Price Tick and Welfare when Assets Trade on a Penny

Jeremy Large*

Jeremy.Large@Economics.Ox.Ac.Uk
All Souls College, University of Oxford, Oxford, OX1 4AL, U.K.

28 March 2006

Abstract

Using a stochastic sequential game, this paper models an electronic limit order book that *trades on a penny*, meaning its bid-ask spread is almost always the price tick size. By deriving flow-equality features of dynamic ergodic equilibrium, it deduces the buy-side's strategy and welfare, while bypassing altogether their complex forecasting problem. Per agent, this welfare decreases with tick size – but is actually invariant to potentially beneficial measures like increased or more consistent trading volumes, more sophisticated trading, or modified queuing rules. Depths adjust producing equilibrium effects which exactly offset these. The paper advocates narrowing the tick, but anticipates resistance from sell-side traders.

JEL classification: C73, G14, G24

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

*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 Market Microstructure Conference participants at the Norwegian Central Bank, Oslo (September 2005). 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 studies efficiency in automated electronic financial markets (limit order books) whose assets 'trade on a penny', meaning that the bid-ask spread is bid down to 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[1] are current examples. Figure 1 illustrates the Vodafone case. Such markets tend to experience very high inside depth (i.e. offered depth at the best quotes), leading to order queueing. Taking the CBOT example, on a typical day in 2004 its median-sized trade consumed less than 1.5 per cent of the available inside depth.[2] For pre-decimalization NYSE stocks Harris (1994) studied such inflated inside depths, and depths' collapse after a cut in tick size was studied in Goldstein and Kavajecz (2000).

I take it as a premise that in many such cases, the exchange has a direct lever over the bid-ask spread, for even a substantial cut in the price tick would leave the asset still trading on a penny, at a reduced spread.[3] Thus there is a clear policy issue to address: would cutting the price tick be more efficient? What parties might resist such a cut? What alternative measures would enhance trader welfare?

The bid-ask spread, though often used as a measure of market quality, does not here impact efficiency directly, as its net effect on the gains from trade is zero. Nor, being largely fixed in this distinctive setting, does it proxy for possibly harmful insider trading (see Kyle 1985, and Glosten and Milgrom 1985). Rather, *depths* drive welfare here: for, in this highly liquid situation no trader fails to transact, so welfare losses emerge primarily in the delay while limit orders effectively queue for execution.

This poses an interesting modelling challenge, since depths depend on past acts and have complex, variable, dynamics. For example, one or the other inside depth falls down to zero each time offered prices change. Nevertheless this can be captured in an unending stochastic sequential game with variable prices and depths, similar to one simulated in

[1] The Chicago Board of Trade.

[2] 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.

[3] Foucault, Kadan, and Kandel (2005) provide conditions (where the asset does not trade on a penny) under which cutting the price tick may not reduce the average spread.

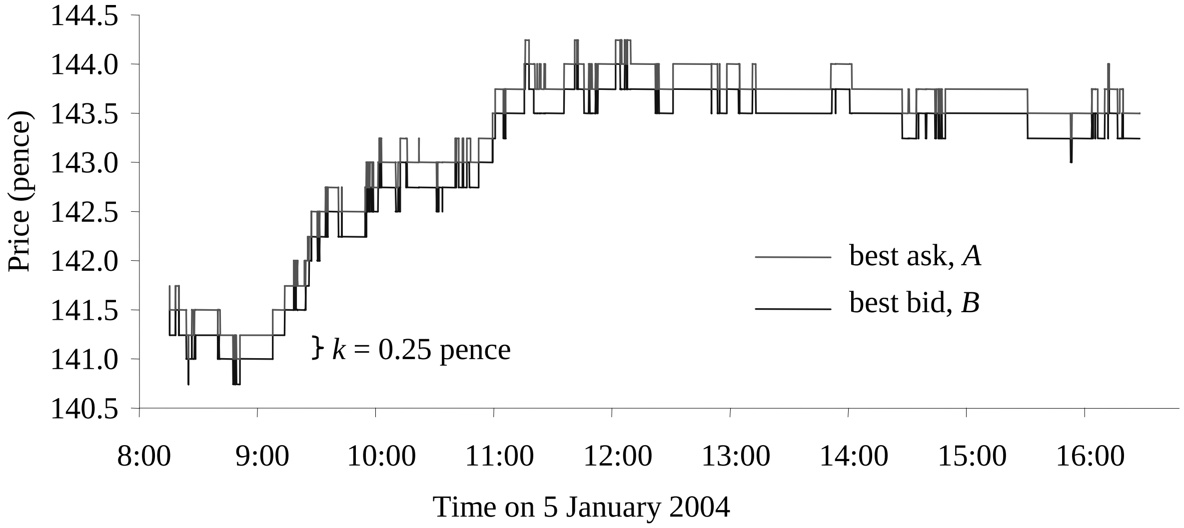

Figure 1: The prices of the best bid and best ask for Vodafone shares on the London Stock Exchange, 5 January 2004. The number of transactions on this day was 14 times higher than the number of changes in the bid (ask). The bid-ask spread was wider than 0.25 pence 3.5 % of the time. The notation, k,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

Goettler, Parlour, and Rajan (2005),[4] and welfare can be derived in closed form.

Welfare implications turn out to depend on a distinction between buy-side investors and sell-side traders. By 'sell-side' is meant: 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 investors, known as the 'buy-side'. In equilibrium cutting the price tick induces lower depths, hence shorter waits, and thereby an increase in buy-side welfare. However, the sell-side prefers to safeguard the value of its queueing expertise with a tick size which is inefficiently bounded away from zero. These results suggest that the tick size should be cut, at least so far as trading is still on a penny, but there may be resistance to such a reform on the sell-side.

On the other hand, under reasonable assumptions buy-side welfare per trader is invariant to a range of other *prima facie* beneficial policies: for example, attracting a larger trading population or more reliable order flow, encouraging more traders to acquire sell-side skills, or making available more detailed information about the book outside the quotes (e.g. Level 2). Within wide bounds it is also invariant to the order queueing rules enforced by the exchange – FIFO or otherwise – as well as to trader impatience. When any of these variables changes, market depth adjusts (with sell-side rents) to produce

[4]Where possible, Goettler *et al.* (2005)'s key notation is used in developing the model.

equilibrium effects which offset any *prima facie* benefits, as it is governed by a constraint that two markets for liquidity clear.

This last formulation follows Foucault, Kadan, and Kandel (2005), which studies the limit order book as a market where suppliers of liquidity (or immediacy) submit limit orders, and demanders of liquidity submit market orders which trade with them. If depths did not adjust, either supply or demand would be over-represented in the order flow. The tick size then impacts welfare as follows: Being the penalty paid for immediate execution, if cut it induces too many market orders, unless depths fall making limit orders *ex ante* more attractive – equivalently, increasing *ex ante* welfare.

This relies on the following simple observation: that seasonality aside, *any limit order book's depth is stationary* (while prices may be non-stationary). But then the volume per unit time of market orders it attracts must be matched on average by an equal flow of uncancelled limit orders. Otherwise, offered 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 uncancelled limit order, and vice versa, they arrive equally frequently over time. As already indicated, this paper interprets it mainly as a competitive market-clearing condition at the bid, and another at the ask, whereby the ergodic distribution of depths, and more generally of bid-ask spreads, adjusts in equilibrium to equate the supply and demand for liquidity.

These intuitions in terms of depths' market-clearing role are bypassed in the core of the paper, which moves straight to a characterization of buy-side welfare. Section 5 analyzes depths' role later and provides predictions in line with 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.

Inside depth's positive dependence on trading volume is observed on the SEHK in Brockman and Chung (1996). See also Danielsson and Payne (2002)[5] 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 was estimated in Harris (1994) and studied in Goldstein and Kavajecz (2000), among other empirical work. None of the above studies is of trading on a penny exclusively.

[5] 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.

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. Cordella and Foucault (1999) explain a channel whereby a large tick on a limit order book can enhance welfare. A simulation experiment of Goettler *et al.* (2005) suggests that the aggregate effect of this can be to diminish welfare, even when trading is not on a penny.

When trading is on a penny welfare is decreasing in depth. This runs counter to the intuition that high depths are desirable since large market orders can trade at or near the quotes, which is nonetheless an important consideration elsewhere – see Jones and Lipson (2001). 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 small tick size.

The paper proceeds as follows: Section 2 sketches the model and introduces the main intuitions. Section 3 fully details the model and equilibrium. Section 4 gives the paper's main results on efficiency and buy-side trader welfare. Section 5 derives average inside depths in a simplified model. The main welfare proof is outlined in Section 6 (details are left to an Appendix). Section 7 relaxes the model's main assumptions, including the requirement that no limit order be posted outside the quotes. Section 8 concludes.

2 Model summary, main result, and intuitions

The paper develops a dynamic trading game inspired by Parlour (1998), adding variable prices. It defines an ergodic equilibrium in a stochastic sequential game, where as in Foucault *et al.* (2005), Goettler *et al.* (2004) and (2005), Hollifield, Miller, and Sandas (2004) 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. Wider aspects of this trade-off are analyzed in Cohen, Maier, Schwartz, and Whitcomb (1981), Chakravarty and Holden (1995) and Handa and Schwartz (1996).

Technically, the model innovates by providing a closed-form expression for welfare 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 – see also Large (2006), which explores this specific contribution.

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 probability π. 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. She aims to maximize her gain from trade, once netted of her implementation shortfall or slippage, and discounted for delay. Limit orders can be cancelled and replaced periodically. 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, π, 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 welfare 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.

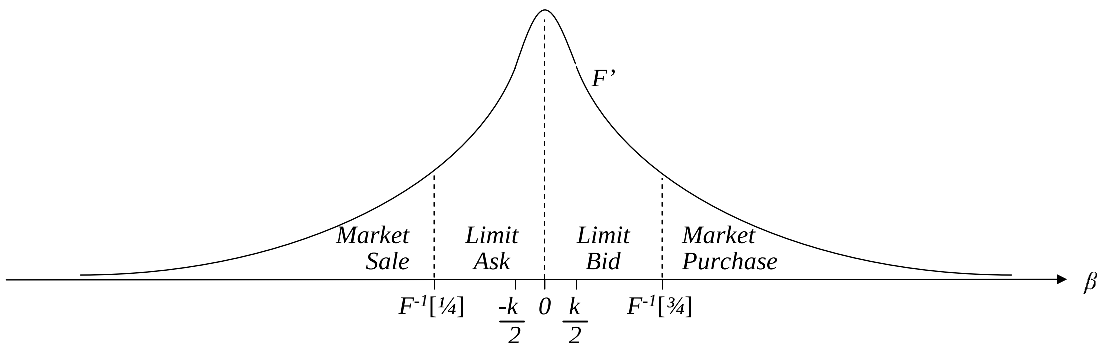

Figure 2: Illustrates the type-space, and the equilibrium strategy of blind traders. The four delimited areas under F' are equal. $F^{-1}\left[\frac{3}{4}\right]$ is indifferent between market purchasing and bidding, and is the 'median buyer'. Likewise, $F^{-1}\left[\frac{1}{4}\right]$ is the 'median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

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 π 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 prefer market orders, which execute immediately, to limit orders in F's two tails. 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 almost exactly half the traders submit limit orders, and half market orders. 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. 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 median types with $\beta \cong \pm F^{-1}\left[\frac{1}{4}\right]$ are when blind indifferent between a limit order and a market order priced one tick less favorably. This fixes all blind traders' expected 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. The main exception is the price tick size, in which buy-side welfare is decreasing. As outlined in the Introduction, cutting the tick size would cause a shortage of limit orders unless on average depths fall to make them *ex ante* more attractive. This involves an increase in welfare.

2.4 Case of arbitrary π

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 welfare is also. The full model takes a realistic parametrization of F and simplifies actions and payoffs. Traders may cancel limit orders only in favor of alternative limit orders. They measure implementation shortfall relative to the infeasible benchmark where they trade immediately at the current mid-quote. Hence there is no winner's curse in limit order execution – for studies of this effect see Foucault (1999) and Handa and Schwartz (1996), as well as Section 7. Conversely, no market order produces negative transaction costs, as some do in the simulation of Goettler *et al.* (2005).

The invariance to π follows approximately for reasonable choices of F. 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). These indifferent cutoff types differ from state to state of the market. They must have as an average both the median buyer, $F^{-1}\left[\frac{3}{4}\right]$, 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 modelled form of F they are both hyperbolic and so coincide) and so $F^{-1}\left[\frac{3}{4}\right]$ 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 essentially these lines.

2.5 The sell-side

The invariance to π 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 $\to 0$ as the price tick, k, $\to 0$. 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 price tick.

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, Ω, is defined in Section 3.2). Subject to a buy-sell symmetry constraint introduced in Section 3.3, 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 π. The content of the signal, b, is defined later in Section 3.2. 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, β, from a distribution on $\mathbb{R}$, whose continuous cumulative density function (CDF), denoted F, is symmetric about 0 and is detailed in Section 3.8. A positive (negative) β predisposes them toward buying (selling). In reality, the buy-side cannot acquire sell-side information and skills simply for the purpose of a single trade, so in the model they commit to π without conditioning on their current trading need, i.e. on their type, β. The next part shows how β enters payoffs.

3.1 Traders' objectives

On discerning β 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 β 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, $-\beta - (m - p)$. Later, the mid-quote as perceived by buyers will differ from that of sellers when the bid-ask spread is wide.

Goettler *et al.* (2005) simulates an equilibrium with a subtly different payoff, where slippage is calculated relative to an unobserved underlying common value of the asset, rather than as here to the mid-quote. It also permits traders to select the size of the order. In a common-value setting with asymmetric information Easley and O'Hara (1987) has also studied the decision on trading volume.

In contrast to this literature, here 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. A consequence of this is to make signals of the asset's long-run common value unimportant to the sell-side – perhaps not unreasonable in this high-frequency setting. Section 7 nevertheless adapts the model to incorporate a stochastic common value, and recovers the paper's results as π and informed trading both decline to zero (interestingly, moderate informed trading appears to *enhance* welfare). Variable trading volumes only enter here via the variable trader arrival process, N.[6]

[6]As N has, subject to a buy-sell symmetry constraint explained in Section 3.3, arbitrary dynamics, it can be interpreted as the reduced-form outcome of a decision about trading intensity conditional on the state of the order book, made *ex ante*.

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.

Definitions Call the highest price of bids in the limit order book B, and the lowest price of asks A, so that $A \geq B + k$. The best bid and offer, 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.

Definition 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.

Definition The signal b is an event on the state space Ω. 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.

Definitions 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.

3.3 Main assumptions

The model restricts traders to replace a limit order only in favor of a limit order, and not of a market order. It therefore abstracts away from the ephemeral supply of liquidity due to 'fleeting' order submission documented in for example 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, λN, so trader arrival intensity is the same in state ω as it is in the state $\tilde{\omega}$. For all ω,

$$N|\omega = N|\tilde{\omega}. \tag{2}$$

3.4 Trader action

Arrival in the market Let Σ 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, ϕ which maps her type, β and signal, b onto an action. Thus for any b and β,

$$\phi(b, \beta) \in \Sigma. \tag{3}$$

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 ϕ' 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 ϕ' is a function such that

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

3.5 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 periodically submits market orders to clear its resulting position. Their timing is self-regulating, so that 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.

3.6 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 $A - B > k$, 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 outstanding limit orders by pricing it between B and A.

Simplifying the case where $A - B > k$ usefully bring the decision problem into the same form as the more prevalent case, $A - B = k$: As previously stated, traders are restricted to bid only at price $A - k$ and submit asks at price $B + k$. Their payoffs are also simplified. For buyers, define m in (1) by $m = \left(A - \frac{k}{2}\right)$. This is the mid-quote which would subsequently prevail if she submitted a bid. Equally, for sellers define $m = \left(B + \frac{k}{2}\right)$. These two definitions coincide in the normal case that $A - B = k$ but not when $A - B > k$. They imply that trading counter-parties agree on the mid-quote, so that the undiscounted gains from trade equal the absolute difference in their β-types.

3.7 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 π, ϕ, and ϕ' maximize expected payoffs given c, (b, β, l) and the belief that all other traders behave according to $\{\pi, \phi, \phi'\}$.

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

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

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

3.8 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. Assume F is symmetric about zero, and has connected, bounded, support $[\beta^-, \beta^+]$. For all $\beta \in (\beta^-, -\frac{k}{2}]$,

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

Furthermore, $F\left(-\frac{k}{2}\right) > \frac{1}{4}$, which implies that more than half of traders can benefit from a market order. The constant η is chosen to ensure that $F(\infty) = 1$. To satisfy (6) there are suitable atoms at β^- and β^+. 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 section first derives some properties of the trader's limit order / market order decision. It then presents a proof that the relevant equilibrium exists, and gives the main welfare result. It concludes by analyzing the sell-side's incentives towards inefficiency regarding the price tick size.

4.1 Trading deadline

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 ρ. They face a common sequence of randomly-timed trading deadlines arriving according to a Poisson process of parameter ρ. 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 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. Suppose she believes that all other traders will follow / have followed some strategy, $\{\pi, \phi, \phi'\}$ and that b is ergodic.

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 these types 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, observing that the case of selling is exactly equal and opposite. A market purchase at price A yields her positive utility equal to β penalized by her price slippage relative to the mid-quote: namely, $\frac{k}{2}$. This gives a total payoff of $\left(\beta - \frac{k}{2}\right)$, 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. $\left(\beta + \frac{k}{2}\right)$.

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

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

She compares this to the payoff from a market order, $\left(\beta - \frac{k}{2}\right)$. 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 β induces her to submit a market purchase. With β in an interval just below the cutoff she bids. This is studied in more detail in Section 6.1.

4.3 Equilibrium existence

This stochastic sequential game has a countable state space, Ω, 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. By redefining the trader action as a choice of cutoffs such as $\bar{\beta}^b$, and initially letting the blind trader act non-strategically according to the strategy depicted in Figure 2, the game may be redefined so that Rieder (1979) applies. Buy-sell symmetry may be added by considering a game where the trader information set is suitably coarsened so that, in short, they cannot distinguish buying from selling. Finally, ergodicity in b may then be inferred from known properties of birth-death processes. If an ergodic distribution exists then the blind trader's best response is well-defined. Theorem 4.2 shows that it in fact corresponds to his assigned strategy. The full proof of equilibrium existence is left to the Appendix, and incorporates elements of Large (2006), as well as Goettler *et al.* (2004) and (2005).

Lemma 4.1 *For any positive π, there exists an equilibrium, $\{c, (\pi, \phi, \phi')\}$.*

Proof. See Appendix. ■

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.

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

$$\Pr[T_A > D]. \tag{8}$$

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 $E[e^{-\rho T_A}]$.

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

$$|\beta| + \frac{k}{2}. \tag{9}$$

4.4 Main welfare result

Definition 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 δ by

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

Thus $\delta \in (0,1)$ describes the 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 welfare per trader, S, is given by:*

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

where

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

Hence buy-side welfare 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 welfare of the blind trader equals the welfare of all 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

$$\Pr[LO] \times E\left[shortfall|\ LO\right], \tag{13}$$

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

$$shortfall = \left(|\beta| + \frac{k}{2}\right)\Pr[T_A > D], \tag{14}$$

since (9) is the utility which is foregone if $T_A > D$. As the events LO and $[T_A > D]$ are independent, expected shortfall is then

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

From Proposition 6.4 she foresees preferring limit orders iff her β 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 β 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 welfare per buy-side trader is invariant to

- trader impatience, ρ,
- 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, π, who acquire the sell-side signal, b,
- whether the sell-side signal b contains Level 1 or Level 2 information, and
- the way that the exchange intervenes to maintain prices in the range $[Mk, Qk]$.

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

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

which lies between $\delta\frac{k}{2}$ and $\frac{k}{2}$. Holding fixed $F^{-1}\left[\frac{3}{4}\right]$ and δ, 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.

4.5 Sell-side rents

For each mixing probability $\pi > 0$ and price tick $k > 0$, there exists a positive fee level, $c(\pi, k)$, 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

$$W(\pi, k) = \pi c(\pi, k) + S(k), \tag{17}$$

where $S(k)$, invariant to π, is buy-side surplus, as given in (11). Note that $c(\pi, k) \leq k$, 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

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

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

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

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 π in the most efficient equilibrium, $\hat{\pi}(k)$, is the one that would be chosen by the exchange to maximize sell-side rents, $\pi c(\pi, k)$. Note $\hat{\pi}(k)$ exists and is greater than zero. Regardless of π, for given k sell-side rents are then at most

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

But this is bounded above by $(E[|\beta|] - S(k))$, which was derived in Theorem 4.2, 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 intuition for the welfare results of this paper rests on the idea that a wide price tick induces high market depths, and so long waits to trade, but depths adjust to offset the welfare effects of a range of other alterations in the trading environment. Yet, it was possible to quantify welfare while bypassing a direct analysis of depths altogether. On this theme Parlour (1998) is informative and insightful about depth dynamics. The purpose of this section is complementary, studying 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 increase in the price tick, trader patience and the trader arrival rate. Therefore, they do indeed adjust in appropriate ways to explain the welfare invariance results of previous sections.

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 $L = L^{bl} + L^{ss}$, 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 ϵ be the expected number of times that a blindly submitted limit order is re-annulled.

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

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

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

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

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 (22) are ϵ and π. The expectation ϵ 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-anulled. 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*

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

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 ϵ 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 $\epsilon \approx 0$,

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

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 outside depth. 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 an shortage of market orders unless depths rose to make limit orders less attractive.
- Average inside depth is increasing in the order flow, $\frac{1}{d}$, 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 HKSE.
- But average inside depth is decreasing in trader impatience, ρ, since impatient traders avoid limit orders, leading to a shortage unless depths fall.

Average inside depth is invariant to the way that the exchange intervenes in the market to maintain prices.

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 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 part of the paper's motivating intuition: namely, that for queue lengths to be ergodic, the average flow of orders joining the queue of bids or the queue of asks should be equal to the flow of orders leaving that queue. This can be interpreted as a statement that two markets clear. The two markets in question are at the bid and at the ask, and match demand and supply for liquidity.

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 μ_b for the probability that a trader submits a market sale given the signal, b,

$$E[\mu_b] = \frac{1}{4}. \tag{25}$$

Proof. See Appendix. ■

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

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

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 β 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 reservation value 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 reservation values place market sales. The condition for the marginal buyer is

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

while the marginal seller is characterized by

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

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 β lower than $\underline{\beta}^b$ make a market sale. Hence μ_b, the probability that a trader of unknown type submits a market sale given signal b, is

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

6.2 Marginal sellers and the median seller

Definition The *median seller* is the trader with reservation value at the $\frac{1}{4}$ quantile of F, i.e. at $F^{-1}\left[\frac{1}{4}\right]$. Likewise, the *median buyer* has reservation value $F^{-1}\left[\frac{3}{4}\right]$.

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,*

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

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:*

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

Proof. From (28),

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

Recall that

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

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, $F^{-1}\left[\frac{1}{4}\right]$. Therefore average execution risk is given by*

$$\Pr[T_A > D] = \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}, \tag{34}$$

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

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

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

Combining this with Lemma 6.3,

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

So it follows that $\underline{\beta}^0 = F^{-1}\left[\frac{1}{4}\right]$, on substituting $F^{-1}\left[\frac{1}{4}\right]$ for β 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) introducing a common value for the traded asset, 3) the effect of linear waiting costs and 4) the model's limitations on trader actions.

7.1 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.

Lemma 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*

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

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

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

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

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 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 β minus their implementation shortfall, or slippage, relative to v, i.e. :

$$\beta - (p - v). \tag{39}$$

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

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

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.[7] However, an equally desirable welfare level could be achieved by simply cutting the price tick size.

[7]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.

7.3 Linear waiting costs as an alternative to impatience

The central model defines, via β, a payoff to trade which varies across agents. Since some agents want to trade more, they are more reluctant to wait for trade. This part recovers the results of the paper in the case where all buyers and sellers may have an equal motive to trade, but some dislike waiting more than others.

Set trader impatience, ρ, to zero. Suppose that traders face a waiting cost per unit time per unit time that their limit order is outstanding. Let this be the absolute value of their type, $|\beta|$. Buyers when they transact pay the quoted price p, and derive a specific value, V, giving an overall payoff of $(V-p)$. This replaces (1). Sellers receive the negative of this payoff. The reservation value V is left unspecified, but consistent with their types, for buyers with $\beta > 0$ let $V \geq Qk$ while for sellers with $\beta \leq 0$ let $V \leq Mk$. Hence waiting costs are the only source of inefficiency. The final adaptation of the model is to change the background distribution, F. Replace (6) so that for some very small negative ι,

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

All other aspects of the model are left 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:*

- *The blind best response is as depicted in Figure 2.*
- *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 welfare is invariant to the average waiting cost per unit time, $E[|\beta|]$.*
- *If $\pi \approx 0$ then average depths are $\frac{k}{2\bar{d}F^{-1}\left[\frac{3}{4}\right]}$. 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 Limitations on the actions of traders

This part returns to the original payoff formulation of (1). The main welfare results of the paper are derived for a simplified setting with restrictions on the traders' actions. 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 in various ways. It recovers the welfare results of the paper in the case where $\pi \downarrow 0$, so that the exchange is almost 'dark'. Then, with discriminatory limit pricing but no cancellation, it provides conditions where Theorem 4.2 is true for any π, 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 just another limit order.

Corollary 7.4 *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 welfare 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 b 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 7.5 *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 enlarged, buy-side welfare is diminished by the enlarged action space provided that some blind types make use of the enlarged action space. Otherwise it is unchanged.*

Proof. See Appendix. ■

However, Corollary 7.4 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 7.4 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. Cordella and Foucault (1999) covers exactly this case and identifies related welfare losses.

Corollary 7.6 *Adapt the main model so that sell-side traders can price limit orders outside the quotes, and outlaw cancellation. Then ergodic equilibrium exists and, if 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 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 market 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 welfare 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 inside depths deter liquidity supply while shallow inside depths attract it, there exists an equilibrium average inside depth to clear the market. Where markets trade on a penny, equilibrium effects then lead to some striking invariances in buy-side trader welfare: 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, depths adjust to offset the expected sensitivity. The main exception is cutting the price tick, which improves the prospects of a market order, thereby overcoming traders' incentives to add their limit orders to an already congested limit order book. This enhances welfare.

This result is shown to be robust to the existence of a common value for the asset (but limited insider trading), and to the presence of linear waiting costs, as well as to 'microstructure speculation' by sell-side traders who place limit orders outside the quotes.

In opposition to buy-side investors, the sell-side has an interest in keeping the price tick strictly away from zero. Yet as the price tick has a monotonic, negative effect on buy-side welfare, it should be cut at least as long as the market still trades 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.

Easley, D. and M. O'Hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics 19*, 69–90.

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. Sandas (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.

Large, J. (2006). Welfare in stochastic sequential games with binary action. Mimeo, University of Oxford.

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.

A Appendix

A.1 Proof of Lemma 4.1

The lemma states that for any mixing probability π, provided that $\pi > 0$, 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 π and that if blind they involuntarily play the strategy depicted in Figure 2. Via the functions $\{\phi, \phi'\}$, they condition their action on the information Π, defined as (b, β, l). Consider the otherwise similar game where players' strategies are cutoffs selected conditional on b. After acting, their type β 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

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

where I is the interval of the type space define 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 also to the original game. Goettler *et al.* (2004) and (2005) previously highlighted the applicability of this theorem in dynamic trading games of this sort.

However, stationarity equilibrium play does not rule out the possibility that the signal, b, 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 Π then bid and if $\tilde{\Pi}$ then ask,
- if $\tilde{\Pi}$ then bid and if Π then ask,
- if Π then market buy and if $\tilde{\Pi}$ then market sale,
- if $\tilde{\Pi}$ then market buy and if Π 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 Π,
- replace the order with a bid if Π 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 whether Π or $\tilde{\Pi}$ obtains, a fact which they do not know. 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 Π 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 $L^B \to \infty$ since time to execution is at most as fast as if all future traders submitted market sales, a time which becomes arbitrarily distant as $L^B \to \infty$. Hence

$$\inf\{\underline{\beta}^b \ : \ b^B = L^B\} \tag{45}$$

converges to $-\frac{k}{2}$ (from below) as $L^B \to \infty$. 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 α. 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. Therefore, from any starting point, A or B

eventually reach all prices in the admissible range, $[Mk, Qk]$. Hence depths at all prices are ergodic.

This proves the existence for any $\pi > 0$ of a buy-sell symmetric and ergodic equilibrium, $\{\phi, \phi'\}$, where blind traders play (possibly sub-optimally) the action dictated in Figure 2. Proposition 6.4 then shows that blind traders' actions thus defined are indeed a best response. The equilibrium implies excess rents for traders with sell-side information relative to a blind trader. Therefore a positive 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 that deadlines arrive independently of the evolution of the market and $\pi < 1$. Then

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

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

$$\frac{E[\text{blind orders annulled by deadline}]}{E[\text{limit orders submitted blindly since last deadline}]}. \tag{47}$$

This is

$$\frac{E[\text{blind orders annulled or re-annulled by deadline}]}{E[\text{limit orders submitted blindly since last deadline}]} - \epsilon, \tag{48}$$

which is

$$\frac{E[L_t^{bl}|\text{deadline at time } t]}{(1-\pi)/2\rho\bar{d}} - \epsilon, \tag{49}$$

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

$$\frac{2\bar{d}\rho\bar{L^{bl}}}{(1-\pi)} - \epsilon. \tag{50}$$

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

$$\epsilon \leq \{\Pr[T_A > D]\}^2. \tag{51}$$

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 ϵ 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 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 μ_b for the probability that a trader submits a market sale given the signal, b,

$$E[\mu_b] = \frac{1}{4}. \tag{52}$$

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 δ_b and σ_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 (δ_b) or supply it via a limit order (σ_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,

$$E[\delta_b] = E[\sigma_b]. \tag{53}$$

But,

$$E[\delta_b] + E[\sigma_b] = 1. \tag{54}$$

Consequently,

$$E[\delta_b] = E[\sigma_b] = \frac{1}{2}. \tag{55}$$

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 Lemma 7.1

The lemma 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

$$E\left[\frac{1}{\Pr[T_A > D|b]}\right]^{-1} = \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}. \tag{56}$$

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

$$E[\underline{\beta}^b] = F^{-1}\left[\frac{1}{4}\right]. \tag{57}$$

Rearranging (28),

$$\frac{\frac{k}{2} - \underline{\beta}^b}{k} = \frac{1}{\Pr[T_A > D|b]}. \tag{58}$$

Then taking expectations,

$$\frac{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}{k} = E\left[\frac{1}{\Pr[T_A > D|b]}\right]. \tag{59}$$

■

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

$$(\Pr[A - B] > k) \downarrow 0. \tag{60}$$

Then the welfare results of Theorem 4.2 are asymptotically true.

Proof. In the limit of such a sequence of equilibria, 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. 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

$$E[\beta - (p - m) + (m - v)], \tag{61}$$

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

$$E[(\beta - (p - m) + (m - v)) \Pr[T_B \leq D|b]], \tag{62}$$

or

$$(\beta - (p - m))E[\Pr[T_B \leq D|b]] + Cov(m - v, \Pr[T_B \leq D|b]\ ). \tag{63}$$

Intuitively, we would expect the covariance in (63) to be negative since if m exceeds v, informed traders will buy, so uninformed bids will execute slowly. However, this depends on the persistence of v. In the limit as $\pi \to 0$ the effect of v becomes negligible, so v becomes independent of $\Pr[T_B \leq D|b]$. In the limit as $(Q - M) \to \infty$, 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. So (63) is, as in the main model, $\left(\beta + \frac{k}{2}\right) \Pr[T_B \leq D]$. 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:

(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 $\frac{k}{2\bar{d}F^{-1}\left[\frac{3}{4}\right]}$. 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:

$$k = |\underline{\beta}^b| E[T_A|b]. \tag{64}$$

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

$$\frac{k}{\underline{\beta}^0} = E\left[\frac{k}{\underline{\beta}^b}\right]. \tag{65}$$

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 $E[T_A] = \frac{k}{F^{-1}\left[\frac{3}{4}\right]}$. Statement (c) implies almost all limit orders are blind. It 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 7.5

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 enlarged, buy-side welfare is diminished by the enlarged action space provided that some blind types make use of the enlarged 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 β 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 β 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.

Overall welfare would therefore be unambiguously enhanced, were it not for an accounting gain when limit orders away from the quotes execute. This is due to a mismatch in the way the midquote is calculated by the buyer and the seller. Take the 'deadline' definition of trader impatience. I net this accounting gain for a fair measure of efficiency. This can be achieved by netting all payoffs to limit order traders in the calculation of welfare by the mid-quote mismatch (if any). All sorts of limit order therefore produce for the trader, as in the original model, a net payoff of $\left(|\beta| + \frac{k}{2}\right)$. However, some blind limit order traders types chose to experience strictly more execution risk and delay than in the original model. So, welfare is lower. ■

A.9 Proof of Corollary 7.6

The corollary states: Adapt the main model so that sell-side traders can price limit orders outside the quotes, and outlaw cancellation. Then ergodic equilibrium exists and, if 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 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. ■