Discriminatory Limit Order Books, Uniform Price Clearing

And Optimality

By

Lawrence R. Glosten

Columbia Business School

lrg2@columbia.edu

October, 2002

This is an extremely preliminary version, and hence I may later disavow any alleged information in the paper. Comments and suggestions would be extremely helpful
Introduction

The answer to the question “Is the electronic limit order book inevitable?” in Glosten (1994) is a qualified “yes.” Theoretically, the quote-based competition in a limit order book mimics the competition that should occur across exchanges. This result does not, however, say anything about the optimality of a Centralized Limit Order Book (CLOB). That is the focus of this paper, and the results generally support the inevitability of a CLOB—if a regulatory authority could choose and protect a single market mechanism it would quite likely choose a limit order book.

This paper takes the point of view that the market design question is most interesting for securities that face potential liquidity difficulties. Hence, problems with asymmetric information, inventory related costs and, potentially, a relatively few number of individuals willing to supply liquidity are all features of the analysis. Asymmetric information played an important part of the analysis in Glosten (1994) whereas, notably, a small number of strategic competing quoters did not. This feature recalls the analysis of Biais and Montmort (2000), who show the existence and characterization of equilibrium in a CLOB with strategic quoters. Like that paper, this paper focuses on some special cases of the environment in order to derive and characterize the equilibrium.

The question being asked in this paper places it in the relatively small literature that addresses the question of market design. It is most closely related to Viswanathan and Wang (VW) (2002) and Roell (1999), both of which examine the welfare properties of a discriminatory (each limit order pays of receives its limit price) CLOB with the equilibrium in a market with a finite number of strategic dealers all trading at the same price (or alternatively, a uniform price limit order book). The notable difference between
this paper and those is that while the distribution of trade sizes is specified exogenously in VW and Roell, this paper derives the equilibrium trade distribution based on the exogenously specified distribution of trader “types.” That is, based on an individual’s type and the terms of trade offered, the agent decides how large a trade to make. As the analysis of VW shows, and this paper confirms, the terms of trade determined by equilibrium in the discriminatory price CLOB are quite different from that in a uniform price clearing. Thus, one might expect the distributions of trade sizes to be different in the two settings. Consideration of elastic trade demand also allows a measure of welfare which includes the quoters. With inelastic trade, the cost to a trader is a benefit to the quoters and hence the total surplus is unaffected.

This paper only indirectly addresses the welfare comparison of a continuous market versus a periodic call. As with the papers cited above the analysis is of the market at a point in time. Conceptually, the market is presumed to consist of a sequence of such equilibria. Notably, the paper does not analyze dynamic trading strategies and the dynamic competition between liquidity suppliers as in Bondarenko (2001).

The outline of the paper is as follows. The next section lays out the economic environment and discusses the measure of welfare to be used. The subsequent section analyzes the optimum market design given this measure of welfare. This is followed by an analysis of equilibrium in a CLOB and a uniform price clearing with the major welfare result. The paper concludes with some observations on the relevance of the results for the regulation and design of markets.

1. The Economic Setting
The model to be analyzed considers the trade in a single security with a risky payoff, X. All of the analysis will be in terms of deviations from the current estimation of the value of the security. Hence, we can take \( E[X] = 0 \). The model considers a moment of time in which a single transaction takes place. Thus, the model is of the “Glosten-Milgrom” type rather than the “Kyle” type in which orders are aggregated. Following a trade, expectations will be updated and the market will continue on with another order.

The world is populated by two types of agents—a large number of potential “market order” users who observe the terms of trade and decide what quantity to buy or sell, and a relatively small number of agents who stand ready to take the other side of the market orders and hence supply liquidity by quoting. To conserve on verbiage, call the two market participant types “traders” and “quoters,” respectively.

A trader observes the terms of trade and determines an optimal trade by setting his or her marginal valuation equal to the marginal price. More specifically, a typical trader of type \( t \) maximizes preferences which are a function of type, quantity and amount spent

\[
U(t,Q,R(Q)) \quad \text{where } R(Q) \text{ is the amount paid to buy } (Q>0) \text{ Q shares or the amount received to sell } (Q<0) \text{ } -Q \text{ shares. Given the terms of trade, } R(.), \text{ the optimal amount to trade } Q(t) \text{ is the solution to (if } Q(t) \text{ is not equal to zero)}
\]

\[
\frac{U_2(t,Q(t),R(Q(t)))}{-U_3(t,Q(t),R(Q(t)))} = V(t,Q(t),R(Q(t))) = R'(Q(t)),
\]

where \( R'(.) \) is the first derivative of \( R(.) \). We shall call \( V(t,Q,R(Q)) \) the marginal valuation of a trader of type \( t \) at the trade \( Q \). For the analysis in this paper, it will be assumed that \( V \) does not depend upon \( R(Q) \) and in that case we will write the condition that determines \( Q(t) \) as \( V(t,Q(t)) = R'(Q(t)) \). In this case, \( V \) is interpretable as the individual’s demand curve for shares. For much of the analysis the special case of a
linear demand curve will be considered: \( V(t,Q(t)) = t - Q(t) \) (the coefficient of -1 on \( Q \) is without loss of generality since any other coefficient can be thought of as changing the units in which \( Q \) is measured). It is assumed that \( V \) is strictly increasing in \( t \).

In general, a trader’s type would involve a specification of all the things that would matter in the portfolio and trading decision—information, existing position in the security, positions in securities with payoffs correlated with this specific security, etc. For tractability this paper assumes that the type is unidimensional. Thus, for example, and drawing from the ubiquitous normal exponential utility example, the type might be given by \( t = E[\text{payoff}|\text{information}] - \text{endowment of shares} \). No one but this agent can know what his or her information is or the endowment of shares, and hence the type of an arriving trader is a random variable \( Z \), a particular realization of which is \( t \). The random variable \( Z \) has a cumulative distribution \( F(.) \) and density \( f(.) \). As will be seen, distributions that satisfy the following will be particularly useful:

\[
\frac{1-F(t)}{f(t)} = a - bt, \text{ for } 0 < t < a/b, a,b>0;
\]

\[
F(t)/f(t) = a + bt, \text{ for } -a/b < t < 0.
\]

For example, \( b=1 \) corresponds to a uniform distribution on \((-a,a)\). Extending the domain of \( b \) to \( b=0 \) corresponds to an exponential distribution. It should be noted that VW use a similar distribution restriction, but the distribution there is the exogenous distribution of quantities. Here it is the exogenous distribution of types.

There are \( N \) identical quoters, supplying liquidity to the market. Supplying liquidity is not costless, however. Specifically, if one of the quoter’s participation in a trade is \( q(t) \), then the cost to supplying liquidity is \( C(t,q(t)) \). Thus, in any symmetric equilibrium the total profit (to all quoters) from a trade from type \( t, Q(t) \), will be \( R(Q(t)) \).
– NC(t,Q(t)/N). It is imagined that this cost arises from two sources. First, there may be trading on private information. Since this private information is included in the type, knowledge of the trader’s type would lead the quoters to revise their expectations concerning the payoff, X, on the security. Of course, quoters do not directly observe type, but having observed a total trade, and knowing that a trader chooses a quantity optimally, the agent’s type can be inferred from the trade. The second source of cost might be thought of as an inventory cost, and a convenient form for this cost is quadratic. Thus, a common specification for the cost function will be:

\[ C(t,q) = e(t)q + \rho q^2/2; \]

Where \( e(t) \) is the revised expectation from seeing a trade from a type t trader:

\[ e(t) = E[X|Z=t]. \]

It is also useful to define the “upper tail” expectation \( E(t) \):

\[ E(t) = E[X|Z>t], t>0, \]

and to note that the derivative of \((1-F(t))E(t)\) is \(-f(t)e(t)\). It is assumed that \(1 > e'(t) > 0\), and \(e(0) = 0\).

There is, of course, a corresponding “lower tail” expectation but it will not be needed in this analysis since the paper will analyze the market for types \(t>0\)—i.e. the paper looks at the offer side. The analysis of the bid side is symmetric.

The measure of welfare to be used in this paper is not uncontroversial. Specifically, the paper will consider a weighted sum of the profits to quoters and the “willingness to pay” (or “consumer surplus”) of the trader averaged over all types t. Thus, if a trader of type t arrives, the quoters receive \(R(Q(t)) – NC(t,Q(t)/N)\), while the surplus to the trader is the integral under his or her demand curve less the amount paid. The total surplus associated with this trader of type t is:
The ex ante welfare is then $E[SUR(Z)]$. Given our assumption about the nature of the individual demand curve, the “willingness to pay” of a trader of type $t$ is merely a monetizing of utility so that it can be compared with the profits of the quoters. What is more controversial is measuring ex ante welfare with the average of the total surplus. In particular, the average willingness to pay is not the same thing as the ex-ante willingness to pay. This measure is used, because it is quite tractable. Those who object, should mentally put quotation marks around the word optimal for the rest of the paper.

The measure allows for different weighting on the quoters and the traders. To allow this seems reasonable. Furthermore, in the derivation of the optimum, maximization of $E[SUR(Z)]$ can be thought of as maximizing trader surplus subject to the quoters earning at least some specified profit level (to cover fixed costs, for example). Choosing the profit level amounts to choosing the weight $w$. With this setup, we can consider the optimal terms of trade in the next section.

### 2. Optimum terms of trade

Consider the simplest case of a linear demand curve, $V(t,q) = t - q$. In this environment trader surplus, at a trade $Q(t)$ is merely $tQ(t) - .5Q(t)^2 - R(Q(t))$, while total quoter surplus is $R(Q(t)) - c(t)Q(t) - .5\rho Q(t)^2/N$. Choosing the optimum terms of trade then consists of choosing the function $R(Q)$ and hence $Q(t)$ via the traders optimality condition to maximize our measure of welfare. It is easier, mathematically, however, to consider the problem as finding the optimal function $Q(t)$ which can then be used to find $R(Q)$. There
are several constraints on the problem. First is the constraint that $R'(Q(t))$ be equal to the trader’s marginal valuation $t - Q(t)$. Second, $Q(t)$ should be non-negative for positive $t$. If this were not the case, then traders would be able to sell at the offer and buy at the bid. However, only quoters are allowed to do this. Thus, we will allow solutions of the form $Q(t) = 0$ for $-t_0 < t < t_0$. This, in effect, allows for the “zero quantity spread” as in Glosten (1994). Third, we will constrain $R(0)$ be zero. To allow this to be positive, for example, would require non-traders to pay for a trade they do not make. Finally, we must have $Q'(t)$ positive if $Q(t)$ is positive and zero if $Q(t)$ is zero for $t$ not equal to $t_0$. This is to ensure that the second order condition holds for the trader’s optimization problem. To see this, note that the second order condition for a trader of type $t$ is $-1 - R''(Q(t)) < 0$, or $R''(Q(t)) > -1$. Differentiating the optimality condition $t - Q(t) - R'(Q(t)) = 0$, shows that $R''(Q(t)) = (1-Q'(t))/Q'(t)$. The constraint above can only be satisfied if $Q'(t) > 0$.

Putting this all together, the welfare maximization problem is:

$$\max \left[ \int_{t_0}^{\infty} f(t) \left[ \frac{1}{2} \left( t - Q(t) \right)^2 - R(Q(t)) \right] + w \left[ R(Q(t)) - e(t)Q(t) - \frac{5 \rho Q(t)^2}{N} \right] \right]$$

s.t. $R'(Q(t)) = t - Q(t), Q(t_0) = 0, t_0$, free, $Q(t) > 0 \Rightarrow Q'(t) > 0$.

Integrate by parts the integral in the maximization. The first term will have the integrand:

$$(1-F(t))[Q(t) + tQ'(t) - Q(t)Q'(t) - R'(Q(t))Q'(t)] = (1-F(t))Q(t)$$

The second term will have the integrand (after substituting for $R'(Q(t))$):

$$w(1-F(t))Q'(t)[t - Q(t) - E(t) - \rho Q(t)/N] =$$

$$= w\left[ (1-F(t))Q'(t) - ((1-F(t))E(t)Q'(t) - [1-F(t)]Q(t)Q'(t)(1+\rho/N).\right]$$

Integrate this expression again by parts (the square brackets surround the “u” term, the second term is “dv”). This yields the integrand:
\[ \text{wf}(t) \{ t \bar{Q}(t) - (1+p/N)\bar{Q}(t)^2/2 - e(t)\bar{Q}(t) - (1-F(t))\bar{Q}(t)(w-1)/(\text{wf}(t)) \}. \]

Maximizing the integral of this merely requires calculus and yields the solution, \( Q_o(t) \):

\[ t - Q_o(t) = e(t) + \rho Q_o(t)/N + ((1-F(t))/f(t))(w-1)/w. \]

In other words, at the optimum, the marginal value to the trader of an additional unit is set equal to the marginal cost of supplying that unit plus a term to ensure the minimum level profit. Solving:

\[ Q_o(t) = \frac{[t - e(t) - (w-1)(1-F(t))/(\text{wf}(t))]}{(1+\rho/N)}. \]

The constraint on the derivative was not used. Since we have in mind a situation in which private information motivates only part of the trade, \( e(t) \) should increase slower than \( t \). For a wide class of distributions, \( (1-F(t))/f(t) \) is non-increasing and hence \( Q_o(t) \) should be increasing. The above also ignores the constraint that the optimum should be non-negative and zero at \( t_0 \). Once the distribution function and \( e(.) \) is specified, \( t_0 \) can be found by setting the expression equal to zero. For example, for \( (1-F(t))/f(t) = a - bt \), and \( e(t) = \alpha t \), the welfare optimum quantity for a trader of type \( t \) is given by \( w > 1 \):

\[ Q_o(t) = B_o t - A_o \text{ for } t > t_0 = A_o/B_o, B_o = (1-\alpha + b(w-1)/w)/(1+p/N), \]

\[ A_o = a(1-w)/[w(1+p/N)]. \]

Except for the case in which \( w \) is less than or equal to one, the optimum involves a small trade spread since \( t_0 > 0 \). This is reminiscent of a CLOB when there is private information. In that case, the small trade spread arises out of quoters’ realization that the first quote will be hit on not only small trades, but large trades as well. Thus the small trade quote recognizes the informational consequences of all sized trades. It is not clear from these expressions, however, whether or not the same logic applies, or whether or not the CLOB will lead to a \( Q(t) \) that is similar to \( Q_o(t) \).
Before going on to the analysis of the CLOB it is useful to consider the aggregate profits to the quoters as a function of the weight, \( w \). Note that the integrand for the quoter profit term is (after integrating by parts):

\[
(1-F(t))Q_o'(t)\{t-Q_o(t) - E(t) - \rho Q_o(t)/N\} = (1-F(t))Q_o'(t)\{(w-1)(1-F(t))/wf(t) + e(t) - E(t)\}.
\]

Since \( E(t) \) exceeds \( e(t) \), a weight of \( w=1 \) would lead to the quoters getting negative profits. This suggests that realistic welfare optima should involve a weight greater than one on the profits, and hence a small trade spread seems likely for the optimum.

The above provided the analysis for the simple case of linear demand and separable cost function with expectation revision and quadratic inventory cost. The more general case is in the following proposition.

**Proposition 1**

Let \( V(t,Q) \) be the demand curve for an individual of type \( t \). Let \( C(t,Q/N) \) be the cost to a single liquidity supplier of providing a quantity \( Q/N \). Then, the welfare optimum quantity purchased by a trader of type \( t \) is the solution to the following equation:

\[
V(t, Q_o(t)) - C_2(t, Q_o(t)/N) - (w-1)(1-F(t))V_1(t, Q_o(t))/(wf(t)) = 0.
\]

Quoter profit is:

\[
\int_{t_0}^{\infty} (1 - F(t)) \{Q_o'(t) - (w-1)(1-F(t))/wf(t)\} V(t, Q_o(t)) - NC_1(t, Q_o(t)/N) dt.
\]

Proof:

Please see appendix.

The above proposition more generally suggests that if we want quoters to have non-negative profits there will have to be a small trade spread. Setting \( w \) less than or equal to one leads to negative quoter profits. Thus we need \( w \) to exceed one. Assuming that \( V(0,0), C_i(0,0) \) are zero it can be seen that \(-(w-1)(1-F(0))/wf(0)\) is not zero for \( w \) larger than one and hence the condition is not satisfied at \((t, Q) = (0, 0)\).
A robust feature of the optimum is that the quantity chosen for the top “type” satisfies marginal value equals marginal cost, and this is independent of the weighting placed on quoter profits. As we will see, this is a feature of the CLOB.

3. Discriminatory CLOB and Uniform Price Clearing

4.1 CLOB

In order to provide the analysis with the minimum complication, I shall describe the equilibrium with the simplest specification—the marginal valuation of a trader is given by \( V(t, Q) = t - Q \) and the cost function for the quoters is given by \( C(t, q) = e(t) - \rho q^2 / 2 \).

The discriminatory limit order book with \( N \) competitors will be considered first.

Let \( 1 - F^*(p) \) be the probability that the next purchase arrival will lead to a stop-out price (highest price) greater than \( p \), and let \( f^* \) be the associated density. Also, let \( e^*(p) \) be the revised expectation of the payoff conditional on the stop-out price being \( p \) and \( E^*(p) \) be the associated upper tail expectation. Consider the problem of quoter number 1. He or she will provide quantity \( q'(p)dp \) at the price \( p \). Thus, the profit to quoter number 1 is:

\[
\int_{p_0}^{\infty} f^*(p) \left\{ \int_{p_0}^{p} q'(s)ds - e^*(p)q(p) - .5\rho q(p)^2 \right\}dp
\]

After integrating by parts, the profit can be expressed as:

\[
\int_{p_0}^{\infty} (1 - F^*(p))(p - E^*(p) - \rho q(p))q'(p)dp
\]

The probability that the stop-out price exceeds a price \( p \) is the probability that a trader’s marginal valuation exceeds \( p \) at the trade \( Q(p) \), the total number of shares offered at the price \( p \) or less. That is, \( 1 - F^*(p) = P\{Z - Q(p) > p\} = 1 - F(p + Q(p)) \). Similarly, \( E^*(p) \) is given by \( E^*(p) = E(p + Q(p)) \). The quoter under consideration considers the quantities supplied
at each price by the other N-1 quoters as given. Thus, \( Q(p) = q(p) + (N-1)q_L(p) \). Thus, from this quota"ers point of view, profits are given by:

\[
\int_{P_0}^{\infty} (1 - F(p + Q(p)))(p - E(p + Q(p)) - \rho q(p))q'(p) \, dp
\]

Maximizing this is a simple calculus in variations problem. The derivative of the integrand with respect to \( q(p) \) is:

\[
q'(p)f(p+Q(p))\{-p + e(p+Q(p)) + \rho q(p)\} - \rho(1-F(p+Q(p)))q'(p).
\]

The derivative with respect to \( q'(p) \) is:

\[
(1-F(p+Q(p)))(p-E(p+Q(p))-\rho q(p)).
\]

After taking the derivative of this latter expression and setting it equal to the first expression, summing over all quoters one gets that the total amount supplied at a price \( p \) or less \( Q(p) \) is given as the solution to the differential equation:

\[
f(p + Q(p))(1 + \frac{N-1}{N} Q'(p))(p - e(p + Q(p)) - \rho Q(p)/N) = 1 - F(p + Q(p))
\]

Recall that \( Q(p) \) is the quantity offered at price \( p \) or less. Thus, \( p \) is the marginal price for a trade of size \( Q(p) \). Now make two changes of variable. First, define the marginal price, by

\[
p = R'(Q(p))
\]

and, define the function \( p(Q) \) by \( Q(p(Q)) = Q \). Evaluating at \( p(Q) \) we have:

\[
f(R'(Q) + Q)(1 + \frac{N-1}{NR''(Q)})(R'(Q) - e(R'(Q) + Q) - \rho Q / N) = 1 - F(R'(Q) + Q).
\]

Now evaluate the above at \( Q_L(t) \) the traders optimum: \( t - Q_L(t) = R'(Q_L(t)) \) and note that the trader’s first order condition implies that \( 1 - Q_L'(t) = R''(Q_L(t))Q_L(t) \). After substituting:

\[
t - Q_L(t) - e(t) - \rho Q_L(t)/N - \frac{N(1-Q_L'(t))(1-F(t))}{N-Q_L'(t)} \frac{1}{f(t)} = 0
\]
Before examining this expression, which looks remarkably like the expression for the optimum, it is useful to get some intuition for how the competition between strategic quoters works in this market. Consider the effect of one quoter adding a small amount \( h \), at the price \( p \). If this quantity transacts at the price \( p \), then the profit per unit is \( p - E(p + Q(p)) - \rho q(p) \). The upper tail expectation is used since this quantity will transact if the stop-out price is \( p \) or larger. The probability of this happening is \((1-F(p+Q(p)))\).

Thus, the effect on expected profits at \( p \) is \((1-F(p+Q(p)))(p-E(p+Q(p)))-\rho q(p))\). However, the addition of \( h \) shares at \( p \) shifts the whole schedule for prices larger than \( p \). Now, in order to have a quantity at price \( s \) picked off, the type has to be \( s+Q(S)+h \) or larger. At each price \( s \), the marginal effect on expected profits is (since \( q'(s)ds \) is offered at \( s \))

\[
(1-F(s+Q(s)+h))\{s-E(s+Q(s)+h)-\rho (q(s)+h)\}q'(s)ds. \quad \text{For } h \text{ small, the effect on profits is:}
\]

\[
hq'(s)f(s+Q(s))\{s + e(s+Q(s)) + \rho q(s)\} - \rho(1-F(s+Q(s)))q'(s)ds. \quad \text{Integrating over all prices larger than } p \text{ provides the total marginal effect of an increase in quantity at a price } p \text{ on the expected profits at all larger prices:}
\]

\[
(1 - F(p + Q(p)))(p - E(p + Q(p)) - \rho q(p)) + \\
+ \int_{p}^{\infty} q'(s)(f(s + q(s))(-s + e(s + Q(s)) + \rho q(s)) - \rho(1 - F(s + Q(s)))
\]

At the optimum, the expected marginal effect at the price \( p \) and all higher prices should be zero. Taking the derivative of the above provides conditions identical to the ones analyzed above.

Consider the case of \((1-F(t))/f(t) = a - bt\), and \( e(t) = \alpha t \). There is a linear solution, given by \( Q_L(t) = B_L t - A_L t > A_L/B_L\), \( A_L = a(1-B_L)/(1+\rho/N)(1-B_L/N)\), and \( B_L \) satisfies the quadratic equation \( B^2 - BN(1+(1-\alpha)/(1+p/N)+b/(1+p/N))+N((1-\alpha)/(1+p/N) + b/(1+p/N)) = 0. \)
Proposition 2, below, provides the condition for general demand curve \( V(t, Q) \) and cost function \( C(t,q) \).

Proposition 2

The equilibrium quantity traded by a trader of type \( t \), \( Q_{t}(t) \) satisfies the following differential equation (the arguments of \( V(t,Q) \) and \( C(t,Q) \) have been suppressed to help legibility:

\[
0 = V - C_{2} - \frac{N(V_{1} + V_{2}Q'(t))}{(NV_{1} + V_{2}Q'(t))} V_{1} \frac{1 - F(t)}{f(t)}
\]

Proof:

As with proposition 1, one can see that \((t,Q) = (0,0)\) does not satisfy the equation and hence the equilibrium in the CLOB looks much like the optimum. It is also interesting to note that if there is a maximum type, \( T \), then \( 0 = V(T,Q(T)) - C_{2}(Q(T)) \), and this is independent of the number of competitors. For the maximum type, marginal valuation is equal to marginal cost of taking the other side of the trade.

4.2 Uniform Price Clearing

Interestingly enough, the analysis of the uniform clearing price equilibrium is far more complicated than the discriminatory CLOB with endogenous trade. It is also far more complicated than the analysis of the uniform price clearing equilibrium with exogenous trade. With exogenous trade, the equilibrium is independent of the trade distribution. This is not the case with endogenous trade. The reason is that with endogenous trade, an agent adding quantity at a particular price has two effects. First, it increases his or her share of the order flow, but it also encourages greater order flow. How much greater order flow depends upon the distribution of the traders’ types.
The analysis is carried out only for the special case of linear demand curves of traders and the special cost function arising out of private information and inventory costs. As before, we have that a typical agent, taking the actions of others as given maximizes \((f^*(p)\) is the endogenously determined distribution for the stop out price and \(e^*(p)\) is the revision in expectations if the stop out price is \(p\)):

\[
\int_0^\infty f^*(p)(pq(p) - e^*(p) - \rho q^2)dp
\]

Notice that in this formulation, \(p\) is the average price rather than the marginal price in the CLOB analysis. Integrating by parts one obtains:

\[
\int_0^\infty (1 - F^*(p))(pq'(p) + q(p) - q' E^*(p) - \rho q(p)q'(p))dp
\]

Now, however, \(1 - F^*(p)\) is given by the following (recall that \(F\) is the distribution of trader type):

\[
1 - F^*(p) = 1 - F(p + Q(p) + \frac{Q(p)}{Q'(p)}) \quad \text{and similarly for } E^*(p), \text{where } Q(p) \text{ is the total quantity offered by all } N \text{ competitors when the stop out price is } p. \text{ Taking the other } N-1 \text{ quantities as given, a typical quote maximizes the above. As before, this is a calculus of variations problem. After finding the first order condition, and then making the same change of variables as in the CLOB analysis, and manipulating the result one obtains that the equilibrium quantity purchases by a trader of type } t \text{ is the solution to the differential equation:}
\]

\[
\frac{N-1}{N}(t - Q(t) - Q(t)P'(Q(t)) - e(t) - \rho Q(t)/N) f(t) = \frac{Q(t)P'(Q(t))}{N} (f(t) + \frac{d}{dt} f(t)(t - Q(t) - e(t) - \rho Q(t)/N))
\]
as well as $P'(Q(t))Q(t) + P(Q(t)) = t - Q(t)$. What makes this expression different than the exogenous trade case is the inclusion of the derivative term on the right hand side of the equation. Without that term, the density of the type would disappear. The important thing to note about this expression is that it implies that in equilibrium there is no zero quantity spread. To see this, suppose that there is a $t^*$ with $Q(t^*)=0$ but $t^*>0$, in which case $R'(0) = P(0) = t^*$. The right hand side becomes zero, while the left hand side is proportional to $(t^*-e(t^*))f(t^*)$. Under the assumptions that we have made about $e(.)$ this expression is positive and hence the first order condition is not satisfied. Thus, there are fundamental differences between the uniform price clearing and the CLOB. For the uniform price clearing competition, equilibrium is tied down by $P(0) = 0$, or price is equal to marginal cost. In the CLOB, the equilibrium is tied down by $V(T,Q(T)) = C_2(T,Q(T))$. Interestingly, the only way that the uniform price clearing and CLOB can both lead to quantities linear in the type, $t$, is if the distribution of types is uniform. These observations motivate the following welfare analysis.

4.3 Welfare Analysis

The next proposition provides a comparison of the welfare optimum and equilibrium in the CLOB. The main result of this paper is that in an environment with linear demand curves and the cost functions considered above the CLOB provides the welfare optimum for a weight $w(N)$ if and only if the welfare optimum takes the form $Q_o(t) = B(w)t - A(w)$ for some constants $B$ and $A$ depending upon the welfare weight, $w$.

**Proposition 3:**

Suppose that a trader of type $t$ has marginal valuation $t-Q(t)$. Further suppose that the cost function for the quoters is of the form $e(t)q(t)+\rho q(t)^2/2$. Then, the optimum for some
weight \( w(N) \) and number \( N \), of participants is implemented by the CLOB with \( N \) competing quoters if and only if the optimum quantity traded by a trader of type \( t \) is affine in \( t \) for all weights \( w \).

**Proof**

Suppose that the optimum is affine in \( t \) for all \( w \). Then examination of the equation that the optimum must satisfy indicates that both \( e(t) \) and \( (1-F(t))/f(t) \) must be affine. But then, there is a linear solution to the CLOB equilibrium. One can verify that the equilibrium and optimum correspond for some weight \( w \).

Suppose that the optimum satisfies the differential equation for the CLOB. One can then verify that \( e(t) - (w-1)(1-F(t))/wf(t) \) must be affine.

A related observation is that uniform price clearing can never be optimal as long as there is private information. If the optimum is of the form \( Q(t) = bt \), then \( w \) must be equal to 1 and it was shown above that this leads to negative profits. This can certainly not be a feature of the equilibrium in the uniform price clearing.

5. *Discussion*

The limit order book form of market (though not centralized) appears to becoming the dominant form of trading throughout the world. This is happening on both a decentralized basis and by regulatory fiat. The prime example of the latter is the adoption by Nasdaq of new order handling rules. This adoption was largely forced, and was the result of alleged non-competitive improprieties on the part of Nasdaq dealers.

Nonetheless, the above analysis suggests that the move by the SEC was a good one—
everyone can be made better off (though it is not clear how the dealers will receive their side payments). It could be argued that Nasdaq is evolving to a hybrid market like the NYSE with an active limit order book and an active dealer. However, my view of Nasdaq is that the dealers are primarily handling orders not considered in this analysis, orders in which there is some amount of non-anonymity. My impression is that the same can be said for London.

Linearity plays an important role in the welfare analysis, which raises the question of how likely it is that equilibrium can be characterized as being linear. At first, the answer would appear to be no. For example, studies have found evidence that there is more price response to medium sized trade than to large or small trades (Peterson, Barclay--). However, that statistical analysis could not break out orders that were crossed upstairs from those transacted on the floor. Theoretical arguments (Seppi) would suggest somewhat less price response from the large upstairs orders. It is also possible that one would see larger trades when the price response to trade is low and somewhat smaller trades when price response is large—i.e. in the statistical analysis one cannot take the trade size as being exogenous to changes in the informational environment (Madhavan).

From a theoretical perspective, there are reasons to think that non-linear price response functions are not stable in that they tend to invite price manipulation trading strategies. If this is true, then there may be an “equilibrium” distribution of “types” so that the price response function is linear.

While the model considered above involved single arrivals, it can be modified to allow for aggregation of trade. In this case, type becomes an aggregate type. If there is
this trivial reinterpretation of the model, then the fact that virtually all exchanges that have an opening clearing use a uniform price procedure is puzzling. I suspect that it may have to do with issues that the above model is not immediately equipped to handle. Perhaps it revolves around the fact that with an aggregation of demand there is uncertainty about what the terms of trade will be. In the single arrival model, the individual trader knows exactly what price he or she will pay for each share sold. This is not the case when orders from a variety of random types are aggregated. This would be an interesting avenue to pursue.

6. Conclusion

This paper provides a model in which to analyze the optimality of various market structures when trade size is determined optimally rather than given exogenously. The main result is that when the optimal market structure involves a linear price function, the CLOB with discriminatory pricing (each limit order pays or receives its quote) implements the optimum. Thus, the CLOB, in some environments, is both “inevitable” and “optimal.” The analysis shows that a uniform price limit order book (each limit order pays or receives the stop out price) will not implement the optimum.