the book. Of course, if the marginal profit e_{sj} is negative to start with (i.e., given $S_j = 0$), then the book stays empty at p_j. At this point, it is helpful to introduce an analogue G_j to H_j in Equation (11). In particular, using the fact that $F(g, \beta_j)[p_j - \mu(g, \beta_j)]$ from Equation (31) is continuous and strictly decreasing in $g \in [0, \beta_j]$ — which implies that its inverse g_j exists and is continuous and decreasing — let G_j be the non-negative "zero" (where possible) of the marginal profit condition

$$G_j \equiv \begin{cases} g_j\left(c/\alpha + F(\beta_j)[\mu(\beta_j) - p_j]\right) & \text{if } e_{sj}(0) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

The new trick with informative orders is that Proposition 13 does not (unlike Proposition 2) guarantee positive depth at each $p_j < p_{max}$, where $S_{j-1} > 0$. Rather it only implies that $S_j > 0$ if and only if G_j (i.e., the potential threshold) is larger than a minimum *floor*,

$$\hat{S}_j^{floor} = \begin{cases} 0 & \text{if } j = 1 \\ \max\left(0, \min(\kappa_j, \beta_{j-1})\right) & \text{otherwise,} \end{cases} \tag{40}$$

reflecting competition from the crowd. In particular, at prices where $e_{sj}(\hat{S}_j^{floor}) < 0$, the specialist strategically sets $\hat{S}_j = \hat{S}_j^{floor}$ to undercut the crowd, which, in turn, leads to $S_j = 0$.[30] However, if $e_{sj}(\hat{S}_j^{floor}) \geq 0$, then, in equilibrium, $S_j > 0$ and inverting the zero profit condition gives $\hat{S}_j = G_j \geq \hat{S}_j^{floor}$.

At p_1 we are done now since $S_1 = \hat{S}_1 = G_1 \geq \hat{S}_1^{floor}$. At prices with positive depth above p_1, we substitute $\hat{S}_j = G_j$ into Proposition 13 to get the corresponding depth

$$S_j = \gamma_j(G_j - Q_{j-1}), \tag{41}$$

where[31]

$$\gamma_j = \begin{cases} \frac{p_j - p_{j-1}}{p_j - v(G_j)} < 1 & \text{if } G_j < \beta_{j-1} \\ 1 & \text{if } G_j \geq \beta_{j-1}. \end{cases} \tag{42}$$

Proposition 14. *In a market with informative order flows, the equilibrium limit book is*

$$S_j = \begin{cases} G_1 & \textit{if } j = 1 \\ \gamma_j(G_j - Q_{j-1}) > 0 & \textit{if } j > 1 \textit{ and } e_{sj}(\hat{S}_j^{floor}) \geq 0 \\ 0 & \textit{if } j > 1 \textit{ and } e_{sj}(\hat{S}_j^{floor}) < 0\ , \end{cases} \tag{43}$$

[30] The reader may wonder what ensures threshold monotonicity $\hat{S}_j \geq \hat{S}_{j-1}$ or even $\hat{S}_j \geq Q_{j-1}$ when $\hat{S}_j = \hat{S}_j^{floor}$. The answer is immediate if $\hat{S}_{j-1} = \hat{S}_{j-1}^{floor}$ since $\hat{S}_j^{floor} \geq \hat{S}_{j-1}^{floor}$. If instead $\hat{S}_{j-1}$ is determined by the zero profit condition $e_{s,j-1}(\hat{S}_{j-1}) = 0$, then we cannot have $\hat{S}_{j-1} > \hat{S}_j = \hat{S}_j^{floor}$ since then $e_{sj}(\hat{S}_{j-1}) \geq e_{s,j-1}(\hat{S}_{j-1})$, which contradicts $\hat{S}_j = \hat{S}_j^{floor}$. This, in turn, implies $S_j = \hat{S}_j^{floor} \geq Q_{j-1}$ since, by definition, $\hat{S}_{j-1} \geq Q_{j-1}$. The advantage of defining $\hat{S}_j^{floor}$ without reference to $\hat{S}_{j-1}$ or Q_{j-1} is that it lets us check whether the book has positive depth at p_j [i.e., if $e_{sj}(\hat{S}_j^{floor}) > 0$] at the outset without needing to know the rest of the equilibrium book.

[31] If $G_j \geq \beta_{j-1}$, then $v(G_j) \geq p_{j-1}$ and $\hat{S}_j = Q_j$ because the specialist cannot profitably undercut the book.

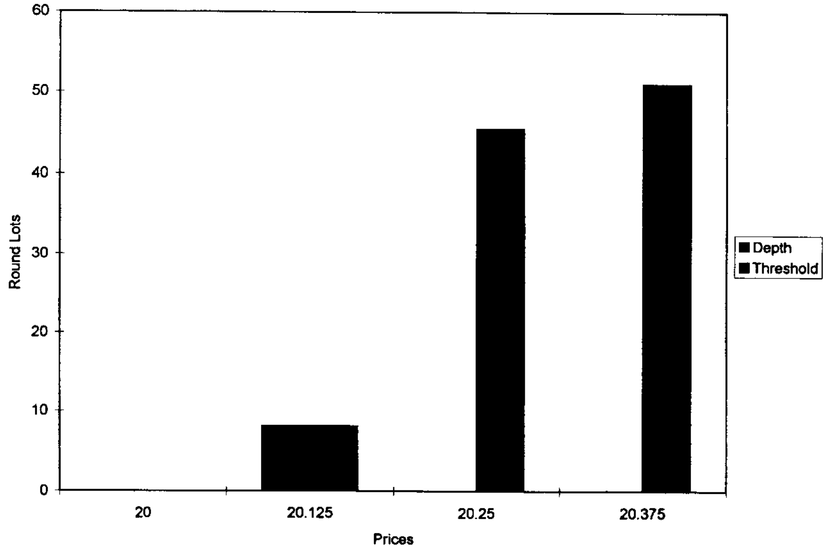

Figure 9
A numerical example of limit orders and thresholds with uniform distributed, informative market orders
This figure gives the equilibrium depths and thresholds at prices $p_1, \ldots, p_3$ for a market in which the updating function is $v(B) = 20 + .00275$, the probability of a market buy $\alpha = 0.5$ where B is uniformly distributed between 0 and 50 round lots, the price grid is centered at $p_0 = v(0) = \$20$ with a constant tick size $\Delta = p_j - p_{j-1} = \$0.125$ equal to the crowd's hurdle rate $r = \$0.125$, and a submission cost $c = \$0.019$ per share.

and the corresponding thresholds are

$$\hat{S}_j = \begin{cases} G_1 & \mathit{if}\, j = 1 \\ G_j & \mathit{if}\, j > 1 \;\mathit{and}\; e_{sj}(\hat{S}_j^{floor}) \geq 0 \\ \hat{S}_j^{floor} & \mathit{if}\, j > 1 \;\mathit{and}\; e_{sj}(\hat{S}_j^{floor}) < 0. \end{cases} \tag{44}$$

The equilibrium is once again unique since (as in Section 3) it is constructed recursively. The solution is also still "almost" analytic — now up to the evaluation of the inverse conditional expectations G_j. Proposition 3 for uninformative orders is clearly a special case of Proposition 14, where $v(B) = v$ and $G_j = H_j$. Although I have not found any closed-form examples, numerical solutions are easily calculated. For example, Figure 9 shows the sell depths and thresholds on an eighths grid centered at $p_0 = \$20$ when volume is uniformly distributed on [0, 50] round lots, $\alpha = 0.5$, $c = .019$, $r = .125$, and the updating function is $v(B) = 20 + .00275B$.

Two special cases deserve comment. If the crowd's reservation profit r is less than the tick size Δ, then the equilibrium is again "competitive" in that $p^*(B)$ is simply the lowest price p_j above $v(B)$.

This follows from Equation (44) and the fact that if $r < \Delta$, then $\max(0, \min(\beta_{j-1}, \kappa_j)) = \max(0, \beta_{j-1})$ [i.e., since $\beta_{j-1} = v^{-1}(p_j - \Delta) < v^{-1}(p_j - r) = \kappa_j$]. The book/crowd sells at p_j to buys $\beta_{j-1} < B \leq \kappa_j$ (i.e., the specialist steps aside), and the book/specialist sells at p_j to buys $\kappa_j < B < \beta_j$ (i.e., to undercut the crowd at p_{j+1}).

At the other extreme, as $r \to \infty$, the crowd becomes irrelevant (i.e., $\lim_{r\to\infty} \kappa_j < 0$ for all j), and only the book and the specialist provide liquidity. In this case, the thresholds $\hat{S}_j$ are determined exclusively by Equation (36).

One complication with informative orders is that the limit book can have "holes" in it (e.g., $S_5 = S_7 = 0$ in Figure 8) if the $\hat{S}_j \geq \hat{S}_j^{floor}$ constraint is binding. Nonetheless, the cumulative depth Q_j from Equation (43) is still a weighted average — now of the $G_1, \ldots, G_j$ at prices with nonzero depths

$$Q_j = \begin{cases} Q_{j-1} & \text{if } S_j = 0 \\ \gamma_j G_j + (1-\gamma_j) Q_{j-1} & \text{if } S_j > 0 \end{cases} \tag{45}$$

or

$$Q_j = \sum_{k \in \mathcal{I} \text{ s.t. } k \leq j} \gamma_k \left[\prod_{i \in \mathcal{I} \text{ s.t. } k < i \leq j} (1 - \gamma_i) \right] G_k, \tag{46}$$

where $\mathcal{I} = \{i \mid e_{si}(\hat{S}_i^{floor}) \geq 0\}$ are the indices corresponding to positive depths in the book.

Once again, this implies bounded cumulative depth in the book across all grids $\mathcal{P}$. However, the possibility of holes complicates the details of the upper bound. In particular, given any valuation $v \geq v(0)$, define $\varphi_v = \operatorname{argmax}\{\varphi \leq v \mid e_\varphi(\hat{S}_\varphi^{floor}) \geq 0\}$ where $\hat{S}_\varphi^{floor}$ is the floor corresponding to $p = \varphi$. In words, φ_v is the highest *potential* price below v (i.e., on or off any particular grid $\mathcal{P}$) where positive depth S_{φ_v} is possible. Finally, let G_{φ_v} solve $e_{\varphi_v}(G_{\varphi_v}) = 0$.

Proposition 15. *Given any value v, the cumulative depth $\sum_{i \text{ s.t. } p_i \leq v} S_i$ on any grid $\mathcal{P}$ is bounded by G_{φ_v}.*

In terms of comparative statics, market liquidity is decreasing in both submission costs c/α and in the informativeness $v(B)$ of orders.

Proposition 16. *If $c'/\alpha' < c/\alpha$ and/or if $v'(B) \leq v(B)$ for all B, then the market with c'/α' and v' has greater marginal liquidity ($\hat{S}_j' \geq \hat{S}_j$ for all j). In addition, the book is deeper ($Q_j' \geq Q_j$) in the c'/α' market, given $v' = v$.*[32]

[32] Perhaps counterintuitively, it may be possible to have *greater* cumulative depths $Q_j > Q_j'$ at some

Unlike in Proposition 4, the volume distribution F's effect on depth is ambiguous now. To see why, consider two distributions where F' first-order stochastic dominates F. Let S_j and $\hat{S}_j$ be the depth and threshold, respectively, at p_j for F and define S'_j and $\hat{S}'_j$ analogously for F'. On the one hand, increasing the probability of large orders x at the expense of small orders y, where $\hat{S}_j < y < x$ reduces the probability of high profit/low loss execution of limit sells (i.e., when $B = y$) and increases the probability of lower profit/higher loss execution (i.e., when $B = x$) at all prices at or below p_j. This has two opposing effects on the depth S'_j relative to S_j. Holding Q'_{j-1} fixed at Q_{j-1}, limit sells at p_j are less profitable [i.e., $e'_{sj}(\hat{S}_j) \leq e_{sj}(\hat{S}_j) \leq 0$], which tends to lower S'_j. However, the reduced profitability of limit sells *below* p_j leads to a lower cumulative depth Q'_{j-1}, which tends to increase S'_j. While the net effect on depth S'_j is unclear given $\hat{S}_j < y < x$, the effect on cumulative depth, $Q'_j \leq Q_j$, is unambiguous.[33]

On the other hand, increasing the probability of large orders x at the expense of small y when $\hat{S}_{j-1} < y < \hat{S}_j < x < \beta_j$ leads to greater depth, $S'_j \geq S_j$, for two reasons. First, it reduces the probability of nonexecution of the marginal limit sell at p_j (i.e., when $B = y$) and increases the probability of profitable execution (i.e., when $B = x$). If $e'_{sj}(\hat{S}_j) > 0$ with this shift, then additional limit sells will be posted at p_j until the new marginal profit is driven to zero. Second, the depth S_j may also increase due to a lower cumulative depth Q'_{j-1} since shifting probability from y to x also reduces (from above) the profitability and thus cumulative depth at prices below p_j. However, while the effect on S'_j is unambiguous, $S'_j \geq S_j$, now the net effect on cumulative depth Q'_j is unclear.

The last result in this section is that the adverse selection problem vis-a-vis informed orders produces a bid/ask spread around $v(0)$ even when the up-front costs c are zero. This is different from the uninformative order case in Section 2, where $c > 0$ is necessary to make "room" for the specialist's market power.

p_j when market orders are more informative. In particular, if there is enough probability in F concentrated just below G'_j then G_j will be only slightly less than G'_j. However, if $v(G_j) \gg v'(G'_j)$, then it is possible from Equation (42) to have γ_j enough larger than γ'_j to offset the effect of the slightly lower G_j in Equation (45). The reason is that the higher $v(G_j)$ makes undercutting limit sells at p_j less attractive to the specialist, as reflected in the larger γ_j.

[33] To see this, suppose $Q'_{j-1} \leq Q_{j-1}$. If $S'_j = 0$, then by definition, $Q'_j = Q'_{j-1} \leq Q_j$. If instead $S'_j > 0$, then from Equation (43), $Q'_j = \gamma'_j G'_j + (1 - \gamma'_j)Q'_{j-1} < \gamma_j G_j + (1 - \gamma_j)Q_{j-1} = Q_j$ since $G'_j < G_j$, $Q_{j-1} < G_j$, and $\gamma'_j \leq \gamma_j$ [from Equation (42)]. In other words, the boost in S'_j from a lower Q'_{j-1} cannot offset the lower G'_j in Q'_j.

Proposition 17. *If* $\lim_{p_1 \downarrow v(0)} F(\beta_1) > 0$ *and* $\lim_{p_1 \downarrow v(0)} \mu(\beta_1) > p_1$*, then no limit sells are posted below* $\varphi_{book} > v(0) + c/\alpha$*, where* φ_{book} *is the lowest value* φ *such that* $e_{p_1}(0) \geq 0$ *when* $p_1 = \varphi$*.*

5. Optimal Market Design (Revisited)

The earlier market design analysis carries over almost directly to informative market orders. This section briefly reconsiders first, the relation between tick size Δ and liquidity and second, the pros and cons of pure limit order versus hybrid specialist/limit order markets.

5.1 Optimal tick size

The key to the liquidity-maximizing tick size Δ_∞ for large uninformative buys B was to choose Δ to maximize the amount by which the specialist had to undercut the crowd's reservation price p_{max}. We could ignore the effect of Δ on inframarginal liquidity because cumulative depth in the book below $v + r$ was bounded and thus unimportant for the average ask premium $A(B, \Delta)$ as $B \rightarrow \infty$. With informative orders, however, larger B's imply higher valuations $v(B)$ and thus higher reservation prices $p_{max}(B)$. If this were to cause an ever-increasing number of limit orders at progressively higher prices to execute, then the weight attached to the inframarginal limit orders in $A(B, \Delta)$ might not go to zero as $B \rightarrow 0$.[34] Fortunately, this is unlikely to be a problem.

Proposition 18. *Fix any finite* $\delta > 0$ *and zero-volume valuation* $v(0)$*. The limiting total depth* $Q_\infty = \sum_{i=1}^{\infty} S_i$ *on any grid* $\mathcal{P}$ *with a tick size* $\Delta \leq \delta$ *is bounded if the hurdle rate* r *is finite and* $c > 0$*.*

If we now take $v(0)$ to be random — as it is from the exchange's perspective when it chooses $\mathcal{P}$ — then regardless of the distribution over possible $v(0)$'s, we have Proposition 19.

Proposition 19. *Given a bounded* Q_∞*, there is for every* Δ *not in* $[\frac{r}{2}, 2r]$ *a tick size* Δ_∞ *in* $[\frac{r}{2}, 2r]$ *that ex ante dominates* Δ *for both large market buys* B *and sells* S *in the limit as* $B, S \rightarrow \infty$*.*

Consider next very small orders. As in Section 4, the key is to find the lowest price from competing liquidity providers that the specialist needs to beat. Given a particular $v(0)$, no limit sells are posted below the φ_{book} from Proposition 17. In addition, the crowd's lowest reservation price is $\varphi_{crowd} = v(0) + r$. Putting these together, define $\varphi_{min} =$

[34] In Proposition 15, cumulative depth is bounded, given a *fixed* upper valuation v. Here, however, the upper valuation is increasing with B so that progressively higher prices are included unless $\lim_{B \rightarrow \infty} v(B)$ is bounded.

$\min(\varphi_{book}, \varphi_{crowd})$. Thus, the specialist cannot ask more than φ_{min} and still undercut his competition on small trades $B \to 0$. Now, taking future $v(0)$'s as random variables — and allowing for possible dependence of F and $\frac{\partial v}{\partial B}$ on $v(0)$ — let $c_{min} = \min_{v(0)}(\varphi_{min}(v(0)) - v(0))$ and $c_{max} = \max_{v(0)}(\varphi_{min}(v(0)) - v(0))$.

Proposition 20. *For every Δ not in $[\frac{c_{min}}{2}, 2c_{max}]$, there is a tick size Δ_0 in $[\frac{c_{min}}{2}, 2c_{max}]$ that ex ante dominates Δ for small market buys B and sells S in the limit as $B, S \to 0$.*

5.2 Hybrid vs. pure limit-order markets.

The trading crowd may still undercut limit orders even in the absence of a specialist. However, since their undercutting is nonstrategic, the Bourse-style equilibrium with informative market orders is still simple.

Proposition 21. *In a pure limit order market with price discovery, the equilibrium limit book is*

$$S_j = \begin{cases} G_1 & \text{if } j = 1 \\ G_j - Q_{j-1} & \text{if } j > 1 \text{ and } e_{sj}(\max(0, \kappa_{j-1})) \geq 0 \\ 0 & \text{if } j > 1 \text{ and } e_{sj}(\max(0, \kappa_{j-1})) < 0, \end{cases} \tag{47}$$

and the corresponding thresholds are

$$\hat{S}_j = \begin{cases} G_1 & \text{if } j = 1 \\ G_j & \text{if } j > 1 \text{ and } e_{sj}(\max(0, \kappa_{j-1})) \geq 0 \\ \max(0, \kappa_{j-1}) & \text{if } j > 1 \text{ and } e_{sj}(\max(0, \kappa_{j-1})) < 0. \end{cases} \tag{48}$$

Aside from the substitution of G's (inverse conditional expectations) for H's (inverse distributions), the relationship between the pure and hybrid equilibria in Proposition 21 and 14 roughly parallels that between Proposition 9 and 3. There are, however, a few subtleties. Since $\max(0, \kappa_{j-1})$ in Equation (48) is less than $\hat{S}_j^{floor}$ in Equation (40), the pure limit order book may have positive depth at prices where the hybrid book is empty. The reason for the lower floor in the pure limit order market is that the crowd undercuts the book at p_j only by selling at p_{j-1}, whereas in the hybrid market, it can preempt the specialist simply by matching him at p_j. Nonetheless, "holes" are still possible in the pure limit book, so that the cumulative depth is

$$Q_j = \begin{cases} G_j & \text{if } S_j > 0 \\ Q_{j-1} < G_j & \text{otherwise}, \end{cases} \tag{49}$$

rather than always G_j [as the parallel with Equation (25) might suggest].

The answer to the market design question of who prefers pure versus hybrid markets is qualitatively the same as in Section 3. Small investors still like the specialist's superior marginal liquidity on small trades (i.e., if $\varphi_{min} > p_1$, he sells at p_{min-1} to sufficiently small B's). Similarly, institutional investors again benefit — given that the pure book is also bounded for the same reasons as in Proposition 18 — from the specialist's undercutting the trading crowd on very large market orders. Some mid-size investors, however, once again prefer the pure limit order market's greater inframarginal depth over the hybrid market's lower (marginal) clean up price.

6. Conclusion

This article has used a single-period microstructure model to study the economics of liquidity provision. Since limit orders are an important competing source of liquidity on the NYSE and other exchanges, it is important to understand how effective such orders are in disciplining the exercise of market power by specialists.

The main contributions of this article are first, a tractable framework in which to study these issues including analytic solutions in some special cases; second, an analysis of large and small investors' preferences over the tick size; and third, an explanation of which types of investors prefer NYSE-style hybrid specialist/limit order markets and which prefer Paris Bourse-style pure limit order markets. The analysis also suggests a number of directions for future work:

1. The Bertrand assumption for value traders could be relaxed. One interesting case to consider is a market with a single strategic value trader (e.g., for low-volume stocks). In this case, the specialist and the value trader both have bilateral (although not identical) market power. Another case is a market where (a) the total number of value traders is random, and (b) they leave immediately after posting their order. Now the market power of a value trader is random, since he submits his orders without knowing whether more limit orders will be submitted after him (in which case, his market power is ex post diminished) or not. The result is a random limit order book depending on the realized number of value traders. Comparisons of pure and hybrid markets would be particularly interesting when the book is random. It is precisely when the limit book is thin — and investors most need someone to undercut the book and the crowd — that the specialist's incentives to offer price improvement are weakest.

2. The effect of randomness in the limit order book on market liquidity could also be studied by adding optimizing traders with liquidity shocks who use limit orders, as well as market orders as in Kumar and Seppi (1994).

3. An important extension is endogenizing the distribution F and (even more ambitiously) the updating rule $v(B)$ by making market orders elastic with respect to price. Market design comparisons could then take feedback effects of the design choice on F and $v(B)$ into account. In particular, the updating rule $v(B)$ in a pure limit order market will differ from that of a hybrid market because of differences in informed trading [i.e., due to their differing non-informational price impacts $A(B)$], and also because of the differential incentives to acquire more precise information when information acquisition is endogenous.

4. On the empirical side, estimating the model and testing its consistency with patterns in specialist participation rates and profits appears promising.

Appendix A

Proof of Proposition 1. Fix volumes B and B' where $B > B'$ and suppose $p^*(B) < p^*(B')$. Let Q and Q' be the total limit orders executed, given these prices, where $Q \leq Q'$. To be optimal, $p^*(B)$ and $p^*(B')$ must each yield higher profits (given B or B') than any other price so that, in particular, we have

$$(B-Q)[p^*(B)-v] > (B-Q')[p^*(B')-v], \tag{50}$$

and

$$(B'-Q')[p^*(B')-v] > (B'-Q)[p^*(B)-v].$$

Dividing through the first equation by $(B-Q)$ and the second by $(B'-Q)$ and comparing gives a contradiction since $\frac{B-Q'}{B-Q} > \frac{B'-Q'}{B'-Q}$. Thus, $p^*(B) \geq p^*(B')$.

Proof of Proposition 2. In equilibrium, if $S_j > 0$, then $e_{sj} = 0$. However, if $S_{j+1} = 0$, then $\hat{S}_{j+1} = \hat{S}_j$, provided that $p_{j+1} < p_{max}$. However, since $p_{j+1} > p_j$, this means $e_{sj+1} > e_{sj} = 0$, which is inconsistent with equilibrium. Thus, we must have $S_{j+1} > 0$ sufficiently large so that $\hat{S}_{j+1} > \hat{S}_j$ and thus $F(B > \hat{S}_j) > F(B > \hat{S}_{j+1})$. Note that for limit orders at p_j to execute, it must be that $p^* \geq p_j$, but for a limit order at p_{j+1} *not* to execute requires (by price priority) that $p^* < p_{j+1}$. Thus, when $\hat{S}_j < B \leq \hat{S}_{j+1}$, it must be that $p^* = p_j$. Since the same logic holds for p_{j-1} and p_j, the relevant price comparisons for determining $\hat{S}_j$ are between p_{j-1} (rather than $p_k < p_{j-1}$) and p_j (rather than $p_k > p_j$) so that $\hat{S}_j = S_{j-1,j}$.

Proof of Proposition 3. From Equation (1), the specialist never sells for less than p_1 or more than p_{max-1} (hence $\hat{S}_{max} = \infty$). Given price, public, and time priority, the marginal limit sell at p_1 goes unexecuted only if $B < S_1$. Hence, S_1 is either zero [i.e., if $e_{s1}(0) \leq 0$ for $F(B > 0) = 1$ because $p_1 \leq v + c/\alpha$] or can be found by inverting $e_{s1}(S_1) = 0$. At $p_j > p_1$, the specialist executes the S_j limit orders if and only if $B > \hat{S}_j = S_{j-1,j}$ from Proposition 2. Hence, S_j again is either zero (i.e., if $p_j \leq v + c/\alpha$) or is found recursively from Equation (9) by setting $e_{sj} = 0$ and inverting to get

$$S_j = \gamma_j \left(H_j - \sum_{i=1}^{j-1} S_i \right). \tag{51}$$

The form of Q_{j-1} follows from recursive substitution of the preceding equation. The form of $\hat{S}_j$ in Equation (15) comes from substituting Equation (51) into Equation (10). Finally, if $p_1 = p_{max}$, then the crowd provides unlimited depth p_1 after the S_1 limit sells are filled.

Proof of Proposition 4. If $F' \geq_{FSD} F$ and/or if $c/\alpha > c'/\alpha'$, then by definition from Equation (11), $H_j' \geq H_j$, and thus $\hat{S}_j' \geq \hat{S}_j$ from Equation (15). Furthermore, since $H_j' \geq H_j$ for all j, a weighted average of the H' values is larger than the average of the H values using the same weights. Since the γ_j weights depend only on the prices p_j and v from Equation (14), we have that $Q_j' \geq Q_j$.

Proof of Proposition 5. From Equation (13), the cumulative depth Q_j below φ on any $\mathcal{P}$ is a weighted average of the intervening H's. Thus, since H_i is increasing in p_i from Equation (11), Q_j is bounded by H_j, the largest H included in the average.

Proof of Proposition 6. For each possible v, the cumulative depth at prices $p_i < v + r$ is bounded from Proposition 5 by $H(\frac{c/\alpha}{r})$ so that as $B \to \infty$, the relative contribution of limit orders goes to zero in $A(\Delta, B)$. Thus, liquidity for each v is primarily determined by $p_{max-1}(v)$ for large buys B and the analogous entry-deterring bid for large sells S. Note that if $\Delta < \frac{r}{2}$, then for each possible v, the corresponding intervals $(v, v + r)$ and $(v - r, v)$ each contain at least *two* prices on the Δ grid. Doubling the tick size to 2Δ eliminates every other price while ensuring (since $2\Delta < r$) that at least one price from the original grid remains. Thus, for each possible v, the entry-deterring buy and sell prices are either unchanged or possibly (one or both) improved by Δ. A fortiori, if the doubled tick 2Δ is itself less than $r/2$, then doubling yet again may further improve liquidity.

Continue doubling until the new tick Δ_B satisfies $\frac{r}{2} \leq \Delta_B < r$. Since this last grid has at *most* two prices (but at least one) from the original Δ grid in each interval, the buy and sell entry-deterring prices in any interval cannot be worse and may possibly be better. A similar procedure involving halving the tick size can be used to improve liquidity, given an excessively coarse grid with $\Delta > 2r$.

Proof of Proposition 7. The proof uses the same halving and doubling logic as in Proposition 6 except that c/α replaces r.

Proof of Proposition 8. Step 1: Fix any value v and price $\varphi > v + c/\alpha$. I first show that the limit of cumulative depth $Q(\varphi, \Delta)$, on a grid $\mathcal{P}_\Delta$ with tick size $\Delta > 0$ such that φ is on $\mathcal{P}_\Delta$, exists as $\Delta \to 0$. Recursive substitution of $p_{j,\Delta} = v + (j - 1 + q)\Delta$ in Equation (13) gives $Q(\varphi, \Delta) = \frac{qH(p_{1,\Delta})+\cdots+H(p_{n,\Delta})}{n-1+q}$. Taking the limit gives

$$\begin{aligned} &\lim_{\Delta\to 0} Q(\varphi, \Delta) \\ &\quad = \lim_{\Delta\to 0}\left\{[qH(p_{1,\Delta}) + \cdots + H(\varphi)]\frac{\Delta}{\varphi - v}\right\} \\ &\quad = \frac{\int_{p=v+c/\alpha}^{\varphi} H(p)dp}{\varphi - v}, \end{aligned} \tag{52}$$

where the first equality follows because $\frac{\Delta}{\varphi-v} = \frac{1}{n-1+q}$, and the existence of the Riemann integral follows from the fact that $H(p) = H(\frac{c/\alpha}{p-v})$ is a continuous (weakly) increasing function of p. The uniqueness of the integral means that the limit does not depend on our choice of the initial tick Δ.

The second part of step 1 is to show that refinements of any tick size $\Delta > 0$ such that $\varphi \in \mathcal{P}_\Delta$ decrease cumulative depth. To see why, consider any two price grids $\mathcal{P}_\Delta$ and $\mathcal{P}_{\Delta/2}$ with tick sizes $\Delta > 0$ and $\Delta/2$ such that φ is on $\mathcal{P}_\Delta$ (and thus $\mathcal{P}_{\Delta/2}$). Let n and n' be the number of prices between v and φ on the two grids and let q and q' be the fraction of a tick between v and the first prices on the two grids, respectively. Each price p_j on $\mathcal{P}_\Delta$ equals some price $p_{k,\Delta/2}$ on $\mathcal{P}_{\Delta/2}$. In particular, if halving Δ adds a new price between v and $v + q\Delta$, then $k = 2j$. If not, then $k = 2j - 1$. The cumulative depths $Q(\varphi, \Delta)$ and $Q(\varphi, \Delta/2)$ can be found recursively from Equation (13). Starting at p_1 on $\mathcal{P}_\Delta$, if $p_{1,\Delta/2} = p_1$, then $Q(p_1, \Delta/2) = H(p_1) = H(\frac{c/\alpha}{p_1-v}) = Q(p_1, \Delta)$. If instead $p_{2,\Delta/2} = p_1$, then we have $Q(p_1, \Delta/2) = \frac{q'H(p_1-\Delta/2)+H(p_1)}{q'+1} \leq H(p_1) = Q(p_1, \Delta)$.

Suppose now by induction that $Q(p_{j-1}, \Delta/2) \leq Q(p_{j-1}, \Delta)$ at price p_{j-1}. On the $\mathcal{P}_\Delta$ grid, the cumulative depth at p_j is $Q(p_j, \Delta) =$

$\gamma_j H(p_j)+(1-\gamma_j)Q(p_{j-1}, \Delta)$, while on $\mathcal{P}_{\Delta/2}$, it is $Q(p_j, \Delta/2) = \gamma_k H(p_j) + (1-\gamma_k)\gamma_{k-1}H(p_j - \Delta/2) + (1-\gamma_k)(1-\gamma_{k-1})Q(p_{j-1}, \Delta/2)$. Notice, however, that $\gamma_j = \frac{1}{j-1+q}$ and $\gamma_k + (1-\gamma_k)\gamma_{k-1} = \frac{2}{k-1+q'}$. If we substitute in $k = 2j$ and $q = \frac{1+q'}{2}$ when $p_{2,\Delta/2} = p_1$ [i.e., since in this case, $p_1 - v = q\Delta = (1+q')\frac{\Delta}{2} = p_{2,\Delta/2} - v$] or $k = 2j-1$ and $q = q'/2$ when $p_{1,\Delta/2} = p_1$ (i.e., since then $p_1 - v = q\Delta = q'\frac{\Delta}{2} = p_{1,\Delta/2} - v$), then we see that $\gamma_j = \gamma_k + (1-\gamma_k)\gamma_{k-1}$ and thus — since $H(p_j - \Delta/2) \leq H(p_j)$ — that $Q(p_j, \Delta/2) \leq Q(p_j, \Delta)$ at each p_j up through and including $p_j = \varphi$. Thus, refinements of any grid $\mathcal{P}_\Delta$ such that $\varphi \in \mathcal{P}_\Delta$ reduce cumulative depth.

Step 2: The specialist's profit $\pi(p^*)$ is maximized in the limit as $\Delta \to 0$ for two reasons. First, shrinking Δ increases the number of possible clean up prices in the set $[v, v + r] \cap \mathcal{P}_\Delta$ from which he chooses. Second, for any order B and fixed price p, shrinking the tick size Δ minimizes $Q(p)$ and thus maximizes the specialist's share $B - Q(p)$ in the limit given $p^* = p$.

Proof of Proposition 9. This follows from recursively setting $e_{sj} = 0$ at prices p_{min} and above, but now taking $\hat{S}_j = \sum_{i=1}^{j} S_i$ in Equation (9).

Proof of Proposition 10. Step 1: From Equation (25), adding a new price p' between p_{j-1} and p_j does not affect the cumulative depth Q_j but does shift $H(p') - H(p_{j-1})$ in depth down from p_j to p', provided $p' > v + c/\alpha$.

Step 2: Given $\min(c/\alpha, r) > \Delta > 0$, the specialist undercuts *hybrid* limit sells at p_{min} by selling up to $H(p_{min}) > 0$ shares at p_{min-1}, which is less than the lowest *pure* limit sell at $v + c/\alpha$. Similarly, on large orders, the specialist sells $B - Q_{max-1}$ at p_{max-1}, while in the pure market, $B - H(p_{max})$ is bought from the crowd at p_{max}. If B is large enough, the lower hybrid clean up price results in a lower average ask premium A, given the bounded depth available in the book in either market.

Proof of Proposition 11. Suppose not, so that $Q_j > \beta_j$ for some p_j. If $S_j > 0$, then in equilibrium, $e_{sj} = 0$, which from Equation (30), implies a nonzero probability of profitable execution for the S_j limit sells, or $F(Z_j^-) > 0$. Since limit sells at p_j are profitably executed only when $B < \beta_j$ [i.e., when $v(B) < p_j$], this implies that for such B's, the specialist must *buy* at p_j (i.e., given the prohibition on front running) to make up the shortfall $Q_j - B$. This is a contradiction, however, since the specialist would be unwilling to buy at a loss. Similarly, if $S_j = 0$, consider the highest price $p_k < p_j$ such that $S_k > 0$. Since $Q_k = Q_j$, we have $Q_k > \beta_j$. Again, for a limit sell at p_k to be profitably executed, we need $B < \beta_k < \beta_j$, which is again a contradiction since it requires the specialist to buy the shortfall $Q_k - B$ at a loss. Lastly, even when $B < Q_j \leq \beta_j$, the specialist still does not *buy* from the book at p_j since doing so again leads to a loss.

Proof of Proposition 12. Suppose not, so that for some $B > B'$ we had $p^*(B') = p_j > p_1$ and $p^*(B) = p_k < p_j$. Consider first large buys $B > \beta_{j-1}$. In this case, neither the specialist nor the crowd are willing to sell at a loss at $p_k \leq p_{j-1}$. Proposition 11 ensures that the same is true of the value traders since $Q_k \leq \beta_{j-1}$. Thus, if $B > \beta_{j-1}$, then $p^*(B) \geq p^*(B')$. Suppose instead $B' < B < \beta_{j-1}$. Once again, the crowd will not sell to B at $p_k < p_j$ because $p^*(B') = p_j$ implies that $B' > \kappa_{j-1}$ and thus that $B > B' > \kappa_{j-1} \geq \kappa_k$. The value traders also cannot be the marginal liquidity provider at p_k. To see this, note that $p^*(B') = p_j$ implies $Q_{j-1} \leq B'$ (i.e., otherwise price priority dictates a clean up price p_{j-1} or lower). Thus, if $Q_{j-1} \leq B' < B$, the cumulative depth in the book is insufficient to be the marginal source of liquidity for B at $p_k \leq p_{j-1}$. Lastly, the single crossing property implies that if $\pi_k(B) \geq \pi_j(B)$, given $p_k < p_j$, then $\pi_k(B'') \geq \pi_j(B'')$ for all $B'' < B$ which contradicts $\pi_j(B') > \pi_k(B')$ [i.e., that $p^*(B') = p_j$].

Proof of Proposition 13. Neither the crowd nor the specialist can profitably undercut limit sells at p_1, so full execution depends simply on having enough buy orders to execute the whole queue. Thus, $\hat{S}_1 = S_1$. This logic generalizes to prices $p_j > p_1$ where $e_{sj}(\beta_{j-1}) \geq 0$. At such prices, value traders will submit limit sells until $\hat{S}_j \geq \beta_{j-1}$. Once again, no one can profitably undercut limit sells at or below p_j, given $B \geq \beta_{j-1}$, so full execution depends simply on there being enough buy orders. Thus, $\hat{S}_j = Q_j$ when $e_{sj}(\beta_{j-1}) \geq 0$.

Now consider prices $p_j > p_1$, where $e_{sj}(\beta_{j-1}) < 0$. There are four cases.

Case 1. If $\beta_{j-1} \leq \kappa_j$, then the specialist will never set $p^*(B) = p_j$ when $\max(Q_{j-1}, \kappa_{j-1}) < B < \beta_{j-1}$ since p_{j-1} profitably undercuts the crowd's p_j. However, once $\beta_{j-1} \leq B$, it is no longer profitable to sell at p_{j-1}. Thus, $\hat{S}_j = \beta_{j-1}$ when $e_{sj}(\beta_{j-1}) < 0$ and $\beta_{j-1} \leq \kappa_j$.

In each of the last three cases, $\kappa_j < \beta_{j-1}$. This implies that $\hat{S}_j \geq \kappa_j$ since the specialist can always use p_{j-1} to profitably undercut the crowd's p_j.

Case 2. If $e_{sj}(\kappa_j) \leq 0$ and $0 < \kappa_j < \beta_{j-1}$, then $\hat{S}_j = \kappa_j$. If $e_{sj}(\kappa_j) = 0$, this is by definition. Consider then $e_{sj}(\kappa_j) < 0$. Notice that if $\hat{S}_j > \kappa_j$, then there are B's such that $\hat{S}_j > B > \kappa_j$, where $p^*(B) = \max(p^* \mid B < \hat{S}_j) = p_k < p_j$. Since p_j already undercuts the crowd when $B > \kappa_j$, and since $S_j = 0$ [i.e., since $e_{sj}(\hat{S}_j) \leq e_{sj}(\kappa_j) < 0$], pricing at p_k just below $\hat{S}_j$ must let the specialist undercut a limit sell $S_i > 0$ at some p_i, where $p_k < p_i \leq p_{j-1}$. Price priority implies then that $\hat{S}_i = \hat{S}_j$ and thus that $0 = e_{si}(\hat{S}_j) < e_{sj}(\hat{S}_j)$ (i.e., since $S_i > 0$), which, however, contradicts $e_{sj}(\hat{S}_j) \leq e_{sj}(\kappa_j) < 0$. Thus, $\hat{S}_j = \kappa_j$.

Case 3. If instead $e_{sj}(\kappa_j) > 0 > e_{sj}(\beta_{j-1})$ and $0 < \kappa_j < \beta_{j-1}$, then $\hat{S}_j > \kappa_j$. In addition, $\hat{S}_j = S_{j-1,j}$. To see why, notice first that if $\hat{S}_j < \hat{S}_{j+1}$, then $\hat{S}_j$ is determined by a crossing point $S_{k,j}$ with some $p_k \leq p_{j-1}$. If $p_k < p_{j-1}$, then it must be that the specialist is using p_k to undercut at least one $S_i > 0$ at some price $p_k < p_i \leq p_{j-1}$. However, since this implies that $\hat{S}_i = \hat{S}_j$, we have $0 = e_{si}(\hat{S}_j) < e_{sj}(\hat{S}_j)$, which is a contradiction. Thus, if $\hat{S}_j < \hat{S}_{j+1}$, then $p_k = p_{j-1}$, so that $\hat{S}_j = S_{j-1,j}$. If instead $\hat{S}_j = \hat{S}_{j+1}$, then for B's such that $\kappa_j < B < \hat{S}_j$, the specialist must be setting $p^*(B) = p_m < p_j$ to undercut at least one $S_i > 0$ at some $p_m < p_i \leq p_j$. However, $\hat{S}_i = \hat{S}_{j+1}$ (from price priority) implies then that $0 = e_{si}(\hat{S}_{j+1}) < e_{sj+1}(\hat{S}_{j+1})$, which is a contradiction. Thus, we must have $\hat{S}_j < \hat{S}_{j+1}$, and thus (from above) $\hat{S}_j = S_{j-1,j}$.

Case 4. If $\kappa_j \leq 0$, then, as in Case 3 (above), the crowd does not constrain the specialist's use of price p_j, and the same argument shows that $\hat{S}_j = S_{j-1,j}$.

Proof of Proposition 14. The proof for S_1 and $\hat{S}_1$ is in the body of the article, so consider prices $p_j > p_1$. Given the threshold floor $\hat{S}_j^{floor}$, if $e_{sj}(\hat{S}_j^{floor}) < 0$, then $S_j = 0$ and $\hat{S}_j = \hat{S}_j^{floor}$. If instead $e_{sj}(\hat{S}_j^{floor}) \geq 0$, then $S_j > 0$, and we can invert the value traders' zero-profit condition $e_{sj}(\hat{S}_j) = 0$ to get $\hat{S}_j = G_j$. Substituting this into Proposition 13 and solving for S_j gives Equation (43).

Proof of Proposition 15. Since the cumulative depth $\sum_{p_i \leq \varphi} S_i$ is a (positive) weighted average of G's corresponding to prices (with positive depth) at or below φ, it is bounded by G_{φ_v}, the largest possible positive G for any grid $\mathcal{P}$.

Proof of Proposition 16. First, note that the e_{sj} are weakly decreasing in the valuation $v(B)$ [due to $\mu(x, y)$ and $\mu(x)$] and in c/α. Thus, the inverse g_j and G_j also inherit this property. Thus, the execution thresholds in Proposition 14 are decreasing in both $v(B)$ and c/α since $\hat{S}_j$ is either G_j or S_j^{floor}, which is also (weakly) decreasing in $v(B)$ via β_{j-1} and κ_j and independent of c/α. Finally, since cumulative depth is just a weighted average of the G_j (where the weight put on the G_j's for higher p_j's is increasing in G_j and thus decreasing in c/α), we have that Q_j is weakly decreasing in c/α.

Proof of Proposition 17. Given limits $\lim_{p_1 \downarrow v(0)} F(\beta_1) > 0$ and $\lim_{p_1 \downarrow v(0)} \mu(\beta_1) > p_1$, we have $\lim_{p_1 \to v(0)} e_{s1}(S_1) < 0$ for all $S_1 > 0$. Thus, no limit orders are posted at p_1 if it is below some $\varphi_{book} > v(0) + c/\alpha$. Furthermore, since prices above p_1 on $\mathcal{P}$ can be undercut, the depth at prices $p < \varphi_{book}$ when $p > p_1$ cannot be greater than when $p = p_1$. Thus, no limit orders are posted at prices below φ_{book}.

Proof of Proposition 18. In Equation (31), set the expected profit per share (given execution) $p_j - \mu(\hat{S}_j, \beta_j)$ equal to the maximum possible profit $p_j - v(B) = \max(\delta, r)$, substitute 0 in for the expected conditional loss (given unprofitable execution) $p_j - \mu(\beta_j)$, and replace the probability of profitable execution, $F(\hat{S}_j, \beta_j)$, with the probability that B exceeds the cumulative depth, $F(Q_j)$. This gives $\alpha F(Q_j)\max(\delta, r) - c \geq e_{sj}(\hat{S}_j)$, which implies, since the upper tail probability $F(Q_j)$ is decreasing in Q_j, that Q_∞ is bounded. Otherwise, the execution probability at sufficiently high prices is too low, given the up-front cost c.

Proof of Proposition 19. If Q_∞ is bounded for each $v(0)$, then the cumulative depth $Q^* = \sum_{p_i \leq p^*(B)} S_i$ is also bounded as $B \to \infty$. Hence, the fraction $\frac{Q^*}{B}$ of B crossed with the book is negligible in the limit, so that for each $v(0)$ the average premium $A(B, \Delta)$ for large B is again determined by p^*. At this point, the same halving and doubling arguments as in Proposition 6 relative to generic price ranges $[v(B), v(B) + r]$ and $[v(S) - r, v(S)]$ complete the proof.

Proof of Proposition 20. The same halving and doubling arguments as in Proposition 7 prove this.

Proof of Proposition 21. This follows from the fact that, without the specialist, the threshold for full execution of limit orders at p_j is

$$\hat{S}_j = \begin{cases} Q_j & \text{if } e_{sj}(\max(0, \kappa_{j-1})) \geq 0 \\ \max(0, \kappa_{j-1}) & \text{if } e_{sj}(\max(0, \kappa_{j-1})) < 0, \end{cases} \tag{53}$$

and from, recursively, (1) inverting the zero profit condition $e_{sj}(\hat{S}_j) = 0$

and solving for S_j from Equation (53) when $e_{sj}(\max(0, \kappa_{j-1})) \geq 0$ or (2) setting $S_j = 0$ when $e_{sj}(\max(0, \kappa_{j-1})) < 0$.

Appendix B

The difficulty with discontinuities in F is that a mass point at a volume b' can lead to situations in which $\lim_{\epsilon \to 0} e_{sj}(\hat{S}_j(S_j - \epsilon)) > 0$ while $e_{sj}(\hat{S}_j(S_j)) < 0$ if, given $S_j > 0$, the threshold $\hat{S}_j(S_j) = b'$. This is because the probability of execution — given our convention that $p^* \geq p_j$ only when $B > \hat{S}_j$ — drops discontinuously at b'. As a result, the equilibrium limit order book cannot always be characterized by a zero-profit condition $e_{sj}(\hat{S}_j(S_j)) = 0$.

If (1) the discontinuities in F are exogenous and (2) market orders B are uninformative, then we simply change our specialist pricing convention (since he is indifferent between the higher and lower prices at $\hat{S}_j$) to one in which he simply charges the *higher* price at $\hat{S}_j$. In addition, we must relax the definition of equilibrium to require just that any depths $S_j > 0$ satisfy $e_{sj}(\hat{S}_j(S_j)) \geq 0$ and $\lim_{\epsilon \to 0} e_{sj}(\hat{S}_j(S_j + \epsilon)) \leq 0$. In other words, the value traders do not necessarily drive expected profits on *submitted* limit orders S_j to zero, but they do — by their competitive behavior — ensure that no expected profits remain for any *additional* limit sells $\epsilon > 0$ beyond S_j. The only change in the analysis this causes is that it is less convenient to calculate H's corresponding to non-positive limiting expected profits than to zero expected profits.

This modification also generalizes to exogenous informative market orders (as in Section 4) as long as *none* of the discontinuities in F occur at any of the κ_j's. This is because — whether or not there are mass points — the specialist strictly prefers undercutting p_j if $\hat{S}_j = \kappa_j < \beta_{j-1}$.

Once F is endogenous, however, we need $p^*(\hat{S}_j) < p_j$ at $p_j > p_1$ (i.e., when the specialist trades) so that the active trader can shave his order to $B = \hat{S}_j$ rather than to the (nonexistent) top of an open set $B < \hat{S}_j$.

References

Amihud, Y., and H. Mendelson, 1986, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, 17, 223–49.

Angel, J., 1992, "Limit versus Market Orders," working paper, Georgetown University.

Biais, B., P. Hillion, and C. Spatt, 1995, "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *Journal of Finance*, 50, 1655–89.

Brennan, M., and A. Subrahmanyam, 1994, "Market Microstructure and Asset Pricing: On the Compensation for Adverse Selection in Stock Returns," working paper, UCLA.

Brown, D., and C. Holden, 1995, "The Design of Limit Orders under a Hybrid Mechanism with Endogenous Depth," working paper, Indiana University.

Byrne, R., 1993, "Discrete Prices in Securities Markets," working paper, Carnegie Mellon University.

Chakravarty, S., and C. Holden, 1995, "An Integrated Model of Market and Limit Orders," *Journal of Financial Intermediation*, 4, 213–41.

Christie, W., and P. Schultz, 1994, "Why Do Nasdaq Market Makers Avoid Odd-Eighth Quotes?," *Journal of Finance*, 49, 1813–40.

Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb, 1981, "Transactions Costs, Order Placement Strategy, and the Existence of the Bid-Ask Spread," *Journal of Political Economy*, 89, 287–305.

Foucault, T., 1993, "Price Formation in a Dynamic Limit Order Market," working paper, Groupe HEC.

Frino, A., and M. McCorry, 1995, "Why are Spreads Tighter on the Australian Stock Exchange than on the NYSE? An Electronic Limit Order Book Versus the Specialist Structure," working paper, University of Sydney.

Glosten, L., 1989, "Insider Trading, Liquidity and the Role of the Monopoly Specialist," *Journal of Business*, 62, 211–35.

Glosten, L., 1994, "Is the Electronic Open Limit Order Book Inevitable?," *Journal of Finance*, 49, 1127–61.

Glosten, L., and P. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogenously Informed Traders," *Journal of Financial Economics*, 21, 123-44.

Grossman, S., and M. Miller, 1988, "Liquidity and Market Structure," *Journal of Finance*, 43, 617–63.

Greene, J., 1996, "The Impact of Limit Order Executions on Trading Costs in NYSE Stocks: An Empirical Examination," working paper, Indiana University.

Harris, L., 1994a, "Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes," *Review of Financial Studies*, 7, 149–78.

Harris, L., 1994b, "Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems," working paper, USC.

Harris, L., and J. Hasbrouck, 1996, "Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy," forthcoming in *Journal of Financial and Quantitative Analysis.*

Hasbrouck, J., 1991, "The Summary Informativeness of Stock Trades: An Econometric Analysis," *Review of Financial Studies*, 4, 571–95.

Hasbrouck, J., 1993, "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement," *Review of Financial Studies*, 6, 191–212.

Hollifield, B., R. Miller, and P. Sandas, 1996, "An Empirical Analysis of an Electronic Limit Order Market," working paper, Carnegie Mellon University.

Kavajecz, K., 1996, "A Specialist's Quoted Depth and the Limit Order Book," working paper, Northwestern University.

Keim, D., and A. Madhavan, 1996, "The Upstairs Market for Large Block Transactions: Analysis and Measurement of Price Effects," *Review of Financial Studies*, 9, 1–36.

Kumar, P., and D. Seppi, 1994, "Limit and Market Orders with Optimizing Traders," working paper, Carnegie Mellon University.

Kyle, A., 1984, "A Theory of Futures Market Manipulations," in R. W. Anderson (ed.), *The Industrial Organization of Futures Markets*, Lexington Books, Lexington, Massachusetts.

Kyle, A., 1985, "Continuous Auctions and Insider Trading," *Econometrica* 53, 1315-35.

Kyle, A., 1992, "Market Failures and the Regulation of Financial Markets," working paper, Duke University.

Leach, J. C., and A. Madhavan, 1993, "Price Experimentation and Security Market Structure," *Review of Financial Studies*, 6, 375–404.

Madhavan, A., and M. Cheng, 1997, "In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets," *Review of Financial Studies*, 10, 175–203.

Madhavan, A., and G. Sofianos, 1994, "Auction and Dealer Markets: An Empirical Analysis of NYSE Specialist Trading," Working Paper No. 94–01, New York Stock Exchange.

Parlour, C., 1994, "Price Dynamics in a Limit Order Market," working paper, Carnegie Mellon University.

Rock, K., 1996, "The Specialist's Order Book and Price Anomalies," forthcoming in *Review of Financial Studies.*

Shapiro, J., 1993, "U.S. Equity Markets: A View of Recent Competitive Developments," Working Paper No. 93-02, New York Stock Exchange.

Sofianos, G., 1995, "Specialist Gross Trading Revenues at the New York Stock Exchange," Working Paper No. 95–01, New York Stock Exchange.

Spiegel, M., and A. Subrahmanyam, 1995, "On Intraday Risk Premia," *Journal of Finance*, 50, 319–39.

Wall Street Journal, 20 Oct. 1994, "U.S. Examines Alleged Price-Fixing on Nasdaq," C1, C18.