A statistical model of a limit order market

Hugh Luckock
University of Sydney

September 2001

Abstract

A model of limit order market is presented, and some of its statistical properties are deduced. Given the underlying supply and demand functions, the analysis yields the stationary probability distributions for the best ask and bid prices in the order book, and for the prices of actual trades. It also predicts the existence of a clearly-defined price window within which all trades take place, thus providing a quantitative explanation for the phenomena of support and resistance. If the bid-ask spread is narrow and the elasticities of supply and demand are not strongly price-dependent, the distributions of quotes and trades within this window are completely determined by a single parameter corresponding to the proportion of market participants who own at least one unit of the asset being traded.

I am grateful to Peter Buchen for comments on a preliminary version of the manuscript, and have also benefited from discussions with Eckhard Platen and Bill Spence.

1 Introduction

Over the last two decades, most securities exchanges have introduced automated trade execution systems that use the continuous double auction (CDA) mechanism to match buyers with sellers during trading hours. [See Domowitz (1993) for a review of this trend.] This mechanism permits traders to submit or to accept buy orders (bids) and sell orders (asks) at any time while the exchange is open, and is characterised by the existence of an electronic order book in which unexecuted or partially executed orders are stored and displayed while awaiting execution.

The CDA mechanism provides a natural multilateral generalisation of the familiar process of haggling between a single buyer and seller, and appears to be well-suited to the needs of modern financial markets. It has been found in laboratory experiments to give very rapid convergence to a competitive equilibrium [as discussed by Smith (1962) and Smith, Williams, Bratton and Vanoni (1982)], and also to yield extremely efficient allocations. [A number of relevant studies are surveyed by Friedman (1993).]

Countless versions of the CDA mechanism can be envisaged, each characterised by the types of orders that may be submitted and the way in which these are handled. In practice most automated exchanges operate primarily as limit order markets, and accept orders of the following basic types:

- Limit orders - orders to buy a specified quantity at a price not exceeding some specified maximum (the bid price), or to sell a specified quantity at a price not less than some specified minimum (the ask price). Unless it can be executed against a pre-existing order, a new limit order joins the queue of asks or bids in the order book and remains there until it is either cancelled or else (after reaching the top of its queue) executed against a subsequent order. The queues are arranged primarily according to price; a high-priced bid takes precedence over a low-priced bid, while a low-priced ask takes precedence over a high-priced ask.

- Market orders - orders to buy or sell a specified quantity at the best price currently available. A market order is executed immediately and as fully as possible. Any unexecuted part may then be converted to a limit order at the same price (as on the Paris Bourse), or else executed at the next best available price (as on the NYSE). Some exchanges (such as Chicago's CME) do not permit market orders, but instead allow traders to "hit the bid" or "take the ask" - i.e. to execute trades against selected bids or asks in the order book. The selected orders are invariably the most competitively priced, and so in most situations these actions have the same effect as market orders.

Although some exchanges utilize hybrid systems, an increasing number (including the Paris, Tokyo, Toronto and Sydney stock exchanges) may be regarded as pure limit order markets. Glosten (1994) argues that exchanges operating in this manner are not only good providers of liquidity in extreme situations, but are also immune to competition from other exchanges.

In a limit order market, transaction prices are determined by the interaction of incoming orders with the limit order book. An understanding of this interaction is therefore required to properly model price formation and the stochastic properties of price fluctuations. This has motivated several empirical studies of real limit order markets, including those by Biais, Hillion and Spatt (1995) and Hamao and Hasbrouck (1995).

A tractable model of the order book would be of considerable value, but the development of such a model has been hindered by the inherent complexity of limit order markets. Recent investigations [such as those by Chakravarty & Holden (1995) and Parlour (1998)] have generally focussed on the strategies available to traders and market makers, and provide limited insight into the statistical properties of the order book.

Instead of attempting to anticipate how traders will behave, the approach taken here is to start by assuming that their combined effect is to generate flows of buy and sell orders with known price distributions. The next step is to determine the resulting distributions of bids and asks in the order book, and the distribution of transaction prices. This approach is similar to that proposed by Garman (1976), although he did not attempt to develop a solvable model except for the simple case in which all bids and asks are made at a single price. It is also related to the approach of Domowitz and Wang (1994), although their results are invalid due to an error in the proof of their main theorem. (This is discussed in Appendix B).

The model considered here is relatively simple, but nonetheless captures the important features of real limit order markets. For example, its structure ensures that competitively priced orders are executed more promptly on average than less competitive orders. Its properties are fully described in Section 2.

For exogenous supply and demand functions, the analysis of Section 3 yields the stationary distributions of the best ask and bid prices in the order book, and the frequency of transactions at any given price. These distributions are generally restricted to a clearly defined price window (referred to as the "competitive window"), between limits that are determined by the form of the supply and demand functions. By way of example, Section 4 deals with a market in which supply and demand have price-independent elasticities. Section 5 considers an application to markets with narrow bid-ask spreads, and Section 6 discusses some of the implications and limitations of the model. Appendices A and B contain the proof of the Theorem presented in Section 3, a detailed discussion of the erroneous theorem of Domowitz and Wang (1994). Appendix C contains an analysis of the problem of determining the supply and demand functions from the order book statistics.

2 The Model

During continuous trading, a limit order at a sufficiently competitive price will be immediately executed against the best available bid or ask just as though it were a market order. Consequently a model of a continuous limit order market does not need to explicitly incorporate markets orders, as these may be represented by very high-priced limit buy orders or very low-priced limit sell orders. This is the approach taken here.

We consider a model limit order market with the following properties:

L1: There are large numbers of potential buyers and sellers acting independently of one another, with each individual only occasionally submitting an order to the exchange. (In particular, there is no single specialist responsible for making a market.) This allows us to regard each order as originating from a different source, and hence unrelated to any other order. The arrival of orders of any specified type, or within any specified price range, will then be a Poisson process.

L2: All orders are for a single unit. Hence, a buyer needs only to specify the maximum price she will pay (the "bid price"), while a seller needs only to specify the minimum price she will accept (the "ask price"). This eliminates the possibility of partial execution, and hence the need for rules governing the handling of partially executed orders.

L3: There is a continuous range of possible prices.

L4: The underlying supply and demand functions are time-independent.

L5: Market participants prepare and submit their orders without making use of detailed information about the current state of the order book. Opportunistic traders are thus unable to take advantage of temporary anomalies in the order book, although they can use their knowledge of its statistical properties. (In real markets this is often the situation faced by retail investors, whose orders may not be relayed promptly enough to allow them to take advantage of order book anomalies.)

L6: Once submitted to the exchange, orders are never cancelled. Though inapplicable to real markets, this assumption seems reasonable here since L4 and L5 preclude the possibility that a change in market conditions may motivate a trader to cancel an order. (In any case, the effect of a non-zero cancellation rate could presumably be approximated by reducing the arrival rate of orders.)

Properties L1 and L2 considerably simplify the problem, and appear indispensible to the following analysis. On the other hand, L3, L4, L5 and L6 are probably inessential, and can hopefully be relaxed in future work. [Note that L1, L2, L4 and L5 are similar to assumptions made in the seminal paper by Garman (1976).]

At a given time t, the order book will contain a queue of unexecuted asks at prices $\alpha_1(t), \alpha_2(t), \ldots$ and another queue of unexecuted bids at prices $\beta_1(t), \beta_2(t), \ldots$ waiting to be matched with incoming orders. These prices are indexed so that α_1 denotes the lowest (i.e. best) ask price, α_2 the second lowest, and so on. Similarly, β_1 denotes the highest (i.e. best) bid price, β_2 the second highest, etc. The prices will therefore satisfy the inequalities

$$\ldots \beta_3(t) \ \leq \beta_2(t) \ \leq \ \beta_1(t) \ < \ \alpha_1(t) \ \leq \ \alpha_2(t) \ \leq \ \alpha_3(t) \ldots . \tag{1}$$

We are primarily interested in the best ask price $\alpha_1(t)$ and best bid price $\beta_1(t)$, but must keep track of all the other unexecuted orders as any of these may eventually rise to the tops of their respective queues.

Whenever a new order is received by the exchange, the contents of the order book are modified according to the following dynamical rules:

R1: If the new order is an ask at the price α, then

- If $\alpha \leq \beta_1$, the new ask is matched with the best unexecuted bid, and the two are executed at the bid price β_1.
- If $\alpha > \beta_1$, no match is possible, and the new ask joins the queue of unexecuted asks in the order book.

R2: If the new order is a bid at the price β then

- If $\beta \geq \alpha_1$, the new bid is matched with the best unexecuted ask, and the two are executed at the ask price α_1.

- If $\beta < \alpha_1$, no match is possible, and the new bid joins the queue of unexecuted bids in the order book.

Given a sequence of orders, these rules completely determine the evolution of the market. Here, however, the arrivals of new bids and asks are regarded as stochastic processes. Assumptions L1 and L4 imply that the arrival of new asks at prices not exceeding x will be a Poisson process with parameter $\lambda_A(x)$, where λ_A is an increasing function. Similarly, the arrival of new bids at prices of at least x will be a Poisson process with parameter $\lambda_B(x)$, where λ_B is a decreasing function. For convenience we suppose that both of these functions are everywhere differentiable and strictly monotonic, so that there is a non-zero probability of receiving a bid or offer within any given price range. In fact $\lambda_A(x)$ and $\lambda_B(x)$ may be interpreted respectively as the supply and demand functions.

Our goal is to determine the statistical properties of an ensemble of order books governed by the dynamical rules R1 and R2, given that the arrivals of asks and bids are Poisson processes as described above. A complete theory would provide a full statistical description of the order book dynamics and the price process. However, the construction of such a theory proves to be a very difficult problem which is beyond the scope of this paper. We will content ourselves with a description which focuses on just a handful of relevant quantities.

A buyer is primarily interested in whether a bid at a particular price x can be matched immediately with a lower-priced ask. Let us therefore define $A(t,x)$ to be the probability that, at time t, the best ask in the order book is executable at the price x;

$$A(t,x) \equiv P[\,\alpha_1(t) \leq x\,]. \tag{2}$$

This is just the cumulative distribution function for the best ask price $\alpha_1(t)$. It will be an increasing function of x, and in general it will also be continuous from the right with respect to x. (It is easy to confirm that right-continuity is unaffected by the arrival of future bids or asks.) Its x-derivative $A_x(t,x)$, if it exists, is the density function for $\alpha_1(t)$. In what follows, we assume that the order book always contains some unexecuted asks at prices above zero and hence

$$A(t,0) = 0, \qquad A(t,\infty) = 1. \tag{3}$$

Similarly, let $B(t,x)$ denote the probability that, at time t, the order book contains at least one bid executable at the price x;

$$B(t,x) = P[\beta_1(t) \geq x]. \tag{4}$$

This is known as the *survivor function* for the highest unexecuted bid price $\beta_1(t)$. It is a decreasing function of x, and is generally continuous from the left. If $B(t,x)$ is differentiable with respect to x, then $-B_x(t,x)$ is the density function for $\beta_1(t)$. We assume that the order book always contains unexecuted bids, and so

$$B(t,0) = 1, \qquad B(t,\infty) = 0. \tag{5}$$

It is useful now to define

$$x_{min}(t) \equiv \inf\{x : B(t,x) < 1\} \tag{6}$$
$$x_{max}(t) \equiv \sup\{x : A(t,x) < 1\}. \tag{7}$$

As we will see, the prices $x_{min}(t)$ and $x_{max}(t)$ have an important interpretation.

Proposition 1. *At any time t, and for any $x > 0$, the following are true:*

$$0 \ \leq\ A(t,x) + B(t,x) \ \leq\ 1 \tag{8}$$

$$x > x_{min}(t) \ \Rightarrow\ 0 \leq B(t,x) < 1 \tag{9}$$
$$x < x_{max}(t) \ \Rightarrow\ 0 \leq A(t,x) < 1 \tag{10}$$
$$x \leq x_{min}(t) \ \Rightarrow\ A(t,x) = 0, \quad B(t,x) = 1 \tag{11}$$
$$x \geq x_{max}(t) \ \Rightarrow\ A(t,x) = 1, \quad B(t,x) = 0 \tag{12}$$

Proof. We suppress the t-dependence of the variables, as this is not relevant. To prove (8), we note that $A(x) + B(x)$ is the probability that either $\alpha_1 \leq x$ or $\beta_1 \geq x$ (since these possibilities are mutually exclusive). For (9,10) note that $A(x), B(x)$ are probabilities and hence non-negative. Also, if $y > x_{min} = \inf\{x : B(x) < 1\}$ then $B(y) < 1$ (since $B(x)$ is decreasing). Similarly, if $y < x_{max} = \sup\{x : A(x) < 1\}$ then $A(y) < 1$ (since $A(x)$ is increasing). For (11), observe that if $y \leq x_{min} = \inf\{x : B(x) < 1\}$ then $B(y) = 1$ (since $B(x)$ is continuous from the left) and so (8) implies $A(y) = 0$. Similarly, if $y \geq x_{max} = \sup\{x : A(x) < 1\}$ then $A(y) = 1$ (since $A(x)$ is continuous from the right) and so (8) gives $B(y) = 0$, proving (12). $\square$

It follows from (11) and (12) that the highest bid price and lowest ask price at a given time t are certain to be found between $x_{min}(t)$ and $x_{max}(t)$:

$$\alpha_1(t), \beta_1(t) \in [x_{min}(t), x_{max}(t)]. \tag{13}$$

We will refer to this interval as the *competitive window*, since unexecuted orders at prices outside this interval are inferior to those inside and are therefore ineligible to be matched with new orders. Conversely, only those unexecuted orders at prices within the competitive window have any chance of being executed against incoming orders.

In principle, if the initial state of the order book is known, it should be possible to use the dynamical rules R1 and R2 to determine the functions $A(t,x)$ and $B(t,x)$ and hence the prices $x_{min}(t)$ and $x_{max}(t)$ at all subsequent times t. Postponing until the next section the question of how this might be done in practice, we note here that this could provide probabilistic answers to many questions about the state of the order book. For example, it could be used to calculate a variety of useful expectation values.

Proposition 2. *At any time t, the following are true:*

(i) The lowest ask price in the order book has expectation

$$E[\alpha_1(t)] = x_{max}(t) - \int_{x_{min}(t)}^{x_{max}(t)} A(t,x)dx. \tag{14}$$

(ii) The highest bid price in the order book has expectation

$$E[\beta_1(t)] = x_{min}(t) + \int_{x_{min}(t)}^{x_{max}(t)} B(t,x)dx. \tag{15}$$

(iii) If $B(t,x)$ is continuous with respect to x at $x = y$, then the expected frequency of trades at prices not exceeding y is

$$\lambda_T(t,y) = \int_{x_{min}(t)}^{y} [\lambda_B(x)dA(t,x) - \lambda_A(x)dB(t,x)]. \tag{16}$$

(iv) The expected overall frequency of trades is

$$\lambda_T(t,x_{max}(t)) = \int_{x_{min}(t)}^{x_{max}(t)} [\lambda_B(x)dA(t,x) - \lambda_A(x)dB(t,x)]. \tag{17}$$

Proof. (i) and (ii) are established by writing the expectations as Riemann-Stieltjes integrals and then integrating by parts. To prove (iii), we note that there are two types of trades; those in which a pre-existing ask is matched with a new bid, and those in which a pre-existing bid is matched with a new ask. The expected frequency of type 1 trades in the price interval

$(x, x + dx]$ is $\lambda_B(x')dA(t, x)$ for some $x' \in (x, x + dx]$, where $dA(t, x) = A(t, x+dx) - A(t, x)$. Hence the expected frequency of such trades at prices not exceeding y is $\int_{x_{min}(t)}^{y} \lambda_B(x)dA(t, x)$. Similarly, the expected frequency of type 2 trades in the price interval $[x, x+dx)$ is $-\lambda_A(x'')dB(t, x)$ for some $x'' \in [x, x+dx)$ and so the expected frequency at prices *strictly* less than y is $-\int_{x_{min}(t)}^{y} \lambda_A(x)dB(t, x)$. This is also the expected frequency of type 2 trades at prices *not exceeding* y, since the continuity of $B(t, x)$ at $x = y$ means that there is vanishing probability of a type 2 trade occuring at exactly this price. Adding the two integrals gives (16). The proof of (iv) immediately follows, since $B(t, x)$ is everywhere continuous from the left, and is also continuous from the right at $x_{max}(t)$ on account of (12). $\square$

3 Stationary Solutions

For these results to be useful, the functions $A(t, x)$ and $B(t, x)$ must be determined. One might hope to do this by deriving and then solving the evolution equations subject to suitable initial conditions, but in fact it proves difficult to write down simple evolution equations for $A(t, x)$ and $B(t, x)$. This is because they contain information only about the distributions of the best ask price and the best bid price, but their time evolution depends also on the distributions of the second best ask and bid prices. Similarly, the time evolution of the distribution functions for the second best prices depends on the distribution functions for the third best prices, and so on.

Without evolution equations for $A(t, x)$ and $B(t, x)$, it is at first unclear how to go about finding any stationary solutions. Fortunately, there is another approach. For each $x > 0$, let $M(t, x)$ denote the number of queued asks executable at the price x (i.e. awaiting execution at prices not exceeding x), and let $N(t, x)$ denote the number of queued bids executable at this price (i.e. awaiting execution at prices of at least x). The functions $M(t, x)$ and $N(t, x)$ then contain a full description of the order book. Their time evolution is governed by the receipt of new bids and offers.

Assuming that the expectations of $M(t, x)$ and $N(t, x)$ are differentiable functions of t and x, we also define

$$u(t, x) \equiv \frac{\partial}{\partial x} E[\, M(t, x)\,], \qquad v(t, x) \equiv -\frac{\partial}{\partial x} E[\, N(t, x)\,]. \tag{18}$$

Then at time t the number of unexecuted asks in the price range $(x_1, x_2]$ and the number of unexecuted bids in the price range $[x_1, x_2)$ are random variables with expectations

$$E[M(t,x_2) - M(t,x_1)] = \int_{x_1}^{x_2} u(t,x)dx \tag{19}$$
$$E[N(t,x_2) - N(t,x_1)] = \int_{x_1}^{x_2} v(t,x)dx \tag{20}$$

respectively. We can think of $u(t,x)$ and $v(t,x)$ as the expected densities of asks and bids in the order book around the price x.

Theorem. *If $A(x)$ is the stationary distribution function for the lowest unexecuted ask price α_1 and $B(x)$ is the stationary survivor function for the highest unexecuted bid price β_1, and $A(x)$ and $B(x)$ are both differentiable at x, then*

$$\frac{\partial}{\partial t}u(t,x) = \lambda'_A(x)[1 - B(x)] - \lambda_B(x)A'(x) \tag{21}$$
$$\frac{\partial}{\partial t}v(t,x) = -\lambda'_B(x)[1 - A(x)] + \lambda_A(x)B'(x). \tag{22}$$

Proof. See Appendix A. □

Corollary. *Suppose that $[x_{min}, x_{max}]$ is the competitive window associated with $A(x)$ and $B(x)$. Then*

$$\frac{\partial}{\partial t}u(t,x) = \lambda'_A(x) > 0 \qquad \forall x > x_{max} \tag{23}$$
$$\frac{\partial}{\partial t}v(t,x) = -\lambda'_B(x) > 0 \qquad \forall x < x_{min}. \tag{24}$$

Proof. The equalities follow directly from the Theorem when one makes use of equations (11) and (12). The inequalities follow from the strict monotonicity of the functions $\lambda_A(x)$ and $\lambda_B(x)$. □

The Corollary makes it clear that there can be no solutions that are stationary outside the competitive window $[x_{min}, x_{max}]$. At prices higher than x_{max}, asks accumulate faster than they can be matched with incoming bids, while at prices lower than x_{min} bids accumulate faster than they can be matched with incoming asks (as shown by equations (23) and (24) respectively). Such orders can be regarded as uncompetitive as they never result in trades, and are of little interest.

As we are concerned with what takes place inside the competitive window, we only need stationary solutions $u(x)$ and $v(x)$ inside this window. The Theorem then implies that $A(x)$ and $B(x)$ must satisfy the coupled first-order linear differential equations

$$A'(x) = \frac{\lambda'_A(x)}{\lambda_B(x)}[1 - B(x)] \tag{25}$$

$$B'(x) = \frac{\lambda'_B(x)}{\lambda_A(x)}[1 - A(x)]. \tag{26}$$

for all $x \in [x_{min}, x_{max}]$. They must also satisfy the boundary conditions

$$A(x_{min}) = 0, \quad B(x_{min}) = 1, \quad A(x_{max}) = 1, \quad B(x_{max}) = 0 \tag{27}$$

that follow from (11) and (12).

This is not a conventional boundary-value problem, as x_{min} and x_{max} are not given but must be deduced along with $A(x)$ and $B(x)$. To proceed, first note that the quantity

$$\kappa \equiv [1 - B(x)]\lambda_A(x) + \lambda_B(x)[1 - A(x)] \tag{28}$$

is independent of x over the competitive window, on account of equations (25) and (26). Putting $x = x_{min}$ and then $x = x_{max}$ gives

$$\kappa = \lambda_B(x_{min}) = \lambda_A(x_{max}). \tag{29}$$

This identity establishes a useful connection between x_{min} and x_{max}. Moreover, making use of the identities (25) and (26) to evaluate the integrals (16) and (17), one finds that the expected frequency of trades at prices not exceeding x is now

$$\lambda_T(x) = \lambda_A(x)[1 - B(x)] \qquad x \in [x_{min}, x_{max}], \tag{30}$$

and the overall frequency of trades is $\kappa = \lambda_T(x_{max})$.

The next step is to define

$$z(x) \equiv \frac{1 - B(x)}{1 - A(x)}. \tag{31}$$

Then (25) and (26) imply that $z(x)$ must satisfy the Ricatti differential equation

$$\frac{dz}{dx} = -\frac{\lambda'_B(x)}{\lambda_A(x)} + \frac{\lambda'_A(x)}{\lambda_B(x)}z^2 \tag{32}$$

everywhere within the competitive window. Forsyth (1929) shows that such an equation admits a family of solutions of the form $z = [cp(x) + q(x)]/[cr(x) + s(x)]$, with each solution characterised by a different value of the constant c. Given $x_{min} \in (0, \infty)$, let $z(x)$ be the particular solution satisfying the initial condition

$$z(x_{min}) = 0. \tag{33}$$

Typically, this solution will have a pole at some point $x_{max} > x_{min}$. (If there is more than one pole on (x_{min}, ∞) we choose x_{max} to be the smallest.) Then $z(x)$ will be an increasing function on the interval $[x_{min}, x_{max})$.

With x_{max} defined in this way, there is no general reason why the value of $\lambda_B(x_{min})$ should coincide with that of $\lambda_A(x_{max})$ as required by equation (29). However, we suppose that x_{min} is specifically *chosen* to ensure this coincidence. We can then define κ by equation (29), and use equation (28) to obtain

$$A(x) = 1 - \frac{\kappa}{\lambda_B(x) + z(x)\lambda_A(x)}, \qquad B(x) = 1 - \frac{\kappa z(x)}{\lambda_B(x) + z(x)\lambda_A(x)} \tag{34}$$

for $x \in [x_{min}, x_{max})$. It is easy to confirm that these functions do indeed satisfy the coupled equations (25,26) as well as the boundary conditions (27), and hence also the inequalities (9,10). Moreover, for all $x \in (x_{min}, x_{max})$ we have $\lambda_B(x) \leq \kappa$ and $\lambda_A(x) \leq \kappa$ and hence

$$1 - A(x) - B(x) \;=\; \frac{\kappa - \lambda_B(x) + z(x)[\kappa - \lambda_A(x)]}{\lambda_B(x) + z(x)\lambda_A(x)} \;\geq\; 0 \tag{35}$$

in agreement with (8).

It has been shown that all trades take place between two limiting prices x_{min} and x_{max}, which are determined by the form of the supply and demand functions. Clearly, x_{min} and x_{max} correspond to what technical analysts refer to as the "support" and "resistance" levels. At prices below the support level x_{min}, buy orders accumulate faster than they can be executed (as shown by (24)) and build up into a solid wall. This prevents any new sell order from being executed at a price less than x_{min}. Similarly, above the resistance level x_{max}, sell orders accumulate to form an impenetrable wall, ensuring that no buy orders are executed at prices higher than x_{max}.

The above procedure determines not only the resistance and support levels x_{min} and x_{max}, but also the stationary distributions of the best bid, the best ask and the transaction price. This is now illustrated with an example.

4 Constant Elasticities of Supply and Demand

Consider a market in which the elasticity of supply and elasticity of demand are both price-independent, with values $\eta_a > 0$ and $\eta_b > 0$ respectively. The supply and demand functions then have the form

$$\lambda_A(x) = kx^{\eta_a}, \qquad \lambda_B(x) = kx^{-\eta_b} \tag{36}$$

for some positive constant k. (We suppose the numeraire has been chosen so that the equilibrium price x_E is exactly 1, and hence $\lambda_A(1) = \lambda_B(1)$.)

If $\eta_b \neq \eta_a$ then the general solution of (32) can be written in the form

$$z(x) = \eta_b x_{min}^{-\eta_a-\eta_b} \frac{(x/x_{min})^{-\eta_b} - (x/x_{min})^{-\eta_a}}{\eta_a(x/x_{min})^{\eta_b} - \eta_b(x/x_{min})^{\eta_a}} \tag{37}$$

with $x_{min} > 0$ to be specified. This function has a zero at x_{min} and diverges as $x \to x_{max} = x_{min}(\eta_a/\eta_b)^{1/(\eta_a-\eta_b)}$. In order that $\lambda_A(x_{max}) = \lambda_B(x_{min}) = \kappa$ as required by (29), it is necessary to choose

$$x_{min} = \left(\frac{\eta_a}{\eta_b}\right)^{-\frac{\eta_a}{\eta_a{}^2-\eta_b{}^2}} \qquad x_{max} = \left(\frac{\eta_a}{\eta_b}\right)^{\frac{\eta_b}{\eta_a{}^2-\eta_b{}^2}} \qquad \kappa = k\left(\frac{\eta_a}{\eta_b}\right)^{\frac{\eta_a\eta_b}{\eta_a{}^2-\eta_b{}^2}}. \tag{38}$$

Equations (34) then give

$$A(x) = 1 - \frac{1}{\eta_a - \eta_b}\left[\eta_a\left(\frac{x}{x_{min}}\right)^{\eta_b} - \eta_b\left(\frac{x}{x_{min}}\right)^{\eta_a}\right] \tag{39}$$

$$B(x) = 1 - \frac{1}{\eta_a - \eta_b}\left[\eta_a\left(\frac{x}{x_{max}}\right)^{-\eta_b} - \eta_b\left(\frac{x}{x_{max}}\right)^{-\eta_a}\right] \tag{40}$$

for $x \in [x_{min}, x_{max}]$. It is easily verified that these functions do indeed satisfy the differential equations (25,26) on this interval, along with the boundary conditions (27).

Observe that $B(x) = A(c^2/x)$ where $c^2 = x_{min}x_{max} = (\eta_a/\eta_b)^{-1/(\eta_a+\eta_b)}$. This demonstrates the presence of a symmetry between the distribution of bids and asks in the order book. Moreover, since $A(c) = B(c)$, the order book is equally likely to contain an ask at a price $\leq c$ or a bid at a price $\geq c$.

The lowest ask price α_1 and highest bid price β_1 have probability density functions $A'(x)$ and $-B'(x)$, both of which are supported on the competitive window $[x_{min}, x_{max}]$. These density functions can be used to calculate expectations of any functions of α_1 or β_1. For example,

$$E[\alpha_1] = \frac{\eta_a \eta_b x_{min} + \sigma x_{max}}{(\eta_a+1)(\eta_b+1)} \qquad (\eta_a \neq \eta_b) \tag{41}$$

$$E[\beta_1] = \frac{\eta_a \eta_b x_{max} - \sigma x_{min}}{(\eta_a-1)(\eta_b-1)} \qquad (\eta_a \neq 1 \neq \eta_b,\ \eta_a \neq \eta_b) \tag{42}$$

where $\sigma \equiv ({\eta_a}^{\eta_a} {\eta_b}^{-\eta_b})^{\frac{1}{\eta_a-\eta_b}}$.

Given any $x \in [x_{min}, x_{max}]$, the frequency of trades at prices less than x can be obtained using (30) and is found to be

$$\lambda_T(x) = \frac{\kappa \eta_b}{\eta_a - \eta_b} \left[\left(\frac{x}{x_{min}} \right)^{\eta_a - \eta_b} - 1 \right] \qquad (\eta_a \neq \eta_b). \tag{43}$$

Since $\kappa = \lambda_T(x_{max})$ is the overall frequency of trades, the density function for the price of an arbitrary trade is

$$\frac{\lambda'_T(x)}{\kappa} = ({\eta_a}^{\eta_a} {\eta_b}^{\eta_b})^{\frac{1}{\eta_a+\eta_b}} x^{\eta_a - \eta_b - 1} \qquad (\eta_a \neq \eta_b) \tag{44}$$

if $x \in [x_{min}, x_{max}]$, or zero otherwise. The transaction price X_T of a randomly chosen trade then has expectation

$$E[X_T] = \frac{\eta_a x_{max} - \eta_b x_{min}}{1 + \eta_a - \eta_b}. \tag{45}$$

For completeness, let us now consider the special case when the elasticities of demand and supply are equal: $\eta_b = \eta_a = \eta$. In this case, the general solution of (32) can be written in the form

$$z(x) = \frac{\eta \ln(x/x_{min})}{x^{2\eta}[1 - \eta \ln(x/x_{min})]}. \tag{46}$$

This function has a zero at x_{min} and diverges at $x_{max} = e^{1/\eta} x_{min}$. Requiring $\lambda_B(x_{min}) = \lambda_A(x_{max}) = \kappa$, we get

$$x_{min} = e^{-1/2\eta}, \qquad x_{max} = e^{1/2\eta}, \qquad \kappa = ke^{1/2} \tag{47}$$

and so equations (34) yield

$$A(x) = 1 - e^{1/2} x^{\eta} \left(\frac{1}{2} - \eta \ln x \right) \qquad x \in [x_{min}, x_{max}] \tag{48}$$

$$B(x) = 1 - e^{1/2} x^{-\eta} \left(\frac{1}{2} + \eta \ln x \right) \qquad x \in [x_{min}, x_{max}]. \tag{49}$$

(Once again, $B(x) = A(c^2/x)$ where now $c^2 = x_{min}x_{max} = 1$.) This time, the expected price of the lowest unexecuted ask works out to be

$$E[\alpha_1] = \frac{\eta^2}{(\eta+1)^2}e^{-\frac{1}{2\eta}} + \frac{\eta e}{(\eta+1)^2}e^{\frac{1}{2\eta}} \tag{50}$$

while the expected price of the highest unexecuted bid is

$$E[\beta_1] = \begin{cases} \frac{\eta^2}{(\eta-1)^2}e^{\frac{1}{2\eta}} - \frac{\eta e}{(\eta-1)^2}e^{-\frac{1}{2\eta}} & \eta \neq 1 \\ \frac{1}{2}e^{\frac{1}{2}} & \eta = 1 \end{cases}. \tag{51}$$

From (30), the expected frequency of trades at prices not exceeding x is

$$\lambda_T(x) = ke^{1/2}\left(\frac{1}{2} + \eta \ln x\right) \qquad \text{for } x \in [e^{-1/2\eta}, e^{1/2\eta}] \tag{52}$$

and so the density function for the price of an arbitrary trade is

$$\frac{\lambda'_T(x)}{\kappa} = \frac{\eta}{x} \qquad \text{for } \; x \in [e^{-1/2\eta}, e^{1/2\eta}]. \tag{53}$$

The transaction price of a random trade is then

$$E[X_T] = 2\eta \sinh(\frac{1}{2\eta}) = 1 + \frac{1}{3!(2\eta)^2} + \frac{1}{5!(2\eta)^4}\dots, \tag{54}$$

which always exceeds the equilibrium price $x_E = 1$.

Figure 1 shows the form of the supply and demand functions $\lambda_A(x)$, $\lambda_B(x)$ for the case $\eta_a = \eta_b = 1$, while figures 2, 3 and 4 show the resulting density functions $A'(x)$, $-B'(x)$ and $\lambda'_T(x)/\kappa$ for the best ask, the best bid and the transaction price. (Note that qualitatively similar graphs are obtained for any positive values of η_a and η_b.)

We conclude this section by observing that the functions $A(x)$ and $B(x)$ given by (39) and (40) are the unique solutions to (25,26) and (27) for *any* supply and demand functions $\lambda_A(x)$ and $\lambda_B(x)$ having the specified elasticities η_a, η_b inside the interval $[x_{min}, x_{max}]$, regardless of their values elsewhere. It follows that the stationary distributions of the transaction prices and the best bid and ask prices are determined entirely by the values of η_a and η_b within the competitive window.

For general supply and demand functions, elasticities may vary slightly with price. However, in most markets price fluctuations are relatively small and so the elasticities can be regarded as approximately constant. In this context, the constant-elasticity example discussed above has broad applicability. Indeed, provided the competitive window is narrow, it is natural to assume that the elasticities remain approximately constant within it and that $A(x)$ and $B(x)$ will be well approximated by expressions (39) and (40).

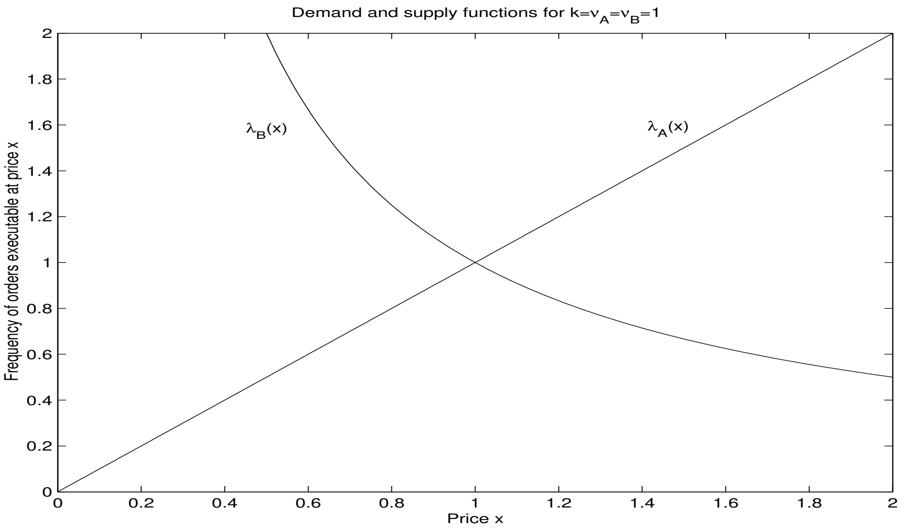

Figure 1: The supply function $\lambda_A(x)$ and demand function $\lambda_B(x)$ with $k = \eta_a = \eta_b = 1$

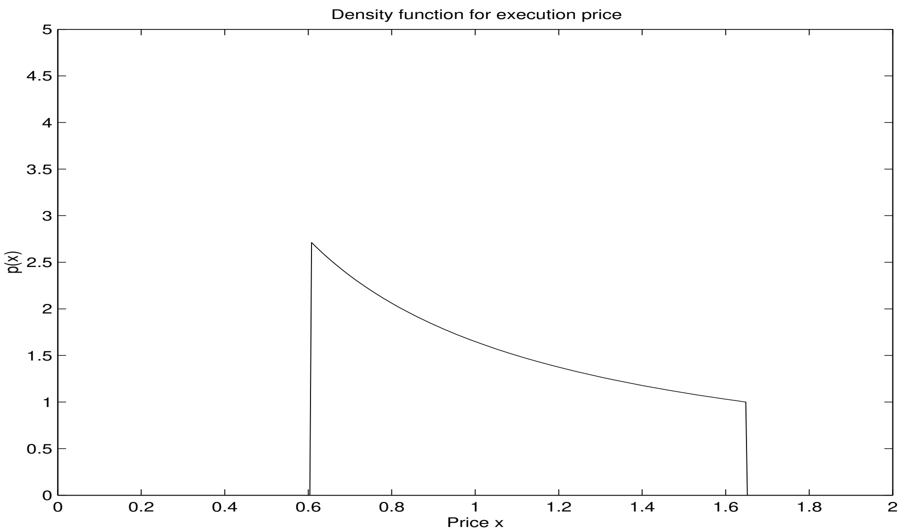

Figure 2: The steady-state probability density function $p(x)$ for the execution price, assuming $\eta_a = \eta_b = 1$

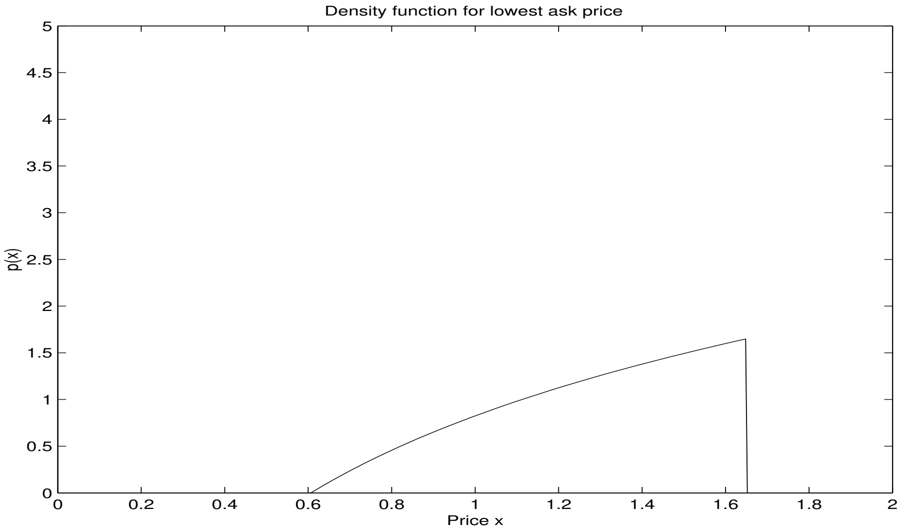

Figure 3: The steady-state probability density function $p(x)$ for the lowest unexecuted ask price α_1, assuming $\eta_a = \eta_b = 1$.

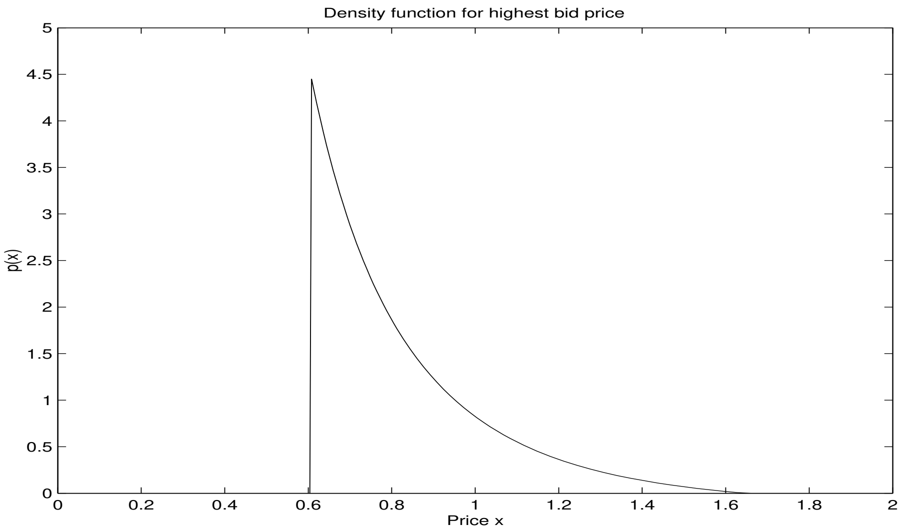

Figure 4: The steady-state probability density function $p(x)$ for the highest unexecuted bid price β_1, assuming $\eta_a = \eta_b = 1$.

5 Narrow Spreads

Real financial markets are characterised by very narrow bid-ask spreads – typically less than 1% of the equilibrium price. An application of the model to such a market should therefore be tailored to give a very narrow competitive window. Within this narrow window it seems reasonable to assume that the elasticities of supply and demand η_a,η_b will hardly vary and may be regarded as constants, and so the analysis of the previous section will be applicable. In particular, the two parameters η_a, η_b (defined here as the values of the elasticities at the equilibrium price) should completely determine the distribution of bids, asks and trades within the window of competition.

It is apparent from (38) that the competitive window $[x_{min}, x_{max}]$ shrinks as η_a and η_b grow. Hence, a narrow window of competition is indicative of large values of η_a and η_b. In order to focus on the features associated with arbitrarily large elasticities, it will helpful to consider the dependence of our results on the *ratio* of the elasticities,

$$\nu \equiv \eta_a/\eta_b. \tag{55}$$

We are interested in the distributions of bids, asks and trades *within* the competitive window as this becomes very narrow, and so it is natural to rescale $[x_{min}, x_{max}]$ to the unit interval $[0, 1]$. The form of equations (39),(40),(48) and (49) suggests that a logarithmic scaling is more appropriate than a linear one, and so we define

$$s(x) \equiv \frac{\ln(x/x_{min})}{\ln(x_{max}/x_{min})}. \tag{56}$$

The competitive window $[x_{min}, x_{max}]$ is then represented by values of s in the interval $s \in [0, 1]$.

Expressing (39) and (40) in terms of s and then differentiating with respect to s, we find that the density functions for $s_\alpha = s(\alpha_1)$, $s_\beta = s(\beta_1)$ and $s_T = s(X_T)$ are respectively

$$p_\alpha(s) \quad = \frac{dA}{ds} = \quad \begin{cases} \frac{\nu\ln(1/\nu)}{(1-\nu)^2}\left(\nu^{-\frac{s}{1-\nu}} - \nu^{-\frac{\nu s}{1-\nu}}\right) & \nu \neq 1 \\ \\ se^s & \nu = 1 \end{cases} \tag{57}$$

$$p_\beta(s) \quad = -\frac{dB}{ds} = \quad \begin{cases} \frac{\nu\ln(1/\nu)}{(1-\nu)^2}\left(\nu^{-\frac{1-s}{1-\nu}} - \nu^{-\frac{\nu(1-s)}{1-\nu}}\right) & \nu \neq 1 \\ \\ (1-s)e^{(1-s)} & \nu = 1 \end{cases} \tag{58}$$

$$p_T(s) \quad = \frac{1}{\kappa}\frac{d\lambda_T}{ds} = \quad \begin{cases} \frac{\ln(1/\nu)}{1-\nu}\nu^s & \nu \neq 1 \\ \\ 1 & \nu = 1 \end{cases} \tag{59}$$

for $s \in [0, 1]$, with each vanishing outside this interval.

Using these density functions, one finds that the expectations of s_α, s_β and s_T are

$$E[s_\alpha] \quad = \quad \begin{cases} \frac{\nu^{-\frac{1}{1-\nu}} - \nu^{-\frac{\nu}{1-\nu}} + \nu - \nu^{-1}}{\ln(1/\nu)} & \nu \neq 1 \\ e - 2 & \nu = 1 \end{cases} \tag{60}$$

$$E[s_\beta] \quad = \quad \begin{cases} 1 - \frac{\nu^{-\frac{1}{1-\nu}} - \nu^{-\frac{\nu}{1-\nu}} + \nu - \nu^{-1}}{\ln(1/\nu)} & \nu \neq 1 \\ 3 - e & \nu = 1 \end{cases} \tag{61}$$

$$E[s_T] \quad = \quad \begin{cases} \frac{1}{\ln(1/\nu)} - \frac{\nu}{1-\nu} & \nu \neq 1 \\ \frac{1}{2} & \nu = 1 \end{cases} \tag{62}$$

while that the overall frequency of transactions is

$$\kappa = k\nu^{-\frac{\nu}{1-\nu^2}} \tag{63}$$

and the value of s corresponding to the equilibrium price is

$$s_E = s(x_E) = \frac{\nu}{1+\nu}. \tag{64}$$

We observe that the density functions $p_\alpha(s)$, $p_\beta(s)$ and $p_T(s)$ (and hence all the expectations) are determined entirely by a single exogenous parameter; namely the ratio of elasticities $\nu = \eta_a/\eta_b$. Figures 5, 6, 7, and 8 show the form of these density functions for various choices of ν.

In fact, the remaining parameter ν has a very simple interpretation. Suppose that a proportion q of all market participants currently own at least one unit of the asset being traded. Suppose also that there are dN market participants who value the asset at a price between x and $x + dx$. At the price x, these individuals would regard the asset as underpriced, but if the price rises to $x+dx$ they will regard it as overpriced. Hence, a price rise from x to $x + dx$ will decrease by dN the number of potential buyers. It will also increase by dN the number who regard selling as a favourable transaction; but since only a fraction q of these actually have anything to sell, there will be an increase of only qdN in the number of potential sellers (provided there is no short-selling). Assuming that potential buyers and seller are equally likely to submit an order to the exchange over a short time interval,

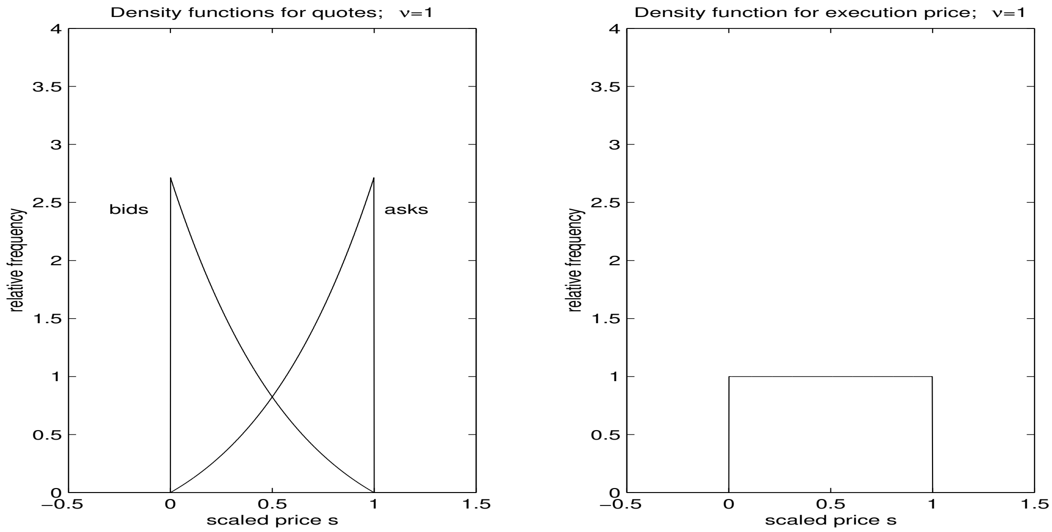

Figure 5: The steady-state density functions for the scaled quote and execution prices, with $\nu = 1$.

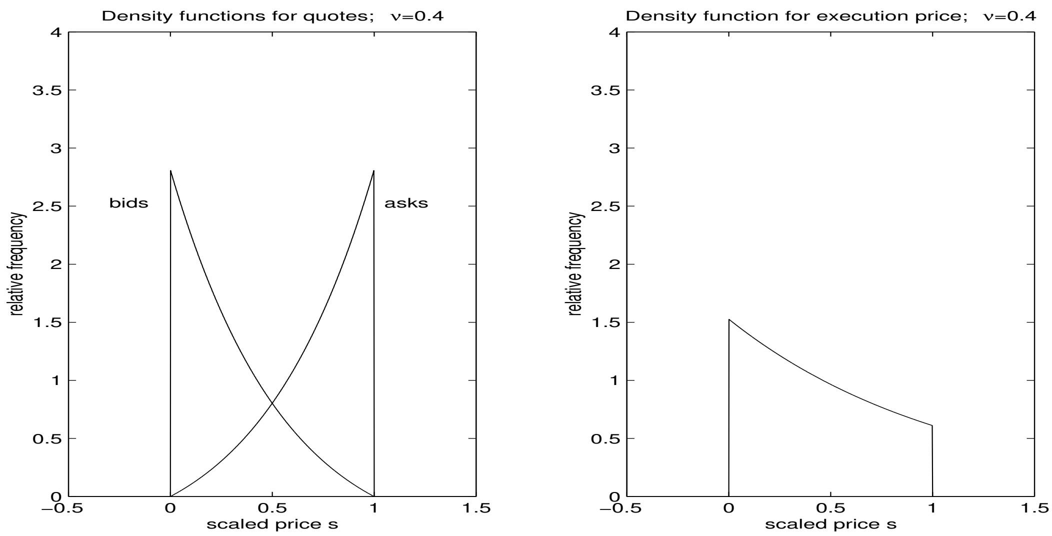

Figure 6: The steady-state density functions for the scaled quote and execution prices, with $\nu = 0.4$.

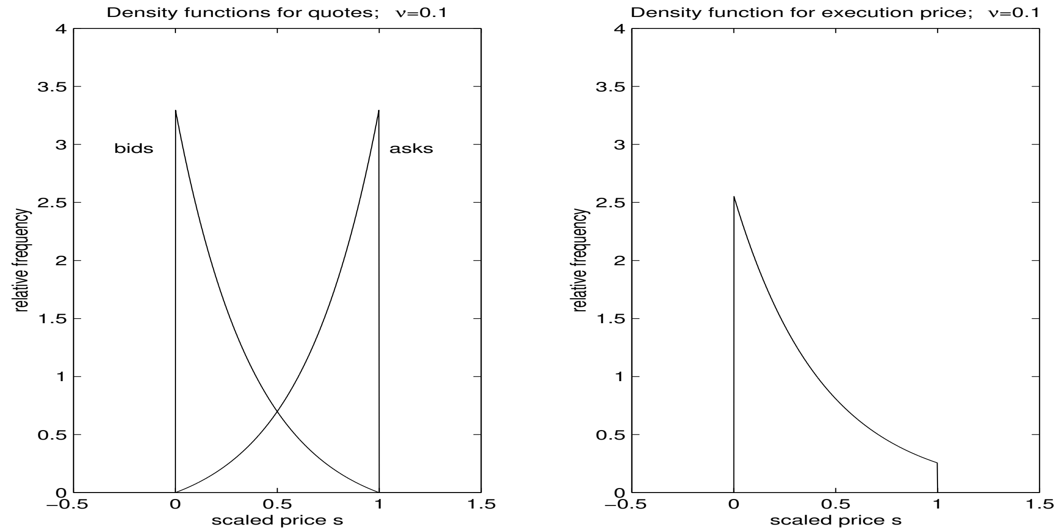

Figure 7: The steady-state density functions for the scaled quote and execution prices, with $\nu = 0.1$.

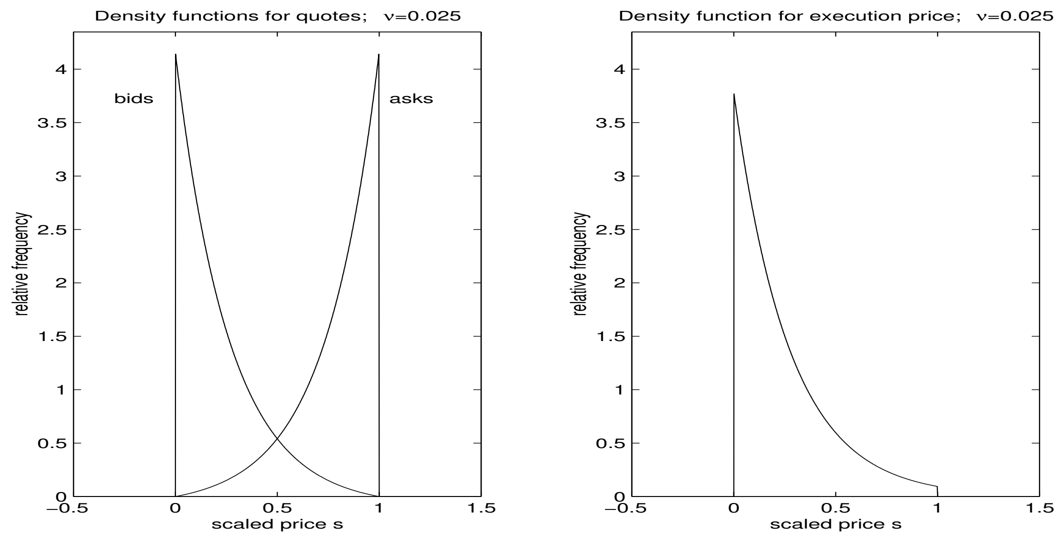

Figure 8: The steady-state density functions for the scaled quote and execution prices, with $\nu = 0.025$.

it follows that the resulting increase $\lambda'_A(x)dx$ in the frequency of sell orders will be smaller by a factor of q than the corresponding decrease $-\lambda'_B(x)dx$ in the frequency of buy orders, and consequently $q = -\lambda'_A(x)/\lambda'_B(x) = \nu$. Thus, the remaining exogenous parameter ν is simply the proportion of the market participants currently holding at least one unit of the asset.

It follows that ν cannot exceed 1, unless there is extensive short-selling, or the traded asset is a foreign currency which is more widely held than the numeraire currency. Except in these two special cases, therefore, trades at the top end of the competitive window ($s = 1$) will always be less frequent than those at the bottom end ($s = 0$) by a factor of $p_T(1)/p_T(0) = \nu$.

All the above results are exact, and do not depend at all on the average elasticity $\eta = (\eta_a + \eta_b)/2$. Indeed, the only apparent significance of this quantity is that it determines the size of the competitive window. For large values of η (such as those occuring in real financial markets), (38) gives

$$x_{min} \simeq 1 - \frac{\nu \ln(1/\nu)}{2\eta(1-\nu)}, \qquad x_{max} \simeq 1 + \frac{\ln(1/\nu)}{2\eta(1-\nu)} \tag{65}$$

and the scaling (56) becomes linear

$$s \simeq \frac{x - x_{min}}{x_{max} - x_{min}}, \qquad x \simeq (1-s)x_{min} + s x_{max}. \tag{66}$$

In the large-η limit, these approximations become exact and s may be regarded simply as the difference (in appropriate units) between a given price and x_{min}.

In practise it is often difficult to determine the elasticities of supply and demand or even the precise equilibrium price, and it may be more convenient to express the density functions in a form that depends only on readily determined quantities such as x_{min}, x_{max} and ν. In the high-elasticity limit where the approximation (66) is exact, we find

$$p_\alpha(x) = \begin{cases} \frac{\mu\nu}{1-\nu}\left[e^{\mu(x-x_{min})} - e^{\mu\nu(x-x_{min})}\right] & \nu \neq 1 \\ \mu^2(x - x_{min})e^{\mu(x-x_{min})} & \nu = 1 \end{cases} \tag{67}$$

$$p_\beta(x) = \begin{cases} \frac{\mu\nu}{1-\nu}\left[e^{\mu(x_{max}-x)} - e^{\mu\nu(x_{max}-x)}\right] & \nu \neq 1 \\ \mu^2(x_{max} - x)e^{\mu(x_{max}-x)} & \nu = 1 \end{cases} \tag{68}$$

$$p_T(x) = \begin{cases} \mu e^{\mu(1-\nu)(x-x_{min})} & \nu \neq 1 \\ 1 & \nu = 1 \end{cases} \tag{69}$$

for $x \in [x_{min}, x_{max}]$, where

$$\mu = \begin{cases} \frac{\ln(1/\nu)}{(1-\nu)(x_{max}-x_{min})} & \nu \neq 1 \\ \frac{1}{(x_{max}-x_{min})} & \nu = 1 \end{cases} . \tag{70}$$

6 Conclusions

Perhaps the most striking prediction of the model is the restriction of market price fluctuations to the competitive window $[x_{min}, x_{max}]$. This feature is not indicative of an absence of supply or demand outside this interval, but arises as a direct consequence of the market microstructure.

It is natural to ask whether the competitive window is merely an artefact of the properties L1-L6 assumed in the construction of the model, or the assumed form of the supply and demand functions, or whether it might also emerge in more realistic models. Although it is not yet possible to give an unequivocal answer to this question, a strong argument can be made that the competitive window should be a generic feature of limit order markets with sufficient liquidity and stability.

While all of the properties L1-L6 were required in the derivation of the mathematical results presented earlier, few of them appear to play an essential role in the basic process leading to the formation of the competitive window; that is, the accumulation of uncompetitive bids and asks into impenetrable walls restricting the range of prices at which new orders can be executed. Provided that order arrival rates remain reasonably steady during the period of interest, and provided the cancellation rate of unexecuted orders is fairly low, it is difficult to imagine what might prevent this process. Indeed, evidence for the existence of competitive windows in real limit order markets is provided by technical analysts, who rely extensively on the existence of support and resistance levels between which price fluctuations are temporarily contained.

The model also predicts the stationary distributions of transaction prices and of the best bid and ask prices. In particular, provided the spread is narrow, it predicts that the form of these distributions within the competitive window should depend on a single parameter which can be identified with the proportion of market participants owning the asset.

These results were derived from the assumptions used to define the model, but again it seems unlikely that their qualitative features would be dramatically affected if certain of these assumptions were relaxed. For example, relaxing assumptions L4 and L6 to allow gradual changes in supply and demand or the cancellation of orders would limit the build-up of unfilled bids and asks at uncompetitive prices and thus blur the edges of the competitive window, but should not otherwise significantly change the form of the distributions. Simlarly, relaxing assumptions L2 and L3 to accommodate variable order quantities and discrete prices should not result in dramatically different distributions.

It is somewhat harder to anticipate how the predictions of the model would be affected if we dropped assumption L5, and allowed opportunistic traders to take advantage of temporary anomalies in the order book. In principle, the removal of this restriction could result in a complicated game-theoretic feedback process between the order book and the order submission rates. However, there are several reasons why this type of opportunism may be less prevalent in real limit order markets than might be imagined:

- on some exchanges, detailed information about the current state of the order book is not disclosed, or is restricted to certain classes of traders;
- increasing numbers of market participants are retail investors who may not have access to fully up-to-date information about the state of the order book, and whose orders may not reach the exchange quickly enough for them to take advantage of temporary order book anomalies.
- in the absence of clear price trends, opportunities arising from order-book anomalies will be relatively infrequent;
- instead of continually monitoring the order book for opportunities, many traders may prefer to submit limit orders that will automatically take advantage of such opportunities as they arise.

For these reasons, it is tempting to speculate that opportunistic responses to order book anomalies may have only a modest effect on real limit order markets. If so, then assumption L5 can be also regarded as relatively innocuous.

However the predictions of the model are certainly influenced by assumption L1, which protects the market from domination by powerful individuals or conglomerates. It is this assumption which makes the statistical approach feasible, and spares us the more difficult task of modelling the actions of the dominant market players. In general, statistical models such as the one presented here will only be applicable to markets that are not dominated by a small number of powerful participants.

Despite this discussion, it rermains unclear how realistic the model is. Hopefully, future work will reveal which of the assumptions L1-L6 are misleading oversimplifications, and which can be safely incorporated into a more realistic market model.

A Proof of Theorem

Suppose a new ask arrives during a brief time interval $[t, t+\Delta t]$. If the new ask price α is less than or equal to the highest unexecuted bid price β_1, these two orders are matched and executed and so, for $x \leq \beta_1$, the number $N(t,x)$ of remaining bids executable at the price x will decrease by one. Alternatively, if $\alpha > \beta_1$, then no match is possible and the new ask takes its place in the queue, increasing by one the number $M(t,x)$ of asks executable at a given price $x \geq \alpha$. For an arbitrary choice of $x > 0$, the changes in the values of $M(t,x)$ and $N(t,x)$ due to the arrival of the new ask are therefore

$$\Delta M = I(x \geq \alpha) I(\alpha > \beta_1), \qquad \Delta N = -I(x \leq \beta_1) I(\alpha \leq \beta_1) \tag{71}$$

where I denotes the indicator function. Similarly, the arrival of a new bid at a price β has the effect

$$\Delta M = -I(x \geq \alpha_1) I(\beta \geq \alpha_1), \qquad \Delta N = I(x \leq \beta) I(\beta < \alpha_1). \tag{72}$$

These expressions precisely encapsulate the dynamical rules R1 and R2.

During an infinitesimal time interval Δt, there is a probability $\lambda_A(x)\Delta t$ of receiving a new ask at a price not exceeding x, and a probability $\lambda_B(x)\Delta t$ of receiving a new bid at a price of at least x. (There is a negligible probability that more than one order will be received.) The expected increases in the quantities M and N over this time interval are hence found to be

$$\begin{aligned} E[\Delta M] &= \{I(x \geq \beta_1)[\lambda_A(x) - \lambda_A(\beta_1)] - I(x \geq \alpha_1)\lambda_B(\alpha_1)\}\Delta t && (73) \\ E[\Delta N] &= \{I(x \leq \alpha_1)[\lambda_B(x) - \lambda_B(\alpha_1)] - I(x \leq \beta_1)\lambda_A(\beta_1)\}\Delta t. && (74) \end{aligned}$$

In the limit $\Delta t \to 0$, this gives

$$\frac{\partial}{\partial t} E[M(t,x)] = E[I(x \geq \beta_1)(\lambda_A(x) - \lambda_A(\beta_1))] - E[I(x \geq \alpha_1)\lambda_B(\alpha_1)] \tag{75}$$

$$\frac{\partial}{\partial t} E[N(t,x)] = E[I(\alpha_1 \geq x)(\lambda_B(x) - \lambda_B(\alpha_1))] - E[I(\beta_1 \geq x)\lambda_A(\beta_1)] \tag{76}$$

where we have now taken expectations with respect to α_1 and β_1 to accommodate situations in which these prices are not known exactly.

We know that $\alpha_1(t)$ has distribution function $A(x)$ while $\beta_1(t)$ has survivor function $B(x)$, and so the last two equations can be written as

$$\frac{\partial}{\partial t}E[M(t,x)] = -\int_0^x [\lambda_A(x) - \lambda_A(b)]dB(b) - \int_0^x \lambda_B(a)dA(a) \tag{77}$$

$$\frac{\partial}{\partial t}E[N(t,x)] = \int_x^\infty [\lambda_B(x) - \lambda_B(a)]dA(a) + \int_x^\infty \lambda_A(b)dB(b). \tag{78}$$

If $A'(x)$ and $B'(x)$ both exist, then differentiating with respect to x gives

$$\frac{\partial}{\partial t}u(t,x) = \lambda'_A(x)[1 - B(x)] - \lambda_B(x)A'(x) \tag{79}$$

$$-\frac{\partial}{\partial t}v(t,x) = \lambda'_B(x)[1 - A(x)] - \lambda_A(x)B'(x) \tag{80}$$

as claimed.

B The Theorem of Domowitz & Wang

In Appendix A of Domowitz and Wang (1994) (hereafter DW), the authors attempt to derive the steady-state conditional probability distribution for the numbers $\{Y_{j+1}, \ldots, Y_R\}$ of unexecuted asks at prices $\{p_{j+1}, \ldots, p_R\}$, given that the highest bid price B^m currently has the value p_j. Their argument is based on the assertion that the steady-state distribution satisifes the following condition:

> "during a small interval Δt, the probability of Y_i $(j+1 \leq i \leq R)$ increasing from y_i to y_i+1 must equal that of Y_i decreasing from $y_i + 1$ to y_i".

In fact this statement is correct only in the absence of conditions on other variables that may be correlated with Y_i. As we will see, it is *not* generally true when conditions are placed on the highest bid price B^m, as then one must also take into account the possibility of changes in B^m.

To see that the quoted assertion is true in the absence of conditions on other variables, we note that only the following instantaneous changes in the value of Y_i are possible:

$$Y_i \to Y_i + 1, \qquad Y_i \to Y_i - 1. \tag{81}$$

Hence, if $\lambda(y_i; y_i \pm 1)\Delta t$ is the probability of observing a transition from $Y_i = y_i$ to $Y_i = y_i \pm 1$ during a short time interval Δt, irrespective of the value of B^m, then the increase of the probability $P(Y_i = y_i)$ over this interval is just

$$\begin{aligned} \Delta P(Y_i = y_i) \;\; = \;\; & \lambda(y_i - 1; y_i)\Delta t \; + \lambda(y_i + 1; y_i)\Delta t \\ & -\lambda(y_i; y_i - 1)\Delta t \; - \lambda(y_i; y_i + 1)\Delta t \qquad \forall y_j \geq 0 \end{aligned} \tag{82}$$

provided we define $\lambda(-1, 0) \equiv \lambda(0, -1) \equiv 0$. Summing from $y_i = 0$ to $y_i = n$ (with $0 \leq n \leq R$) then gives

$$\Delta P(Y_i \leq n) = \lambda(n + 1; n)\Delta t - \lambda(n; n + 1)\Delta t \tag{83}$$

and so for the steady-state distribution we have

$$\lambda(n; n + 1)\Delta t = \lambda(n + 1; n)\Delta t \tag{84}$$

in agreement with the quoted assertion.

If $\mathcal{H}_n$ denotes the subset of the sample space in which $Y_i \leq n$, then this argument can be summarised by the following two observations:

- the steady-state probability of a transition from $\mathcal{H}_n$ to $\overline{\mathcal{H}}_n$ must be the same as the probability of a transition from $\overline{\mathcal{H}}_n$ to $\mathcal{H}_n$

- a transition $\mathcal{H}_n \to \overline{\mathcal{H}}_n$ necessarily corresponds to a change in the value of Y_i from y_n to $y_n + 1$, while a transition $\overline{\mathcal{H}}_n \to \mathcal{H}_n$ corresponds to a change in the value of Y_i from $y_n + 1$ to y_n.

The second observation reflects the fact that the possible values of Y_i form a one-dimensional lattice, with transitions only possible between neighbouring points on this lattice. Hence, there is only one way out of $\mathcal{H}_n$ and one way into $\mathcal{H}_n$; and in the steady-state distribution there must be the same frequency of transitions in both directions.

The situation is more complicated when the highest bid price B^m is also specified, as now we must consider transition rates between neighbouring points in the two-dimensional lattice of possible values of Y_i and B^m. There is more than one way into any subset of this lattice, and more than one way out, and so the balance between flow in and flow out does not provide a simple equality between a pair of transition rates. Consequently, the quoted assertion does not hold.

More explicitly; when conditions are placed on the value of B^m, we must take into account the possibility of transitions

$$B^m \to B^{m\prime} \quad (B^{m\prime} \neq B^m). \tag{85}$$

in addition to those described in (81). Let $\lambda(y_i, p_j; y_i \pm 1; p_j)\Delta t$ denote the probability that, during a short time interval Δt, the value of Y_i changes from y_i to $y_i \pm 1$ while B^m has the value p_j. Also, let $\lambda(y_i, p_j; y_i; p'_j)$ denote the probability that, during this interval, the value of B^m changes from p_j to p'_j while Y_i has the value y_i.

The increase of the joint probability $P(\, Y_i \, = \, y_i \, , \, B^m \, = \, p_j \,)$ over this time interval is then just

$$\begin{aligned}\Delta P(Y_i = y_i, B^m = p_j) \quad &= \quad \lambda(y_i - 1, p_j; y_i, p_j)\Delta t \,+ \lambda(y_i + 1, p_j; y_i, p_j)\Delta t \\ &\quad -\lambda(y_i, p_j; y_i - 1, p_j)\Delta t \,- \lambda(y_i, p_j; y_i + 1, p_j)\Delta t \\ &\quad + \sum_{j' \neq j} \lambda(y_i, p'_j; y_i, p_j)\Delta t - \sum_{j' \neq j} \lambda(y_i, p_j; y_i, p'_j)\Delta t.\end{aligned} \tag{86}$$

This time, summing from $y_i = 0$ to $y_i = n$ (with $0 \leq n \leq R$) gives

$$\begin{aligned}\Delta P(Y_i \leq n, B^m = p_j) \quad &= \quad \lambda(n+1, p_j; n, p_j)\Delta t - \lambda(n, p_j; n+1, p_j)\Delta t \\ &\quad + \sum_{i=0}^{n} \sum_{j' \neq j} [\lambda(y_i, p'_j; y_i, p_j) - \lambda(y_i, p_j; y_i, p'_j)]\Delta t\end{aligned} \tag{87}$$

and so, for the steady-state distribution, the time-independence of the joint probability $P(Y_i \leq n, B^m = p_j)$ implies that

$$\lambda(n, p_j; n+1, p_j)\Delta t = \lambda(n+1, p_j; n, p_j)\Delta t + E(n, j)\Delta t \tag{88}$$

where we have defined

$$E(n, j) \equiv \sum_{i=0}^{n} \sum_{j' \neq j} [\lambda(y_i, p'_j; y_i, p_j) - \lambda(y_i, p_j; y_i, p'_j)]. \tag{89}$$

In Appendix 2 of DW, it is implicitly assumed that the quantity $E(n, j)$ vanishes. However this assumption is unfounded. Indeed, one can easily construct counter-examples where $E(n, j)$ takes non-zero values.

The non-vanishing of $E(n, j)$ implies that the probability of Y_i $(j + 1 \leq i \leq R)$ increasing from y_i to $y_i + 1$ while B^m has the value p_j is *not* generally equal to that of Y_i decreasing from $y_i + 1$ to y_i while $B^m = p_j$. The assertion quoted above is therefore *untrue* when conditions are placed imposed on B^m as well as Y_i, which is the case in Appendix 2 of DW. The proof of Theorem

1 in DW is therefore fallacious. Moreover, since this theorem was used to prove all the other theorems in this paper and in Bollerslev, Domowitz and Wang (1997), they are also invalid.

One might attempt to correct the error in DW by including the non-vanishing quantity $E(n, j)$ in the expressions dereived therein. Continuing to borrow the notation of DW, let $\lambda^a[i, j]$ denote the arrival rate of new asks at the price p_i, given that $B^m = p_j$; let $\mu^a[i, j]$ denote the cancellation rate for asks at the price p_j, given that $B^m = p_j$; let $\gamma^b[i, j]$ denote the arrival rate of hits (i.e. orders to buy at the price p_i if this is the current lowest ask) given that $B^m = p_j$, and let $\lambda^b[i, k]$ denote the arrival of rate of new bids at the price p_i, given that the lowest ask price A^m currently has the value p_k. Then the transition rate of the pair (Y_i, B^m) from the values $(y_i + 1, p_j)$ to the new values (y_i, p_j) is given by

$$\begin{aligned} \lambda(y_i + 1, p_j; y_i, p_j) \quad = \quad & P(Y_1 = \ldots = Y_{i-1} = 0, Y_i = y_i + 1, B^m = p_j)\lambda^b[i, i] \\ & + P(Y_1 = \ldots = Y_{i-1} = 0, Y_i = y_i + 1, B^m = p_j)\gamma^b[i, j] \\ & + \mu^a[i, j](y^i + 1)P(Y_i = y_i + 1, B^m = p_j). \end{aligned} \tag{90}$$

The first term on the right-hand side represents the rate at which asks at the price p_i are crossed against newly received bids at the same price, while the second term represents the rate at which asks at the price p_i are picked off by hits, and the third term represents the rate at which asks at the price p_i are currently being cancelled.

Similarly, the transition rate of the pair (Y_i, B^m) from the values (y_i, p_j) to the new values $(y_i + 1, p_j)$ is given by

$$\lambda(y_i, p_j; y_i + 1, p_j) = \lambda^a[i, j]P(Y_i = y_i, B^m = p_j) \tag{91}$$

where the term on the right-hand side represents the arrival rate of new bids at the price p_i.

Dividing both sides of equation (88) by the probability $P(B^m = y_j)$ and using the expressions obtained above for $\lambda(y_i, p_j; y_i + 1, p_j)$ and $\lambda(y_i + 1, p_j; y_i, p_j)$, we then obtain the identity

$$\begin{aligned} & \lambda^a[i, j]\, \Delta t\, P(Y_i = y_i | B^m = p_j) \\ & \quad = (\lambda^b[i, i] + \gamma^b[i, j])\, \Delta t\, P(Y_1 = \ldots = Y_{i-1} = 0, Y_i = y_i + 1 | B^m = p_j) \end{aligned}$$

$$
+\mu^a[i,j]\,(y^i+1)\,\Delta t\,P(Y_i = y_i + 1 | B^m = p_j)
$$

$$
+\frac{E(y_i,j)\,\Delta t}{P(B^m = y_j)}. \tag{92}
$$

This is the corrected form of the first equation displayed in Appendix 2 of DW, from which the last term had been omitted. Unfortunately the inclusion of this correction term considerably complicates the remainder of the analysis attempted by these authors, and invalidates their results.

C The Inverse Problem

This Appendix addresses the problem of whether it is possible to deduce the form of the supply and demand functions $\lambda_B(x)$ and $\lambda_A(x)$ from knowledge of the distributions of the best ask and bid prices.

Suppose one knows the equilibrium distribution functions $A(x), 1-B(x)$ of the highest unexecuted bid and the lowest unexecuted ask on some price interval $[x_{min}, x_{max}]$ where all trades take place. It is convenient to define

$$
F(x) \equiv 1 - A(x), \qquad G(x) = 1 - B(x). \tag{93}
$$

According to equations (27), these functions must satisfy the boundary conditions

$$
\begin{array}{ll} F(x_{min}) = 1, & G(x_{min}) = 0 \\ F(x_{max}) = 0, & G(x_{max}) = 1. \end{array} \tag{94}
$$

while equations (25,26) imply that they must also satisfy the Neumann boundary conditions

$$
F'(x_{min}) = 0, \qquad G'(x_{max}) = 0 \tag{95}
$$

If $\lambda_T(x)$ denotes the average frequency of trades at prices less than x, then (30) and (28) immediately give

$$
\lambda_A(x) = \frac{\lambda_T(x)}{G(x)}, \qquad \lambda_B(x) = \frac{\kappa - \lambda_T(x)}{F(x)} \tag{96}
$$

where we can take $\kappa = \lambda_T(x_{max})$. Requiring these expressions to satisfy equations (25,26) leads to the consistency condition

$$
\frac{d}{dx}\left(\frac{\lambda_T(x)}{F(x)G(x)}\right) = -\frac{\kappa F'(x)}{F(x)^2 G(x)} \tag{97}
$$

which can be integrated to give

$$\lambda_T(x) = \frac{F(x)G(x)\lambda_T(x_0)}{F(x_0)G(x_0)} + \kappa F(x)G(x) \int_{x_0}^{x} \frac{1}{G(s)} \left(\frac{d}{ds} \frac{1}{F(s)} \right) ds \tag{98}$$

for any $x, x_0 \in (x_{min}, x_{max})$. It then follows from (96) that

$$\lambda_A(x) = \kappa F(x) \left\{ \frac{\rho(x_0)}{F(x_0)G(x_0)} + \int_{x_0}^{x} \frac{1}{G(s)} \left(\frac{d}{ds} \frac{1}{F(s)} \right) ds \right\} \tag{99}$$

$$\lambda_B(x) = \kappa G(x) \left\{ \frac{1-\rho(x_0)}{F(x_0)G(x_0)} + \int_{x_0}^{x} \frac{1}{F(s)} \left(\frac{d}{ds} \frac{1}{G(s)} \right) ds \right\} \tag{100}$$

where $\rho(x_0) \equiv \lambda_T(x_0)/\kappa$ is the average proportion of trades that take place at prices less than x_0. Thus, we see that full determination of the supply and demand functions $\lambda_A(x)$ and $\lambda_B(x)$ on the interval $[x_{min}, x_{max}]$ requires knowledge not just of the functions $F(x)$ and $G(x)$, but also the two parameters κ and $\rho(x_0)$.

Alternatively, taking $x_0 \to x_{min}$ in the expression for $\lambda_A(x)$ and $x_0 \to x_{max}$ in the expression for $\lambda_B(x)$, we can write

$$\lambda_A(x) = F(x) \left\{ \lambda_A(x_{min}) - \kappa \int_{x_{min}}^{x} \frac{F'(s)}{G(s)F(s)^2} ds \right\} \tag{101}$$

$$\lambda_B(x) = G(x) \left\{ \lambda_B(x_{max}) + \kappa \int_{x}^{x_{max}} \frac{G'(s)}{F(s)G(s)^2} ds \right\} \tag{102}$$

where the three quantities $\lambda_A(x_{min})$, $\lambda_B(x_{max})$ and κ are constrained to satisfy the identity

$$\lambda_A(x_{min}) + \lambda_B(x_{max}) = \kappa \left\{ 1 + \int_{x_{min}}^{x_{max}} \left(\frac{1}{F(s)} - 1 \right) \frac{d}{ds} \left(\frac{1}{G(s)} - 1 \right) ds \right\} \tag{103}$$

which can be shown to follow from (28).

References

Biais, B., Hillion, P. and Spatt, C. (1995): "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse", *The Journal of Finance*, 50, 1655-1689

Bollerslev, T., Domowitz, I. and Wang, J. (1997): "Order flow and the bid-ask spread: An Empirical Probability Model of Screen-Based Trading", *Journal of Economic Dynamics and Control*, 21, 1471-1491

Domowitz, I. (1993): "Automating the Continuous Double Auction in Practice: Auotmated Trade Execution Systems in Financal Markets", in *The Double Auction Market: Institutions, Theories and Evidence*, ed. by D. Friedman and J. Rust. (Addison-Wesley), pp. 3-25

Domowitz, I. and Wang, J. (1994): "Auctions as Algorithms: Computerized trade execution and price discovery", *Journal of Economic Dynamics and Control*, 18, 29-60

Forsyth, A.R. (1929), *A Treatise on Differential Equations (sixth edition)* (MacMillan and Co, London)

Friedman, D. (1993A): "The Double Auction Market Institution: A Survey", in *The Double Auction Market: Institutions, Theories and Evidence*, ed. by D. Friedman and J. Rust. (Addison-Wesley), pp. 3-25

Garman, M.B. (1976): "Market Microstructure", *Journal of Financial Economics*, 3, 257-275

Glosten, L. (1994): "Is the Electronic Open Limit Order Book Inevitable?", *Journal of Finance*, 49, 1127-1161

Hamao, Y. and Hasbrouck, J. (1995): Securities Trading in the Absence of Dealers: Trades, and Quotes on the Tokyo Stock Exchange", *The Review of Financial Studies*, 8, 849-878

Parlour, C. (1998): "Price Dynamics in Limit Order Markets", *The Review of Financial Studies*, 11, 789-816

Smith, V.L. (1962): "An Experimental Study of Competitve Market Behaviour", *The Journal of Political Economy*, 70, 111-137

Smith, V., A. Williams, W.K. Bratton, and M. Venoni (1982): "Competitive Market Institutions: Double Auctions vs. Sealed Bid Offers", *American Economic Review*, 72, 58-77