Pro-Rata Matching in One-Tick Markets

Jonathan Field
AHL, Man Investments
JField@ManInvestments.com
Sugar Quay, London, EC3R 6DU, U.K.

Jeremy Large*
Oxford-Man Institute of Quantitative Finance,
University of Oxford, and
AHL, Man Investments
Jeremy.Large@Economics.ox.ac.uk
Eagle House, Walton Well Road, Oxford, OX2 6ED, U.K.

15 March 2012

Abstract

We identify electronic markets which match counterparties using a pro-rata rule, and whose bid-ask spread is bid down to the price tick size practically always. These futures markets' depths at the quotes on average exceed mean market order size by two orders of magnitude, while order cancellation rates (the probability of an offered lot being cancelled) significantly exceed 96 per cent.

To help explain these facts we then develop a theoretical model where the pro-rata rule creates strategic complementarities in the choice of limit order size, causing traders to risk overtrading by submitting over-sized limit orders, most of which they expect to cancel.

JEL classification: G12; G13

Keywords: Limit Order Book; Pro-Rata; Tick Size; Bid-Ask Spread; Depths.

*Corresponding author. We are grateful to Anthony Ledford, Andrew Patton, Robert Almgren and Patrik Sandås for great support and comments. We also thank participants at the conference on the Industrial Organisation of Securities Markets in Frankfurt (June 2008) and seminar participants at the Oxford-Man Institute for Quantitative Finance (November 2007). Please note that the views expressed herein are those of the authors and not necessarily of Man Investments.

1 Introduction

This paper studies the pro-rata order-matching rule for electronic limit order books, which is in force across a range of futures markets. It argues that the choice of matching rule, for example whether to match orders on a pro-rata basis rather than on the more common price-time basis, is an important policy variable that can profoundly affect market microstructure.

To make this case the paper models and identifies empirically a setting where the order-matching rule is influential. Specifically, we study one-tick markets, i.e. markets whose price tick size is large so that the bid-ask spread does not normally exceed it. We develop a simple theoretical model which highlights how in one-tick markets the queuing of limit orders has distinctive importance compared to elsewhere, so that queuing conditions stemming from a pro-rata order-matching rule can have a range of substantial effects including very high depths.

This is borne out in our 2007 dataset. Scanning over a large majority (by value traded electronically) of all futures markets of that period, we identify four one-tick pro-rata markets. In all four cases total average depths, adding together both the best bid and the best ask, exceed the mean market order size by multiples of over 170 (and rise to a multiple of 600 in the case of Short Sterling). Furthermore, the probability that any given offered lot is withdrawn before it trades approaches certainty: it is significantly greater than 96 per cent (and exceeds 98 per cent on Eurodollar and Euribor). Finally, the bid-ask spread is infrequently larger than the price tick size: significantly less that 3 per cent of the time.

Our theoretical model predicts the existence of these large inside depths and cancellation rates in one-tick pro-rata markets. This is because the pro-rata algorithm, by matching limit orders to each countervailing market order in proportion ('pro-rata') to their sizes, creates strategic complementarities in the order-size decisions of liquidity suppliers, which are absent under the price-time rule.[1] Therefore, when faced with one

[1] The pro-rata matching rule is in some respects similar to book-building on primary equity markets. Section 3.3 gives a full account of it. For a comprehensive review of strategic complementarities, see Cooper (1999), see also Cooper and John (1988) and Vives (1990).

another's limit orders, in equilibrium traders submit bids and asks of greater size than the quantity that they wish to trade, thereby over-inflating market depths.

As they over-offer, liquidity suppliers in the model accept a risk of overtrading. We show that overtrading risk is an important factor in forestalling an unending order-size 'arms race', where ever larger limit orders are submitted and depths explode. Indeed, substantial overtrading risk (which we define with reference to a Pareto distribution of market order sizes) is sufficient to ensure that equilibrium exists. However, if such overtrading risk is relatively weak, then equilibrium over-offering can be of large magnitude. Furthermore, we show that as the population of traders on the market becomes large then, rather than these strategic distortions attenuating towards a competitive outcome, they actually become inefficiently accentuated.

This explains the high cancellation rates in our data: over-offering limit order submitters often meet their trading objectives while parts of their limit orders remain unfilled (i.e. unexecuted). These parts can then be cancelled, and may constitute a large majority of the volume that was initially offered for trade.

To assess the extent of overtrading risk we study quantiles of the distribution of market order sizes across the four identified one-tick pro-rata markets. We find that the distributions have a fairly long upper tail. On all four markets, their upper 1 percentile exceeds the mean by more than ten times. Given the pro-rata rule, this implies that, when a market sale (purchase) arrives at a deep book, there is probability greater than 1 per cent that it is so large as to cause all bids (asks) at the quotes to execute a size that exceeds expectations by a factor of over 10. While significant, this is not an overwhelming tail, which may help to explain the large extent of over-offering we observe on these markets.

Our model can be better understood in the context of Glosten (1994), which also describes a pro-rata limit order book, while modeling market order submission more completely than us. In Glosten (1994) liquidity suppliers compete depth up to such a level that they obtain zero expected payoff. While they profit by trading with small market orders because such orders carry less information, they suffer when picked-off by larger market orders – breaking even on average.

Because in Glosten (1994) players' limit orders directly substitute for and crowd-out one another, Glosten (1994) does not model the strategic complementarities that can drive traders to inflate high depths to much greater heights, simply because they are already high. These are helpful, in supplement to the intuition in Glosten (1994), for explaining our data. In particular, they can help to explain the two orders of magnitude that separate average inside depths from trade sizes.

In order to bring these strategic complementarities into focus, our model leaves aside the effects due to information asymmetry in Glosten (1994) (see also Kyle 1985, Glosten and Milgrom 1985 and Handa and Schwartz 1996), and postulates a simple motive on the part of each liquidity supplier to trade a target quantity. We can then measure both over-offering and overtrading directly against this target quantity.

Consequently, our model relates to theories of execution uncertainty (see Cohen, Maier, Schwartz and Whitcomb 1981), or about the costs of waiting to trade (see Foucault 1999, Foucault, Kadan, and Kandel 2005). Compared to prior work, waiting costs take novel forms in our setting: in particular, to reduce the cost of waiting, limit order submitters bear *over*-execution uncertainty.

Other important antecedents to this work are in Bernhardt and Hughson (1996), Kadan (2006) and Seppi (1997), which present theories of tick size, inside depths (i.e. depths at the best quotes) and welfare in dealer and hybrid markets. Consequences of discreteness for bias in volatility estimation are treated in Ball (1988), Gottlieb and Kalay (1985) and Harris (1990). Ways of overcoming such bias are discussed in Delattre and Jacod (1997), Oomen (2006), Zeng (2003) and Large (2005). To our knowledge, there are not yet empirical market microstructure studies in the literature of cases where bid-ask spreads scarcely ever differ from the tick size, although there are theoretical analyses in Parlour (1998) and Large (2008). A change in order matching rules is studied empirically in Frino, Hill, and Jarnecic (2000).

The paper proceeds as follows: Section 2 observes a wide range of futures markets and identifies a subset with the market microstructure that interests us. These markets are analyzed empirically in some detail Section 3, in order to establish the properties around

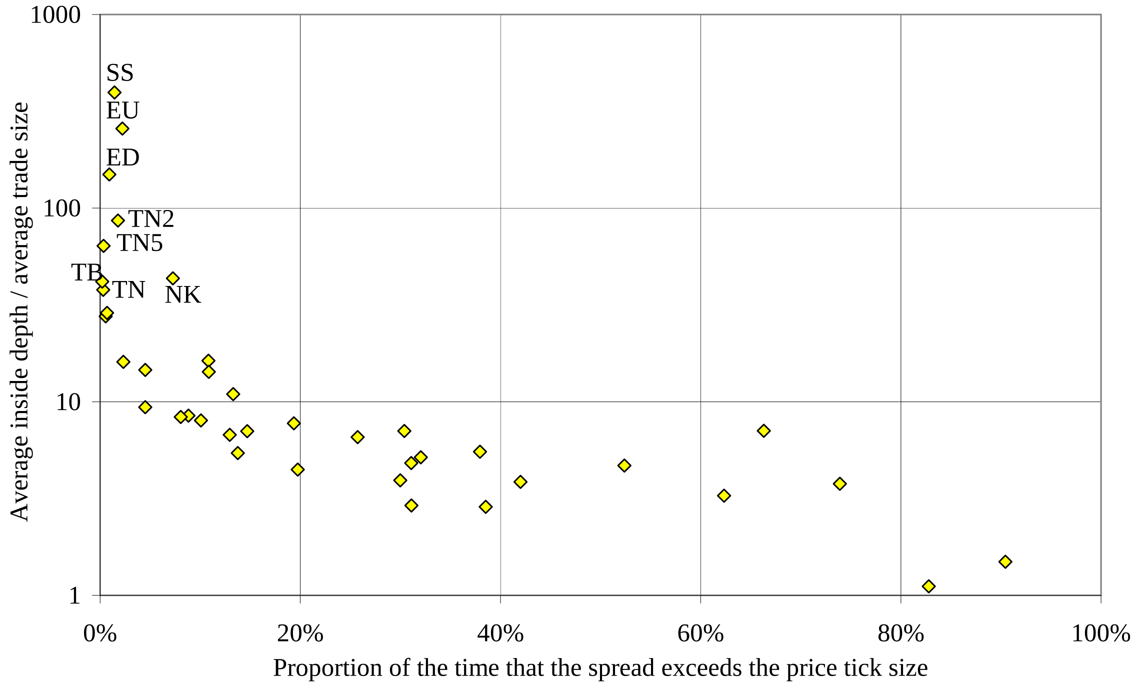

Figure 1: A chart plotting the 40 observed futures markets. Those markets of greatest interest to us are labeled. One observation fell below the x-axis, at coordinates (94%,0.7).

cancellations, depths and spreads outlined in the Introduction. Section 4 develops and analyzes our model of strategic complementarities in limit order submission size. Finally, Section 5 concludes.

2 Search for one-tick pro-rata markets

We begin our analysis by identifying 100 underlying instruments with liquid futures contracts, such as physical commodities, financial assets or indices. In initial investigations, we studied a period from 23 November to 11 December 2006. However, to reduce risks of over-fitting, following Biais, Hillion, and Spatt (1995) we select a fresh period to report our results. We find little change. This period is from 16th to 20th April 2007.

Futures contracts (futures) on a given instrument are available at multiple delivery dates (normally at spacings of a month or a quarter), and each such future is traded and cleared separately, with its own market microstructure. For each of the 100 instruments we select the most actively traded delivery date over our observed period. For this purpose we proxy market activity by the number of changes to depths or prices at the best quotes on the electronic limit order book per minute, calling this measure 'updates per minute'. We confine our attention to trading on limit order books, rather than on trading floors or pits.

Of these 100 electronic limit order books, 40 had on average more than 10 updates per minute. Since our interest is in fairly liquid markets, we select these for closer analysis.[2] The retained set of 40 markets comprises six currency futures, nine government bond futures, three short-term interest rate futures (STIRs), 20 stock index futures from 12 countries (including 'E-mini' products), and two commodity futures.

Based on a simple criterion regarding deviation from median daily prices, outliers in our data were censored.[3] Each market was observed continuously between an appropriate fixed start and end time of day, so as to include more than 90 per cent of official opening hours on its primary exchange. Table 1 documents our observation start and end times for eight markets of particular interest.

2.1 A comparison of the top-40 futures markets

Figure 1 plots the 40 retained futures markets along two central dimension for this study. On the vertical axis it shows the average of the mean depth at the best bid and the mean depth at the best ask, when divided by the average trade size: in brief, the 'average inside depth / average trade size'.[4] To catch the substantial disparity between markets, a logarithmic scale is used. On the horizontal axis is shown the proportion of the time that the bid-ask spread exceeds the price tick size.

Figure 1 indicates that high depths at the quotes are associated with bid-ask spreads near the price tick size, a relationship that was investigated in Harris (1994) and Goldstein and Kavajecz (2000). For our purposes the most interesting part of the chart is its top-left corner, where inside depths are very high and the bid-ask spread infrequently exceeds the price tick.

[2] The busiest market in this subsample had on average 484 updates per minute.

[3] Applied to cases where prices were greater than 25% away from the daily mid. Instances of crossed markets, where the price of the best bid exceeded the best ask, were also excluded.

[4] This statistic is half of the related statistic quoted in the Introduction, which regarded the bid and ask depths added together.

Among the 40 markets, the four deepest (on the measure reported in Figure 1) comprise the two-year US Treasury Note future (labeled TN2 in the graph), together with the three short-term interest rate futures (STIRs): Short Sterling (SS), Euribor (EU), and Eurodollar (ED). STIR futures contracts differ from stock index futures, for example, in that they operate on a cash-settlement basis with no underlying traded security. An Appendix explains how they are created and administered.

These four markets match orders using a pro-rata matching algorithm. As they have bid-ask spreads which equaled one price tick more than 96 per cent of the time (a statement we will later make statistically rigorous), we term them *one-tick pro-rata* markets. We select them for more detailed analysis.

The next four deepest markets are the Nikkei (Osaka) stock index future (NK), the 5-year and 10-Year US Treasury Note futures (TN5 and TN), and the US Treasury Bond future (TB). At the time of our data, these markets operated on a price-time basis. For comparative purposes, we provide the same data on their microstructure. All eight mentioned markets experience on average more than 80 updates per minute.

3 One-tick pro-rata market microstructure

This section studies the market microstructure of the one-tick pro-rata markets in more detail. They are the two-year US Treasury Note future, Short Sterling, Euribor, and Eurodollar. Formal inference is done to establish confidence bounds for the statistics on the vertical and horizontal axes of Figure 1, and the scope is widened to include their cancellation rates and their depth distributions.

Table 1 provides summary data about the four one-tick pro-rata markets. In generating this data we observe every change in the prices and depths of the limit order book at the best quotes, as well as the time of that change, to the nearest second. We do not use data about the shape of the limit order book outside the best quotes.[5]

[5] Although we do not study prices outside of the best quotes, informal investigation of the one-tick pro-rata markets indicates that the limit order book is in fact typically almost empty outside the quotes.

	Eurodollar	Euribor	Short Sterling	2yr T-Notes	5yr T-Notes	10yr T-Notes	T-Bills	Nikkei
Exchange	CME	LIFFE	LIFFE	CBOT	CBOT	CBOT	CBOT	Osaka
Market organization	pro-rata				price-time			
Start of observation (GMT)	14:00	08:00	08:00	13:30	13:30	13:30	13:30	01:12
End of observation (GMT)	19:30	15:54	15:51	19:45	19:45	19:45	19:45	07:00
Updates per minute	308	161	101	213	472	355	339	86
Market orders per hour	299	259	335	270	648	1,409	925	1,500
Mean depth at best bid (lots)	12,162	6,811	19,573	5,172	4,801	3,019	1,706	422
standard error	*963*	*690*	*3,930*	*365*	*316*	*150*	*94*	*83*
Mean depth at best ask (lots)	10,595	7,518	18,805	5,418	4,485	2,810	1,664	425
standard error	*543*	*692*	*2,239*	*562*	*330*	*157*	*94*	*87*
Mean market order size (lots)	76	28	46	61	73	77	40	10
standard error	*2.8*	*1.1*	*2.0*	*3.6*	*1.6*	*1.1*	*0.7*	*0.2*
Upper 1% quantile, trade size	873	300	805	750	750	1,000	426	135
when divided by the mean	*11.5*	*10.8*	*17.6*	*12.2*	*10.3*	*13.0*	*10.6*	*13.8*
Mean size of limit bids (lots)	157	189	227	133	43	40	22	9
standard error	*1.7*	*3.8*	*5.0*	*2.4*	*0.2*	*0.3*	*0.1*	*0.2*
Mean size of limit asks (lots)	163	172	239	136	40	38	22	10
standard error	*1.7*	*3.4*	*5.6*	*2.6*	*0.2*	*0.2*	*0.1*	*0.3*
Propn. of time that the bid-ask spread is the price tick size	99.1%	97.8%	98.6%	98.2%	99.7%	99.7%	99.8%	92.7%
Price Tick Size	0.005	0.005	0.01	0.0078	0.0156	0.0156	0.0312	10
Proportion of offered lots that are cancelled at the quotes	98.2%	98.9%	97.4%	97.3%	90.7%	76.5%	84.2%	44.2%
standard error	*0.1%*	*0.2%*	*0.5%*	*0.4%*	*0.4%*	*0.8%*	*0.7%*	*2.8%*

Table 1: Summary information about four short-term interest rate futures markets.

3.1 Inside depths, and their comparison with market order size

Table 1 reports the average depth observed at the best bid, and that at the best ask. In italics below these averages is shown a standard error. Both bid and ask depths range from about 5,000 lots in the case of 2-Year US Treasury Note futures to over 19,000 lots in the case of Short Sterling (the unit of quantity here is the 'lot').[6]

The figures reporting depths in Table 1 are based on sampling the data periodically, rather than at every available fresh observation of depth. This accommodates a procedure to estimate standard errors. Specifically, the observed period is broken into a large number of successive durations: 50 for each trading day.[7] Rather than sampling the data at evenly spaced times, we prefer to sample in business time, see Oomen (2006) and Hansen and Lunde (2006). This controls for intraday seasonal and other fluctuations in market activity and volatility. To proxy for the passing of business time, we observe the rate at which fresh limit orders are added at the quotes. The successive durations of time are chosen to contain (as closely as possible subject to rounding limitations) equal numbers of newly arrived lots offered for trade at the quotes.

Inside depths were observed at the start of each duration. On the assumption that once sampled in business time the resulting time series of inside depths is stationary, standard errors may be constructed for the mean observation, using the well-known technique due to Newey and West (1987). This is done with 24 lags. To compare with this, the average size of all market orders was found, and Newey-West standard errors were again calculated with 24 lags. Average inside depths dramatically exceed average market order size.

Figures 3 and 4 present time-weighted histograms of the inside depths observed on, respectively, the four one-tick pro-rata markets of principal interest in this study, and the four one-tick price-time markets described in Table 1. The eight histograms have similar features within each group of four, but differ markedly across groups. These markets are all one-tick markets with large inside depths. However, within this set, there is a clear difference in the inside depth distributions between the pro-rata and price-time priority

[6]See Table 2.

[7]Alternative values to 50 were evaluated, and results are similar.

markets.

While the four one-tick price-time markets in Figure 4 exhibit a single, pronounced, mode near to the mean, Figure 3 tells a different story. The one-tick pro-rata markets in Figure 3 all have modes that are greatly exceeded by the mean together with long upper tails, indicating high variability in offered depths.

3.2 Cancelled limit orders

Market participants are informed whenever an offer to trade ceases to be available to them. Such events also appear in our data, as reductions in offered depths. There are two possible causes of such a reduction: either some outstanding bids or asks are cancelled, or they successfully execute during a transaction. However, the two causes cannot be distinguished by market participants, or in our data.

Nevertheless, because we observe frequently-updated (but, by a few seconds, slightly delayed) trading volume data we can calculate the proportion of such depth reductions at the quotes, that were due to cancellation rather than to trades. We call this the *cancellation rate*. It is given by

$$\textit{cancellation rate} \ = 1 - \frac{\text{trading volume in lots}}{\text{number of offered lots removed at the best quotes}} \quad . \tag{1}$$

Under assumptions, the cancellation rate is interpretable as the empirical probability that a lot is cancelled before it trades. The assumptions are as follows: a) we assume conservatively that *no* lots are cancelled at stale prices outside the quotes. Because these markets are deep at the quotes, we also assume that b) no lots are traded against, which just previously were priced outside the quotes. Once a lot has been offered for trade on the exchange, it has one of two fates: either it is traded, or it is cancelled. Under assumptions a) and b) it can only trade or be cancelled while priced at the quotes. This implies that the cancellation rate is the empirical frequency with which offered lots are cancelled.

To do inference on the size of the cancellation rate, we calculate it repeatedly and separately for each period in the set of consecutive, irregular-length periods that were used to sample inside depths, and then take the average. Newey-West standard errors for this average are provided (with 24 lags), on the assumption that this sequence of cancellation rates is a stationary time series.

Table 1 indicates that on all of the four one-tick pro-rata markets, cancellation rates exceeded 96 per cent at a (one-sided) confidence level of 0.05.

3.3 The pro-rata matching algorithm

The next Section will offer an explanation for such high cancellation rates and depths involving the pro-rata matching algorithm. The pro-rata algorithm differs importantly from the price-time algorithm used widely on equity and other markets. Both on pro-rata and on price-time markets, the sequence in which limit orders execute is determined according to price priority. However, evidently price priority must be supplemented by a further criterion to allocate counterparties among limit orders that are equally priced.

Unlike on a price-time market, on a pro-rata market all limit orders at the same price trade simultaneously against each countervailing market order. The market order is typically too small to trade with them all in their entirety. So, the trade is shared among the limit orders in proportion to their sizes. For example, suppose that traders A and B have both submitted limit asks at the best ask, whose price is \$10. A offers 100 lots for sale, while B offers 50 lots. If a counterparty buys three lots at \$10, then the order matching algorithm breaks this purchase into two parts, allocates a trade of two lots to trader A, and one lot to trader B.

There are some further details of lesser importance: additional rules are needed to distribute 'odd lots' arising due to quantity discreteness; and 'bid-up' orders have special rights. An Appendix explains these terms and gives a precise account of the pro-rata algorithm, as well as some further examples, as implemented by Euronext.liffe.

Some exchange officials argue that the pro-rata algorithm encourages liquidity supply on markets with low volatility, such as short-term interest rate futures (STIRs). The pro-rata algorithm has similarities to practices seen in long-standing futures trading pits, where, possibly because time priority is hard to establish, market orders are often shared-out among a number of distinct competing liquidity suppliers.

4 A model of over-offering

The previous Section described four markets where

1. the order-matching rule is pro-rata,
2. the bid-ask spread is infrequently greater than the price tick size,
3. inside depths are some orders of magnitude greater than typical trade sizes, and
4. most lots offered for trade at the best quotes, in the form of limit orders, are withdrawn before they actually trade.

This section presents a theoretical model of equilibrium behavior to explain the latter two of these features (3 and 4), while taking as given the former two (1 and 2).

4.1 Model intuition

For concreteness, this part develops some of the intuitions that the model will formalize. The model is specified in the next Section, and then equilibrium behavior is characterized.

We have observed that inside depths can be of larger order of magnitude than the average trade size. For example, on the Short Sterling future, average inside depths at the bid (ask) exceed the average trade size by more than 300 times. Assuming that the quantity per transaction intended by a limit order submitter is of similar magnitude to the average trade size, it is therefore of smaller order of magnitude than average inside depths.

The pro-rata rule distributes each market order among all competitively-priced counterparty limit orders in quantities proportional to their sizes. Therefore in the Short Sterling example a 'truth-telling' limit order of quantity equal to the average trade size, submitted when inside depths are near the mean, executes in expectation only 1/300 of its volume with each market order (and adverse rounding rules detailed in the Appendix may make that proportion even lower). However, a limit order of much larger size may execute a more satisfactory quantity for its submitter.

This leads to a number of conjectures: perhaps, to get an adequate 'slice' of countervailing market orders, a trader may submit limit orders of larger quantity than their true trading intention. Because of equilibrium effects, this might create escalating feedback in offered depths. Furthermore, because traders are offering more than they intend to trade, they may have a high propensity to cancel their limit orders once they have achieved their trading objectives.

To assess these conjectures, we now study equilibrium depths in a theoretical model.

4.2 Model specification

Suppose that there are exactly n traders who wish to sell a future, or some other asset, which is traded on a limit order book. Because the price tick is large, the bid-ask spread and cost of immediacy are also large. Therefore the traders submit asks, rather than immediately selling cheaply with a market sale. Furthermore prices change slowly so that the best ask, which is the upper limit of the bid-ask spread, is effectively fixed for the duration of the game. Under these conditions the limit order book presents a queuing problem to traders where the order-matching rules, because they determine traders' progress in the queue, are central.

Traders each hold an identical inventory in the asset, so that each of them would ideally desire to sell S^* units at the prevailing best ask price. Because of uncertainty, they cannot be sure to trade S^*: rather, they trade a different amount which is determined in equilibrium: call the amount that trader i actually does trade f_i. For simplicity let traders have symmetric, quadratic loss, so that trader i seeks to minimize

$$E[(f_i - S^*)^2]. \tag{2}$$

Each of the n traders has one decision to take: trader i must choose a quantity Q_i to offer for sale at the best ask. We consider a simultaneous game where all quantity decisions are irrevocably taken at the same moment. However, because of idiosyncratic latencies in transmitting orders to the electronic exchange, we will assume that they are implemented sequentially.

After the players have submitted their asks to the market, one market purchase arrives exogenously and trades with one or more of the submitted asks. Its size, M, is random of commonly known distribution. After that, all remaining limit orders in the market are cancelled and the game ends. Thus, trader i cancels a limit order of size $(Q_i - f_i)$, which is the number of units of the asset that she offers for sale, less the part that actually executes (i.e. fills).

Write

$$D = \sum_{i=1}^{n} Q_i.$$

So, D is the offered inside depth at the best ask which is available for trade with the market purchase of size M. Assume that M may be greater than S^*.

We will study two settings: in the first, the exchange implements a price-time matching rule; while in the second an alternative pro-rata rule is employed.

Pro-rata rule Leaving aside issues of quantity discreteness and consequent rounding, the pro-rata rule states that a total traded quantity of $\min(M, D)$ is divided between agents $\{i = 1, ..., n\}$ in the ratio $\{Q_1 : Q_2 : ... : Q_n\}$. Thus f_i, the amount that agent i trades, is

$$f_i = \frac{Q_i}{D} \min(M, D). \tag{3}$$

Write Q_{-i} for the sum of the quantities chosen by the agents other than agent i, so that $D = (Q_i + Q_{-i})$. Making f_i's dependence on Q_i and Q_{-i} explicit, we may write

$$f_i(Q_i, Q_{-i}) = \min\left(\frac{Q_i M}{Q_i + Q_{-i}}, Q_i\right). \tag{4}$$

Price-time rule Recall that because of idiosyncratic latencies there is no probability of order submissions arriving simultaneously at the exchange. Because of the symmetry of the model, for notational convenience we may assume that players arrive in the order: player 1, then player 2, then player 3, ..., n, but that players are not aware of this order. The price-time rule states that a quantity $f_1 = \min(M, Q_1)$ is allocated to player 1; that $f_2 = \min(M - f_1, Q_2)$ is allocated to player 2; and so forth.

Comment It is possible that the random size of the market order, M, exceeds the available inside depth at the best ask, D. In such cases we leave undetermined the question, what happens to the remaining part of the market order, of size $(M - D)$ lots: the model is invariant whether it is cancelled automatically, or the market order 'walks the book' finding unmodeled asks at higher prices than the best ask to trade with.

Definition 1 *Let $\hat{Q}$ be the best response function, so that $\hat{Q}(Q_{-i})$ minimizes expected loss, (2), conditional on Q_{-i} as given by (4).*

Because there is a concave loss function, taking First Order Conditions we have that if $Q_i = \hat{Q}(Q_{-i})$ then

$$E\left[2(f_i - S^*)\frac{\partial f_i}{\partial Q_i}\right] = 0. \tag{5}$$

Definition 2 (equilibrium) *In **equilibrium**, for all $i \leq n$, player i selects*

$$Q_i = \hat{Q}(Q_{-i}).$$

Therefore, provided that $\hat{Q}$ is weakly increasing, $Q_{-i} = (n-1)Q_i$.

4.3 Model analysis and equilibrium characterization

If $Q_{-i} = 0$, then, regardless of the matching rule $f_i = \min(M, Q_i)$. Consequently, it is best for a trader to select $Q_i = S^*$. This observation gives the following lemma.

Lemma 4.1 *Faced with $Q_{-i} = 0$, it is optimal for a trader to be truth-telling, so that:*

$$\hat{Q}(0) = S^*. \tag{6}$$

So if only trader i is supplying liquidity at the best ask, she is happy to submit an ask of truth-telling size S^*. Likewise, we have the following Proposition:

Proposition 4.2 *Suppose that the matching rule is price-time. Then in the unique equilibrium each agent submits an order of size S^*.*

Proof. Writing $R_i = M - \sum_{j=1}^{i-1} f_j$, player i receives $f_i = \min(R_i, Q_i)$. Minimizing $E[(f_i - S^*)^2]$ is therefore achieved by setting $Q_i = S^*$. This is strongly optimal for player 1 and at least weakly optimal for other players. Since all players ascribe positive probability to being player 1, they choose S^*. ■

Because of the absence of discretionary pricing, as well as of asymmetric information, equilibrium is simple in the price-time case. Traders express their true intended volume of trade in their choice of limit order size. Matters are however otherwise in the pro-rata case:

Proposition 4.3 *Under a pro-rata matching rule this game exhibits* ***strategic complementarities****, whereby the best response function is increasing:*

$$\frac{\partial \hat{Q}(Q_{-i})}{\partial Q_{-i}} > 0. \tag{7}$$

Proof. See Appendix. ■

Proposition 4.3 formalizes the principle that under a pro-rata rule, when facing great depth at the ask (bid) an asker (bidder) has an incentive to *increase* the size of her limit order. Jointly with Lemma 4.1, this entails that their best response functions to Q_{-i} are monotonically increasing, starting from S^*. This is illustrated in Figure 2.

Consequently, players i prefers to **over-offer** by offering a larger limit order than the quantity S^* which is really desired, whenever other players wish to submit limit orders of any size.

Definition 3 (substantial overtrading risk) *The market has* ***substantial overtrading risk*** *if market order sizes follow a Pareto distribution, so that for $x > 1$*

$$Pr[M \geq x] = \frac{1}{x^{\alpha}},$$

with $\alpha \in (0, 1)$, so that M has a heavy upper tail.

So, substantial overtrading risk is modeled by giving M a well-known distribution with a fat upper tail (it may well be that this condition, which is sufficient for our results, could be weakened). This places a lower bound on the danger of overtrading which arises from over-offering. The following Lemma infers a resulting condition, which Proposition 4.5 uses to deduce equilibrium existence.

Lemma 4.4 *Suppose that the market has a pro-rata matching rule and substantial overtrading risk. Then there exists* $\epsilon > 0$ *such that:*

$$\frac{\partial \hat{Q}}{\partial Q_{-i}} < (1-\epsilon)\frac{\hat{Q}}{Q_{-i}}, \qquad \textit{so that} \qquad \frac{\partial \log \hat{Q}}{\partial \log Q_{-i}} < 1-\epsilon. \tag{8}$$

Thus the elasticity of $\hat{Q}$ *with respect to* Q_{-i} *is less than unity.*

Proof. See Appendix. ■

Proposition 4.5 *Suppose that the market has a pro-rata matching rule and substantial overtrading risk. Then a unique equilibrium exists which is symmetric. Traders choose to submit asks of equal size such that there is over-offering: i.e.* $Q_i > S^*$.

Proof. We noted that equilibrium has the property that $Q_{-i} = (n-1)Q_i$. Figure 2 shows the monotonic increasing best response function intersecting with the line of gradient $1 : (n-1)$. From Proposition 4.3 and Lemma 4.4,

$$0 < \frac{\partial \log \hat{Q}(Q_{-i})}{\partial \log Q_{-i}} < 1-\epsilon. \tag{9}$$

It follows from this that such intersection always exists and is unique. This point of intersection describes a unique and symmetric equilibrium where $Q_i > S^*$. ■

4.4 Comparative Statics

Combining Propositions 4.2 and 4.5 gives the following result:

Corollary 4.6 *On a market with substantial trading risk, equilibrium depths are greater under the pro-rata rule than under the price-time rule.*

To understand how strategic complementarities may compound offered depths, we may consider the effect of increasing the number of liquidity suppliers:

Corollary 4.7 *Under a pro-rata matching rule, let the number of players,* n*, increase while holding fixed other primitives of the model. Then the equilibrium ratio by which each player over-offers (i.e.* $\frac{Q_i}{S^*}$*) rises, as does depth per player (i.e.* $\frac{D}{n}$*).*

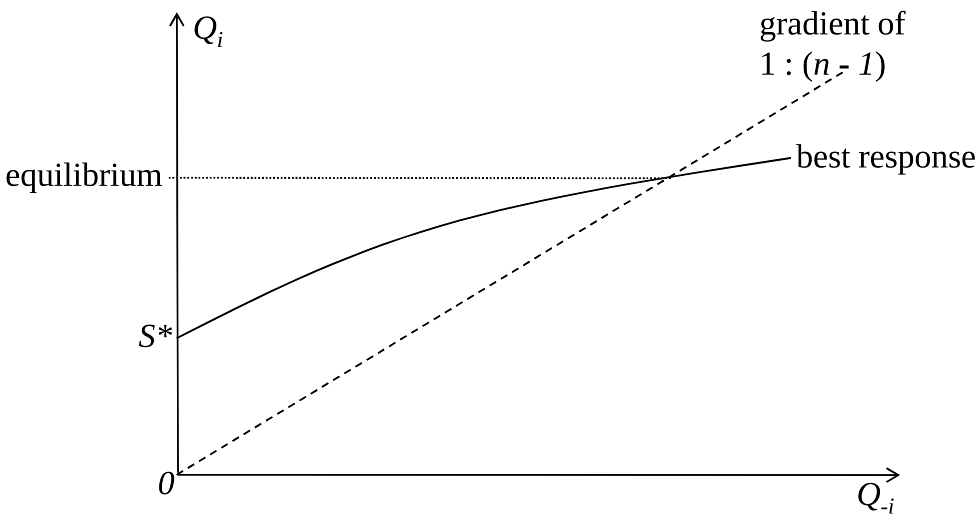

Figure 2: Illustration of equilibrium in the pro-rata market game.

Proof. Note that the function $\hat{Q}(\cdot)$ is invariant to n. From Lemma 4.4 it is monotonic increasing. Therefore its intersection with the line of gradient $1 : (n-1)$, depicted in Figure 2, rises as n increases, resulting in an equilibrium where each trader over-offers more. ■

So in a market with a large number of independent traders, over-offering and over-trading risk are particularly accentuated. Strategic interaction among competing limit order suppliers creates escalating feedback in limit order size and depths to a greater degree, the more players there are. Furthermore, the effect of these interactions on depths is increasing, even after correcting for the increased number of traders by considering depth per trader. Finally, if the amount by which players overtrade is unbounded in the limit as $n \to \infty$, then expected loss also becomes large without bound as $n \to \infty$. This follows from the fact that the Pareto distribution from which market order sizes are drawn has infinite first and higher moments.

The next Proposition outlines a further consequence regarding cancellation rates:

Proposition 4.8 *Under a pro-rata matching rule, as the number of players, n, increases the equilibrium amount that each trader cancels increases, in the sense of first order stochastic dominance.*

Proof. Player i cancels a quantity $\max(Q_i - f_i, 0)$. In symmetric equilibrium this is $\max\left(Q_i - \frac{M}{n}, 0\right)$. However, by Lemma 4.4 $\left(Q_i - \frac{M}{n}, 0\right)$ is increasing in n. ■

In the model, once a bid or ask has attracted the intended volume of trade, its remainder is cancelled. Order cancellation becomes greater when there are many limit order suppliers because of two effects: as the number of players n rises, first player i's limit order executes with a smaller proportion of the market order; and second each player is inclined to offer a greater quantity for trade, more of which is therefore cancelled in the event of no trade.

4.5 Over-execution risk and large market orders

The model identifies substantial overtrading risk as the element which causes depths on one-tick pro-rata markets to be bounded. It is therefore helpful to check that in our data we see evidence of overtrading risk. In a pro-rata setting, overtrading by liquidity suppliers can be identified quite easily, by looking for abnormally large market orders in the historical order flow.

To gather evidence of overtrading risk on the four one-tick pro-rata markets, Table 1 reports the upper 0.01 quantile of their trade size distributions. This exceeded the mean size by a factor of between 10.8 (in the case of Euribor) and 17.6 (in the case of Short Sterling). When a market purchase arrives that is 17.6 times larger than the expectation, and asks at the quote are over-offering by at least 17.6 times, then those asks over-execute relative to (unconditional) expectations by a potentially unwelcome factor of 17.6. So, the market shows similar properties to *substantial overtrading risk*. However, given the very high observed depths it seems likely that this overtrading risk is not found by traders to be overwhelming.

5 Conclusion

We identify a type of market where the choice of order-matching rule significantly affects the quality of the market microstructure. Specifically, using rich tick-level data from a majority (by traded volume) of futures markets worldwide, we find a set of one-tick pro-rata markets which exhibit extremely high offered depths at the quotes; and a likelihood of withdrawing limit orders before they trade which approaches certainty.

We describe in a simple theoretical model a way in which the pro-rata order-matching rule can explain these extreme features, where the more common price-time rule cannot. High offered depths arise due to the submission of oversized limit orders, brought about by a strategic interaction among liquidity suppliers because of the pro-rata rule. High cancellation rates arise because traders cancel the parts of such oversized orders whose sole purpose is to attract from the pro-rata matching algorithm a greater allocation of market order flow.

The topic of order-matching rules presents a large number of openings for future research. Direct study of policy experiments where matching rules have been changed from price-time to pro-rata, or back, or to a hybrid (a current example of which is called 'time-pro-rata') such as in Frino, Hill, and Jarnecic (2000) could offer additional insight. Determining if the interactions described in this paper exclude less-specialized clienteles of investors from supplying liquidity at these markets, thereby damaging market liquidity, would also be important, and would be of interest to policymakers and the designers of exchanges. Understanding if order-matching rules can alter competition between exchanges would be a further promising approach.

Finally, we would argue that if the order-matching rule can have profound consequences in one-tick markets, then it can also have effects elsewhere. This offers promise for studies of the pro-rata order-matching rule under other conditions, such as for example a smaller price tick size.

References

Ball, C. A. (1988). Estimation bias induced by discrete security prices. *Journal of Finance 43*, 845–865.

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

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.

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.

Cooper, R. (1999). *Coordination Games: Complementarities and Macroeconomics.* Cambridge, New York and Melbourne: Cambridge University Press.

Cooper, R. and A. John (1988). Coordinating coordination failures in Keynesian models. *Quarterly Journal of Economics 103*, 441–463.

Delattre, S. and J. Jacod (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli 3*, 1–28.

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.

Frino, A., A. Hill, and E. Jarnecic (2000). An empirical analysis of price and time priority and pro rata trade execution algorithms in screen-traded markets. *Journal Of Derivatives 7*, 41–48.

Glosten, L. R. (1994). Is the electronic limit order book inevitable? *Journal of Finance 49*, 1127–1161.

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.

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.

Gottlieb, G. and A. Kalay (1985). Implications of discreteness of observed stock prices. *Journal of Finance 40*, 135–153.

Handa, P. and R. Schwartz (1996). Limit order trading. *Journal of Finance 51*, 1835–1861.

Hansen, P. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics 24*, 127–281. The 2005 Invited Address with Comments and Rejoinder.

Harris, L. (1990). Estimation of stock variances and serial covariances from discrete observations. *Journal of Financial and Quantitative Analysis 25*, 291–306.

Harris, L. (1994). Minimum price variations, discrete bid-ask spreads, and quotation sizes. *Review of Financial Studies 7*, 149–178.

Kadan, O. (2006). So who gains from a small tick size? *Journal of Financial Intermediation 15*, 32–66.

Kuprianov, A. (1986). Short-term interest rate futures. *Economic Review* (Sep/Oct), 12–26. available at http://ideas.repec.org/a/fip/fedrer/y1986isep-octp12-26nv.72no.5.html.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica 53*, 1315–1336.

Large, J. (2005). Estimating quadratic variation when quoted prices jump by a constant increment. Nuffield College Economics Group working paper W05.

Large, J. (2008). A market-clearing role for inefficiency on a limit order book. Forthcoming, *Journal of Financial Economics.*

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite heteroskedasticity and autocorrelation consistent variance covariance matrix. *Econometrica 55*, 703–708.

Oomen, R. A. C. (2006). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics 3*, 555–577.

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

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

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics 19*, 305–321.

Zeng, Y. (2003). A partially observed model for micromovement of asset prices with Bayes estimation via filtering. *Mathematical Finance 13*, 411–444.

A Definition of STIRs

Short-term interest rate futures (STIRs) are described in Kuprianov (1986). They are never settled by delivery of an underlying security. Rather, when a contract's expiry date is reached, it is settled by the exchange making a final margin call on traders ('cash settlement'). The final settlement price of a futures contract is a linear function of a filtered average – say, Y – of the annualized interest rates offered on deposits, by a representative panel of banks. This is calculated by an independent agency.

Futures contracts are sold in lots. The size of a lot is set by the exchange. In the case of Eurodollar futures with a maturity of 3 months, the exchange levies by margin call a cost of $25 per lot for owners of the contract, when Y falls by a basis point.

With this rule, owning a lot is commonly interpreted as being like synthetically owning a deposit of $1,000,000 at a bank. The reasoning for this is that one basis point of annualized interest on a principal of $1,000,000 corresponds to $100 per year, or $25 over three months: which, as outlined, is the cost to owners of a lot when Y falls by a basis point. Table 2 records the nominal values per lot of the STIRs we study here.

B Account of the pro-rata trading algorithm

The pro-rata trading algorithm was introduced at Euronext.liffe in August 1999, to replace the traditional price-then-time priority algorithm on some of its markets. The algorithm determines the exchange's order-matching rules. Thereby, it controls the time-to-execution of limit orders.

Contract Unit of Trading	(Nominal Contract Size)
Three Month Euro (EU) contract	EUR 1,000,000
Three Month Sterling (SS) contract	GBP 500,000
Three Month US Dollar (ED) contract	USD 1,000,000

Table 2: STIR nominal values.

This account concentrates on the case where a market sale trades with some of the assembled bids in the order book. The case of a market purchase trading with asks, is exactly similar. The market sale specifies a quantity to be sold. Typically this quantity is exceeded by the total quantity currently bid. So the issue is as follows: how should the market sale be allocated?

The order matching algorithm applies commonly-known rules to determine this. Possibly universally, it imposes the rule that bids can only trade with a market order when all higher-priced bids have traded. This is known as price priority. (Equally, asks only trade when all lower-priced asks have traded.) The remaining issue concerns bids at the same price, and the sequence in which they trade. This is where the pro-rata algorithm differs from the price-time priority algorithm.

In broad terms, a market order is allocated among equally-priced bids in proportion to the size of those bids. However, because trades can only occur in integer numbers of lots, there are details around rounding (points 4 and 5 below). An exception is also made for price-improving bids (point 1).

1. If the bid which raised the best bid to its current level is still unexecuted, it executes first. However, there are maximum and minimum limits to its size.

2. The market sale's remaining volume, if any, is allocated to the remaining bids in proportion to their size. Fractions of a lot a rounded down, unless they are below 1, in which case they are rounded up to 1.

3. If this results in exactly the right number of lots being traded, the algorithm ends.

4. If this results in fewer lots being allocated than the market order specifies, the remainder lots are allocated to the largest bid. Time priority is enforced in the event a tie-break, where the largest orders are equally sized.

5. If this results in more lots being allocated than the market order specifies, lots are withheld from the smallest orders. Time priority is enforced in the event a tie-break, where small orders are equally sized (e.g. size 1).

To clarify this further, I work through some examples. In all the examples: The bid which raised the best bid to its current level has already executed. Consequently, it is out of the picture. The total offered volume is 80 lots. A market order arrives, requiring 8 lots.

Example 1 : Two bids at the same price. Bid A : 50 lots. Bid B : 30 lots.

A market order for 8 lots arrives. This is fully executed with A and B. Bid A gets a quantity of $50/80 * 8 = 5$ lots. Likewise, Bid B gets a quantity of 3 lots.

Example 2 : Two bids at the same price. Bid A : 51 lots. Bid B : 29 lots.

A market order for 8 lots arrives. Bid A gets a quantity of $51/80 * 8 = 5.1$ lots. Likewise, Bid B gets a quantity of 2.9 lots. However, only whole numbers of lots can be traded. Point 2 applies. So Bid A gets 5 lots, and Bid B gets 2 lots. However, only 7 of the 8 lots ordered have now been traded. Point 4 applies. So, A trades 6 lots.

Example 3 : Four bids at the same price. Bid A : 51 lots. Bid B : 27 lots. Bid C : 1 lot. Bid D : 1 lot, and D arrived before C.

A market order for 8 lots arrives. Bid A is allocated a quantity of $51/80 * 8 = 5.1$ lots. Likewise, Bid B gets a quantity of 2.7 lots. Bids C and D are allocated 0.1 lots. Only whole numbers of lots can get traded. Point 2a applies. So Bid A gets allocated 5 lots, and Bid B gets 2 lots. Finally, Bids C and D get 1 lot. However, 9 lots have been allocated, but only 8 lots are required by the market sale. Point 5 applies. So a lot is withheld from the smaller of C and D. But, C and D are of the same size, so the lot is withheld from C.

C Proof of Proposition 4.3

When $M > D$, so that there is a very large trade, $(f_i - S^*)\frac{\partial f_i}{\partial Q_i}$ is equal to $(Q_i - S^*)$. This is **invariant** to Q_{-i}. On the other hand, when $M \leq D$ matters are more complicated:

$$(f_i - S^*)\frac{\partial f_i}{\partial Q_i} = \left(\frac{Q_i M}{D} - S^*\right)\frac{Q_{-i}M}{D^2}. \tag{10}$$

Note that this quantity is **decreasing** in Q_{-i}. Also note that when $M \lesssim D$ is very close to D but below, then $(f_i - S^*)\frac{\partial f_i}{\partial Q_i}$ is $(Q_i - S^*)\frac{Q_{-i}}{D}$, which is less than $(Q_i - S^*)$. Combining these three remarks,

$$\frac{\partial}{\partial Q_{-i}}E\left[2(f_i - S^*)\frac{\partial f_i}{\partial Q_i}\right] < 0. \tag{11}$$

Starting from a case where Q_i is a best response to Q_{-i}, a small increase in Q_{-i} implies that

$$E\left[2(f_i - S^*)\frac{\partial f_i}{\partial Q_i}\right] < 0, \tag{12}$$

or

$$\frac{\partial}{\partial Q_i}E\left[(f_i - S^*)^2\right] < 0, \tag{13}$$

so that to minimize loss, Q_i should be raised.

D Proof of Lemma 4.4

Starting from a case where Q_i is a best response to Q_{-i}, consider a small increase in Q_{-i} to rQ_{-i} (where $r > 1$) and a *proportionate* increase in Q_i to rQ_i. We will consider the change in the quantity

$$r\frac{\partial}{\partial Q_i}E\left[(f_i - S^*)^2\right] \quad \text{which may be written} \quad rE\left[2(f_i - S^*)\frac{\partial f_i}{\partial Q_i}\right], \tag{14}$$

which is zero when $r = 1$. The next part is directed towards proving that it becomes positive as r rises above unity.

From (10), we see that whenever $M < D$, we have that $2r(f_i - S^*)\frac{\partial f_i}{\partial Q_i}$ is invariant to r. However, when $M > D$, we have that $f_i = Q_i$ so that

$$2r(f_i - S^*)\frac{\partial f_i}{\partial Q_i} \tag{15}$$

rises from $2(Q_i - S^*)$ to $2r(rQ_i - S^*)$ as r rises above unity.

Also note that when $M \lesssim D$ is very close to D but below, then $2(f_i - S^*)\frac{\partial f_i}{\partial Q_i}$ is $2(Q_i - S^*)\frac{Q_{-i}}{D}$, which is less than $2(Q_i - S^*)$. So, the rate of change in (14) with a small increase in r is, when evaluated at $r = 1$,

$$2(2Q_i - S^*)Pr(M > D) - Df_M(D)2(Q_i - S^*)\left(1 - \frac{Q_{-i}}{D}\right) \tag{16}$$

which is

$$2(2Q_i - S^*)Pr(M > D) - 2Q_i(Q_i - S^*)f_M(D). \tag{17}$$

This is positive if

$$\frac{Pr(M > D)}{f_M(D)} > \frac{Q_i(Q_i - S^*)}{2Q_i - S^*} \tag{18}$$

We now prove that (18) is true, by using distributional properties of M. From Proposition 4.3, $Q_i > S^*$, so that

$$\frac{Q_i(Q_i - S^*)}{2Q_i - S^*} < \frac{Q_i^2}{Q_i} = Q_i < D. \tag{19}$$

So, the inequality follows if there exists $\eta > 0$ such that for any x (including $x = D$),

$$\frac{Pr(M > x)}{f_M(x)} > x(1 + \eta), \tag{20}$$

which follows if $Pr(M > x) = \frac{1}{x^\alpha}$, setting $\eta = \left(\frac{1}{\alpha} - 1\right)$.

Therefore, for r just greater than 1 we have that

$$\frac{\partial}{\partial Q_i} E\left[(f_i - S^*)^2\right] > 0, \tag{21}$$

and it would be better for player i to reduce Q_i. Thus $\hat{Q}(rQ_{-i}) < r\hat{Q}(Q_{-i})$ for $r \gtrsim 1$ so that:

$$\left.\frac{\partial}{\partial r}\left(\frac{\hat{Q}(rQ_{-i})}{rQ_{-i}}\right)\right|_{r=1} < 0. \tag{22}$$

This gives

$$\frac{\partial \hat{Q}}{\partial Q_{-i}} < \frac{\hat{Q}}{Q_{-i}}. \tag{23}$$

Because $\eta > 0$ we can further find a positive bounding value $\epsilon > 0$, such that the lemma is true.

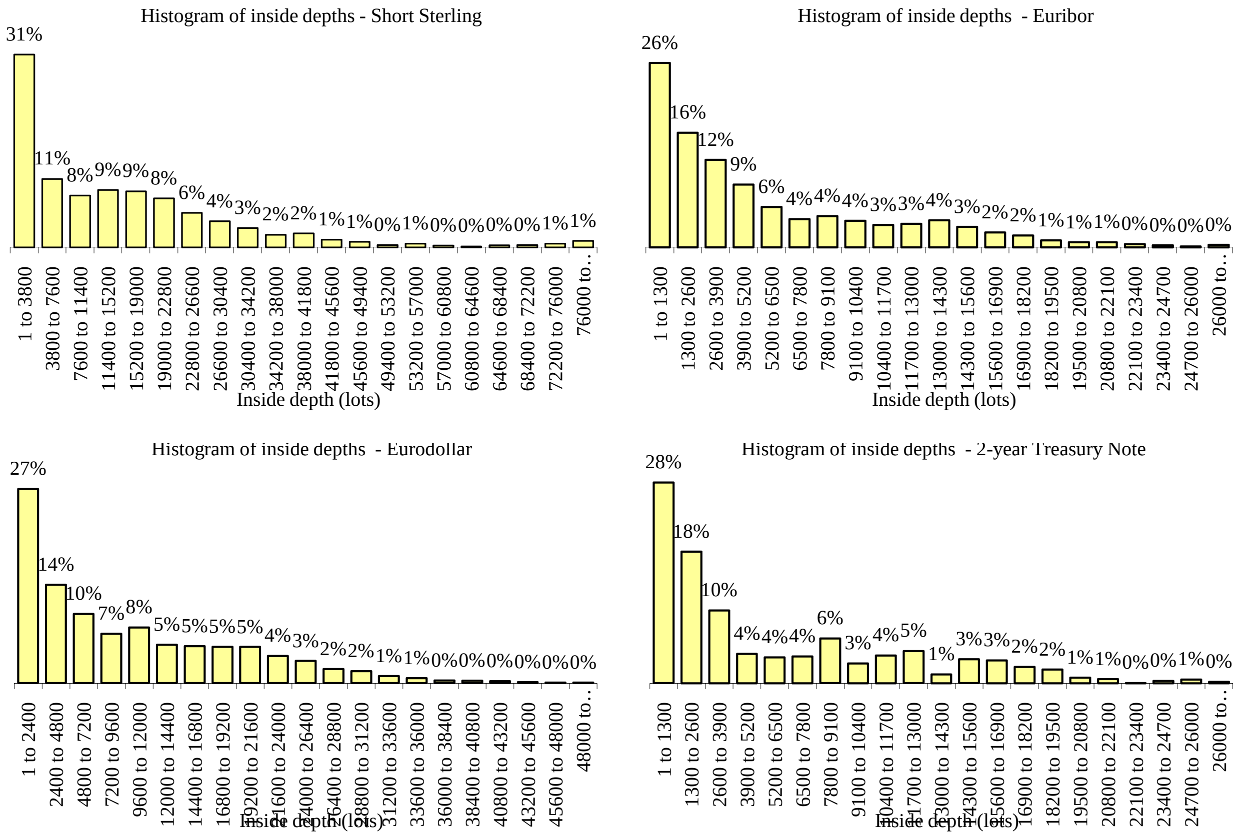

Figure 3: These histograms are derived by averaging the histogram of inside depths at the best bid, and the histogram of inside depths at the best ask. Except for the right-most bin in each histogram, bins are equally sized.

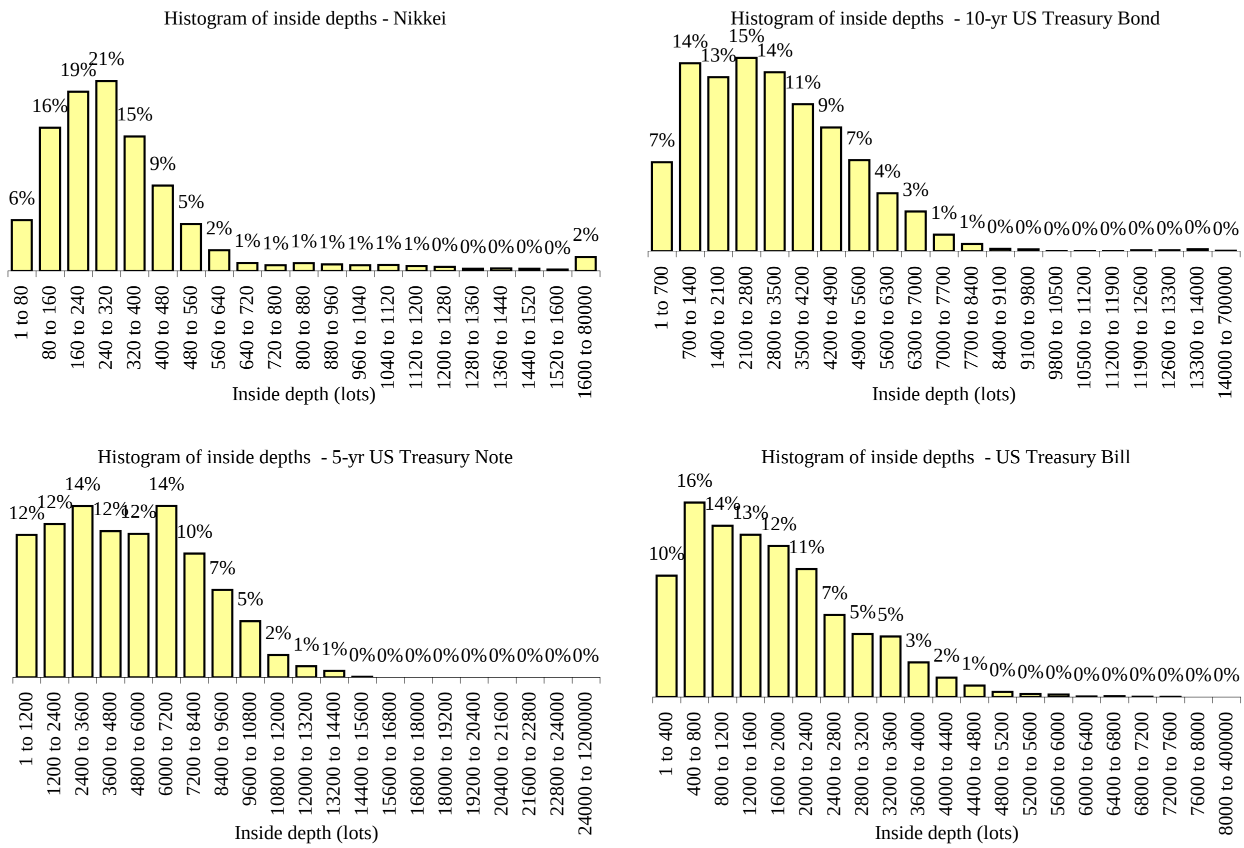

Figure 4: These histograms are derived by averaging the histogram of inside depths at the best bid, and the histogram of inside depths at the best ask. Except for the right-most bin in each histogram, bins are equally sized.