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

Jeremy Large*

JEREMY.LARGE@ECONOMICS.OX.AC.UK
All Souls College and the Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, OX1 4AL, U.K.

2 January 2008

Abstract

Limit order markets with stationary dynamics attract equal volumes of market orders and uncanceled limit orders, equalizing the supply and demand for liquidity and immediacy. To maintain this balance, market orders must share any benefit obtained by limit order traders from more efficient trading conditions, such as better order queuing policies. Therefore an efficient market places a low price on immediacy, producing small bid-ask spreads. Furthermore, when price-discreteness leads to a mainly-constant spread, cutting the price tick raises surplus. This is modeled with a stochastic sequential game, using stationarity considerations to bypass direct analysis of traders' intricate market forecasts.

JEL classification: C73, G14, G24

Keywords: stochastic sequential game, stationary equilibrium, 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, Albert Menkveld, Meg Meyer, Christine Parlour, Hyun Shin and Gunther Wuyts for insightful discussions of this work, as well as to anonymous referees. I thank 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

Electronic limit order books have become a dominant mode of exchange on liquid financial markets. Where the trading day is sufficiently long and uninterrupted, the order book microstructure has time to evolve into a stationary distribution. This paper explores the constraints stationarity imposes. Stationarity is useful because it implies that depths rise as often as they fall. Therefore, over time traders choose to submit equally many market orders as limit orders (leaving canceled orders aside). Following Foucault, Kadan, and Kandel (2005), one may think of limit orders as the supply of liquidity or immediacy, and market orders as the demand; and understand this as a 'liquidity market-clearing' condition. This paper models a dynamic setting where, by concentrating on this condition, traders' equilibrium strategies can be characterized. This leads to results about welfare, market inefficiency, and average depths, as functions of the tick size and other institutional features.

This setting has two important features: first, following the literature of Foucault et al. (2005), Goettler, Parlour, and Rajan (2005), Parlour (1998), Ohta (2006) and Rosu (2006), it contains only liquidity trading. This means that traders' idiosyncratic impulses to trade the asset are insensitive to its fundamental value. However, as in Foucault (1999) and Parlour (1998) traders are strategic, dynamically choosing between limit orders and market orders by trading-off delay against slippage.

Second, the market exhibits tight bid-ask spreads because of a large price tick size, so that the spread is practically always constant and equal to the price tick size. This feature, first modeled in Parlour (1998), is found in Field and Large (2007) in fixed-income limit order markets, including ten-, five- and two-year US Treasury Bond futures, as well as short-term interest rate futures such as Euribor, Short Sterling, and Eurodollar. Until early 2007, Vodafone on the London Stock Exchange (depicted in Figure 1) was another example.

In the model, the tick size fixes market efficiency at a single level consistent with stationarity. Consequently, without changing the tick size, adaptations to the trading environment intended to improve welfare fail. For example, Field and Large (2007) notes that some of the above futures markets use a price-time matching rule for limit orders, while others prefer a pro-rata rule. Recently, exchanges have experimented with a third, hybrid, order matching rule for Euribor and Short Sterling.[1] How should the order-matching rule be chosen? Underlying equilibrium responses vary with the choice of order-matching rule; nevertheless, I show

[1]See www.euronext.com/fic/000/022/595/225955.pdf (28 June 2007). Section 6.2 below explains the pro-rata order-matching rule.

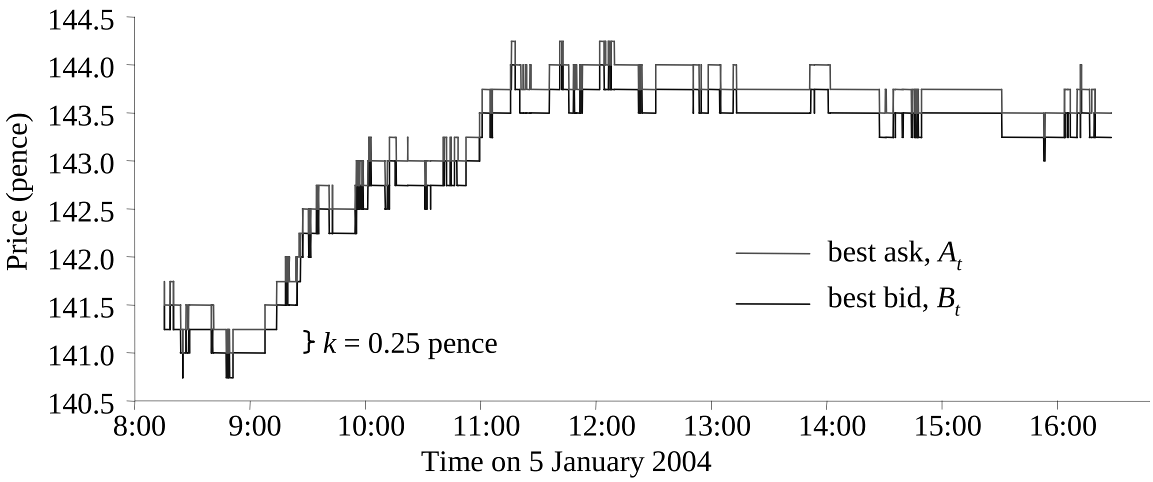

Figure 1: The prices of the best bid and best ask for Vodafone shares on the London Stock Exchange, 5 January 2004. The number of transactions on this day was 14 times higher than the number of changes in the bid (ask). The bid-ask spread was wider than 0.25 pence, only 3.5 per cent of the time. The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

that because market dynamics are in any case stationary, market efficiency is invariant. I also prove similar invariances with respect to the quality of trader information, and the rate of trader arrival.[2]

A useful intuition for the invariance results is as follows. As a benchmark, consider a stationary equilibrium satisfying the liquidity market-clearing condition. Suppose that changing a primitive of the model, such as the queuing rules, resulted in another stationary equilibrium with greater surplus per agent but the same bid-ask spread. As market orders' slippage would not thereby change, nor, in the current setting, would their payoff. Thus the supposed gains in surplus would be enjoyed only by limit order-submitters. So, the market would be oversupplied with agents who prefer to submit limit orders, which in turn would accrue indefinitely, violating the stationarity of offered depths in equilibrium. This is a contradiction. So, the new equilibrium offers no greater surplus. A converse argument shows that it also offers no less.

Since the market only contains liquidity traders, their surplus equals those gains from trade that are not wasted 1) as limit orders await their counterparties; or 2) in matching failures when the asset is bought (sold) by a trader with an excessively low (high) willingness to pay. The waste is the market's inefficiency, which is identified, via this analysis, by the 'liquidity market-clearing' condition. Inefficiency adjusts to a market-clearing level when the market microstructure is stationary.

[2]A later section proves a similar invariance with respect to average depths, under conditions.

On markets like the one in Figure 1, where the bid-ask spread is normally equal to the price tick size, the exchange can directly alter the bid-ask spread by adjusting the price tick. In fact, in February 2007, the LSE cut Vodafone's tick size by 60%. My results support such a policy: in the model, the price tick size has a strong positive effect on market inefficiency and should be cut, at least so far as the bid-ask spread continues to approximate the price tick size. The mechanism for this works as follows: Being equal to the bid-ask spread, the tick size is the penalty paid for the advantage of immediate execution. Hence, cutting it encourages too many market orders to sustain stationarity, unless at the margin limit orders also become ex ante more attractive – in aggregate, increasing ex ante surplus.

This mechanism is well-illustrated by a comparison in Goettler et al. (2005) of two simulated limit order markets, which are identical except that the tick size is halved in the second simulation (see their Tables XIII and XIV). The cut in the tick diminishes the average quoted spread (although it need not: see Seppi 1997 on non-monotonicities). This benefits market orders. Equally, limit orders trade at worse prices; but are compensated for this in stationary equilibrium with lower execution risk, and a shorter wait to trade. As execution risk is lower, the aggregate effects of the smaller price tick are greater welfare and less inefficiency.[3]

The adjustment of inefficiency to market-clearing levels can be understood partly as an adjustment in average depths, since (all other things being equal – in particular the order arrival rate) high depths are symptoms of inefficient congestion as limit orders queue to trade. The paper analyzes depths' role in this and provides empirical predictions: Average depths at the best bid and ask ('inside depths') increase with

- order arrival rate (in the sense of Foucault et al. 2005: i.e. limit orders plus market orders per unit time), which proxies for trading volume, and
- the price tick size.

Depths' positive dependence on trading volume is observed on the Stock Exchange of Hong Kong in Brockman and Chung (1996). On the other hand, Lee, Mucklow, and Ready (1993) finds that at times of high traded volume, NYSE specialists reduce their choice of quoted inside depths. Kavajecz (1999) also focuses on the NYSE. Foucault et al. (2005) and Rosu (2006) also predict more aggressive limit orders when the order arrival rate rises (or impatience falls). However, this shows up in a narrower bid-ask spread rather than in higher depths.

Depths' positive dependence on the price tick size was estimated for NYSE stocks in Harris

[3]In Goettler et al. (2005), as here, welfare is decreasing in average depth. This runs counter to the intuition that high depths are desirable since unusually large market orders can trade at or near the quotes, which nonetheless is an important consideration in some cases – see Jones and Lipson (2001).

(1994), and was corroborated in a suite of contemporary and subsequent papers on recent cuts in price tick size, see the survey in Goldstein and Kavajecz (2000). Among others, important papers in this area are Ahn, Cao, and Choe (1996) and (1998), Bessembinder (2000), Jones and Lipson (2001), Porter and Weaver (1997), Ronen and Weaver (2001), and Van Ness, Vann Ness and Pruitt (2000). Field and Large (2007) finds that when a market's pricing grid is very coarse, its average inside depths can exceed the average trade size by a large multiple. My prediction holds when the tick size is large; and in this respect complements Seppi (1997), which predicts that depths are increasing in the tick size when the tick size is small.

Kadan (2006) also predicts a decline in liquidity supply, coupled with an increase in welfare, when the tick size falls: A decreasing tick size (in a competitive market) shifts rents from liquidity suppliers to a separate population of liquidity demanders, who consequently trade greater volumes, boosting welfare. By contrast, in the current model welfare per (unit) trade improves when the tick size falls, whether or not volumes rise. Empirically, the effect of cuts in tick size on volumes has varied depending on the case: the literature on this subject is well surveyed in Goldstein and Kavajecz (2000).

Cordella and Foucault (1999) and Foucault et al. (2005) describe a countervailing advantage of a large tick size, namely that it helps to make the market more resilient. Rosu (2006) also treats market resiliency, in cases of a small or absent price tick. Questions of market resiliency, which are outlined in empirical work in Biais, Hillion, and Spatt (1995), Coppejans, Domowitz, and Madhavan (2004) and Large (2007), are not raised by the current model, because it describes a highly resilient 'tight' market, where the spread equals its minimum tick size most of the time. Likewise, issues in Seppi (1997) regarding variation in trader size offer a complementary insight into the effects of changing the tick size.

Specifically, the model is a stationary 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 (2006), traders arrive at a limit order book sequentially and may alter its state. Prevailing prices, spreads and depths fluctuate stochastically with the sequence of market orders, limit bids and limit asks. Agents trade one unit, and then leave. They face a simple trade-off between 1) a market order, and 2) 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), Chakravarty and Holden (1995), Foucault (1999) and Handa and Schwartz (1996). A number of simplifications are made to the action space: hidden or iceberg orders are not available; and orders may not be canceled or submitted speculatively behind the best quotes.

Modeling a sequence of traders who share a stationary decision problem can yield clean analytic solutions about complex limit order book dynamics. This approach is developed in Foucault (1999), Foucault et al. (2005) and Rosu (2006) (Parlour 1998, by contrast, is robust to non-stationarity). For example, Foucault (1999) ensures stationarity by introducing a random end-date into the model. However, a stationary decision problem can be consistent with a non-stationary – even explosive – market state. This is the case in Foucault (1999), where the underlying price is an unbounded random walk. Goettler et al. (2004) and (2005) go a different way, exploiting the ergodicity and (near) stationarity of the market state around the asset's consensus value, in order to design market games for simulation, where unlimited information about optimal strategies can be gathered in time.[4]

This paper uses the stationarity of the market state to derive equilibrium best responses analytically, without needing to tackle the intricate market forecasting problems solved by traders in equilibrium. To do this, it introduces a subset of 'uninformed' traders who do not observe the market state, because they have not invested (at a cost of $x > 0$) in the necessary technology. Informed traders' behavior has complicated contingencies, which we cannot derive. However, the best response of uninformed traders can be identified: As they give ergodic weights to the various market states that they might encounter, they rationally adopt the average trader behavior. But average trader behavior is pinned-down by the market-clearing condition.

Viewed ex post, the liquidity market-clearing condition is a necessary accounting identity: that for every market order there is an executed limit order, and vice versa. Consequently, it exists in earlier dynamic models. But its effects differ, depending on the treatment of cancelation. In Foucault (1999) limit orders cancel automatically if they don't execute in one period; in Ohta (2006) they have two periods to execute. This guarantees market-clearing trivially. On the other hand, in Foucault et al. (2005) cancelation is not permitted. Therefore, ex ante, traders choose to submit equal volumes of market and limit orders, and our model's key driving mechanism is clearly present. That paper finds that, as here, wide spreads are a source of inefficiency,[5] but it focuses on cases where wide spreads are strategically posted. Goettler et al. (2004) and (2005), and Rosu (2006), are intermediate cases with some cancelation.

Other important theoretical antecedents to this work are in Anshuman and Kalay (1998),

[4]Because of its fixed initial value, the market state is not strictly stationary in Goettler et al. (2005).
[5]And they find another source: when impatient traders submit limit orders. See their Section 2.2.3.

Bernhardt and Hughson (1996) and Portniaguina, Bernhardt, and Hughson (2006), which study tick size, inside depths and welfare in markets with dealers. A good overview of this literature is available in Kadan (2006). In the prominent works of Kyle (1985) and Glosten and Milgrom (1985), market dynamics are studied when some agents obtain elsewhere superior information about the underlying asset's future cash flows. This introduces a set of important economic issues about the winner's curse, or picking-off risk in limit order execution, in the sense developed *inter alia* in Foucault (1999) and Handa and Schwartz (1996). These issues are somewhat orthogonal to this paper's interest in the matching properties of the market.[6] In the current setting, by contrast, some traders do have superior information, but only about future order flow, which they infer rationally from the current market state.

The paper proceeds as follows. Section 2 details the model and equilibrium, and previews the paper's main results. Section 3 proves the main Theorems about strategies and surplus (detailed derivations are left to an Appendix). Section 4 interprets these results. Section 5 proves that equilibrium exists. Section 6 investigates adaptations to the model, including the effect of introducing order cancelation. With empirical consequences in mind, Section 7 derives average inside depths in a simplified model and draws out some comparative statics. Section 8 concludes.

2 The Model

One agent (or player, or trader) arrives at each time $t \in \mathbb{N}$. The duration between successive times t and $(t+1)$ is the constant $\Delta > 0$.[7] The agent makes two actions, then leaves the game. First she chooses whether to acquire information about the market, and second she submits an order. The player has a type, β_t, which I introduce later. I first describe the market state, which she can learn if she wishes.

Let $k > 0$ be the price tick size. All admissible prices are multiples of k. Prices $\underline{B}$ and $\bar{A}$ are lower and upper inclusive bounds on prices respectively. They are multiples of

[6] Arguably, in this setting moderate asymmetric information about fundamentals *enhances* trading surplus. It deters limit orders because of the winner's curse (see Foucault 1999 and Handa and Schwartz 1996). But it leaves expected market order slippage unchanged. So stationarity requires that this deterrent to limit orders be offset: and the offsetting effects may enhance efficiency. This may help in thinking about why inside information is scarcely detrimental to welfare in the comparative simulations of similar models in Goettler, Parlour, and Rajan (2004).

[7] In an earlier version of this paper, the duration between successive times t and $(t+1)$ was an exponentially distributed random variable independent of $\mathcal{F}_t$. This assumption leaves all results of the paper unchanged, except for the existence result in Proposition 5.1. A problem arises in the proof of existence because the assumption implies an uncountable state space.

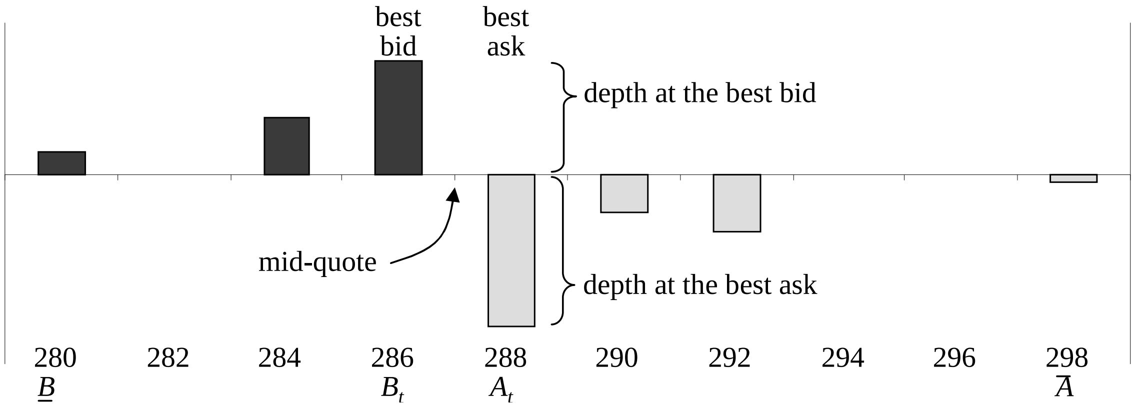

Figure 2: An example of a market state at time t is given. Columns indicate depths at various prices. The price tick size, k, is 2. The state variable, χ_t, is $\chi_t = (6, 0, 15, 30, -40, -10, -15, 0, 0, -2)$: these 10 values are the heights of the columns. Positive elements of χ_t indicate unfilled *bids* in the order book; while negative entries (with a light shade) are unfilled *asks*. The fixed lower price limit, $\underline{B}$, is 280; and $\bar{A}$ is 298. The best bid, B_t, is 286; and $A_t = 288$. So the bid-ask spread, $(A_t - B_t)$, is 2, which equals k. The depth at the best bid, $L_t^{B_t}$, is 30; and $L_t^{A_t} = 40$. Slippage is calculated relative to the mid-quote of 287.

k. Mathematically, they close the model in an important way. However, since they may be arbitrarily far apart, their economic importance is comparatively limited.

There is a stochastic process $\chi = \{\chi_t : t \in \mathbb{N}\}$, where χ_t is a state variable describing the shape of the limit order book. It is a vector of integers, each indicating the depth at some price between $\underline{B}$ and $\bar{A}$. The first element in χ_t is the depth at $\underline{B}$, the second is the depth at $(\underline{B} + k)$, and so on inductively until $\bar{A}$. Therefore χ_t has length $(1 + (\bar{A} - \underline{B})/k)$. If an element of χ_t is positive there are unfilled bids at that price. If it is negative, there are unfilled asks. An example to work through is shown in Figure 2.

Definition 1 *For all prices p such that $\underline{B} < p < \bar{A}$, define*

$$L_t^p = \left| \ \chi_t^{\left(1+\frac{p-\underline{B}}{k}\right)} \ \right|, \tag{1}$$

using χ_t^i to mean the i'th element of χ_t.

Say that L_t^p is the ***market depth*** *at price p at time t.*

Definition 2 *Let the* ***best ask****, A_t, be the lowest price with an unfilled ask, or $\bar{A}$ if there are no asks. Let the* ***best bid****, B_t, be the highest price with an unfilled bid, or $\underline{B}$ if there are no bids.*

Only states with $A_t > B_t$ arise. We may call $L_t^{A_t}$ the 'depth at the best ask', and $L_t^{B_t}$ the 'depth at the best bid': they are the *inside depths*. I will operationalize the concept of stationary microstructure, which was heavily leant upon in the Introduction, by finding a game where χ **is stationary**.

2.1 Competitive fringe of traders

As in Foucault *et al.* (2005) and Rosu (2006) there is a competitive fringe of traders standing ready to sell and buy when prices rise above, or fall below, the range $(\bar{A}, \underline{B})$. By transacting with them, agents can always sell immediately at at least $\underline{B}$, and buy immediately at at most $\bar{A}$, even if the limit order book is empty (when $\chi_t = \mathbf{0}$).[8] So fringe traders periodically hold a positive inventory in the asset. Even if this eventuality is extremely infrequent, the detailed analysis of the game keeps track of this inventory, and checks it reverts to zero in finite expected time.

To model this without more notation, I formally define $L_t^{\underline{B}} = \chi_t^1$ and $L_t^{\bar{A}} = -\chi_t^{(1+(\bar{A}-\underline{B})/k)}$. In principle this implies that depths at the two extreme prices $\underline{B}$ and $\bar{A}$ can go negative. 'Negative depth' will have the natural interpretation that the fringe traders hold positive inventory in the asset.

2.2 Acquiring order book information

The first of player t's two actions is taken before she goes to the market. She decides whether to invest in a costly technology permitting her to condition her behavior on up-to-date prices and depths, as summarized by χ_t. The investment has a cost of $x > 0$, representing the computing and physical effort of arranging to gather, organize, interpret and act on timely information.

In fact, player t can mix over this choice. So she chooses a probability π_t from the interval $[0, 1]$. With probability π_t she observes χ_t, but she pays a cost $x > 0$ to do so. Otherwise she does not observe χ_t and pays no cost.

Notice that the choice whether to invest in the technology is made ex ante, even before the trader learns anything about her trading objective and preferences. This is realistic whenever, as is usually the case on financial markets, traders cannot develop exchange-specific technology and skills only for the purpose of a single trade.

[8] A vector of zeros will be denoted by $\mathbf{0}$, in bold.

2.3 Choice of order

On arriving at the market, the trader knows χ_t if she paid x for order book information. She now performs the second of her two actions, which is to select one of four orders of unit quantity: a *bid, ask, buy* or *sell. Buy* and *sell* are market orders which execute immediately, while *bid* and *ask* are limit orders, which execute sequentially according to a price-then-time matching rule.

Assumption 1 Bids *and* asks *cannot be submitted behind the best quotes.*

So, *bids* cannot be submitted at prices below B_t, and *asks* cannot be priced above A_t. This helpfully reduces the number of order types to only four. Even though an *ask* or *bid* is initially submitted at the most aggressive price available, it may become stale, as before execution its price may come to differ substantially from either A_t or B_t as they change.

The assumption is most plausible for those assets where the bid-ask spread is normally equal to the price tick size, with high offered depths at B_t and A_t. Then, empirically, the most aggressive price possible (as limited by the large price tick size) is chosen by a wide range of trader types. An analysis of limit order submission outside the quotes is given in Cordella and Foucault (1999) in a model of a limit order book with bid-ask spreads routinely wider than the price tick.

2.4 Payoffs

In choosing her order, player t trades off slippage against delay. Slippage is the amount by which her price falls short of the mid-quote, m_t, which is the average of A_t and B_t. So her slippage is

$$\textit{direction} \times (p - m_t),$$

where *direction* is $+1$ if she buys, and -1 if she sells, and p is her trade price. Evidently, slippage can be negative. It also fluctuates with the bid-ask spread, $(A_t - B_t)$. However a benefit of the current setting is that, while stochastic, the bid-ask spread is predominantly constant and equal to the price tick size, k. So given Assumption 1, we have that predominantly

$$\textit{slippage} = \begin{cases} \frac{k}{2} & : \text{order is a } \textit{buy or sell} \\ -\frac{k}{2} & : \text{order is a } \textit{bid or ask} \end{cases} . \tag{2}$$

Just when the bid-ask spread is momentarily wider than k, (2) is incorrect. However, to simplify the problem, in the formulation of payoffs I adopt (2) as the definition of *slippage* at all times.

Player t's trade-off between slippage and delay depends on her type, β_t, which is drawn randomly from a symmetric distribution on the interval $[-\beta^+, \beta^+]$, independently of all other random variables known at time t. The CDF of this distribution is called F.

The agent's payoff from trading is

$$direction \times \beta_t - slippage. \tag{3}$$

Thus $\beta_t > 0$ gives player t a motive to buy, while $\beta_t < 0$ gives a motive to sell. The trade's payoff is penalized additively by the slippage. The benefit of trading with a limit order, rather than a market order, is to enjoy negative slippage. However, a player waits to trade if she uses a limit order, and the payoff in (3) is discounted at rate $\rho > 0$. The rate is the same for all traders; but a player with large $|\beta_t|$, who will gain more from a trade, finds waiting more costly. So, a large $|\beta_t|$ should indicate a preference for market orders.

In reality, limit orders typically wait at most a few minutes to execute, not long enough for impatience, as usually understood, to become important. So I do not include impatience here. Rather, the rate ρ captures opportunity and capital costs outside the model that the agent incurs in waiting to trade. One way of introducing ρ is to endow each player with an imminent, unobserved, idiosyncratic liquidity event or deadline.[9]

The payoff to player t is summarized in $\Pi_t(\beta_t)$:

Definition 3 *Let $\Pi_t(\beta_t)$ be the value of (3) when discounted at rate ρ until execution, minus x, the cost of the trading technology, if she invested in it.*

Observe that we may exclude a fifth action of *no order*, which would result in zero payoff. Since limit orders have negative slippage, for any player type at least one out of *bid* or *ask* has positive expected payoff[10].

2.5 Order submission: via a trading algorithm

To formalize order submission I introduce a trading algorithm. This establishes the notation for cutoffs, c_t, which will be central to the solution (and it simplifies the proof that equilib-

[9]On this interpretation the deadline arrives exogenously as the first event of a Poisson process with parameter ρ. Execution after the deadline results in a zero payoff. It is possible to formulate an adapted version of the model with the same qualitative properties, where, as in Foucault, Kadan, and Kandel (2005), rather than having a discount rate, traders undergo a constant waiting cost per unit time that their limit order remains in the market before trading.

[10]Although somewhat specific to the current payoff structure, this reflects a general property of limit order markets: that trading with a limit order 'wins the spread'.

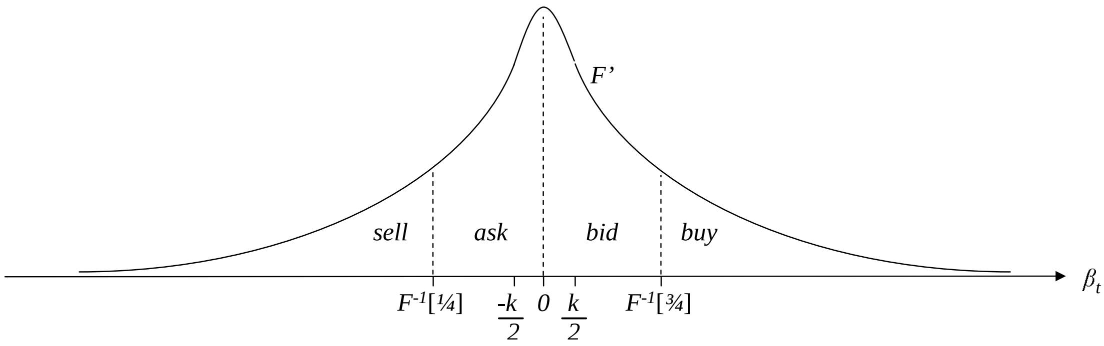

Figure 3: The distribution of the type, β_t, and (dotted lines) the three equilibrium cutoffs for traders who do not observe χ_t. The four delimited areas under F' are equal to one another and equal to $\frac{1}{4}$.

rium exists). Analytically it implies nothing new: Section 3.2 will show that changing the information set can yield an observationally identical game without a trading algorithm.

Instead of submitting the order herself, player t programs the algorithm up-front, *before she learns* β_t. Conditional on the signal χ_t iff she observes it, she selects a cutoff vector c_t.[11] Let $\hat{\mathbb{R}}$ be $\mathbb{R} \cup \{-\infty, \infty\}$, the closure of $\mathbb{R}$. Then

$$c_t \text{ is a vector of three weakly increasing elements in } \hat{\mathbb{R}}^3,$$

and each element of c_t is a cutoff determining the order. The algorithm is fast enough to act upon β_t. It compares c_t to β_t, and submits an order on that basis. Without yet knowing β_t, the trader can use c_t to determine which order will be made for each possible value of β_t.

Rule for the trading algorithm On behalf of the agent, the trading algorithm submits an order to the limit order book, denoted or_t. It automatically selects $or_t = or(\beta_t, c_t)$, where

- $or(\beta_t, c_t)$ is *sell* if β_t is less than all **three** elements of c_t.
- $or(\beta_t, c_t)$ is *ask* if β_t is less than exactly **two** elements of c_t.
- $or(\beta_t, c_t)$ is *bid* if β_t is less than exactly **one** elements of c_t.
- $or(\beta_t, c_t)$ is *buy* if β_t is less than **none** of the elements of c_t.

It will become apparent why or_t is *sell* when β_t is very low, while or_t is *buy* when β_t is very high; and why *ask* and *bid* are at intermediate realizations of β_t. The reasoning for this goes back to the early work of Parlour (1998).

[11] At some notational expense, I could have it that in advance she programs the algorithm with a suitable c_t for every state of the market that it might encounter.

2.6 Discussion previewing the results

To look ahead: I will be interested in an equilibrium where all agents choose c_t identically given their signal of χ_t, and where the dynamics of buying and selling are alike. I will identify the choice of c_t when player t receives *no* signal, and, with this, derive market efficiency. Theorem 3.3 will show that

$$\text{when player } t \text{ is uninformed of } \chi_t, \quad c_t = \begin{pmatrix} F^{-1}\left(\frac{1}{4}\right) \\ 0 \\ F^{-1}\left(\frac{3}{4}\right) \end{pmatrix}. \tag{4}$$

This is illustrated in Figure 3. Traders with β_t below $F^{-1}\left(\frac{1}{4}\right)$ submit a market sale, and trade immediately; traders with $\beta_t \in [F^{-1}\left(\frac{1}{4}\right), 0)$ submit a limit ask; and so on.

I briefly review the intuition for the Theorem's result, and what may be inferred from it. We start from the liquidity market-clearing condition discussed in the Introduction. For this game, the condition requires that over time equally many limit orders as market orders are submitted to the exchange. Therefore, *bids*, *asks*, *buys* and *sells* are, on average, equally likely to be submitted by any trader of unknown type. When selecting from these four options 'blindly', i.e. without knowledge of χ_t, a trader can do no better than to ascribe ergodic weights to each possible state χ_t that they might encounter. In consequence, the uninformed trader faces an average of the problems faced by better-informed traders. Her response to this average problem is equal to the average response: and therefore *also* involves submitted *bids*, *asks*, *buys* and *sells* with equal probability. This best response is given in Figure 3, and is such as to break the type space into four quartiles of equal probability.

It is thereby possible to identify the best response to an absence of limit order book information, while bypassing the out-of-equilibrium reasoning for that best response almost entirely. Such reasoning is intricate and involved, for it reviews each possible market state; and for that state, it considers the market's likely subsequent evolution. Tackling it head-on might be infeasible. Furthermore, the derivation makes scanty use of the details of the market's dynamics, and is therefore invariant to many features, including Δ, the rate of trader arrival, ρ, the discount rate, and the order queuing rules imposed by the exchange.

Figure 3 implies certain indifference conditions for the types $F^{-1}\left(\frac{1}{4}\right)$ and $F^{-1}\left(\frac{3}{4}\right)$, which are enough to identify the inefficiency of the market. As a result, market efficiency has the same invariances as those just mentioned of Figure 3. In fact, the only policy variable that it depends on is the price tick size, k. The price tick fixes the bid-ask spread, and thereby holds market inefficiency at a liquidity market-clearing level.

Intuitions for these results were presented in the Introduction; and Section 3 provides rigorous proofs, extensions and conditions.

2.7 Shape of the CDF, F

As Section 6.1 will argue, the above analysis is approximately correct for reasonable shapes of F. However, it is only precisely right for a certain family of CDFs, F, of β_t. Specifically, suppose that for all $\beta \in S = (-\beta^+, -\frac{k}{2}]$,

$$F(\beta) = \frac{\eta}{\frac{k}{2} - \beta}. \tag{5}$$

The constant η is chosen to ensure that $F(\infty) = 1$. To satisfy (5) there are suitable atoms at $-\beta^+$ and β^+. This truncation of F simply ensures that moments exist. It should not be too extreme: suppose that $\beta^+ > \frac{k}{1-e^{-\rho\Delta}}$. However, it should also allow some dispersion: suppose $F\left(-\frac{k}{2}\right) > \frac{1}{3}$. A typical form of F is depicted in Figure 3.

2.8 State transitions

When an order is submitted to the limit order book, this changes the order book shape. A limit order adds to the already-assembled offered depth; while a market order, by trading with an extant limit order, diminishes the depth. So there is a sequence of transitions in the state variable, χ_t, induced by order submission. Indeed, the state χ_{t+1} is measurable with respect to χ_t and or_t. Specifically, χ_{t+1} is equal to χ_t, except that

- if $or_t = sell$, then its depth at B_t is 1 less (so $L_{t+1}^{B_t} = L_t^{B_t} - 1$);
- if $or_t = buy$ then its depth at A_t is 1 less;
- if $or_t = bid$ then depth at p is 1 more: where $p = (A_t - k)$ unless $L_t^{B_t} < 0$, when $p = B_t$;
- if $or_t = ask$ then depth at p is 1 more: where $p = (B_t + k)$ unless $L_t^{A_t} < 0$, when $p = A_t$;

The latter two bullet points implement Assumption 1, that *bids* and *asks* are not submitted outside the quotes. They also include special rules for when the competitive fringe of traders have positive inventory and prices are at upper or lower limits, leading to $L_t^{A_t} < 0$ or $L_t^{B_t} < 0$. On those occasions, the *bid* or *ask* executes immediately.

2.9 Equilibrium

Let the initial condition, χ_0, be a random variable with a distribution Σ.

Definition 4 *Define player t's overall action to be a_t, so $a_t = (\pi_t, c_t)$.*

Let's say that the (Bernoulli) random variable $Z_t \in \{0, 1\}$ is 1 if player t observes the market state χ_t, and zero otherwise. Adopt the convention that if $Z_t = 0$, then $Z_t\chi_t$ is the scalar, 0.

Definition 5 *Define player t's signal to be* $Z_t\chi_t$.

So if $Z_t\chi_t = 0$ player t may infer that $Z_t = 0$, and that she has no information about the current limit order book, as summarized by χ_t. But if $Z_t = 1$ then player t's signal $Z_t\chi_t$ is just χ_t.

Definition 6 *A strategy is a function which maps signals,* $Z_t\chi_t$*, into actions,* a_t*. It is a pair* $(\pi, c(\cdot))$*, where* $\pi \in [0, 1]$ *and c takes as its argument signals of the form* $Z_t\chi_t$*.*

In a stationary Markov equilibrium all traders have the same strategy. Equilibrium is a probability space $(\Omega, \mathcal{F}, P)$, together with a strategy and the primitives of the model. Although formally there is no mixing over strategies, the action π_t parameterizes the distribution of the random variable Z_t. In this way we may interpret agent t as randomizing her choice whether to learn χ_t.

Definition 7 *Let* $E^t_{\pi^\dagger, c^\dagger}$ *indicate the expectations operator when for all* $s \neq t$, $a_s = (\pi, c(Z_s\chi_s))$*, the equilibrium action; but player t may be different: for* $a_t = (\pi^\dagger, c^\dagger)$*.*

Equilibrium properties:

1. Players choose actions $a_t = (\pi, c(Z_t\chi_t))$. (So $\pi_t = \pi$ and $c_t = c(Z_t\chi_t)$.)

2. When $\pi_t = \pi$ and $c_t = c(Z_t\chi_t)$, player t acts in the way which maximizes the conditional expectation of $\Pi_t(\beta_t)$, given other players' strategies. Precisely,

$$E^t_{\pi^\dagger, c(Z_t\chi_t)}[\Pi_t(\beta_t)] \tag{6}$$

   exists and is maximized when $\pi^\dagger = \pi$. And for all signals z,

$$E^t_{\pi, c^\dagger}[\Pi_t(\beta_t) | Z_t\chi_t = z] \tag{7}$$

   exists and is maximized when $c^\dagger = c(z)$. $\square$

As mentioned in the Introduction, I will be exclusively concerned with equilibria where the state variable process, χ, is stationary. Although this stationarity property makes equilibrium existence harder to prove, the benefit is substantial: If the state variable were not stationary, then the decision problem of 'blind' traders would not be stationary, and their best response would not be constant (it would depend on t, the time since the game began). As was previewed in Section 2.6, if their decision problem is constant then it is tractable.

I also concentrate on *buy-sell symmetric* equilibrium, where buyers and sellers behave alike, but symmetrically to one another. Buy-sell symmetry is precisely defined in an Appendix. It is a natural constraint in this symmetric environment where buyers and sellers face congruent payoffs and symmetric model primitives. However, it is simply a desirable equilibrium selection, although it may be that all equilibria in this game are buy-sell symmetric.

3 Equilibrium characterization

The characterization of equilibrium will proceed in four parts. After a Section introducing some notation, Section 3.2 establishes that, when ignorant of the limit order book (i.e. when $Z_t = 0$), traders face a particular *average* problem. In parallel, Section 3.3 deduces from stationarity the average best response.

After these preliminary steps, Section 3.4 provides the main theorems of the paper. The reader can move directly to those, if desired. It shows that the best response, namely $c(0)$, to the average problem just mentioned, is this average best response, which depends in fact on the quartiles of F – as depicted in Figure 3. This identifies $c(0)$ and indirectly thereby surplus per agent.

The results of this Section are interpreted in Section 4, while the fairly involved proof of equilibrium existence is left to Section 5.

3.1 Notation and definitions

Definition 8 *If player t submits a* bid *or an* ask*, then let T_t^B and T_t^A respectively be their times to execution.*

Then trader t's payoff, $\Pi_t(\beta_t)$ is given by

$$\Pi_t(\beta) = -xZ_t + \begin{cases} \beta - \frac{k}{2} & : or_t = buy \\ -\beta - \frac{k}{2} & : or_t = sell \\ \left(\beta + \frac{k}{2}\right)\ e^{-\rho T_t^B} & : or_t = bid \\ \left(-\beta + \frac{k}{2}\right)\ e^{-\rho T_t^A} & : or_t = ask \end{cases}. \tag{8}$$

An expectation without sub- or superscripts, E, will mean equilibrium expectations. For out-of-equilibrium expectations, it will clarify some expressions to suppress the symbol π in $E^t_{\pi,c^\dagger}$, by writing $E^t_{c^\dagger}$ for $E^t_{\pi,c^\dagger}$.

Definition 9 *Let c_{ask} be the cutoff function such that the player is sure to submit an* ask*, so*

$$c_{ask} = (-\infty, \infty, \infty)', \tag{9}$$

and define c_{sell}, c_{bid}, and c_{buy} similarly.

I will often use the out-of-equilibrium expectations operator $E^t_{c_{ask}}$ which describes what would happen if player t were forced to submit an *ask*.

Definition 10 *Let the* expected discount factor associated with limit order execution *be*

$$E^t_{c_{ask}}[e^{-\rho T^A_t}], \tag{10}$$

which, in a buy-sell symmetric equilibrium, is equal to $E^t_{c_{bid}}[e^{-\rho T^B_t}]$.

An important step in the characterization of welfare in this game, will be to identify the expected discount factor associated with limit order execution.

3.2 Out-of-equilibrium reasoning when $Z_t = 0$

In equilibrium, we can consider the case of a player, t, who could observe β_t and Z_t, and on this basis select the order, or_t (*bid*, *ask*, etc.), conditioning her selection on her signal $Z_t\chi_t$. So, she is able to act directly on the exchange, rather than via a trading algorithm. She is not thereby constrained by buy-sell symmetry (which is defined precisely in Definition 15 of the Appendix). Note that the stochastic time to execution were she to submit an *ask* is independent of her type, β_t. So, if β_t is very low, she prefers to act urgently with a *sell*. Over a suitable region her payoff is monotonic in β_t so at some cutoff she switches and prefers an *ask*. A similar argument applies to buyers, implying another cutoff. So the player would choose a cutoff strategy, implementable through a choice of c_t. Therefore, the structure in the trading algorithm's use of the cutoffs in c_t described in Section 2.5, is in fact not a constraint to optimal order submission.

Denote the first element of a vector, v, by v^1. The analysis will be concerned mainly with c^1_t and $c^1(Z_t\chi_t)$. A player t who knows that $\beta_t = c^1(Z_t\chi_t)$ is at a cutoff where she is indifferent between submitting an *ask* and a *sell*, on observing $Z_t\chi_t$. This can be formalized as follows: from (8),

$$\left(-c^1(Z_t\chi_t) - \frac{k}{2}\right) = \left(-c^1(Z_t\chi_t) + \frac{k}{2}\right) E^t_{c_{ask}}[e^{-\rho T^A_t}|Z_t\chi_t]. \tag{11}$$

Proposition 3.1 *In a buy-sell symmetric equilibrium where* $\pi \in (0, 1)$

$$F(c^1(0)) = E[F(c^1(Z_t\chi_t))]. \tag{12}$$

Proof. See Appendix. ■

The proof of Proposition 3.1 follows the reasoning about out-of-equilibrium paths, of a marginal cut-off type who is indifferent between submitting an *ask*, and submitting a *sell*; and who does not observe the state χ_t. It shows that she faces the average of the problems faced by players who observe χ_t, and therefore uses an average cutoff. Because of the form of the indifference condition (11), this average is hyperbolic rather than, more familiarly, arithmetic; and it can be represented in terms of the CDF, F. The proof makes no reference to the details of the dynamic equilibrium, in particular to the distribution of T_t^A.

The next part shows how stationarity in fact identifies $E[F(c^1(Z_t\chi_t))]$: for, were it higher or lower, depths would be non-stationary.

3.3 The consequences of stationarity

The following Lemma states that in equilibrium *sells, buys, bids* and *asks* are equally prevalent to one another in the order flow. This formalizes one of the paper's main intuitions: namely, that if depths are stationary then the average number of executed limit orders per unit time equals the average number of market orders. As discussed in the Introduction, this can be interpreted as a 'liquidity market-clearing' condition.

Definition 11 *Let* $D_t = \sum_{p=M}^{Q} L_t^{pk}$*. So,* D_t *is the total market depth at time* t*.*

Lemma 3.2 *In a buy-sell symmetric equilibrium where* $\pi \in (0, 1)$*, for any* t

$$\Pr[or_t = sell] = \Pr[or_t = ask] = \Pr[or_t = bid] = \Pr[or_t = buy] = \frac{1}{4}. \tag{13}$$

Proof. Since $\{\chi_t\}$ is stationary with a first moment, $\{D_t\}$ is also. D_{t+1} always differs from D_t by exactly 1. So, D_t rises as often as it falls. As it falls exactly when $or_t \in \{sell, buy\}$, for all t

$$\Pr[or_t \in \{sell, buy\}] = \frac{1}{2}. \tag{14}$$

In a buy-sell symmetric equilibrium, χ and $-rev(\chi)$ have the same distribution. So, *buy* and *sell* are equally prevalent orders. And, *bid* and *ask* are equally prevalent orders. ■

Since *sells*, or market sales, are submitted by those players who draw β_t below $c^1(Z_t\chi_t)$,

$$\Pr[or_t = sell | Z_t\chi_t] = F(c^1(Z_t\chi_t)). \tag{15}$$

Given Lemma 3.2 this implies (by the Law of Iterated Expectations) that

$$E[F(c^1(Z_t\chi_t))] = \frac{1}{4}. \tag{16}$$

This equation is central. Think of $F^{-1}\left(\frac{1}{4}\right)$ as the median seller. Then (16) shows that the median seller is $F^{-1}E[F(c^1(Z_t\chi_t))]$. So, because of stationarity, in particular of the stationarity of offered depths, the median seller is (just like $c^1(0)$ although for quite different reasons), a deformed average of the set of marginal types, $\{c^1(Z_t\chi_t)\}$ (it might have been a simple arithmetic average if F had had a uniform CDF). In the light of this, it is perhaps a natural thought that the median seller would be roughly indifferent between a *sell* and an *ask* in the average of their problems, as faced by traders with $Z_t = 0$. In fact, features of F imply that this deformation is the same as the one in Proposition 3.1, so that this natural thought is exactly true. This is formalized in Theorem 3.3.

3.4 The main theorems

Theorem 3.3 *In a buy-sell symmetric equilibrium where* $\pi \in (0,1)$*,*

$$c(0) = \begin{pmatrix} F^{-1}\left(\frac{1}{4}\right) \\ 0 \\ F^{-1}\left(\frac{3}{4}\right) \end{pmatrix}. \tag{17}$$

Proof. From Proposition 3.1 and (16), $c^1(0) = F^{-1}\left(\frac{1}{4}\right)$. The theorem then follows from the buy-sell symmetry of $c(0)$. ■

The derived best response is illustrated in Figure 3.

Theorem 3.3 shows that the types who have $\beta_t = \pm F^{-1}\left[\frac{1}{4}\right]$, are, when $Z_t = 0$, indifferent between 1) a limit order, and 2) a market order which incurs k in extra slippage. This actually identifies the *expected discount factor associated with limit order execution*, namely $E^t_{c_{ask}}[e^{-\rho T^A_t}]$, for, from (11),

$$\left(-F^{-1}\left(\frac{1}{4}\right) - \frac{k}{2}\right) = \left(-F^{-1}\left(\frac{1}{4}\right) + \frac{k}{2}\right) \; E^t_{c_{ask}}[e^{-\rho T^A_t}], \tag{18}$$

so that

$$E^t_{c_{ask}}[e^{-\rho T^A_t}] = 1 - \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}. \tag{19}$$

Furthermore, it leads to a characterization of each agent's expected payoff:

Definition 12 *Let* δ *be defined by*

$$\delta = \frac{E[|\beta_t| \;:\; |\beta_t| < F^{-1}\left[\frac{3}{4}\right]]}{F^{-1}\left[\frac{3}{4}\right]}, \tag{20}$$

so that δ *is a measure of* F*'s relative dispersion within its interquartile range,* $\left[F^{-1}\left[\frac{1}{4}\right], F^{-1}\left[\frac{3}{4}\right]\right]$.

Theorem 3.4 *In a buy-sell symmetric equilibrium where* $\pi \in (0,1)$ *the ex ante expected payoff, or surplus, of player* t *is given by*

$$E[\Pi(\beta_t)] = E(|\beta_t|) - \frac{k}{2}\left(\frac{\frac{k}{2} + \delta F^{-1}\left[\frac{3}{4}\right]}{\frac{k}{2} + F^{-1}\left[\frac{3}{4}\right]}\right). \tag{21}$$

Proof. See Appendix for the full proof. From Theorem 3.3, any player t with $Z_t = 0$ foresees ex ante, that $or_t \in \{bid, ask\}$ iff β_t is in the interquartile range of F, and so

$$\Pr\{or_t \in \{bid, ask\}|Z_t = 0\} = \frac{1}{2}. \tag{22}$$

Furthermore, the cases where β_t is in the interquartile range of F are exactly the cases where her payoff to trading is discounted due to delay. But, the expected discount factor for these cases is known from (19). Combining these observations, her ex ante surplus follows after algebra.

Were it not for the cost x, this would be exceeded by the surplus of those for whom $Z_t = 1$; but, since players mix between $Z_t = 0$ and $Z_t = 1$, in fact x absorbs exactly this excess. Ex ante surplus is therefore equal to expected surplus conditional on $Z_t = 0$. ■

4 Interpretation of the result and comparative statics

Theorem 3.4 shows that surplus per agent in this market is strictly less than $E[|\beta_t|]$. This bound, $E[|\beta_t|]$, has the following interpretation. Consider an ideal market where no trader waits to trade, and each trade has the property that the buyer, say player t, had $\beta_t \geq 0$ while the seller, say s, had $\beta_s \leq 0$. Then the trading surplus achieved with each trade would be

$$E[(\beta_t - \beta_s)|\beta_t \geq 0; \beta_s \leq 0], \tag{23}$$

or $2E[|\beta_t|]$. Per trader, the surplus would be $E[|\beta_t|]$.

Both of these ideal conditions are violated by the electronic limit order book in practice. Theorem 3.4 quantifies the resulting shortfall in surplus, or inefficiency per trader; and shows the shortfall to be a necessary condition for stationarity. It is

$$\frac{k}{2}\left(\frac{\frac{k}{2} + \delta F^{-1}\left[\frac{3}{4}\right]}{\frac{k}{2} + F^{-1}\left[\frac{3}{4}\right]}\right). \tag{24}$$

Given F and k, (so that the bid-ask spread is held fixed) this shortfall or inefficiency is invariant to changes in the other primitives of the model which may freely alter both

- (*Order Flow*) the duration between trader arrivals, Δ, and
- (*Quality of Trader Information*) the endogenous proportion of traders, π, who observe χ_t.

In fact, inefficiency per trader is a little under half the price tick size: which is natural, for half of all traders opt to wait for a counterparty by submitting a limit order, and these traders expect the wait to be less costly than paying up the price tick to cross the bid-ask spread, and trading immediately.

4.1 Benefits to cutting the price tick size

Theorem 3.4 shows that the inefficiency per trader 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 (5) it is not strictly possible to hold F fixed in this thought experiment where k changes).

In (24), the half-spread bounds the inefficiency per trader above. An explanation for this is that exactly half of those traders drawing $Z_t = 0$ wait inefficiently before trading; and when a trader does wait, she finds paying the spread ex ante weakly costlier than waiting.

Cutting the tick size affects surplus per player because it makes market orders strictly more attractive. All other things remaining equal, some types would be induced to submit market orders, who previously preferred limit orders. This would violate Theorem 3.3 and imply a shortage of limit orders over time, non-stationary depths and no equilibrium. In equilibrium, therefore, endogenous frictions adjust to restore the types in $c(0)$ to their respective indifference conditions as illustrated in Figure 3 (for example, average depths fall). Hence, perhaps counter-intuitively, like market orders limit orders also become ex ante more attractive when the bid-ask spread falls – which necessarily involves an increase in ex ante more attractive when the bid-ask spread falls – which necessarily involves an increase in ex ante welfare per

The Introduction compared this mechanism to a similar one in Kadan (2006). A countervailing consideration, regarding resiliency, is developed in Cordella and Foucault (1999) and Foucault et al. (2005).

5 Existence of equilibrium

Goettler et al. (2005) points out that a result of Rieder (1979) can be used to prove the existence of equilibrium in this context. To apply his result directly here, however, initially we must make some restrictions:

1. Suppose that the initial condition is fixed at, say, $\chi_1 = (1, 0, 0, ..., 0, -1)$.

2. Suppose that $\pi \in (0, 1)$ is fixed and exogenous.

3. Suppose that traders are restricted to buy-sell symmetric choices of c_t.

4. Suppose that $c(0)$ is fixed at its value in Theorem 3.3.

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 (as c_t inhabits a compact space). His result applies to this restricted game (although β_t, whose space is not countable, requires technical attention).

While this shows the existence of a stationary sequential equilibrium, it does not guarantee that χ_t is a stationary random variable with a first moment. However, as $\pi > 0$, all orders are with positive probability observed by some traders. This creates negative feedback in depths, because it is unattractive to traders to add their limit order to a deep book: many therefore submit a market order, which clears depths away.

More formally, an Appendix lays out the proof that there exists some large number L^* such that when inside depths deviate from zero by more than L^*, they drift towards zero; and that this drift can be bounded uniformly away from zero. Thus, a drift condition, as expressed in Tweedie (1976) and (1983), is satisfied. In a generalized sequential game Large and Norman (2008) use drift condition results in Tweedie (1983) to show that this implies that the state variable χ_t is ergodic and has a first ergodic moment.

In this equilibrium, the above four restrictions can be relaxed sequentially:

1. As equilibrium is stationary, the initial condition, χ_1, can be altered within its recurrent set without changing equilibrium strategies. Large and Norman (2008) show that it can therefore be set to a stochastic value drawn from the ergodic distribution of χ_t; and that the state process (χ_t) is then stationary.

2. There exists a value of x which exactly compensates 'blind' traders for being uninformed of χ_t. Setting x to this value implies that traders are willing to mix over Z_t, so that π is at a (weakly) optimal value.

3. Trader t's best response to a buy-sell symmetric market is to act in a buy-sell symmetric way herself, so it is no restriction to impose that c_t is buy-sell symmetric.

4. Theorem 3.3 applies here, so fixing the value of $c(0)$ above was not restrictive.

Therefore:

Proposition 5.1 *For all $\pi \in (0,1)$, there exists an x, such that a buy-sell symmetric equilibrium exists where $\pi_t = \pi$ for all t.*

6 Comments on the model and alternative formulations

This Section provides some comments about adaptations to the model which weaken or change leading assumptions. It addresses the effect of cancelation on the model, the consequences of a change in the queuing rules enforced by the exchange, and the robustness of the model to the functional form of F.

6.1 The functional form of F

Proposition 3.1 depends importantly on the specification of the functional form of F in Section 2.7. On this Proposition hangs the main Theorems 3.3 and 3.4. It is therefore important to understand the sensitivity of the model to this tractability assumption. This Subsection address the issue in two ways. First, it consider the case where the CDF of β_t is a uniform distribution, F_u, and consider the modification this brings to the results. Second, it shows that given any symmetric distribution for β_t with support $[-\beta^+, \beta^+]$, Theorems 3.3 and 3.4 are true in the limit of a sequence of equilibria such that $\pi \to 0$.

6.1.1 A uniform distribution for β_t

Consider a buy-sell symmetric equilibrium where $\pi \in (0,1)$, and where β_t are i.i.d. random variables distributed uniformly between $-\beta^+$ and β^+. Call the CDF of β_t, F_u. Let F, not now a primitive of the model, be a function satisfying the conditions in Section 2.7. Then, tracing through its proof, it can be seen that Proposition 3.1 is still true: $F(c^1(0)) = E[F(c^1(Z_t\chi_t))]$. Equally, Lemma 3.2 is also still true: the four order types are equally prevalent in the order flow. However, it follows from Lemma 3.2 now that, in contrast to (16), $E[F_u(c^1(Z_t\chi_t))] = \frac{1}{4}$. Rearranging,

$$c^1(0) - F_u^{-1}\left(\frac{1}{4}\right) = F^{-1}\left(E[F(c^1(Z_t\chi_t))]\right) - F_u^{-1}\left(E[F_u(c^1(Z_t\chi_t))]\right), \tag{25}$$

So $c^1(0)$ differs from the median seller, $F_u^{-1}\left(\frac{1}{4}\right)$, by the quantity on the right-hand-side of (25). This has the following interpretation: $F_u^{-1}\left(E[F_u(c^1(Z_t\chi_t))]\right)$ is the arithmetic mean of the finite elements of $\{c^1(Z_t\chi_t)\}$. On the other hand, $F^{-1}(E[F(c^1(Z_t\chi_t))])$ is a distorted, hyperbolic, mean of them. Therefore, insofar as the hyperbolic mean approximates the arithmetic mean, Theorem 3.3 is approximately true.

6.1.2 A limit as $\pi \to 0$

Suppose now that β_t are i.i.d. random variables, distributed symmetrically about zero with support $[-\beta^+, \beta^+]$. Call the CDF of β_t simply F. Consider a sequence of buy-sell symmetric equilibria with $\pi \in (0,1)$ of a sequence of games where in the limit $\pi \to 0$, so that as the sequence progresses fewer and fewer traders observe χ_t. This is somewhat related to the idea of an exchange 'going dark', a phenomenon which has occurred, see Hendershott and Jones (2005).

Note that in the limit,

$$F(c^1(0)) - \Pr[or_t = sell] \to 0, \tag{26}$$

for in a market where almost all traders observe $Z_t\chi_t = 0$, the events $[or_t = sell]$ and $[\beta_t \leq c^1(0)]$ are almost the same. But Lemma 3.2 is true of every equilibrium in the sequence. Hence

$$c^1(0) \to F^{-1}\left(\frac{1}{4}\right). \tag{27}$$

Therefore, in the limit Theorem 3.3 is true. So, in the limit $F^{-1}\left(\frac{1}{4}\right)$ is indifferent between $[or_t = ask]$ and $[or_t = sell]$. Therefore, in the limit (19) is true. Finally, the proof of the paper's main welfare result, Theorem 3.4 is true as stated, for the limit as $\pi \to 0$.

6.2 Other queuing rules for limit orders

In the model, as almost universally in reality, limit orders execute in sequence according to price priority. When, however, they are equally priced, price priority must be supplemented by a further criterion: in the model they execute in the order in which they were submitted.

Field and Large (2007) exhibits an alternative to this time priority (or FIFO) rule: Some markets operate on a *pro rata* system, whereby all limit orders at the same price trade simultaneously against each countervailing market order. The market order is normally too small to trade with them all in their entirety. So, the trade is shared among the limit orders in proportion to their size (The exchange uses rounding rules and a coin-flip to distribute 'odd lots' arising due to quantity discreteness).

This can be implemented within the current model as follows: Suppose that an independent, fair coin-flip is used to allocate each *sell* (*buy*) to the extant bids at B_t (asks at A_t).

How would this change the results of the model?

Since T_t^A and T_t^B are not now measurable with respect to the history of orders, the proof of equilibrium existence does not follow as stated in Section 5. However, existence aside, the characterizations of equilibrium in Theorems 3.3 and 3.4 are true. Tracing through the proofs of these Theorems, it can be seen that no distributional properties of T_t^A and T_t^B are invoked. So the proofs remain valid under the altered distribution.

We may conclude, subject to a concern about equilibrium existence, that trader surplus is unchanged by the altered queuing rules. An intuition for this was given in the Introduction.

6.3 Cancelation

This Section discusses an adaptation to the model where periodically limit orders are canceled involuntarily. This modeling feature is present in Goettler *et al.* (2005) and Hollifield *et al.* (2004), but is replaced with an endogenous cancelation decision in the simulations in Goettler *et al.* (2004). Deliberate, endogenous, cancelation of limit orders, while often the more realistic case, is harder to analyze in this framework, and would be an interesting exercise for future work.

Adapt the model so that there exists $\gamma \in [0, 1)$, such that at each time t, each extant bid and ask is independently removed from the limit order book with probability γ, without payoff to its submitter.

Although such involuntary cancelation may seem to be detrimental to traders, in fact it is advantageous for trader surplus. When $\gamma = 0$ we revert to the original game. Consider a buy-sell symmetric equilibrium where $\gamma > 0$. As some limit orders are now removed before they trade, stationarity implies that initially limit orders are voluntarily submitted *more* frequently than market orders, and therefore more frequently than in equilibrium when $\gamma = 0$. Since setting $\gamma > 0$ does not change the payoff to a market order, if more limit orders are to be submitted, they must become more attractive to traders in this equilibrium when $\gamma > 0$, than when $\gamma = 0$. This implies that there is a higher surplus per agent.

So, the dominant effect of involuntary cancelation is the positive externality for *other* limit orders when a limit order is removed, causing them to trade sooner.[12]

[12] A full proof of this (which would unfortunately require an involved reformulation of the model) proceeds via the intermediate result that if $\gamma > 0$ then $c^1(0) < F^{-1}\left(\frac{1}{4}\right)$.

7 Average depths

As mentioned in the Introduction, surplus per agent may be thought of as the gains from trade that are not lost 1) in the delay as limit orders await their counterparties; or 2) in matching failures when the asset is bought (sold) by a trader with an excessively low (high) willingness to pay. The first of these, waiting, is closely correlated with market depths. Therefore the identification of market inefficiency in Theorem 3.4 may also be informative about average depths.

This Section studies the stationary distribution of 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 depth: and shows that (to a first order approximation) this does indeed adjust in ways consistent with the invariances implied by Theorems 3.3 and 3.4. The Section is in this way complementary to the study in Parlour (1998) of depths' transitions from one state to the next. The special case of the model that will be analyzed has these features:

- The market is a 'two-tick' market, where $\bar{A} = \underline{B}+k$, so that there are only two admissible prices for trade: $\underline{B}$ and $\bar{A}$.
- In equilibrium, the market is fairly efficient, so that $e^{-\rho T_t^A} \approx 1$.
- In equilibrium $\pi \approx 0$, so that inefficiencies arising due to mismatched counterparties scarcely arise.[13]

If the market is not two-tick, then limit orders risk becoming 'stale' if A_t and B_t move far away before they execute. Their submitters may be compensated for this 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 to order books of low depths. A characterization of average inside depth under these generalizations is beyond the current scope.

Definition 13 *Define* $L_t^{A_t+}$ *for* $\max\{L_t^{A_t}+1, 0\}$. *So* $L_t^{A_t+}$ *is a measure of depth (where the inventory of the competitive fringe of traders outside the model is excluded). Define* $L_t^{B_t+}$ *likewise.*

[13]Theorem 3.3 shows that when $Z_t = 0$, player t sells iff $\beta_t < 0$. Therefore, when $\pi \approx 0$ the inefficiencies arising due to badly-matched counterparties are almost zero. More precisely, as in Section 6 one may think of a sequence of equilibria such that $\pi \to 0$.

Recall (19), which gave the equilibrium expected discount factor associated with a limit order:

$$1 - E^t_{c_{ask}}[e^{-\rho T^A_t}] = \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}.$$

Via a Taylor expansion, we have that

$$1 - e^{-\rho T^A_t} = \rho T^A_t + O_p\left((\rho T^A_t)^2\right). \tag{28}$$

As $e^{-\rho T^A_t} \approx 1$, $(\rho T^A_t)^2$ is small, and

$$1 - E^t_{c_{ask}}[e^{-\rho T^A_t}] \approx \rho E^t_{c_{ask}}[T^A_t]. \tag{29}$$

$E^t_{c_{ask}}[T^A_t]$ is the expected time until $L^{A_t+}_t$ future *buys* have been submitted. The duration between the consecutive arrivals of traders is Δ. Consider a sequence of equilibria where $\pi \to 0$. In the limit, the probability of each future trader submitting a *buy* is $\frac{1}{4}$. Hence in the limit

$$E^t_{c_{ask}}[T^A_t] = 4\Delta E[L^{A_t+}_t]. \tag{30}$$

Combining (19), (28) and (30), when $\pi \approx 0$

$$E[L^{A_t+}_t] \approx \frac{1}{4\Delta\rho}\left(\frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}\right). \tag{31}$$

Although this analysis has considered the waiting time of an *ask* at A_t, an equivalent analysis and the same approximation hold when A_t is replaced with B_t.

Comparative Statics On the basis of (31) some comparative statics can be derived for the equilibrium average (positive) depths at the best ask, as captured by $E[L^{A_t+}_t]$, which might also be called the average inside depths:

- Average inside depth is increasing in the tick size, k. This is because, as discussed in the Introduction, a greater tick size would cause an 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), and related predictions are made in Kadan (2006) and Seppi (1997).
- Average inside depth is increasing in the order arrival rate, $\frac{1}{\Delta}$. An increased order arrival rate, would clear a given depth faster. All other things being equal, this would attract too many limit orders to sustain stationarity. To offset this, equilibrium depths rise. This corresponds to a finding in Brockman and Chung (1996) for the Stock Exchange of Hong Kong.[14]

[14]See also Danielsson and Payne (2002) which, at a 20-second frequency, gives nuanced evidence on depths and order flow dynamics for Reuters FX markets. It does not fully isolate inside depths.

- But average inside depth is decreasing in trader impatience, ρ, since impatient traders avoid limit orders, violating the stationarity of χ_t, unless induced not to by low depths.

The latter two predictions are consistent with a finding in Foucault (1999), Foucault et al. (2005) and Rosu (2006): namely, that spreads narrow when impatience, ρ, falls, or when the time between traders, Δ, falls. If narrower spreads and higher inside depths are both observable effects of aggressive limit order submission, then their sensitivities to Δ and ρ should be alike.

8 Conclusion

The market-clearing condition is a central means of analyzing prices without modeling market microstructure details directly. In focusing on these details, this paper rediscovers market-clearing in a different context: in a minimal stationarity property of microstructure dynamics, meaning that a limit order book induces in expectation equal numbers of market orders and uncanceled limit orders over time. To understand the consequences of this, a limit order market is modeled where the bid-ask spread is almost always constant. Its stationarity permits us to derive welfare, and the best response of particular traders, while bypassing altogether their difficult problem of forecasting the dynamic order book. Applying this approach to stochastic sequential games with binary action, such as queuing, may offer an intriguing opening for future research.

Equilibrium effects lead to striking invariances in trader surplus: with respect to changes in the sophistication of trading (π); in the rate of order flow (Δ); and in limit order queuing rules. In every such case, the market's endogenous inefficiency does not vary, for its anticipated sensitivity to the change is perfectly offset by endogenous equilibrium adjustments, for example in average depths.

But trader surplus is decreasing in the tick size. Cutting the tick size lets the bid-ask spread decline, improving the prospects of a market order, and thereby overcoming traders' incentives to add their limit orders to an already-congested limit order book. Trader surplus is thereby enhanced.

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

References

Ahn, H., C. Cao, and H. Choe (1996). Tick size, spread and volume. *Journal of Financial Intermediation 5*, 2–22.

Ahn, H., C. Cao, and H. Choe (1998). Decimalization and competition among stock markets: Evidence from the Toronto Stock Exchange cross-listed securities. *Journal of Financial Markets 1*, 51–87.

Anshuman, V. and A. Kalay (1998). Market making with discrete prices. *Review of Financial Studies 11*(1), 81–109.

Bernhardt, D. and E. Hughson (1996). Discrete pricing and the design of dealership markets. *Journal of Economic Theory 71*, 148–182.

Bessembinder (2000). Tick size, spreads, and liquidity: An analysis of Nasdaq securities trading near ten dollars. *Journal of Financial Intermediation 9*, 387–407.

Biais, B., P. Hillion, and C. Spatt (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance 50*, 1655–1689.

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.

Coppejans, M., I. Domowitz, and A. Madhavan (2004). Resiliency in an automated auction. Working Paper, Duke University.

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.

Field, J. and J. Large (2007). Pro-rata matching and the order book dynamics of fixed income futures. Working Paper, University of Oxford.

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 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.

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. (2006). 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. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets 10*, 1–25.

Large, J. and T. Norman (2008). Ergodic equilibria in monotone stochastic sequential games. 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.

Ohta, W. (2006). Quote competition in limit order markets. Working Paper, University of Tokyo.

Parlour, C. (1998). Price dynamics in limit order markets. *Review of Financial Studies 11*, 789–816.

Porter, D. and D. Weaver (1997). Decimalization and market quality. *Financial Management 26*, 5–26.

Portniaguina, E., D. Bernhardt, and E. Hughson (2006). Hybrid markets, tick size and investor trading costs. *Journal of Financial Markets 9*, 433–447.

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.

Ronen, T. and D. Weaver (2001). Teenies' anyone? *Journal of Financial Markets 4*, 231–260.

Rosu, I. (2006). A dynamic model of the limit order book. Working Paper, University of Chicago Graduate School of Business.

Seppi, D. (1997). Liquidity provision with limit orders and strategic specialists. *Review of Financial Studies 10*, 103–150.

Tweedie, R. L. (1976). Criteria for classifying general Markov chains. *Advances in Applied Probability 8*, 737–771.

Tweedie, R. L. (1983). The existence of moments for stationary Markov chains. *Journal of Applied Probability 20*, 191–196.

Van Ness, B., R. Van Ness, and S. Pruitt (2000). The impact of the reduction in tick increments in major U.S. markets on spreads, depth, and volatility. *Review of Quantitative Finance and Accounting 15*, 153–167.

A Appendix

A.1 Definition of a buy-sell symmetric equilibrium

Definition 14 *For any vector v, let $rev(v)$ be v with its elements reversed. Say a function f is buy-sell symmetric if for any argument, v, $f(-rev(v))$ and $rev(f(v))$ exist, and*

$$f(-rev(v)) = -rev(f(v)). \tag{32}$$

Definition 15 *An equilibrium is buy-sell symmetric if c is buy-sell symmetric.*

Note that $-rev(\chi_t)$ represents a limit order book where the unfilled bids and asks in the order book at time t have been 'swapped'. Put loosely, Definition 15 states that a buyer responds to χ_t just as a seller would respond to $-rev(\chi_t)$. This has the consequence that the (ordered) pair of processes (χ, β) has the same distribution as $(-rev(\chi), -\beta)$.

A.2 Proof of Proposition 3.1

Note that in equilibrium

$$E^t_{c_{ask}}[e^{-\rho T^A_t} | Z_t \chi_t = 0] = E^t_{c_{ask}}[e^{-\rho T^A_t}], \tag{33}$$

for the fact that $Z_t = 0$, provides no information about the state of the process χ. The lower bound imposed on β^+ in Section 2.7 ensures that all finite cutoffs $c^1(Z_t\chi_t)$ are identified by (11) and within the support of β_t. Rearranging (11), using the definition of F,

$$F(c^1(Z_t\chi_t)) = \frac{\eta}{k}\left(1 - E^t_{c_{ask}}[e^{-\rho T^A_t} | Z_t\chi_t]\right). \tag{34}$$

Taking expectations in (34),

$$E^t_{c_{ask}}[F(c^1(Z_t\chi_t))] = \frac{\eta}{k}\left(1 - E^t_{c_{ask}}[e^{-\rho T^A_t}]\right). \tag{35}$$

From (33) and (34), the right-hand side of this is $F(c^1(0))$. So the Proposition follows on observing that since the distribution of $Z_t\chi_t$ is unchanged by a change in c_t,

$$E[F(c^1(Z_t\chi_t))] = E^t_{c_{ask}}[F(c^1(Z_t\chi_t))]. \tag{36}$$

A.3 Proof of Theorem 3.4

The expected payoff ex ante to player t is $E[\Pi(\beta_t)]$. Since $\pi \in (0,1)$, the player is indifferent to the realization of Z_t. So $E[\Pi(\beta_t)] = E[\Pi(\beta_t)|Z_t = 0]$. The right hand side of this is, from Theorem 3.3 and (8), and exploiting the buy-sell symmetry,

$$\frac{1}{2} \quad E \quad \left(\beta_t - \frac{k}{2} \quad |Z_t = 0, or_t = buy\right) \tag{37}$$

$$+\frac{1}{2} \quad E \quad \left(\left[\beta_t + \frac{k}{2}\right] \quad e^{-\rho T_t^B} \quad |Z_t = 0, or_t = bid\right). \tag{38}$$

Rewriting, this is

$$\frac{1}{2} \quad E \quad (\beta_t \quad |Z_t = 0, or_t = buy) - \frac{k}{4} \tag{39}$$

$$+\frac{1}{2} \quad E \quad (\beta_t \quad |Z_t = 0, or_t = bid) + \frac{k}{4} \tag{40}$$

$$-\frac{1}{2} \quad E \quad \left(\left[\beta_t + \frac{k}{2}\right] \quad (1 - e^{-\rho T_t^B}) \quad |Z_t = 0, or_t = bid\right). \tag{41}$$

The first two lines of this sum to $E(|\beta_t|)$. So, the whole expression is

$$E\left(|\beta_t|\right) - \frac{1}{2}\left[\delta F^{-1}\left(\frac{3}{4}\right) + \frac{k}{2}\right] \; E\left[1 - e^{-\rho T_t^A}|Z_t = 0, or_t = ask\right], \tag{42}$$

and, finally, from (19),

$$E\left[1 - e^{-\rho T_t^A}|Z_t = 0, or_t = ask\right] = 1 - E_{c_{ask}}^t[e^{-\rho T_t^A}] = \frac{k}{\frac{k}{2} - F^{-1}\left[\frac{1}{4}\right]}. \tag{43}$$

A.4 Stationarity of the state, χ_t

This section discusses how to use drift condition results in Large and Norman (2008), which applies results in Tweedie (1983), in order to prove the stationarity of the state χ_t when in stationary Markov equilibrium under restrictions 1-4 of Section 5.

From Tweedie (1983) we can infer the following condition:

Condition A.1 *Let W_t, defined for $t \in \mathbb{N}$, be a Markov process with state space contained in $\mathbb{Z}$, where for all W_t the conditional support of $(W_{t+1} - W_t)$ is $\{-1, 0, 1\}$. Suppose that there exists a threshold $L^* \in \mathbb{N}$ and $\eta \in \mathbb{R}^+$ such that*

$$E[|W_{t+1}| \;\; | \;\; W_t] \;\; \leq \;\; |W_t| - \eta$$

whenever $|W_t| > L^*$. Then a distribution may be given to W_0 so that W_t is stationary.

An intuition for this condition is that the process reverts toward zero outside $[-L^*, L^*]$, so that L^* and $-L^*$ are positive recurrent. It is now applied in various cases when W_t is a function of the offered depth at time t.

A.4.1 First, let $W_t = \min(\chi_t^1, 0)$.

So W_t is non-zero only in the case of a negative depth at B_t, when $B_t = \underline{B}$. Then 1) no player, knowing that the bid depth is negative, would submit a *buy* (since a *bid* would have superior slippage and also execute immediately) and 2) if $\beta_t = 0$ player t would submit a *bid*. So $\Pr[or_t = bid|Z_t = 1] > \frac{1}{2}$, while only types below $-\frac{k}{2}$ can prefer to *sell*, so $\Pr[or_t = sell|Z_t = 1] \leq F\left(-\frac{k}{2}\right)$. Players with $Z_t = 0$ submit *bids* and *sells* with equal probability. Hence in the case of a negative depth at B_t *bid* pressure strictly exceeds *sell* pressure, and there is upward drift to the negative depth, back towards zero.

A.4.2 Second, let $W_t = \min(L_t^{A_t+}, L_t^{B_t+})$.

Recall that $L_t^{A_t+}$ is $\max\{L_t^{A_t} + 1, 0\}$, and $L_t^{B_t+}$ is $\max\{L_t^{B_t} + 1, 0\}$. So W_t is the lesser of the two inside depths, provided it is positive. For any natural number L, let $LA(L, t) \subset \Omega$ be the set $\{\omega \in \Omega : \ W_t = L, or_t = ask\}$ and consider the quantity

$$\sup\{E[e^{-\rho T_t^A}|\chi_t, or_t](\omega) \ : \ \omega \in LA(L, t)\}. \tag{44}$$

For all t this quantity converges to zero as $L \to \infty$ since time to execution is at most as fast as if all future traders submitted *buys*, a time which becomes arbitrarily distant as $L \to \infty$. Hence from (11), $\inf\{c^1(\chi_t) \ : W_t = L\}$ converges to $-\frac{k}{2}$ (from below) as $L \to \infty$, and so $\inf\{F(c^1(\chi_t)) \ : W_t = L\}$ converges to $F(-\frac{k}{2})$, which is $> \frac{1}{3}$. Hence there exist L^* and $\alpha > \frac{1}{3}$ such that for all $L > L^*$,

$$\Pr[or_t = sell|Z_t = 1, W_t = L] > \alpha > \frac{1}{3}. \tag{45}$$

There is an equivalent argument on the other side of the book, and that *bid*, *ask*, *sell* and *buy* are of course mutually exclusive. Hence there also exist L^* and $\eta > 0$ such that for all $L > L^*$,

$$\begin{aligned}
\Pr[or_t = sell|Z_t = 1, W_t = L] &> \frac{1}{3} + \eta, \\
\Pr[or_t = buy|Z_t = 1, W_t = L] &> \frac{1}{3} + \eta, \\
\Pr[or_t = bid|Z_t = 1, W_t = L] &< \frac{1}{3} - \eta, \\
\Pr[or_t = ask|Z_t = 1, W_t = L] &< \frac{1}{3} - \eta.
\end{aligned}$$

Therefore there exists an $\eta^* > 0$ such that if $W_t > L^*$,

$$E[W_{t+1}|W_t, Z_t = 1] < W_t - \eta^*. \tag{46}$$

But as, when $Z_t = 0$, *bid*, *ask*, *sell* and *buy* are played with equal probability,

$$E[W_{t+1}|W_t, Z_t = 0] = W_t. \tag{47}$$

Combining the last two equations using the Law of Iterated Expectations,

$$E[W_{t+1}|W_t, Z_t = 1] < W_t - \pi\eta^*. \tag{48}$$

□

A.4.3 Third, let $W_t = |L_t^{A_t+} - L_t^{B_t+}|$

So, disregarding negative depths, W_t is the absolute difference between the depth at the best bid, and the depth at the best ask. As the equilibrium is buy-sell symmetric, consider w.l.o.g. the case where $L_t^{A_t+} > L_t^{B_t+}$. Let $c^2(\chi_t)$ be the second element of $c(\chi_t)$. Then,

$$\left(\frac{k}{2} - c^2(\chi_t)\right) E^t_{c_{ask}}[e^{-\rho T_t^A}|\chi_t] = \left(\frac{k}{2} + c^2(\chi_t)\right) E^t_{c_{bid}}[e^{-\rho T_t^B}|\chi_t]. \tag{49}$$

Now, $E^t_{c_{bid}}[e^{-\rho T_t^B}|\chi_t]$ is at most as great as $e^{-\rho\Delta L_t^{B_t+}}$ since time to execution is at most as fast as if all future traders submitted *sells*. And, $E^t_{c_{ask}}[e^{-\rho\Delta T_t^A}|\chi_t]$ is larger than $e^{-\rho L_t^{A_t+}F(-\beta^+)}$, since for all χ_t there is probability of at least $F(-\beta^+)$ of a trader arriving who submits a *buy*. Using these inequalities,

$$\left(\frac{k}{2} - c^2(\chi_t)\right) e^{-\rho\Delta L_t^{A_t+}F(-\beta^+)} > \left(\frac{k}{2} + c^2(\chi_t)\right) e^{-\rho\Delta L_t^{B_t+}} \tag{50}$$

so

$$e^{-\rho\Delta(L_t^{A_t+}F(-\beta^+) - L_t^{B_t+})} > 1 + \frac{2c^2(\chi_t)}{\left(\frac{k}{2} - c^2(\chi_t)\right)}. \tag{51}$$

Under the maintained assumption that $L_t^{A_t+} > L_t^{B_t+}$,

$$W_t F(-\beta^+) < L_t^{A_t+} F(-\beta^+) - L_t^{B_t+}. \tag{52}$$

So,

$$\sup\left\{\frac{2c^2(\chi_t)}{\left(\frac{k}{2} - c^2(\chi_t)\right)} : W_t = L\right\} < e^{-\rho\Delta L F(-\beta^+)} - 1, \tag{53}$$

and, via an increasing transformation,

$$\sup\left\{F\left(c^2(\chi_t)\right) : W_t = L\right\} < F\left(\frac{\frac{k}{2}\left(e^{-\rho\Delta L F(-\beta^+)} - 1\right)}{2 + e^{-\rho\Delta L F(-\beta^+)} - 1}\right), \tag{54}$$

so that

$$\sup\left\{F\left(c^2(\chi_t)\right) : W_t = L\right\} \to F\left(-\frac{k}{2}\right), \tag{55}$$

as $L \to \infty$. As $\inf\{F(c^1(\chi_t)) \ : W_t = L\}$ converges to $F(-\frac{k}{2})$ from below,

$$\lim_{L\to\infty} \Pr[or_t = ask|Z_t = 1, W_t = L] = 0. \tag{56}$$

But for all χ_t there is probability of at least $F(-\beta^+)$ of a trader arriving who submits a *buy*.

Therefore there exists an L^* and an $\eta^* > 0$ such that if $W_t > L^*$,

$$E[W_{t+1}|W_t] < W_t - \eta^*. \tag{57}$$

□

A.4.4 Conclusion

The three previous parts showed that $\min(L_t^{B_t}, 0)$, $\min(L_t^{A_t}, 0)$ and $\min(L_t^{A_t+}, L_t^{B_t+})$ are ergodic with first ergodic moment, and therefore can be given initial distributions so that they are stationary with first moment. Then, $\min(L_t^{B_t}, L_t^{A_t})$ can also; and the third part showed that $|L_t^{A_t+} - L_t^{B_t+}|$ can also. Combining these results, we have that $L_t^{B_t}$ and $L_t^{A_t}$ can be given initial distributions so that they are stationary with first moment, so that they are positive recurrent at zero. Therefore the support of the random variables A_t and B_t is $\{\underline{B}, \underline{B} + k, ..., \bar{A} - k, \bar{A}\}$, implying that χ_t can be given an initial distribution so that it is stationary with first moment.