

# A Market-Clearing Role for Inefficiency on a Limit Order Book

Jeremy Large\*

JEREMY.LARGE@ECONOMICS.OXFORD.AC.UK

*All Souls College, University of Oxford, Oxford, OX1 4AL, U.K.*

14 July 2006

## Abstract

Using a stochastic sequential game in ergodic equilibrium, this paper models limit order book trading dynamics. It deduces investor surplus and some agents' strategies from depth's stationarity, while bypassing altogether agents' intricate forecasting problems. Market inefficiency adjusts to induce equal supply and demand for liquidity over time. Consequently, at a given bid-ask spread surplus per investor is invariant to faster, more regular or more sophisticated trading, or modified queuing rules: apparent improvements are offset as inefficiency adjusts back to market-clearing levels. Moreover, investor surplus decreases with the spread. In the model, price discreteness fixes the spread at the tick size. Narrowing the tick is beneficial, but may be resisted by sell-side traders.

*JEL classification:* C73, G14, G24

*Keywords:* stochastic sequential game, ergodic equilibrium, market microstructure, limit order book, market depths, bid-ask spread.

---

\*This paper previously appeared under the name: *Price Tick and Welfare when Assets Trade on a Penny*. I am grateful to Thierry Foucault, Meg Meyer, Christine Parlour and Hyun Shin for insightful conversations about this work, as well as to seminar participants at HEC Paris (December 2005) and Warwick Business School (November 2005). I also thank participants at the Dauphine/Banque de France conference, Paris (June 2006). I thank the Bendheim Center for Finance for accommodating me at Princeton University, and the FMG for accommodating me at the London School of Economics, during part of the writing. I gratefully acknowledge financial support from the US-UK Fulbright Commission and the Economic and Social Research Council, UK.

# 1 Introduction

This paper analyzes a natural game-theoretic model of trading dynamics on an electronic limit order book. While equilibria have previously been simulated, realistic specifications are intricate and have so far been viewed as analytically intractable. Nevertheless, the paper derives a subset of equilibrium strategies, yielding a simple expression for investor surplus. Above all, the solution exploits the repetitive stationarity of market microstructure over time. This stationarity implies that although traders can choose strategically between making and accepting offers to trade, and condition this choice on rich information about the order book, the resulting supply of liquidity (of offers to trade) is competitive, in the sense that it obeys a market-clearing condition involving an equality of intensities. It has been recognized that this has far-reaching consequences for market microstructure.<sup>1</sup> However, market-clearing has often been built into definitions of equilibrium, relying on an exact law of large numbers which would suppress important dynamics here.

Market-clearing implies that the order book's bid-ask spread *prices* (the right to forego) the frictional costs of making a public offer to trade. In fact, the half-spread bounds frictions per trader above, since exactly half of all traders undergo trading frictions, and for this half the spread is *ex ante* costlier than the frictions. In the absence of adverse selection or order submission fees, these frictions are identically the market's inefficiencies. So the spread is a measure of the market's endogenous shortfall in welfare, versus an ideal first-best.

In practice, institutional details such as price discreteness impede the bid-ask spread from adjusting on its own to market-clearing levels. Rather, the *inefficiency itself* also adjusts to clear the market for liquidity. Focusing on this phenomenon, the paper models the case where the spread is actually fixed at the price tick size. It defines an ergodic equilibrium in a stochastic sequential game, where some agents choose not to observe (for a fee) the market state. It infers from ergodicity a set of quantiles giving these agents' best response, without tackling their associated, complex, forecasting problem. This leads to a concise characterization of investor surplus, depending only on the spread. Inefficiency per trader is a little under half the spread.

**Trading on a penny** Assets are said to trade on a penny when the bid-ask spread is bid down to equal its regulatory minimum, the price tick size (a cent, five cents etc.), almost all the time. BNPParibas equity on Euronext Paris, Vodafone on the LSE, and 10-year US Treasury Bond Futures at CBOT<sup>2</sup> are current examples. Figure 1 shows the Vodafone case. The average number of offers to trade that are simultaneously displayed at the best bid and ask prices, tends

---

<sup>1</sup>See for example Duffie, Gârleanu, and Pedersen (2005), Vayanos and Wang (2003) and Weill (2003).

<sup>2</sup>The Chicago Board of Trade.

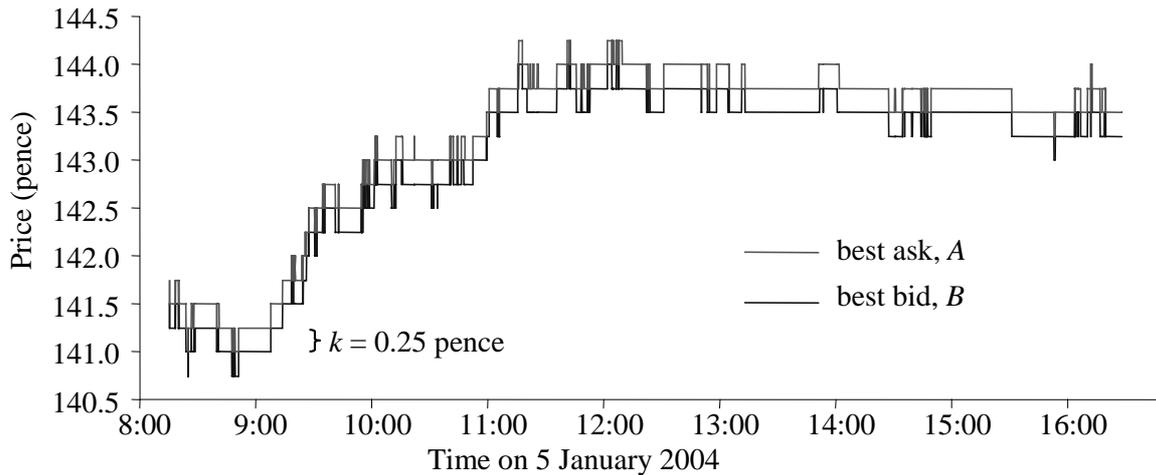


Figure 1: The prices of the best bid and best ask for Vodafone shares on the London Stock Exchange, 5 January 2004. The number of transactions on this day was 14 times higher than the number of changes in the bid (ask). The bid-ask spread was wider than 0.25 pence 3.5 per cent of the time. The notation,  $k$ , indicates the price tick size, and bid-ask spread.

here to be large. More succinctly, inside depth is high – an effect noted in Harris (1994). In the CBOT example, on a typical day in 2004 the median-sized trade consumed under 1.5 per cent of the available inside depth.<sup>3</sup> Moreover, depths have complex dynamics with substantial variation: for example, one or the other inside depth falls entirely to zero each time the best bid or ask price changes. Nevertheless, all this can be described by a partially-tractable stochastic sequential game, similar to ones simulated in Goettler, Parlour, and Rajan (2004) and (2005).<sup>4</sup>

**Policy implication** In this setting, the exchange can directly alter the bid-ask spread by adjusting the price tick, so far as this would leave the asset still trading on a penny. So, would cutting the price tick be more or less efficient? What parties might resist a cut? What alternative measures would enhance investor welfare? In fact, in the model the price tick size has a uniquely strong effect on welfare and should be cut, at least so far as trading is still on a penny.

**The sell-side and the buy-side** Surpluses in this game turn out to depend on a distinction between buy-side investors and sell-side traders. Sell-side traders are experts in buying or selling one or a few assets, who execute trades for a fee on behalf of a variety of non-specialized client

<sup>3</sup>For its average-sized trade, this figure was 5 per cent. Bid-ask spreads were exactly the price tick over 99 per cent of the time (based on three days in July / August 2004). For BNPParibas, the figure for the median trade was less than 4 per cent, while for its average-sized trade (including out-of-hours trading), the figure was 9.5 per cent (Based on the first three trading days in September 2004, source: TickPlus Data). If BNPParibas' price falls below 50 Euros, its tick size will fall and it may cease to trade on a penny.

<sup>4</sup>Where possible, Goettler *et al.* (2005)'s key notation is used in developing the model.

investors, known as the buy-side.<sup>5</sup> It is the surplus of buy-side investors which is derived in a closed form expression. The only observable quantity this refers to is the spread, in which it is decreasing. However, the sell-side prefers to safeguard the value of its queueing expertise with a tick size which is inefficiently bounded away from zero. So there may be resistance on the sell-side to an efficient policy of cutting the price tick and spread.

The average surplus of buy-side investors is invariant to a range of *prima facie* beneficial policies: for example, attracting more traders who trade faster, more patiently or more regularly; encouraging them to acquire sell-side skills; or making available richer order book data (e.g. Level 2). Within wide bounds it is invariant to order queueing rules (FIFO or otherwise). When these features change, the market's endogenous inefficiency adjusts to produce equilibrium effects which offset any *prima facie* benefits, as it is governed by a market-clearing constraint.

**Consequences of the market microstructure's stationarity** This formulation in terms of 'market-clearing' follows *inter alia* Foucault, Kadan, and Kandel (2005), which studies the limit order book as a dynamic market where suppliers of liquidity (or immediacy) submit limit orders, and demanders of liquidity submit market orders which trade with them.<sup>6</sup>

Now, seasonality aside, *a limit order book's depth is stationary* (while its prices may be non-stationary). So the volume per unit time of market orders it attracts must be matched on average by an equal flow of uncanceled limit orders. Otherwise, depths would deplete or explode arbitrarily in finite time. This must be true separately at the bid and at the ask. It can also be seen as an accounting identity: that since for every market order there is an uncanceled limit order, and vice versa, they arrive equally frequently over time.

This observation amounts to a competitive market-clearing condition at the bid, and another at the ask, whereby the ergodic distribution of market states adjusts in equilibrium to induce from traders equal supply and demand for liquidity over time. Section 2 will cash this out theoretically by showing how stationary depths are consistent with only one best response from buy-side investors who submit their orders 'blind' – that is, without conditioning their action on current prices and depths. It follows that in fact the distribution of market states adjusts to produce a market-clearing level of inefficiency.

So the tick size impacts welfare as follows: Being equal to the bid-ask spread, it is the penalty paid for immediate execution. Hence if cut it encourages too many market orders, unless inefficiency falls making limit orders *ex ante* more attractive – equivalently, increasing *ex ante*

---

<sup>5</sup>This semantic distinction is blurred in practice. For example, some hedge funds are viewed as being sell-side, though they delegate trade execution to others (I'm grateful to Richard Lyons for pointing this out).

<sup>6</sup>Bids and asks together constitute 'limit orders'. When a trader submits a 'market order' she makes an immediate sale or purchase by accepting the terms of trade publicly offered by a bid or an ask, respectively.

welfare. By contrast, altering the trading population, the queuing rules, or the quality of trader information leaves the effective spread, and therefore also the value of submitting a market order, unchanged. It therefore can have little effect on the value of a limit order – or supply and demand for immediacy would not equate. If it does not affect the *ex ante* value of either a limit order or a market order, nor does it affect *ex ante* welfare per trader or market efficiency.

**Empirical predictions** In the model, inefficiency arises when limit orders wait for a counterparty. Waiting increases in depth, which therefore emerges as the key observable market-clearing feature of the market state. This is a corollary not a foundation of the main theoretical analysis, which moves straight to a characterization of buy-side surplus. Section 5 analyzes depths' role later and provides predictions in line prior empirical findings: that when a limit order book trades on a penny, average inside depths increase with

- trading volume or order intensity (limit order plus market order flow), and
- the price tick size.

Both these results accord with intuition, once depth is seen as a market-clearing variable. Inside depth's positive dependence on trading volume is observed on the SEHK in Brockman and Chung (1996). See also Danielsson and Payne (2002)<sup>7</sup> and Lee, Mucklow, and Ready (1993) which, like Kavajecz (1999), focuses on the hybrid NYSE exchange. Inside depth's positive dependence on the price tick size on the NYSE was estimated in Harris (1994) and studied in Goldstein and Kavajecz (2000). None of the above studies is of trading on a penny exclusively.

**Related market microstructure theory** Important theoretical antecedents to this work are in Bernhardt and Hughson (1996), Kadan (2005) and Seppi (1997) which study tick size, inside depth and welfare in dealer and hybrid markets. Foucault, Kadan, and Kandel (2005) explain a channel whereby a large tick size on a limit order book can enhance welfare. A simulation experiment of Goettler *et al.* (2005) suggests that the aggregate effect of a large tick can be to diminish welfare, even when trading is not on a penny. In their model, as here, welfare is decreasing in depth. This runs counter to the intuition that high depths are desirable since unusually large market orders can trade at or near the quotes. Nonetheless this may be an important consideration in some cases – see Jones and Lipson (2001).<sup>8</sup>

---

<sup>7</sup>At a 20 second frequency Danielsson and Payne (2002) gives nuanced evidence on depths and order flow dynamics for Reuters FX markets. It does not fully isolate inside depths.

<sup>8</sup>The intuition for sell-side preferences also differs from previous studies: before the decimalization of the NYSE floor exchange as reported in Harris (1997), which unlike here was not expected to leave markets trading on a penny, the opportunity legally to 'front-run' outsiders' limit orders via small price improvements inside the quotes may have drawn constituencies on the sell-side to a smaller tick size. Domowitz and Wang (1994) have analyzed the electronic limit order book as the length of a queue.

In a literature beginning with Kyle (1985) and Glosten and Milgrom (1985), exchange dynamics are studied when some agents obtain elsewhere superior information about the asset’s fundamental value. This is particularly realistic for many equities and corporate bonds. Here, by contrast, sell-side traders have superior information only about future order flow, which they infer rationally from the current market state. Therefore there is no winner’s curse, or picking-off risk in limit order execution: for studies of this effect see Foucault (1999) and Handa and Schwartz (1996). Later, Section 7 introduces moderate adverse selection about a fundamental value to the asset, which it finds may in fact *enhance* trader surplus here.

The paper proceeds as follows: Section 2 sketches the model and gives the main intuitions. Section 3 details the model and equilibrium. Section 4 gives the main results on investor surplus. Section 5 derives average inside depths in a simplified model. The central proof is in Section 6 (details are in an Appendix). Section 7 relaxes leading assumptions. Section 8 concludes.

## 2 Model summary, main result, and intuitions

The model is an ergodic equilibrium in a stochastic sequential game, where as in Foucault *et al.* (2005), Goettler *et al.* (2004) and (2005), Hollifield, Miller, and Sandås (2004), Parlour (1998) and Rosu (2004), impatient traders arrive at a limit order book sequentially and may alter its state. They face a simple trade-off between a market order, and a limit order at a better price which must wait for execution. Other aspects of this trade-off are analyzed in Cohen, Maier, Schwartz, and Whitcomb (1981) and Chakravarty and Holden (1995).

Technically, the modeling innovates by providing a closed-form expression for investor surplus in this dynamic setting (with order cancellation), and generalizing to endogenously fluctuating trader arrival intensities, as advocated by Foucault *et al.* (2005). The use of ergodicity constraints to infer best responses in a stochastic sequential game without solving the forecasting or decision problem, is (to the best of my knowledge) a new approach here.

### 2.1 Model summary

Equally impatient traders arrive according to a stochastic point process, initially in complete ignorance of market prices and depths. Because they cannot condition on the current microstructure of the market, at this stage identify them as ‘buy-side investors’. However, the monopolistic exchange *ex ante* offers each arriving trader order book information for a fee. Define trading as ‘sell-side’ if it is thus informed. The trader accepts with mixing probability  $\pi$ . This is the only source of asymmetric information. In the manner of Parlour (1998), the trader then draws an

independent type  $\beta \in \mathbb{R}$  giving a private, signed, gain from trade, from a continuous distribution  $F$  which is symmetric about zero.

Responding optimally to current depths and prices if she bought sell-side skills, and (as it will be termed throughout the paper) ‘blind’ otherwise, she places an order of unit quantity, choosing from among a bid or an ask, or a market purchase or sale. She may not submit ‘speculative’ limit orders outside the best quotes. Limit orders can be cancelled and replaced periodically. The trader aims to maximize her gain from trade, which is discounted for delay once it has been netted of her implementation shortfall or slippage. This stochastic sequential game is detailed in full in Section 3.

## 2.2 Main result

For any  $\pi > 0$  the existence of an equilibrium is proven where all buy-side investors use the same mixing probability,  $\pi$ , and which is entirely symmetric in buying and selling, with sufficient negative feedback in depths for them to be ergodic. The detail of this complex, dynamic equilibrium is not solved for, but a closed-form expression for buy-side investor surplus is derived, and refers only to the price tick, denoted  $k$ , relative to two measures of the dispersion in  $F$ . It decreases in  $k$ . This result is given in Theorem 4.2.

## 2.3 Intuition in the case where $\pi \downarrow 0$

Though it is later proven more generally, the central welfare result can quickly be deduced in the limiting case where  $\pi \downarrow 0$ : that is, in the case where the exchange goes dark. Hendershott and Jones (2005) study such an event on the Island ECN for Nasdaq in 2002. Suppose that  $\pi$  is very small, so that the vast majority of buy-side investors trade blind. More extreme values of  $|\beta|$  face higher stakes and so greater absolute aversion to execution delay. Hence blind traders who draw  $\beta$  in  $F$ ’s two tails prefer market orders, which execute immediately, to limit orders. Having foregone all informative updating, they will not cancel any limit order they submit. As remarked in the Introduction, the ergodic equilibrium must be such that exactly half the traders submit limit orders, and half market orders. Almost all of them are blind. Thus in a buy-sell symmetric and ergodic equilibrium, the tails where market orders are preferred blind are (almost exactly) the exterior of  $F$ ’s interquartile range (as they contain half the probability weight). The equilibrium strategy of blind traders is illustrated in Figure 2.

Given the continuity in the traders’ decision problem, at the cutoffs in Figure 2 the median buyer and seller, who have  $\beta \cong \pm F^{-1} \left[ \frac{1}{4} \right]$ , are when blind indifferent between a limit order and a market order priced one tick less favorably. This fixes all types of blind traders’ expected

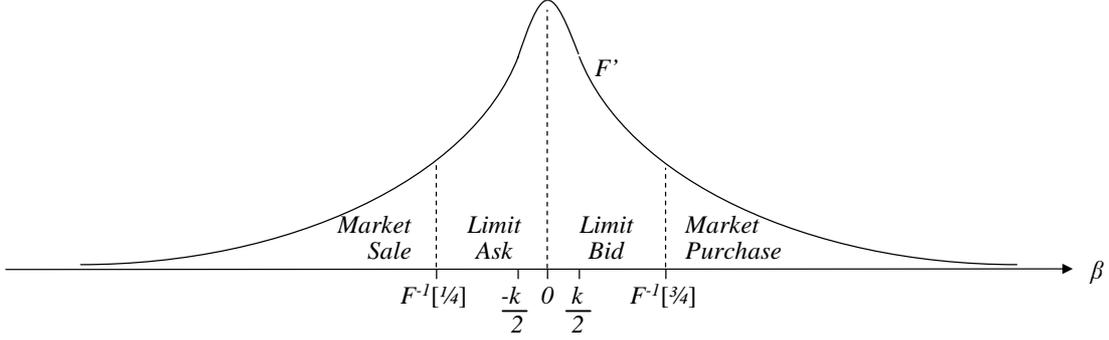


Figure 2: Illustrates the type-space, and the equilibrium strategy of blind traders. The four delimited areas under  $F'$  are equal.  $F^{-1} \left[ \frac{3}{4} \right]$  is indifferent between market purchasing and bidding, and is the ‘median buyer’. Likewise,  $F^{-1} \left[ \frac{1}{4} \right]$  is the ‘median seller’.

discount factor due to the delay in limit order execution. *Ex ante*, they also know there is a  $\frac{1}{2}$  chance they will choose to experience this discount factor via a limit order. Combining these last two deductions, their *ex ante* surplus then follows trivially. Were it not for exchange fees, this would be exceeded by that of the small minority who acquire sell-side information; but in fact the fees absorb exactly this excess. Buy-side welfare is therefore simply the blind traders’ *ex ante* surplus.

This argument, and so its conclusion, is invariant to many relevant policy variables, with the exception of the price tick size, in which buy-side surplus is decreasing. Cutting the tick size affects welfare because, uniquely, it makes market orders strictly more attractive. All other things being equal, the blind median buyer and seller would switch to market orders. This would imply a shortage of limit orders over time, non-stationary depths and no equilibrium. In equilibrium, therefore, endogenous frictions adjust to restore the median buyer and seller to indifference (for example, average depths fall). Hence limit orders also become *ex ante* more attractive – which necessarily involves an increase in *ex ante* welfare.

Note in passing that if  $\pi = 0$  then no feasible equilibrium exists because all traders are blind, and depths follow a non-stationary random walk with at best zero drift.<sup>9</sup>

## 2.4 Intuition in the case of arbitrary $\pi$

The blind trader’s best response as depicted in Figure 2 is in fact exact and invariant to  $\pi > 0$  under the conditions of the model. Repeating the argument of the previous paragraphs, buy-side trader surplus is also. The full model takes a realistic parametrization of  $F$  and introduces some identifying restrictions on actions and payoffs: Traders may cancel limit orders only in favor

<sup>9</sup>The pathological case where depths are implausible leads to explosive spreads. This emerges in the detailed model as causing the exchange to hold an explosive position in the asset (see Section 3.6).

of alternative limit orders. They measure implementation shortfall relative to the infeasible benchmark where they trade immediately at the current mid-quote.

The invariance to  $\pi$  follows approximately for reasonable choices of  $F$ , even if they violate our parametric assumption (given in Section 3.9). The intuition for this is as follows: for each possible state of the market, buyers with types above some indifferent cutoff prefer market buys over bidding (as in Figure 2, a special case where the cutoff is  $F^{-1}[\frac{3}{4}]$ ). These indifferent cutoff types differ from state to state of the market. They must have as an average both the median buyer,  $F^{-1}[\frac{3}{4}]$ , and the equivalent cutoff in the blind state. The former average holds so that markets for liquidity clear, the latter because she faces the average of their problems. These two averages may not be arithmetic, but are generically close (indeed for the form of  $F$  as given in Section 3.9 they are both hyperbolic and so coincide) and so  $F^{-1}[\frac{3}{4}]$  is roughly indifferent in the blind problem. The same argument holds for the sellers, and the welfare characterization follows directly from the blind strategy as before. The paper's main proof proceeds along these lines.

## 2.5 The sell-side

The invariance to  $\pi$  means that the sell-side pays for itself, in the sense that for a given price tick the fees charged for its services equal the efficiency gain it produces in equilibrium. This is bounded above by the buy-side welfare shortfall compared to first best, which  $\rightarrow 0$  as the price tick,  $k$ ,  $\rightarrow 0$ . Therefore sell-side traders prefer a price tick bounded away from zero. Their incentives conflict with efficiency considerations and the interests of the buy-side: in some cases they may resist a cut in the price tick or bid-ask spread.

## 3 A dynamic limit order book model

The main intuitions having been laid out in the last Section, this Section gives a precise definition of the model. Time, denoted  $t$ , is continuous and runs on indefinitely from 0. Risk-neutral buy-side investors arrive at the market one by one according to a point process of stochastic intensity  $\{\lambda N_t : t \geq 0\}$ . The parameter  $\lambda > 0$  is a constant, but  $N$  is a process capturing fluctuating trader numbers, which inhabits the fixed real interval  $[N^-, N^+] \subset \mathbb{R}^{++}$ .

Foucault *et al.* (2005) advocates letting  $N$  vary as a function of the current market state, since some states might attract order flow, while others repel it (the market state space,  $\Omega$ , is defined in Section 3.2). Subject to a buy-sell symmetry constraint introduced in Section 3.4, allow arbitrary functional dependence of this sort, and suppose it is common knowledge. At any given time  $t$ ,  $N_t$  determines whether trading is fast or slow in the market. There is no heterogeneity in the order flow only in the special case that  $N$  is constant.

Traders all discount future payoffs at a rate  $\rho > 0$  (Section 7 presents an alternative specification where time preferences stem from linear waiting costs – the main results of the paper are unchanged). They arrive in complete ignorance of the market and are informed of a price  $c$  that must be paid for the right to receive sell-side information,  $b$ , about the limit order book on entry to the market, and on any subsequent re-entry. They buy the signal  $b$  with some probability  $\pi$ . The content of the signal,  $b$ , is defined later in Section 3.3. With probability  $(1 - \pi)$  a trader does not acquire  $b$  and subsequently just trades ‘blindly’ in the market. This is currently practiced by some hedge funds. In expectation the exchange gathers  $c\pi$  from the trader in fees. The amount  $c\pi$  will be called ‘sell-side rent’, because it is the value extracted from buy-side investors in return for sell-side skills and information.

Traders then independently draw a private motive to trade the asset,  $\beta$ , from a distribution on  $\mathbb{R}$ , whose continuous cumulative density function (CDF), denoted  $F$ , is symmetric about 0 and is detailed in Section 3.9. A positive (negative)  $\beta$  predisposes them toward buying (selling). In reality, the buy-side cannot acquire sell-side information and skills only for the purpose of a single trade, so in the model they commit to  $\pi$  without conditioning on their current trading need, i.e. on their type,  $\beta$ . The next part shows how  $\beta$  enters payoffs.

### 3.1 Traders’ objectives

On discerning  $\beta$  each successive trader places an order of (once normalized) unit volume if they wish. The price of the order must be an integer multiple of the exchange’s minimum admissible price increment,  $k$ , and must be in the interval  $[Mk, Qk]$ , for potentially large natural numbers  $M$  and  $Q > M$ . The order must be one of four types: a buy limit order (bid), a buy market order, a sell limit order (ask), or a sell market order. At execution, traders who buy at price  $p$  receive a payoff equal to the private gain  $\beta$  minus their implementation shortfall, or slippage, relative to the mid-quote,  $m$ , i.e. :

$$\beta - (p - m). \tag{1}$$

Sellers at price  $p$  receive the negative of this,  $-\beta - (m - p)$ . So the criterion determining trader behavior and welfare is implementation shortfall as it is commonly calculated in practice: i.e. relative to a measure of the mid-quote. No market order enjoys negative transaction costs, as some do in the simulation of Goettler *et al.* (2005). Later, the mid-quote as perceived by buyers will differ from that of sellers when the bid-ask spread is wide.

## 3.2 Trading dynamics

A market order executes immediately, but can only be submitted at prices where a countervailing offer to trade (i.e. a limit order) is still outstanding. Limit orders may be submitted at any price, but have to wait for execution until a future trader submits a market order at their price. If a market order is submitted at a price where there is more than one countervailing limit order, it trades with one of them according to an allocation rule that gives weak preference to orders submitted earlier. This excludes a LIFO rule. Typically, one can have in mind the usual, queueing, case where the market order transacts with the limit order that was first submitted. However, an alternative ‘coin-flip’ rule is also admissible.

## 3.3 Definitions

- Denote by  $B$  the highest price of bids, and by  $A$  the lowest price of asks, so that  $A \geq B + k$ . The best bid and ask,  $B$  and  $A$ , follow a stochastic process over time.
- Denote by  $b^B$  the number of outstanding bids at price  $B$ ; and by  $b^A$  the number of outstanding asks priced at  $A$ .
- The state  $\omega \in \Omega$  contains the price of every outstanding limit order in the book, and the complete order in which they were submitted. It also contains any position in the asset held by the exchange (exchange interventions in the market are described in Section 3.6).
- The signal  $b$  is an event on the state space  $\Omega$ . It contains  $\{A, B, b^A, b^B\}$  and in addition gives anonymous information about the number of bids and asks outstanding at prices outside the quotes. One can remain agnostic whether this additional information is complete (the Level 2 case), partial, or empty (Level 1). Set  $b = 0$  if the trader has no information rights.
- To capture buy-sell symmetry, let  $\tilde{b}$  be the signal  $b$  with bids and asks ‘swapped’: for  $n \in \mathbb{N}$  let it contain as many bids (asks) at price  $nk$  as  $b$  contains asks (bids) at price  $(M + Q - n)k$  (inverting the interval  $[Mk, Qk]$ ). Define the ‘inverted’ state  $\tilde{\omega}$  similarly.

## 3.4 Main assumptions

The model restricts traders to replace a limit order only in favor of a limit order, and not of a market order. Since traders cannot ‘give up’ supplying liquidity by switching to a market order, this rules out a rationale for the the ephemeral supply of liquidity due to ‘fleeting’ order submission documented in Hasbrouck and Saar (2002). Moreover, once  $A$  or  $B$  change, extant

limit orders at the time of the change may no longer be cancelled. This circumvents questions of time-consistency due to moves in the mid-quote,  $m$ , in calculating slippage.

The model specifies that traders may only bid at price  $A - k$ , which normally equals  $B$ , and submit asks at price  $B + k$ , which is normally  $A$ . Section 7 relaxes this restriction and shows that welfare then falls: so the restriction forestalls a tragedy of the commons.

Finally, place a buy-sell symmetry condition on the trader arrival intensity,  $\lambda N$ , so trader arrival intensity is the same in state  $\omega$  as it is in the state  $\tilde{\omega}$ . For all  $\omega$ ,

$$N|\omega = N|\tilde{\omega}. \quad (2)$$

### 3.5 Trader action

**Arrival in the market** Let  $\Sigma$  be the trading action space for traders on arrival at the market, containing the five items

- market purchase at  $A$ ,
- market sale at  $B$ ,
- bid at  $A - k$  (which is normally equal to  $B$ ),
- ask at  $B + k$  (which is normally equal to  $A$ ),
- no action.

Each trader acts according to a function,  $\phi$  which maps her type,  $\beta$  and signal,  $b$  onto an action. Thus for any  $b$  and  $\beta$ ,

$$\phi(b, \beta) \in \Sigma. \quad (3)$$

**Re-entry in the market** On submitting a market order, a trader leaves the market for ever. On the other hand, traders who submit limit orders may update them intermittently until the best quotes next change. Until either  $A$  or  $B$  changes each such trader re-enters the market periodically according to an independent idiosyncratic Poisson process of intensity  $\gamma > 0$ . On any re-entry, she may cancel the limit order and simultaneously submit a replacement bid or ask. The function  $\phi'$  governs the decision whether and how to cancel and replace an order. When a trader re-enters the market, she observes the current  $b$ , together with the price, time priority at that price, and trade direction of her outstanding limit order, summarized by the variable  $l$ . Then  $\phi'$  is a function such that

$$\phi'(b, \beta, l) \in \{\text{no cancel, replace with bid at } A - k, \text{ replace with ask at } B + k\}. \quad (4)$$

### 3.6 Intervention by the exchange

The exchange exerts a mild stabilizing effect on prices at the limits of the interval  $[Mk, Qk]$ . At all times it maintains a single ask at price  $Qk$  (which queues for execution in the normal way). It replaces the ask whenever it executes. It likewise maintains a bid priced at  $Mk$ . It thereby pegs prices so that  $A \leq Qk$  and  $B \geq Mk$ , and may accumulate a position in the asset when its limit orders execute.

The exchange submits market orders as rapidly as possible to clear its resulting position in the asset. Through these operations the exchange is not a net consumer or supplier of liquidity: As is normal, for liquidity provision it relies purely on selfishly-motivated traders. This means that in ergodic equilibrium, over any time interval the expected number of limit orders submitted by the exchange equals the expected number of market orders it submits.

### 3.7 Dynamics of wide spreads

Should  $b^A$  fall to zero,  $A$  increases to the lowest remaining asking price; while if  $b^B$  falls to zero,  $B$  decreases to the price of the highest bid. When the spread is  $k$  a bid at price  $B$  increases  $b^B$  by 1. A market purchase at price  $A$  reduces  $b^A$  by 1. Respectively equivalent effects are made by sellers. The model is intended for a market which trades on a penny, so that  $A - B = k$  almost all the time.

When however  $A - B > k$ , bids and asks can improve on the best quotes. For example, a price-improving bid increases  $B$ , and resets  $b^B$  to 1. This is a necessary mechanism for price levels to change. The incentives to narrow thus the bid-ask spread are high, since the submitter of a limit order is attracted to this unusual opportunity to ‘jump the queue’ of already posted limit orders by pricing it between  $B$  and  $A$ . When trading is on a penny this opportunity is on average very rapidly taken, for only fleetingly is  $A - B > k$ .

These observations suggest a particular formulation of the case where  $A - B > k$ , which usefully brings the decision problem into the same form as the more prevalent case,  $A - B = k$ . This formulation has two parts. First, as previously stated, let traders be restricted to bid only at price  $A - k$  and submit asks at price  $B + k$ , even when  $A - B > k$ . Second, for a buyer, define  $m$  in (1) by  $m = (A - \frac{k}{2})$ . This is the mid-quote which would subsequently prevail if she submitted a bid. Equally, for sellers define  $m = (B + \frac{k}{2})$ . These two definitions coincide in the normal case that  $A - B = k$  but not when  $A - B > k$ . They imply that trading counterparties agree on the mid-quote, so that the undiscounted gains from trade equal the absolute difference in their  $\beta$ -types.

### 3.8 Definition of equilibrium

An equilibrium is a set  $\{c, (\pi, \phi, \phi')\}$  defined as follows.

**Exchange fees** The exchange sets a fee level,  $c$ , for order book information. It sets  $c$  to make buy-side traders just indifferent *ex ante* to acquiring the signal  $b$ .

**Symmetric best response** The actions dictated by  $\pi$ ,  $\phi$ , and  $\phi'$  maximize expected payoffs given  $c$ ,  $(b, \beta, l)$  and the belief that all other traders behave according to  $\{\pi, \phi, \phi'\}$ .

**Buy-sell symmetry** Define the following condition on  $\phi$  and  $\phi'$ : let  $\tilde{\phi}$  be  $\phi$  with buying and selling swapped. Then equilibrium is buy-sell symmetric if for any suitable elements  $b$ ,  $\beta$  and  $l$ ,

$$\tilde{\phi}(\tilde{b}, -\beta) = \phi(b, \beta) \quad \text{and} \quad \tilde{\phi}'(\tilde{b}, -\beta, \tilde{l}) = \phi'(b, \beta, l). \quad (5)$$

**Ergodicity** The initial limit order book shape, an element of  $\Omega$ , is drawn from a long-run ergodic distribution of order book shapes, which exists. This is like imagining that equilibrium has already played out for a long time.

### 3.9 The distribution of private motives to trade

Putting the following functional form on the CDF,  $F$ , delivers closed form solutions in this complex dynamic trading game. As discussed in Section 2.4, this functional form implies that the average best response involve a particular hyperbolic average, which is the same as one characterizing the best response to the average problem. Assume  $F$  is symmetric about zero, and has connected, bounded, support  $[\beta^-, \beta^+]$ . For all  $\beta \in (\beta^-, -\frac{k}{2}]$ ,

$$F(\beta) = \frac{\eta}{\frac{k}{2} - \beta}. \quad (6)$$

Furthermore,  $F(-\frac{k}{2}) > \frac{1}{4}$ , which implies that more than half of traders can benefit from a market order. The constant  $\eta$  is chosen to ensure that  $F(\infty) = 1$ . To satisfy (6) there are suitable atoms at  $\beta^-$  and  $\beta^+$ . For given  $k$  there remains some flexibility in the form of  $F$ . A typical form of  $F$  is depicted in Figure 2.

## 4 The main result, and other aspects of equilibrium

This stochastic sequential game has a countable state space,  $\Omega$ , a finite action space, and a countable number of players. Were it not for the uncountable (but compact) type space,  $[\beta^-, \beta^+]$ ,

and the presence of blind trading, a result of Rieder (1979) would give the existence of a stationary equilibrium under these conditions. Nevertheless:

**Lemma 4.1** *For any  $\pi \in (0, 1]$ , if  $\beta^+$  is large enough there exists an equilibrium,  $\{c, (\pi, \phi, \phi')\}$ .*

**Proof.** The full proof of equilibrium existence is left to the Appendix, and incorporates elements Goettler *et al.* (2004) and (2005). It considers an observationally-equivalent game with coarser information sets, and uses known properties of birth-death processes. ■

The requirement that  $\pi > 0$  means that all orders are eventually observed by some traders with positive probability. This ensures sufficient negative feedback in depths for them to be stationary. Thus, the need for sell-side traders emerges here as a precondition for the limit order book to behave in an orderly, stationary, fashion.

The rest of this Section presents the main equilibrium welfare result and some corollaries. To prepare the intuition for the result, some notation is first introduced (namely,  $D$ ,  $T_A$ , and  $T_B$ ). However, the reader can also move straight to the main Theorem 4.2. The section then closes by analyzing the sell-side's incentives towards an inefficient price tick size.

## 4.1 Trading deadline, $D$

All traders prefer trading early than late due to a discount rate  $\rho > 0$ . It aids intuition to rephrase this intertemporal preference in a way which produces a strategically equivalent game. The adaptation involves deadlines, and is adopted for the rest of the paper.

Set traders' discount rate to zero, not  $\rho$ . They face a common sequence of randomly-timed trading deadlines arriving according to an independent Poisson process of parameter  $\rho$ . Limit orders which are outstanding at the time of a deadline deliver zero utility to the submitter when they later execute, even if cancelled and replaced. Say that the deadline *annuls* the limit order. Market orders are unaffected. On re-entering the market traders do not know if their limit order has been annulled. When considering a representative trader's decision problem, the random time until the next deadline to occur is denoted  $D$ .

## 4.2 The offer / accept decision

Consider the equilibrium best response of a trader who arrives at the market at time  $t$ . The trader receives a signal of the order book state, namely  $b$ .

### 4.2.1 Case of small $|\beta|$

If the trader has a low private motive to trade, so that  $|\beta| \leq \frac{k}{2}$ , her reservation value for the asset is in the range between the quotes,  $[B, A]$  and a market order gives a negative payoff. She

therefore submits a bid or an ask, which she may at later re-entries replace with asks or bids. So types of this sort invariably supply liquidity.

#### 4.2.2 Case when $|\beta|$ is not small

Now confine attention to a trader with reservation value outside  $[B, A]$ . Consider the case where she draws  $\beta > \frac{k}{2}$  and is a buyer (observe that the case of selling is exactly equal and opposite). A market purchase at price  $A$  yields her positive utility equal to  $\beta$  penalized by her price slippage relative to the mid-quote: namely,  $\frac{k}{2}$ . This gives a total payoff of  $(\beta - \frac{k}{2})$ , which is positive. So she certainly submits an order. However, a bid priced at  $A - k$  may offer greater expected utility than a market order because if it is not annulled it pays off  $k$  more: i.e.  $(\beta + \frac{k}{2})$ .

**Definition** Define the random variables  $T_B$  and  $T_A$  as the respective durations from  $t$  until execution were the trader to submit a bid or an ask and not cancel. These have a joint distribution with  $D$  conditional on  $b$ , and unconditionally provided  $b$  is ergodic.

Suppose that it is a best response for the trader to submit a bid. She will have no incentive to replace her bid with another limit order at least until  $A$  moves. Later adjustments are not permitted. She correctly infers from  $\{\pi, \phi, \phi'\}$  the joint distributions of the random variables  $D$  with  $T_B$  and  $T_A$ . So her expected payoff to submitting a bid is given by

$$\left(\beta + \frac{k}{2}\right) \Pr[T_B \leq D|b]. \quad (7)$$

She compares this to the payoff from a market order,  $(\beta - \frac{k}{2})$ . Just as shown in Goettler *et al.* (2005) and Hollifield *et al.* (2004), there therefore exists a cutoff, say  $\bar{\beta}^b$ , above which her type  $\beta$  induces her to submit a market purchase. With  $\beta$  in an interval just below the cutoff she bids. This is studied in more detail in Section 6.1.

**Definition** In equilibrium the *ex ante* limit order execution risk perceived by traders is defined by

$$\Pr[T_A > D]. \quad (8)$$

Buy-sell symmetry implies that this also equals  $\Pr[T_B > D]$ . It is the probability of annulment of a randomly-submitted limit order, or one submitted by a trader who acquires no limit order book information, meaning it is also equal to  $\Pr[T_A > D|b = 0]$ . It exceeds the average execution risk undertaken by sell-side traders for they avoid limit orders when  $b$  suggests that execution risk is high. The expected discount factor applied by blind traders to future limit order payoffs is  $\Pr[T_A \leq D]$ , or equivalently  $E[e^{-\rho T_A}]$ .

Due to buy-sell symmetry, in equilibrium blind traders buy iff  $\beta > 0$ , so their gain from trade from an unannulled bid or an ask is

$$|\beta| + \frac{k}{2}. \quad (9)$$

### 4.3 Main welfare result

**Remark** The first best benchmark case for welfare is where all traders transact immediately at time 0 and at the mid-quote. The surplus per trader is then  $E(|\beta|)$ .

This surplus is unattainable in a continuous auction since not all traders appear simultaneously at the market, and therefore some traders must wait.

**Definition** The trader type space divides symmetrically into the four quartiles of  $F$ , as depicted in Figure 2. Define  $\delta$  by

$$\delta = \frac{E[|\beta| : |\beta| < F^{-1} \left[ \frac{3}{4} \right]]}{F^{-1} \left[ \frac{3}{4} \right]}. \quad (10)$$

Thus  $\delta \in (0, 1)$  describes the relative dispersion of  $F$  within its interquartile range. If  $\delta \approx 0$  then dispersion within the range is low, but if  $\delta \approx 1$  then dispersion is high.

**Theorem 4.2** *Suppose that the support of  $F$  is broad enough that for all  $b$ , market sales occur with positive probability. Then in equilibrium, the buy-side's expected surplus per trader,  $S$ , is given by:*

$$S = E(|\beta|) - \frac{1}{2} \Pr[T_A > D] \left( \frac{k}{2} + \delta F^{-1} \left[ \frac{3}{4} \right] \right), \quad (11)$$

where

$$\Pr[T_A > D] = \frac{k}{\frac{k}{2} + F^{-1} \left[ \frac{3}{4} \right]}. \quad (12)$$

Hence buy-side surplus per trader depends only on  $F$  and  $k$ .

**Proof.** The proof relies on Proposition 6.4, which implies that blind traders follow the strategy depicted in Figure 2. This proposition is developed and proven in Section 6, and intuition is provided in Section 2.

Since the *ex ante* cost  $c$  is set to make traders indifferent to buying the sell-side signal,  $b$ , the surplus of the blind trader equals the surplus of all buy-side traders. This holds even in an equilibrium where  $\pi = 1$ , for then the price  $c$  is set so that traders are indifferent about deviating to the outside option of not acquiring the signal  $b$  and trading blind. The blind trader's shortfall in *ex ante* surplus, relative to  $E(|\beta|)$ , can be written

$$\Pr[LO] \times E[\text{cost of delay} | LO], \quad (13)$$

where  $LO$  is the event that she chooses to submit a limit order on discerning  $\beta$ , and

$$\text{cost of delay} = \left( |\beta| + \frac{k}{2} \right) \mathbb{I}_{[T_A > D]} \quad (14)$$

(where  $\mathbb{I}$  is the Indicator Function), since (9) is the utility which is foregone if  $T_A > D$ . As  $\beta$  and  $[T_A > D]$  are independent, and  $\Pr[T_A > D | LO] = \Pr[T_A > D]$ , expected shortfall is

$$\Pr[LO] \times \Pr[T_A > D] \times E \left[ |\beta| + \frac{k}{2} \mid LO \right]. \quad (15)$$

From Proposition 6.4 she foresees preferring limit orders iff her  $\beta$  is in the interquartile range of  $F$ . From this she deduces that  $\Pr[LO]$  is  $\frac{1}{2}$ , and she calculates that the expected  $|\beta|$ , conditional on  $LO$ , is  $\delta F^{-1} \left[ \frac{3}{4} \right]$ . Finally, if she draws  $\beta$  on the boundary of the interquartile range, she must be just indifferent between market and limit orders. From this indifference condition she calculates the execution risk in a limit order,  $\Pr[T_A > D]$ , as given in Proposition 6.4. ■

**Comparative statics** While all the following variables affect trading dynamics, nevertheless surplus per buy-side trader is invariant to

- trader impatience,  $\rho$ ,
- the rate of trader arrival,  $\lambda E[N]$ ,
- the nature and variability of the trader number process  $N$ ,
- the abolition of time-priority in limit order execution in favor of alternatives e.g. a coin-flip rule,
- the proportion of traders,  $\pi$ , who acquire the sell-side signal,  $b$ , and
- whether the sell-side signal  $b$  contains Level 1 or Level 2 information.

Theorem 4.2 shows (simplifying the expression) that the shortfall in buy-side surplus per trader relative to the first best,  $E(|\beta|)$ , is

$$\frac{k}{2} \left( \frac{\frac{k}{2} + \delta F^{-1} \left[ \frac{3}{4} \right]}{\frac{k}{2} + F^{-1} \left[ \frac{3}{4} \right]} \right), \quad (16)$$

which lies between  $\delta \frac{k}{2}$  and  $\frac{k}{2}$ . Holding fixed  $F^{-1} \left[ \frac{3}{4} \right]$  and  $\delta$ , it is increasing in the price tick,  $k$  (given the relationship of  $F$  to  $k$  in (6) it is not strictly possible to hold  $F$  fixed in this thought experiment where  $k$  changes).

The half-spread bounds (16) above, since the blind trader waits before trading exactly half the time, and when she does wait, she finds paying the spread *ex ante* costlier than waiting.

## 4.4 Sell-side rents

For each mixing probability  $\pi > 0$  and price tick  $k > 0$ , there exists in equilibrium a positive fee level,  $c$ , chosen by the exchange to make buy-side traders indifferent to acquiring sell-side information. Thus, the total efficiency per trader of the equilibrium,  $W(\pi, k)$ , is

$$W(\pi, k) = \pi c + S(k), \quad (17)$$

where  $S(k)$ , invariant to  $\pi$ , is buy-side surplus, as given in (11). Note that  $c \leq k$ , since in her blind outside option, a trader's expected losses are at most  $\frac{k}{2}$  relative to trading at the mid-quote – but as an informed trader she gains at most  $\frac{k}{2}$ . So for all  $k$

$$\lim_{\pi \rightarrow 0} W(\pi, k) = S(k). \quad (18)$$

It is therefore legitimate to reason that for given  $k$  the sell-side pays for itself, as its rents equal the efficiency gain it produces, in that

$$\pi c = W(\pi, k) - W(0^+, k). \quad (19)$$

Recall that  $W(\pi, k)$  is bounded above by  $E[|\beta|]$ , the first best outcome where all traders arrive at the market simultaneously. Fixing  $k$ , the value of  $\pi$  in the most efficient equilibrium,  $\hat{\pi}(k)$ , is the one that would be chosen by the exchange to maximize sell-side rents,  $\pi c$ . Note  $\hat{\pi}(k)$  exists and is greater than zero. For given  $k$  and all  $\pi$  sell-side rents are then at most

$$(W(\hat{\pi}(k), k) - S(k)). \quad (20)$$

But this is bounded above by  $(E[|\beta|] - S(k))$ , which is (16), and  $\rightarrow 0$  as  $k \rightarrow 0$ . Therefore the sell-side's optimal tick size,  $\hat{k}_{ss}$ , is in the interval  $(0, \infty]$ , and it does not have an interest in minimizing  $k$ .

## 5 Average depths

An important intuition for the properties of this model rests on the idea that average depths adjust to a market-clearing level, offsetting the effects of changing the trading environment. Yet, it was possible in Theorem 4.2 to find the blind best response, and quantify buy-side surplus, while bypassing a direct analysis of depths altogether.

This section studies the ergodic distribution of inside depths in a special case to understand better their sensitivity to parameters in the trading environment. It concentrates simply on their first moment, namely average inside depths: and shows that they do indeed adjust in appropriate ways to explain Theorem 4.2. It is therefore complementary to the study in Parlour (1998) of depths' transitions from one state to the next.

## Definitions

- Let  $\bar{d}$  denote the average duration between trader arrivals.
- Let  $L^{bl}$  and  $L^{ss}$  denote the stochastic processes over time giving the market depth due to the blind submission of limit orders, and sell-side submission, respectively.
- Let  $L = L^{bl} + L^{ss}$ , be total depth.
- Let  $\bar{L} = E[L_t]$  for any time  $t$ , so it is average depth. Define  $\bar{L}^{bl}$  and  $\bar{L}^{ss}$  likewise.
- When a limit order persists for the arrival of two or more deadlines, and so is annulled for a second, or subsequent time, call this being ‘re-annulled’.
- Let  $\epsilon$  be the expected number of times that a blindly submitted limit order is re-annulled.

**Proposition 5.1** *Suppose  $\pi < 1$ . Then*

$$\Pr[T_A > D] = \frac{2\bar{d}\rho\bar{L}^{bl}}{(1-\pi)} - \epsilon. \quad (21)$$

*Therefore, under, in addition, the conditions of Theorem 4.2,*

$$\bar{L}^{bl} = \frac{(1-\pi)}{2\bar{d}\rho} \left( \epsilon + \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]} \right). \quad (22)$$

**Proof.** This follows from law of large numbers on considering the evolution of the market over a long interval of time. See Appendix. ■

The only endogenous variables on the RHS of (47) are  $\epsilon$  and  $\pi$ . The expectation  $\epsilon$  covers two eventualities: that a limit order at the best quotes is re-annulled, and the eventuality that a limit order becomes stale, meaning that the best quotes move away to other prices, and it lingers in the limit order book where it is re-annulled. The latter eventuality is of lesser interest, while for low execution risk the former of these is typically small. In the light of these comments, I now provide conditions under which  $\epsilon \approx 0$ .

**Lemma 5.2** *Let order flow and trader numbers,  $N\lambda$ , be constant. Suppose that  $Q = M + 1$ . This implies that the market is two-tick, so that the exchange maintains or pegs constant  $A$  and constant  $B = A + k$ . Then*

$$\epsilon \leq \{\Pr[T_A > D]\}^2. \quad (23)$$

*Therefore, if execution risk (given by  $\Pr[T_A > D] = \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}$ ) is small, then  $\epsilon$  is second order.*

**Proof.** See Appendix. ■

If the market is two-tick, then limit orders' prices never become stale. In the case discussed in Section 2, where  $\pi \approx 0$ , so that only a small minority of traders obtain limit order book information and almost all trade blind, it follows that  $\bar{L} \approx \bar{L}^{bl}$  and so, setting  $\epsilon \approx 0$ ,

$$\bar{L} \approx \frac{1}{2\rho\bar{d}} \frac{k}{\left(\frac{k}{2} + F^{-1}\left[\frac{3}{4}\right]\right)}. \quad (24)$$

**Comparative Statics** On the basis of (24) some comparative statics can be derived for average depths,  $\bar{L}$ . As all depth is at the inside quote when prices are constant it is prudent to view these as valid for the inside depth, but not necessarily for the depth of the limit order book outside the quotes. Recall that  $L$  counts up both bids and asks. From (24):

- Average inside depth is increasing in the tick size,  $k$ , since, as discussed in the Introduction, a greater tick size would cause a shortage of market orders unless depths rise to make limit orders less attractive. This effect was observed in Goldstein and Kavajecz (2000) and Harris (1994).

- Average inside depth is increasing in the order flow,  $\frac{1}{\bar{d}}$ , since a given depth clears faster in a faster market, meaning that limit orders would execute faster, be more attractive and therefore be in excess – unless depths rise. This corresponds to a finding in Brockman and Chung (1996) for the SEHK.

- But average inside depth is decreasing in trader impatience,  $\rho$ , since impatient traders avoid limit orders, leading to a shortage unless depths fall.

If  $A$  and  $B$  are not held constant by the exchange, then limit orders risk becoming stale; and their submitters are compensated by lower inside depths. Where  $\pi \gg 0$ , so that a substantial number of traders obtain order book information, average depths decline where such traders avoid adding to deep order books, but rise where traders are drawn to supply liquidity by the option value in the possibility of resubmission. A characterization of average inside depth under these generalizations is beyond the current scope.

## 6 Proof of the Main Theorem 4.2

The following Lemma shows that in equilibrium market sales, market purchases, bids and asks are equally prevalent in the order flow. This formalizes one of the paper's main intuitions: namely, that if depths are stationary then the average number of uncanceled limit orders per unit time equals the average number of market orders. As discussed in the Introduction, this can be interpreted as meaning that the markets for liquidity (or immediacy) at the bid and at

the ask clear.

**Lemma 6.1** *In equilibrium the ergodic probability that any given order is a market sale is  $\frac{1}{4}$ . This is also the probability it is respectively a market purchase, a bid, or an ask. Therefore, writing  $\mu_b$  for the probability that a trader submits a market sale given the signal,  $b$ ,*

$$E[\mu_b] = \frac{1}{4}. \quad (25)$$

**Proof.** See Appendix. The proof uses the fact that limit orders can only be cancelled in favor of other limit orders. ■

## 6.1 Cost of immediate, and delayed, execution

The arguments of this subsection follow Hollifield *et al.* (2004) and Goettler *et al.* (2005) in a special case. On arrival at the market, traders with  $\beta > \frac{k}{2}$  prefer to purchase than sell, but must decide between making a bid and a market purchase. Sellers with  $\beta < -\frac{k}{2}$  face a symmetric problem. The buyers' decision can be formulated as selecting the minimum of two costs defined relative to the unattainable benchmark payoff,  $(\beta + \frac{k}{2})$ , which would be gained by purchasing immediately at price  $(A - k)$ . A market order incurs a cost of immediate execution, while a limit order incurs a cost of delayed execution.

The cost of delayed execution is the trader's payoff at  $(A - k)$ , namely  $(\beta + \frac{k}{2})$ , multiplied by the probability that execution delay causes the trader to miss the next deadline. Thus it is

$$\left(\beta + \frac{k}{2}\right) \Pr[T_B > D|b]. \quad (26)$$

The cost of immediate execution is simply the price tick,  $k$ . If the cost of immediate execution, the spread, is less than the cost of delayed execution, the trader chooses to place a market order, otherwise she places a limit order.

For buyers, the cost of delayed execution in (26) increasing linearly in  $\beta$  but is equal to zero when  $\beta = -\frac{k}{2}$ . Therefore for each  $b$  there exists a trader type  $\bar{\beta}^b > -\frac{k}{2}$  who is indifferent between bids and market purchases, which will be termed the 'marginal buyer'. All types with higher  $\beta$  prefer market purchases to bids and thus place market purchases. Similarly, there is a 'marginal seller',  $\underline{\beta}^b < \frac{k}{2}$ , for whom all types with lower  $\beta$  place market sales. The condition for the marginal buyer is

$$\left(\bar{\beta}^b + \frac{k}{2}\right) \Pr[T_B > D|b] = k, \quad (27)$$

while the marginal seller is characterized by

$$\left(\frac{k}{2} - \underline{\beta}^b\right) \Pr[T_A > D|b] = k. \quad (28)$$

From (27) and (28), it follows that  $\bar{\beta}^b$  is greater than  $\frac{k}{2}$ , and  $\underline{\beta}^b$  is less than  $-\frac{k}{2}$ . Thus for any  $b$  there are buyers and sellers who would gain from a market order, but prefer to submit limit orders. All types and only those types with  $\beta$  lower than  $\underline{\beta}^b$  make a market sale. Hence  $\mu_b$ , the probability that a trader of unknown type submits a market sale given signal  $b$ , is

$$\mu_b = F(\underline{\beta}^b). \quad (29)$$

## 6.2 Marginal sellers and the median seller

**Definition** The *median seller* is the trader with  $\beta$  at the  $\frac{1}{4}$  quantile of  $F$ , so  $\beta = F^{-1}[\frac{1}{4}]$ . Likewise, the *median buyer* has  $\beta = F^{-1}[\frac{3}{4}]$ .

In what follows, the argument concentrates on the median seller, although an identical analysis holds for the median buyer.

**Lemma 6.2** *Marginal sellers (buyers) are distributed around the median seller (buyer): precisely,*

$$F^{-1}\{E[F[\underline{\beta}^b]]\} = F^{-1}\left[\frac{1}{4}\right]. \quad (30)$$

**Proof.** (25) in Lemma 6.1, and (29), together imply that  $E[F[\underline{\beta}^b]] = \frac{1}{4}$ . ■

Lemma 6.2 shows that, via market-clearing, it is possible *ex ante* to identify a particular deformed average of all the marginal sellers, namely the median seller. This identification makes no reference to the details of the dynamic equilibrium, or the process  $N$ . The next part exploits the intuition that the median seller, being an average of the marginal sellers is, naturally, *close in type* to the marginal seller in the average problem. Indeed, for the chosen form of  $F$ , Proposition 6.4 will show they coincide.

## 6.3 The blind marginal buyer and seller

Define the ‘blind’ marginal buyer and seller, called  $\bar{\beta}^0$  and  $\underline{\beta}^0$ , as the marginal buyer and seller when  $b = 0$ . As stated in (28), the blind marginal seller perceives an equal cost of delayed execution as of immediate execution *ex ante*. At the time she submits an order, her expected payoff therefore depends directly on  $\Pr[T_A > D]$ , the *ex ante* limit order execution risk. A similar account applies for the blind marginal buyer.

**Lemma 6.3** *The blind marginal seller is a hyperbolic average of the marginal sellers:*

$$\frac{k}{\frac{k}{2} - \underline{\beta}^0} = E\left[\frac{k}{\frac{k}{2} - \underline{\beta}^b}\right]. \quad (31)$$

**Proof.** From (28),

$$\Pr[T_A > D|b] = \frac{k}{\frac{k}{2} - \underline{\beta}^b}. \quad (32)$$

Recall that

$$\Pr[T_A > D] = E\{\Pr[T_A > D|b]\}. \quad (33)$$

The proposition follows on substituting (32) in (33). ■

## 6.4 Concluding result

Lemma 6.2 used ergodicity to show that the marginal sellers have as an average the median seller. Lemma 6.3 showed they have as a hyperbolic average the blind marginal seller, as she faces an average of their decision problems. The concluding proposition of this Section is proven by showing that these averages are both hyperbolic, and so identical. This bypasses the need to evaluate  $\Pr[T_A > D|b]$  directly – a complex problem depending on many possible future paths of the market state.

**Proposition 6.4** *Suppose that the support of  $F$  is broad enough that for all  $b$ , market sales occur with positive probability. Then in equilibrium the marginal seller ex ante,  $\underline{\beta}^0$ , coincides with the median seller,  $F^{-1}[\frac{1}{4}]$ . Therefore average execution risk is given by*

$$\Pr[T_A > D] = \frac{k}{\frac{k}{2} - F^{-1}[\frac{1}{4}]}, \quad (34)$$

and, given buy-sell symmetry, blind traders follow the strategy depicted in Figure 2.

**Proof.** For all  $b$ ,  $\underline{\beta}^b \in (\beta^-, -\frac{k}{2}]$ . So, (6) applies to  $\underline{\beta}^b$  for all  $b$ . So from Lemma 6.2,

$$E\left[\frac{\eta}{\frac{k}{2} - \underline{\beta}^b}\right] = \frac{1}{4}. \quad (35)$$

Combining this with Lemma 6.3,

$$\frac{\eta}{\frac{k}{2} - \underline{\beta}^0} = \frac{1}{4}. \quad (36)$$

So it follows that  $\underline{\beta}^0 = F^{-1}[\frac{1}{4}]$ , on substituting  $F^{-1}[\frac{1}{4}]$  for  $\beta$  in (6). ■

Proposition 6.4 shows that the dynamics of market depths adjust in equilibrium to make the median seller just indifferent between committing *ex ante* to make a market sale, and an ask. A similar condition is satisfied for the median buyer. Therefore, equilibrium has the property that those who trade blind follow the strategy depicted in Figure 2 – opting for market sales, market purchases, bids and asks equally frequently. Incidentally, it then follows from Lemma 6.1 that sell-side traders also plump for each order type equally frequently.

With Proposition 6.4 given, the proof of the Main Theorem 4.2 follows as stated.

## 7 Sensitivity to key assumptions

This Section relaxes modelling assumption in four areas, in many cases recovering the results of the model in modified form. These areas are 1) the functional form of  $F$ ; 2) whether traders' payoffs depend on a common value for the traded asset, 3) whether traders have linear waiting costs and 4) the model's limitations on trader actions.

### 7.1 Robustness to the distribution $F$

Essential to the main welfare results is the assumption that  $F$  has the distribution defined in (6). The next proposition therefore provides a robustness check, showing that the markets for liquidity also clear *ex ante*, though only approximately, in the case that  $F$  is uniform. The invariance properties of interest then follows approximately.

**Corollary 7.1** *Suppose  $F$  has a uniform distribution symmetric about zero. Suppose that the support of  $F$  is broad enough that for all  $b$ , market sales occur with positive probability. Then the harmonic mean of execution risk is given by*

$$E \left[ \frac{1}{\Pr[T_A > D|b]} \right]^{-1} = \frac{k}{\frac{k}{2} - F^{-1} \left[ \frac{1}{4} \right]}. \quad (37)$$

**Proof.** See Appendix. The proof is similar to that for Lemma 6.3. ■

In so far as the harmonic mean approximates the mean, Corollary 7.1 implies that

$$\Pr[T_A > D] \approx \frac{k}{\frac{k}{2} - F^{-1} \left[ \frac{1}{4} \right]}. \quad (38)$$

Therefore, under this distributional assumption the median seller is *approximately* equal to the blind marginal seller. The next part returns to the true distribution of  $F$ .

### 7.2 Robustness to the calculation of implementation shortfall

In the main model, implementation shortfall is calculated relative to the current mid-quote. While this may accord with practice, it shuts down the possibility of a winner's curse in limit order execution (see Foucault 1999 and Handa and Schwartz 1996 among other papers). This section adapts the model to incorporate this, using an approach that is related to Goettler, Parlour, and Rajan (2004).

Assume the signal  $b$ , if informative, also contains a signal of the common value of the traded asset,  $v$ . The common value  $v$  follows an exogenous stationary stochastic process over time within the interval  $(Mk, Qk)$ . Define its 'buy-sell mirror', by  $\tilde{v} = (M + Q)k - v$ . Then assume that  $\tilde{v}$

and  $v$  have the same distribution over time. At execution, traders who buy at price  $p$  receive a payoff equal to the private gain  $\beta$  minus their implementation shortfall, or slippage, relative to  $v$ , i.e. :

$$\beta - (p - v). \quad (39)$$

So this replaces  $m$  with  $v$  in (1). As in (1), sellers at price  $p$  receive the negative of this.

**Corollary 7.2** *Under this generalization to include a common value, the blind best response as given in Figure 2 is played in the limit of a sequence of equilibria where  $\pi \downarrow 0$ , so that the market goes dark.*

*Suppose in addition that  $(Q - M) \rightarrow \infty$  (so that the exchange almost never intervenes), and trading becomes ever more on a penny, so*

$$\Pr[(A - B) > k] \downarrow 0. \quad (40)$$

*Then the welfare results of Theorem 4.2 are asymptotically true.*

**Proof.** See Appendix. ■

This corollary studies a limiting case where asymmetric information converges to zero, and it recovers the main welfare results of the paper. More generally, it is expected that moderate asymmetric information *enhances* welfare if the asset trades on a penny. It deters blind limit order submission because of the winner's curse. But it does not deter blind market order submission since the bid-ask spread is fixed at  $k$ . So for there not to be a shortage of limit orders in the order flow, equilibrium inside depths must fall, and with them the delay before limit order execution. Although this would transfer trading rents to insiders, it would enhance welfare.<sup>10</sup> However, an equally desirable welfare level could be achieved by simply cutting the price tick size.

### 7.3 Linear waiting costs as an alternative to impatience

The central model defines, via  $\beta$ , a payoff to trade which varies across agents. Since some agents want to trade more, they are more reluctant to wait for trade. But heterogenous intertemporal preferences can be introduced in another way: This section studies the case where all buyers and sellers may have an equal motive to trade, but some dislike more than others the experience of waiting to trade with a limit order in the market. For example, Cordella and Foucault (1999) cite the cost of monitoring the limit order. In fact the adaptation does little to change the paper's main results.

---

<sup>10</sup>This may help explain the rather neutral effects of adverse selection on welfare in the comparative simulations of Goettler, Parlour, and Rajan (2004), in so far as they resemble trading on a penny.

**Adaptations to the main model** Set trader impatience,  $\rho$ , to zero. However, suppose that traders face a type-specific waiting cost per unit time that their limit order waits before trading. Let this be the absolute value of their type,  $|\beta|$ . Let the sign of  $\beta$  determine whether the trader is a buyer or a seller: specifically, if  $\beta > 0$ , let the trader's reservation value, say  $V$ , be a higher price than any in the model, so  $V > Qk$  and she is a buyer. Conversely, if  $\beta \leq 0$ , let  $V$  be less than  $Mk$ , so she certainly sells. Buyers when they transact have an overall payoff of  $(V - p)$ . This replaces (1). Sellers receive the negative of this payoff. The final adaptation of the model is to change the background distribution,  $F$ . Replace (6) so that for some very small negative  $\iota$ ,

$$\beta \in (\beta^-, \iota] \rightarrow F(\beta) = -\frac{\eta}{\beta}. \quad (41)$$

All other aspects of the model are left unchanged (making  $\iota < 0$  ensures that  $F$  is integrable).

The following corollary shows that under these adaptations, the main properties of the model are unchanged.

**Corollary 7.3** *Under these adaptations to the model, suppose that in equilibrium the support of  $F$  is broad enough that for all  $b$ , market sales occur with positive probability. Then for  $\iota$  near enough to zero:*

- *The blind best response is as depicted in Figure 2.*
- *Waiting costs are the only source of inefficiency.*
- *The expected waiting costs per blind trader are  $\frac{k\delta}{2}$ . So the welfare results of the paper are qualitatively unchanged. Moreover buy-side surplus is invariant to the average waiting cost per unit time,  $E[|\beta|]$ .*
- *If  $\pi \approx 0$  then average depths are  $\frac{k}{2dF^{-1}[\frac{3}{4}]}$ . So the comparative statics on depths in Section 5 are also unchanged, except that they refer to entire depth, not just inside depth.*

**Proof.** See Appendix. ■

## 7.4 The effect of involuntary cancellation

This Section returns to the original payoff formulation of (1). It considers the case where limit orders are periodically cancelled *involuntarily*. This modelling feature is present in Goettler *et al.* (2005) and Hollifield *et al.* (2004), but is replaced with an endogenous cancellation decision, as here, in Goettler *et al.* (2004). While exogenous cancellation may seem to be detrimental to traders, in fact it is not.

**Corollary 7.4** *In the main model, suppose that over time, some limit orders are involuntarily cancelled. Cancellations occur one-at-a-time according to an adapted point process. The limit*

order which is cancelled on any given time is determined according to a coin-flip rule. Then in equilibrium, buy-side surplus per trader is enhanced compared to the expression in Theorem 4.2.

**Proof.** For markets for liquidity to clear when some limit orders are cancelled,  $E[\mu_b] < \frac{1}{4}$  (c.f. Lemma 6.1). So certain types lying at quantiles strictly outside the interquartile range must be, when blind, indifferent between submitting a limit order and a market order. In the original model, these types strictly prefer a market order. So, to make them indifferent, blind limit order execution risk has *declined* relative to the original model. Correspondingly, blind surplus has increased. ■

**Intuition** Involuntary cancellation is comparable to annulment by a deadline,  $D$ . Like annulment risk, cancellation risk has no direct effect on the desirability of a limit order, because of the way that markets for liquidity clear. However, it confers a positive externality on those limit orders which are *not* cancelled or annulled. They execute faster because they are not delayed by previously cancelled orders, while they would be obstructed by previously annulled orders.

Deliberate cancellation of limit orders, when in favor of market orders, appears to be harder to analyse in this context. Its effect on welfare is hard to sign.<sup>11</sup>

## 7.5 Robustness to the limitations on the actions of traders

Recall that the main welfare results of the paper are derived for a simplified setting with restrictions on the traders' actions. Specifically, they may only replace cancel limit orders in favor of other limit orders, not market orders. Furthermore, all limit orders must be placed at best prices.

This section relaxes these for sell-side traders alone in Corollaries 7.5 and 7.7; and for all traders in Corollary 7.6. Corollaries 7.5 and 7.6 recover the welfare results of the paper in the case where  $\pi \downarrow 0$ , so that the exchange is almost 'dark', though in the latter case an effect in Cordella and Foucault (1999) appears. With discriminatory limit pricing but no cancellation, Corollary 7.7 provides conditions where Theorem 4.2 is true for any  $\pi$ , and the results of the paper follow.

**Definition** Say that a trader's action space is expanded if she may submit limit orders at any price, not just the best prices, and she may cancel her limit order in favor of a market order, not

---

<sup>11</sup>Deliberate cancellation introduces an option value to limit order submission which is unavailable to blind traders. So, it transfers rents from the buy-side to the sell-side (and the proof of Corollary 7.4 does not hold). On the other hand, it 'roughly doubles' cancellation's positive externality, since limit orders cancelled in favor of market orders add to the demand for liquidity.

just another limit order.

**Corollary 7.5** *Expand the action space of sell-side traders only. The welfare results of Theorem 4.2 are true in the limit of a sequence of equilibria as  $\pi \downarrow 0$ , so that the market is almost dark.*

**Proof.** For the markets for liquidity to clear, types  $\pm F^{-1} \left[ \frac{1}{4} \right]$  are indifferent, when blind, between a market order and a limit order priced one tick less favorably. Hence, the expected discount factor blind traders associate with the execution of a limit order submitted at the best quotes is as in Proposition 6.4. Figure 2 represents blind equilibrium strategies. Buy-side trader surplus follows as in the proof of Theorem 4.2. ■

The next corollary addresses the case where blind traders' action spaces are also expanded. Note that they still never cancel a limit order, because their beliefs about  $\omega$  do not change over time. Thus, at a cancellation opportunity, they wish to maintain their outstanding limit order, because its time to execution has become less since they chose to submit it. So this added flexibility does not confer an advantage. However, blind traders might also now submit limit orders away from the quotes. Does this enhance or detract from welfare?

Section 5 showed that a wide price tick produces greater depths. It induces traders to wait longer for limit order execution. The next lemma shows that if traders were further permitted to bid and ask *outside* the quotes, some would choose to tolerate an even long wait, resulting in further welfare losses. This effect was noted in Cordella and Foucault (1999) and Foucault *et al.* (2005). The next corollary shows that it defeats a countervailing consideration here: that submitting an ask above the best offer, rather than at it, leads to a shorter wait for those asks that are submitted later at the best offer – so if those asks are on average more urgent than it, welfare might be enhanced when it exits their queue.

**Corollary 7.6** *Suppose that  $\pi \approx 0$ , so that the market is almost dark. In comparison to the original game, equilibrium buy-side surplus is diminished by the expanded action space provided that some blind types make use of the expanded action space. Otherwise it is unchanged.*

**Proof.** See Appendix. ■

However, Corollary 7.5 showed that if only the sell-side's action space is expanded, the efficiency gains of sorting limit orders by urgency are entirely absorbed by its fees when  $\pi \approx 0$ . This is perhaps a realistic case: it would be bold for an uninformed type to submit a limit order away from the quotes. The section concludes by providing conditions under which the results of Corollary 7.5 hold for all  $\pi > 0$ .

**Definition** In equilibrium, define a signal,  $b$ , to *preclude sell-side orders at best* if, knowing  $b$  no type of trader would submit a bid priced  $(A - k)$ , or knowing  $b$  no type of trader would submit an ask at  $(B + k)$ .

Wherever a market trades on a penny, states of the market that preclude sell-side orders at best satisfy a stringent condition and are therefore infrequent. If no type at all would, given  $b$ , add liquidity at the bid, or at the ask, depths must be very high at the quote, and low outside the quotes. If the market does not trade on a penny, this condition may well be violated. Foucault *et al.* (1995) covers exactly this case (presaged in Cordella and Foucault (1999)) and identifies related welfare losses.

**Corollary 7.7** *Adapt the main model so that sell-side traders can price limit orders outside the quotes, and outlaw cancellation. Suppose in addition that the exchange intervenes to avert states which preclude sell-side orders at best (closing its position so that its net liquidity supply is zero).*

*Then ergodic equilibrium exists and the results of Theorem 4.2 are true.*

**Proof.** See Appendix. ■

## 8 Conclusion

The market-clearing condition is a central means of analyzing prices without modelling market microstructure details directly. In focusing on these details, this paper rediscovers market-clearing in a different context: as a minimal stationarity requirement on microstructure dynamics, meaning that equal supply and demand for liquidity, or immediacy, are induced from traders over time. This stationarity is harnessed to infer the best response and so the surplus of ‘blind’ traders, while bypassing altogether their difficult forecasting problem. Applying this approach to other stochastic sequential games with binary action, such as queueing and extensions thereof, offers an exciting opening for future research.

Standard intuitions about market-clearing are thereby shown to be pertinent: in the context of the model it is legitimate to reason that since deep depths deter liquidity supply while shallow depths attract it, there exists an equilibrium average inside depth to clear the market. Such equilibrium effects lead to striking invariances in buy-side trader surplus: with respect to the prevalence of sophisticated sell-side traders; with respect to weak or uncertain order flow; with respect to trader impatience; with respect to order queueing rules. In every such case, the market’s endogenous inefficiency adjusts to offset the anticipated sensitivity. The main exception is cutting the price tick, which alone improves the prospects of a market order, thereby overcoming

traders' incentives to add their limit orders to an already congested limit order book. Investor surplus is thereby enhanced. This result is unchanged if traders have a common value for the asset (with limited insider trading), or linear waiting costs. It is also robust to 'microstructure speculation' by sell-side traders who place limit orders outside the quotes.

This provides an intuition for using the quoted or effective spread as a measure of market quality: for it shows that wide spreads penalize not only market orders, but also, via equilibrium effects, limit orders. Extending this insight to other forms of market microstructure institution, which would involve introducing a market-clearing role for bid-ask spreads as well as for market inefficiency and depths, represents another research avenue of considerable interest.

In opposition to buy-side investors, sell-side traders have an interest in keeping the price tick strictly away from zero, since they thrive on the complicated queueing problems of deep markets. Yet as the price tick size has a monotonic, negative effect on buy-side investor surplus, it should be cut at least as long as the market would still trade on a penny.

## References

- Bernhardt, D. and E. Hughson (1996). Discrete pricing and the design of dealership markets. *Journal of Economic Theory* 71, 148–182.
- Brockman, P. and D. Chung (1996). An analysis of depth behavior in an electronic order-driven environment. *Journal of Banking and Finance* 51, 1835–1861.
- Chakravarty, S. and K. Holden (1995). An integrated model of market and limit orders. *Journal of Financial Intermediation* 4, 213–241.
- Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb (1981). Transaction costs, order placement strategy, and existence of the bid-ask spread. *Journal of Political Economy* 89, 287–305.
- Cordella, T. and T. Foucault (1999). Minimum price variation, time priority and quote dynamics. *Journal of Financial Intermediation* 8, 141–173.
- Danielsson, J. and R. Payne (2002). Liquidity determination in an order driven market. Working Paper, London School of Economics.
- Domowitz, I. and J. Wang (1994). Auctions as algorithms: computerized trade execution and price discovery. *Journal of Economic Dynamics and Control* 18, 29–60.
- Duffie, D., N. Gârleanu, and L. Pedersen (2005). Over-the-counter markets. *Econometrica* 73, 1815–1847.

- Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets* 2, 99–134.
- Foucault, T., O. Kadan, and E. Kandel (2005). Limit order book as a market for liquidity. *Review of Financial Studies* 18, 1171–1217.
- Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Goettler, R., C. Parlour, and U. Rajan (2004). Information acquisition in a limit order market. Working paper Carnegie Mellon University.
- Goettler, R., C. Parlour, and U. Rajan (2005). Equilibrium in a a dynamic limit order market. *Journal of Finance* 60, 2149–2192.
- Goldstein, M. and K. Kavajecz (2000). Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. *Journal of Financial Economics* 56, 125–149.
- Handa, P. and R. Schwartz (1996). Limit order trading. *Journal of Finance* 51, 1835–1861.
- Harris, L. (1994). Minimum price variations, discrete bid-ask spreads, and quotation sizes. *Review of Financial Studies* 7, 149–178.
- Harris, L. (1997). Decimalization: a review of the arguments and evidence. Working Paper, University of Southern California.
- Hasbrouck, J. and G. Saar (2002). Limit orders and volatility on a hybrid market. Working Paper, New York University.
- Hendershott, T. and C. Jones (2005). Island goes dark: Transparency, fragmentation, and regulation. *Review of Financial Studies* 18, 743–793.
- Hollifield, B., R. A. Miller, and P. Sandås (2004). Empirical analysis of limit order markets. *Review of Economic Studies* 71, 1027–1063.
- Jones, C. and M. Lipson (2001). Sixteenths: Direct evidence on institutional execution costs. *Journal of Financial Economics* 58, 253–278.
- Kadan, O. (2005). So who gains from a small tick size? *Journal of Financial Intermediation* 15, 32–66.
- Kavajecz, K. (1999). A specialist’s quoted depth and the limit order book. *Journal of Finance* 54, 747–771.
- Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica* 53, 1315–1336.

- Lee, C., B. Mucklow, and M. Ready (1993). Spreads, depths and the impact of earnings information: an intraday analysis. *Review of Financial Studies* 6, 345–374.
- Parlour, C. (1998). Price dynamics in limit order markets. *Review of Financial Studies* 11, 789–816.
- Rieder, U. (1979). *Equilibrium Plans in Non-Zero-Sum Markov Games*. In O. Moeschlin and D. Pallaschke (Eds.), *Game Theory and Related Topics*, pp 91-101. North Holland Publishers.
- Rosu, I. (2004). A dynamic model of the limit order book. PhD Thesis, MIT.
- Seppi, D. (1997). Liquidity provision with limit orders and strategic specialists. *Review of Financial Studies* 10, 103–150.
- Vayanos, D. and T. Wang (2003). Search and endogenous concentration of liquidity in asset markets. Working Paper, MIT.
- Weill, P.-O. (2003). Liquidity premia in dynamic bargaining markets. Working Paper, Stanford University.

## A Appendix

### A.1 Proof of Lemma 4.1

The lemma states that for any  $\pi \in (0, 1]$ , if  $\beta^+$  is large enough there exists an equilibrium,  $\{c, (\pi, \phi, \phi')\}$ .

**Proof.** Consider any  $\pi > 0$ . Suppose, to begin with, that traders involuntarily acquire sell-side information with probability  $\pi$  and that if blind they involuntarily play the strategy depicted in Figure 2. Set  $\beta^+ = \infty$ . Via the functions  $\{\phi, \phi'\}$ , all other traders condition their action on the information  $\Pi$ , defined as  $(b, \beta, l)$ . Consider the otherwise similar game where players' strategies are cutoffs on the compact set  $\mathbb{R} \cup \{-\infty, \infty\}$  selected conditional on  $b$ . After choosing the cutoffs, their type  $\beta$  is revealed. The dual game evolves as the original game. The payoff to a limit order is suitably averaged over the types that select it so that for example the expected payoff to a bid is

$$E \left[ \left( \beta + \frac{k}{2} \right) | \beta \in I \right] \Pr[T_B \leq D|b]. \quad (42)$$

where  $I$  is the interval of the type space chosen by the trader such that  $\beta \in I \rightarrow$ (submit bid). Rieder (1979) proves the existence of a stationary equilibrium with Markov strategies when there are a countable number of players, a countable state space, and a compact action space. This proof applies to the dual game thus defined, and so its result also applies to the original game.

Goettler *et al.* (2004) and (2005) previously highlighted the applicability of this theorem in dynamic trading games of this sort. Note that the set of cutoffs,  $\bar{\beta}^b$  and  $\underline{\beta}^b$ , is bounded above and below. The equilibrium is unchanged by taking  $\beta^+$  as any upper bound of that set.

The stationary equilibrium thus derived is consistent with the possibility that the state,  $\omega$ , is a non-stationary random variable over time. The existence of a buy-sell symmetric equilibrium also remains to be proven.

**Buy-sell symmetry** To find a buy-sell symmetric equilibrium, now consider the otherwise similar trading game where sell-side information sets are coarsened to the unordered pairs  $\{\Pi, \tilde{\Pi}\}$ , where  $\tilde{\Pi} = (\tilde{b}, -\beta, \tilde{l})$  (let  $\tilde{0} = 0$ ). Then traders' information is buy-sell symmetric. Furthermore, suppose that the trader action space contains, at market entry,

- if  $\Pi$  then bid and if  $\tilde{\Pi}$  then ask,
- if  $\tilde{\Pi}$  then bid and if  $\Pi$  then ask,
- if  $\Pi$  then market buy and if  $\tilde{\Pi}$  then market sale,
- if  $\tilde{\Pi}$  then market buy and if  $\Pi$  then market sale,
- no action.

At subsequent re-entries (cancellation opportunities) the action space contains

- do not cancel the order,
- replace the order with a bid if  $\tilde{\Pi}$  and with an ask and if  $\Pi$ ,
- replace the order with a bid if  $\Pi$  and with an ask and if  $\tilde{\Pi}$ .

This defines a variant on the trading game where traders have less information about the market because they cannot distinguish buying from selling. Their action is made conditional on the disjunction of  $\Pi$  and  $\tilde{\Pi}$ : but they do not know which of  $\Pi$  and  $\tilde{\Pi}$  obtains. As in the original trading game, a Markov-perfect and stationary equilibrium exists. The joint distributions

$$(T_A, D)|b \text{ and } (T_B, D)|\tilde{b} \tag{43}$$

are equal by construction. Hence, traders would never prefer a different action to that dictated by this equilibrium strategy if they knew which of  $\Pi$  or  $\tilde{\Pi}$  obtained. The equilibrium is therefore a buy-sell symmetric equilibrium in the original game.

**Ergodicity** Consider a buy-sell symmetric equilibrium. For any natural number  $L^B$ , consider the quantity

$$\sup\{\Pr[T_B \leq D|b] : b^B = L^B\}. \tag{44}$$

This quantity converges monotonically to zero as  $L^B \rightarrow \infty$  since time to execution is at most as fast as if all future traders submitted market sales, a time which becomes arbitrarily distant as

$L^B \rightarrow \infty$ . Hence

$$\inf\{\underline{\beta}^b : b^B = L^B\} \quad (45)$$

converges to  $-\frac{k}{2}$  (from below) as  $L^B \rightarrow \infty$ . Repeat this argument for the best ask.

Hence there exists  $L^*$  and  $\alpha < \frac{1}{4}$  such that whenever the depth at a best quote is greater than  $L^*$ , the probability that any given sell-side trader lengthens it is less than  $\alpha$ . Thus the probability that any given trader lengthens it is less than  $\alpha\pi + \frac{1}{4}(1 - \pi)$ , which is less than  $\frac{1}{4}$ . Hence, whenever the depth at a best quote exceeds  $L^*$  the probability that its next change is an increase is bounded away from and below  $\frac{1}{2}$ . And, the probability of a decrease is similarly  $> \frac{1}{2}$ . Standard properties of birth-death processes now imply that market depths at best prices have an ergodic distribution. Therefore with probability 1 they are fully depleted in finite time. Hence from any starting point,  $A$  or  $B$  eventually reach all prices in the admissible range,  $[Mk, Qk]$ . So depths at all prices are ergodic.

Now suppose for a contradiction that in equilibrium, to maintain prices in the admissible range,  $[Mk, Qk]$ , the exchange's position in the asset is non-stationary. Therefore, its position is explosive. Hence it immediately executes against all limit orders. Therefore limit orders are strictly more desirable than market orders, and all strategic traders submit limit orders. So over time, more limit orders than market orders are submitted by traders (taking together both blind and strategic traders). This is a contradiction because then in finite time with probability 1, the exchange *does* close its position in the asset.

This proves the existence for any  $\pi > 0$  of a buy-sell symmetric and ergodic equilibrium,  $\{\phi, \phi'\}$ , where blind traders are non-strategic and play the action dictated in Figure 2. Proposition 6.4 then shows that provided  $\beta^+$  is large enough blind traders' actions thus defined are indeed a best response in this equilibrium. The equilibrium implies excess rents for traders with sell-side information relative to a blind trader. Therefore a non-negative fee for order book information  $b$  exists making traders just indifferent to acquiring it. The exchange can set  $c$  equal to this fee. So defined, the set  $\{c, \{\pi, \phi, \phi'\}\}$  is an equilibrium in the full game. ■

## A.2 Proof of Proposition 5.1

The proposition states: suppose  $\pi < 1$ . Then

$$\Pr[T_A > D] = \frac{2\bar{d}\rho\bar{L}^{bl}}{(1 - \pi)} - \epsilon. \quad (46)$$

Therefore, under, in addition, the conditions of Theorem 4.2,

$$\bar{L}^{bl} = \frac{(1 - \pi)}{2\bar{d}\rho} \left( \epsilon + \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]} \right). \quad (47)$$

**Proof.** Deadlines recur throughout time. Thinking of some representative deadline, execution risk is

$$\frac{E[\text{blind orders annulled by deadline}]}{E[\text{limit orders submitted blindly since last deadline}]}. \quad (48)$$

This is

$$\frac{E[\text{blind orders annulled or re-annulled by deadline}]}{E[\text{limit orders submitted blindly since last deadline}]} - \epsilon, \quad (49)$$

which is

$$\frac{E[L_t^{bl} | \text{deadline at time } t]}{(1 - \pi)/2\rho\bar{d}} - \epsilon, \quad (50)$$

where  $\frac{2\bar{d}}{1-\pi}$  is the expected time between successive blind limit orders. If the deadline is independent of the market, this may be further simplified to

$$\frac{2\bar{d}\rho\bar{L}^{bl}}{(1 - \pi)} - \epsilon. \quad (51)$$

### A.3 Proof of Lemma 5.2

The lemma states: Let order flow and trader numbers,  $N\lambda$ , be constant. Suppose that the market is two-tick, so that the exchange maintains or pegs constant  $A$  and constant  $B = A + k$ . Then

$$\epsilon \leq \{\Pr[T_A > D]\}^2. \quad (52)$$

Therefore, if execution risk (given by  $\Pr[T_A > D] = \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}$ ) is small, then  $\epsilon$  is second order.

**Proof.** The *ex ante* probability of being re-annulled is  $\Pr[T_A > D]$  multiplied by the probability of an order being annulled conditional on already having been so. Just after its first annulment, such an order is no less likely to trade soon than if it were then submitted. Therefore this latter conditional probability does not exceed  $\Pr[T_A > D]$ . The lemma follows. ■

### A.4 Proof of Lemma 6.1

The lemma states: In equilibrium the ergodic probability that any given order is a market sale is  $\frac{1}{4}$ . This is also the probability it is respectively a market purchase, a bid, or an ask. Therefore, writing  $\mu_b$  for the probability that a trader submits a market sale given the signal,  $b$ ,

$$E[\mu_b] = \frac{1}{4}. \quad (53)$$

**Proof.** Let  $L$  be the total number of outstanding and unmatched limit orders in the market. On market entry, submitting a market order decreases  $L$  by 1, while a limit order increases  $L$  by 1. Traders do not change  $L$  on re-entry to the market, since if they cancel a limit order they replace it with another. Let  $\chi_b$  and  $\sigma_b$  be the respective probabilities that on arriving at the market for the first time, and gathering information  $b$  (with possibly  $b = 0$ ), a trader draws a type inducing her to demand liquidity via a market order ( $\chi_b$ ) or supply it via a limit order ( $\sigma_b$ ). Recall that the net excess supply of liquidity due to the regulating activities of the exchange is on average zero. Therefore, as  $L$  is a stationary random variable, the ergodic probability that any given trader increases  $L$  by 1 is equal to the probability that she decreases  $L$  by 1. That is,

$$E[\chi_b] = E[\sigma_b]. \quad (54)$$

But,

$$E[\chi_b] + E[\sigma_b] = 1. \quad (55)$$

Consequently,

$$E[\chi_b] = E[\sigma_b] = \frac{1}{2}. \quad (56)$$

Due to buy-sell symmetry in equilibrium, half of all a limit orders are bids, and half are asks. Half of all market orders are buys, and half are sells. The lemma follows. ■

## A.5 Proof of Corollary 7.1

The corollary states : Suppose  $F$  has a uniform distribution symmetric about zero. Suppose that the support of  $F$  is broad enough that for all  $b$ , market sales occur with positive probability. Then the harmonic mean of execution risk is given by

$$E \left[ \frac{1}{\Pr[T_A > D|b]} \right]^{-1} = \frac{k}{\frac{k}{2} - F^{-1} \left[ \frac{1}{4} \right]}. \quad (57)$$

**Proof.** For all  $b$ ,  $\underline{\beta}^b \in (\beta^-, -\frac{k}{2}]$ , where  $\beta^-$  is the lower bound of  $F$ 's support. So

$$E[\underline{\beta}^b] = F^{-1} \left[ \frac{1}{4} \right]. \quad (58)$$

Rearranging (28),

$$\frac{\frac{k}{2} - \underline{\beta}^b}{k} = \frac{1}{\Pr[T_A > D|b]}. \quad (59)$$

Then taking expectations,

$$\frac{\frac{k}{2} - F^{-1} \left[ \frac{1}{4} \right]}{k} = E \left[ \frac{1}{\Pr[T_A > D|b]} \right]. \quad (60)$$

■

## A.6 Proof of Corollary 7.2

The corollary states: Under this generalization to include a common value, the blind best response as given in Figure 2 is played in the limit of a sequence of equilibria where  $\pi \downarrow 0$ , so that the market goes dark.

Suppose in addition that  $(Q - M) \rightarrow \infty$  (so that the exchange almost never intervenes), and trading becomes ever more on a penny, so

$$\Pr[(A - B) > k] \downarrow 0. \quad (61)$$

Then the welfare results of Theorem 4.2 are asymptotically true.

**Proof.** In the limit of a sequence of equilibria where  $\pi \downarrow 0$ , as the markets for liquidity clear, types  $\pm F^{-1}[\frac{1}{4}]$  are indifferent, when blind, between a market order and a limit order priced one tick less favorably. So Figure 2 represents limiting blind equilibrium strategies. Redefine  $m = \frac{A+B}{2}$ . As the equilibrium is buy-sell symmetric, the expectation of  $(m - v)$  is zero. The expected payoff to a market purchase at  $p = A$  is

$$E[\beta - (p - m) - (m - v)], \quad (62)$$

which is, as in the main model,  $(\beta - \frac{k}{2})$ . The expected payoff to a bid at  $p = A - k$  is

$$E[(\beta - (p - m) - (m - v)) \Pr[T_B \leq D|b]], \quad (63)$$

or

$$(\beta - (p - m))E[\Pr[T_B \leq D|b]] + Cov(v - m, \Pr[T_B \leq D|b]). \quad (64)$$

Intuitively, the covariance in (64) would typically be negative since if  $v$  exceeds  $m$ , informed traders will buy, so uninformed bids will execute slowly. However, this depends on the persistence of  $v$  – which is not specified. In the limit as  $\pi \rightarrow 0$  the effect of  $v$  becomes negligible, so  $v$  becomes independent of  $\Pr[T_B \leq D|b]$ . In the limit as  $(Q - M) \rightarrow \infty$ , the exchange almost never intervenes, so the mid-quote  $m$  becomes independent of  $\Pr[T_B \leq D|b]$ .

Combining the last two remarks,  $Cov(m - v, \Pr[T_B \leq D|b]) = 0$  in the limit. The corollary also implies that  $(m - p) \rightarrow \frac{k}{2}$ . So (64) is  $(\beta + \frac{k}{2}) \Pr[T_B \leq D]$ , as in the main model. Hence, the expected discount factor blind traders associate with the execution of a limit order submitted at the best quotes is as in Proposition 6.4. Buy-side trader welfare follows as in the proof of Theorem 4.2. ■

## A.7 Proof of Corollary 7.3

The Corollary states: under these adaptations to the model, suppose that in equilibrium the support of  $F$  is broad enough that for all  $b$ , market sales occur with positive probability. Then for  $\iota$  near enough to zero:

(a) The blind best response is as depicted in Figure 2.

(b) The expected waiting costs per blind trader are  $\frac{k\delta}{2}$ . So the welfare results of the paper are qualitatively unchanged.

(c) If  $\pi \approx 0$  then average depths are  $\frac{k}{2dF^{-1}[\frac{3}{4}]}$ . So the comparative statics on depths in Section 5 are also unchanged, except that they refer to *entire* depth, not just inside depth.

**Proof.** In any equilibrium, traders follow cutoff strategies, and for all  $b$ , the marginal seller with type  $\underline{\beta}^b$  has equal cost of delayed execution as cost of immediate execution:

$$k = |\underline{\beta}^b|E[T_A|b]. \quad (65)$$

(marginal buyers are similar). Substituting in the identity  $E[T_A] = E[E[T_A|b]]$  using the last equation,

$$\frac{k}{\underline{\beta}^0} = E \left[ \frac{k}{\underline{\beta}^b} \right], \quad (66)$$

provided that  $\iota$  is near enough to zero. Hence  $-\frac{\eta}{\underline{\beta}^0} = E \left[ -\frac{\eta}{\underline{\beta}^b} \right]$ . Lemma 6.2 then implies that  $\underline{\beta}^0 = F^{-1} \left[ \frac{1}{4} \right]$ . From buy-sell symmetry, (a) is true. Statement (b) then follows from an analogous argument to the proof of Theorem 4.2, where  $shortfall = |\beta|E[T_A]$ . Then the average blind waiting time of a limit order is  $E[T_A] = \frac{k}{F^{-1}[\frac{3}{4}]}$ . Statement (c) concerns the case where almost all limit orders are blind. Statement (c) follows, as in Section 5, on noting that the average waiting time of a limit order is the average depth multiplied by the average duration between limit order arrivals. ■

## A.8 Proof of Corollary 7.6

The corollary states: Suppose that  $\pi \approx 0$ , so that the market is almost dark. When equilibrium welfare is compared to that in the game where all traders' action space is expanded, buy-side welfare is diminished by the expanded action space provided that some blind types make use of the expanded action space. Otherwise it is unchanged.

**Proof.** For the markets for liquidity to clear, the median buyer and seller are, blind, indifferent between market order and limit order. Traders with  $\beta$  in  $F$ 's interquartile range submit limit orders when blind. As some of these trader types strictly prefer to submit limit orders at the quotes, the median buyer and seller (who face greater stakes but the same execution risks)

do, weakly. Hence, the expected discount factor associated with execution at the best quotes is as in Theorem 4.2. Figure 2 represents in partial detail blind equilibrium strategies. Traders (with  $\beta$  near zero) who now submit blindly away from the quotes expect higher utility than they would if their action space were not expanded. The others expect the same utility.

An accounting gain arises when limit orders submitted outside the best quotes ultimately execute. This is due to a mismatch in the way the mid-quote is calculated by the buyer and the seller. We net this accounting gain for a true measure of efficiency. Hence in the calculation of blind welfare, net the mid-quote mismatch (if any) off of all payoffs to limit order traders. All limit orders thereby netted provide on execution a payoff of  $(|\beta| + \frac{k}{2})$  (as in the original model).

Suppose that some blind traders types chose to submit limit orders outside the best quotes. These experience strictly more execution delay than in the original model. So, welfare is lower.

■

## A.9 Proof of Corollary 7.7

The corollary states: Adapt the main model so that sell-side traders can price limit orders outside the quotes, and outlaw cancellation. Suppose in addition that the exchange intervenes to avert states which preclude sell-side orders at best (closing its position so that its net liquidity supply is zero).

Then ergodic equilibrium exists and the results of Theorem 4.2 are true.

**Proof.** The proof of equilibrium existence goes through unchanged. Let  $\underline{\beta}^b$  be defined as the marginal type with signal  $b$  in the (market sale) / (ask at best offer) decision. Then (by averaging their decision problems)  $F(\underline{\beta}) = E[F(\underline{\beta}^b)]$ . Furthermore, as markets for liquidity clear in this buy-sell symmetric setting,  $\frac{1}{4} = E[\mu_b]$ . Therefore

$$F(\underline{\beta}) = \frac{1}{4} + E[F(\underline{\beta}^b) - \mu_b].$$

With signal  $b$ , if any type wishes to submit an ask at the quote, then  $\underline{\beta}^b$  does weakly. So, for all  $b$ ,  $\mu_b = F(\underline{\beta}^b)$ . The proposition follows. ■