No. 2008/40

Pro-Rata Matching and One-Tick Futures Markets

Jonathan Field and Jeremy Large

Center for Financial Studies
Goethe-Universität Frankfurt ■ House of Finance
Grüneburgplatz 1 ■ 60323 Frankfurt ■ Deutschland

Telefon: +49 (0)69 798-30050
Fax: +49 (0)69 798-30077
http://www.ifk-cfs.de ■ E-Mail: info@ifk-cfs.de

Center for Financial Studies

The *Center for Financial Studies* is a nonprofit research organization, supported by an association of more than 120 banks, insurance companies, industrial corporations and public institutions. Established in 1968 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center´s Research Fellow Program or members of one of the Center´s Research Projects.

If you would like to know more about the *Center for Financial Studies*, please let us know of your interest.

Prof. Dr. Jan Pieter Krahnen

Prof. Volker Wieland, Ph.D.

Center for Financial Studies
House of Finance ■ Goethe-Universität
Grüneburgplatz 1 ■ 60323 Frankfurt am Main ■ Deutschland

Telefon: +49 (0)69 798-30050
Fax: +49 (0)69 798-30077
http://www.ifk-cfs.de ■ E-Mail: info@ifk-cfs.de

CFS Working Paper No. 2008/40

Pro-Rata Matching and One-Tick Futures Markets*

Jonathan Field[1] and Jeremy Large[2]

May 27, 2008

Abstract:
We find and describe four futures markets where the bid-ask spread is bid down to the fixed price tick size practically all the time, and which match counterparties using a pro-rata rule. These four markets' offered depths at the quotes on average exceed mean market order size by two orders of magnitude, and their order cancellation rates (the probability of any given offered lot being cancelled) are significantly over 96 per cent. We develop a simple theoretical model to ex- plain these facts, where strategic complementarities in the choice of limit order size cause 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.

* Please note that the views expressed herein are those of the authors and not necessarily of Man Investments. We are grateful to Anthony Ledford and Andrew Patton for their support and comments, as well as to seminar participants at the Oxford-Man Institute for Quantitative Finance (November 2007).

1 Man Investments, Sugar Quay, London, EC3R 6DU, U.K.; e-mail: JField@ManInvestments.com

2 corresponding author: Oxford-Man Institute of Quantitative Finance, and Man Investments; Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, OX1 4EH, U.K.; e-mail: Jeremy.Large@Oxford-Man.ox.ac.uk

1 Introduction

This paper finds and studies one-tick futures markets, where the price tick size is large and normally equal to the bid-ask spread. It begins by observing over a large majority (by value traded electronically) of all futures markets, the positive relationship noted in Harris (1994) and Goldstein and Kavajecz (2000), and predicted in Large (2007), between tick size and offered depths at the best quotes. It then concentrates on a subset of markets exhibiting this relationship in an extreme form, with very high depths at the best quotes, together with spreads which are scarcely ever wider than the price tick size.

The four most accentuated cases (in depth terms) turned out to be the leading short-term interest rate (STIR) futures markets in the US (Eurodollar), Europe (Euribor) and the UK (Short Sterling), together with the two-year US Treasury Note future. These four outlier markets had an interesting commonality: they all matched orders on a pro-rata basis, rather than on the more common price-time basis. We report unusual features of trading dynamics on these *one-tick pro-rata* markets, and develop a simple theoretical model to explain these.

In all four of our example *one-tick pro-rata* cases, the bid-ask spread is larger than the price tick size only infrequently: significantly less that 3 per cent of the time. 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, on all four markets 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).

To explain these findings we develop a theoretical model which predicts great inside depths and cancellation rates in equilibrium because of the pro-rata matching algorithm. The pro-rata algorithm matches limit orders to each countervailing market order in proportion (pro-rata) to their sizes; this 'rations' the market order out, as it is typically too small to trade with them all in their entirety, in a manner reminiscent of book-building on primary equity markets.[1]

[1]Section 3.3 gives a full account of the pro-rata rule.

In our model this creates strategic complementarities in the quantity decisions of liquidity suppliers.[2] Therefore, when faced with one another's limit orders, traders' equilibrium strategy is to submit bids and asks of greater size than the quantity that they wish to trade. As they over-offer, liquidity suppliers accept a risk of overtrading. We show that such overtrading risk is an important factor in forestalling an unending order-size 'arms race', where ever larger limit orders are submitted. Indeed, substantial overtrading risk ensures the existence of equilibrium. However, if such overtrading risk is relatively weak and/or if traders are numerous, then over-offering can be of large magnitude.

This helps to explain 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 on these markets.

So some of the effects on liquidity supply due to the costs of waiting to trade (see Foucault 1999, Foucault, Kadan, and Kandel 2005) or to execution uncertainty (see Cohen, Maier, Schwartz and Whitcomb 1981) appear to take novel forms in one-tick pro-rata markets, because of the large price tick size and the pro-rata matching rule. In particular, it seems that to reduce the cost of waiting, limit order submitters bear *over*-execution uncertainty.

[2]For a comprehensive review of strategic complementarities, see Cooper (1999), see also Cooper and John (1988) and Vives (1990).

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 (2007).

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 in some detail Section 3, in order to establish the empirical properties around cancellations, depths and spreads outlined in the Introduction. Section 4 presents a simple theoretical model to account for these findings. 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.

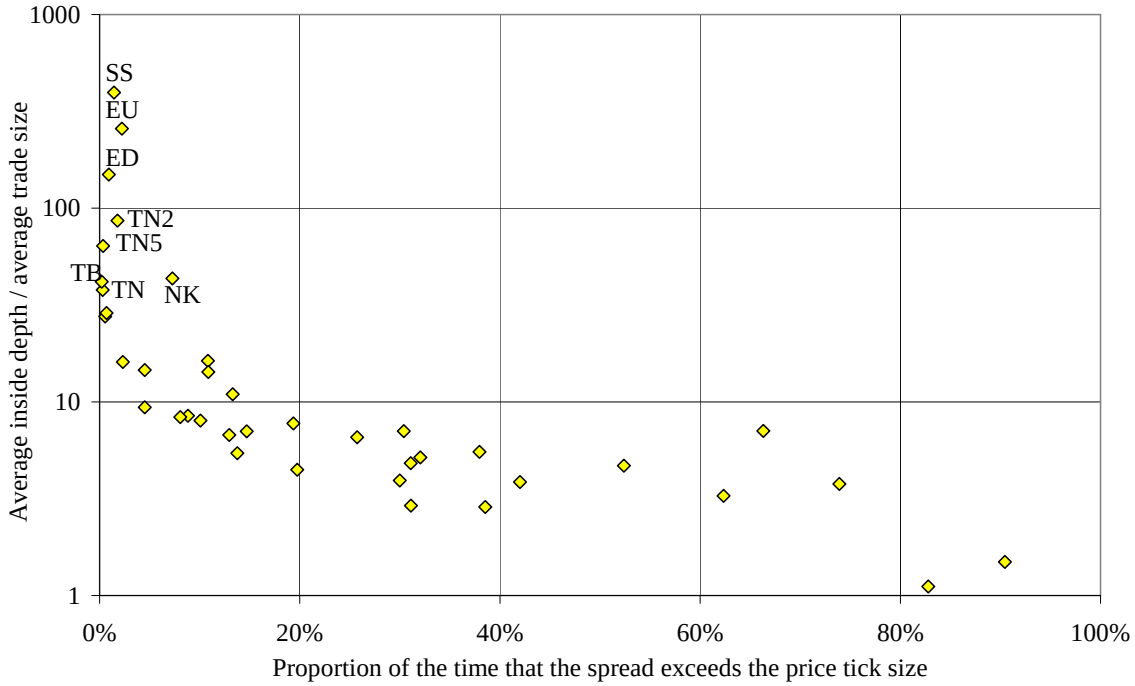

Figure 1: A chart plotting the 40 observed futures markets. Those markets of greatest interest to us are labeled. One observation fell below the x-axis, at coordinates (94%,0.7).

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.[3] The retained set of 40 markets comprises six currency futures, nine government bond futures, three short-term interest rate futures (STIRs), twenty 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.[4] 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 floor exchange. Table 1 documents our observation start and end times for eight markets of particular interest.

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

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

[4]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.

'average inside depth / average trade size'.[5] 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.

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 are distinctive because 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 Bill future (TB). These markets operate on a price-time basis. While they are not therefore in our main focus, we provide data on their microstructure as well. All eight mentioned markets had average depths at the best bid (ask) which exceed the average trade size by a factor of more than 30. They all experience on average more than 80 updates per minute.

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

	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 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, and Short Sterling, Euribor, Eurodollar and 2-Year US Treasury Note futures. 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.[6]

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').[7]

The figures 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.[8] 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

[6] 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.

[7] See Table 2.

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

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 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 and long upper tails, indicating high variability in offered depths.

3.2 The propensity of offered lots to be cancelled

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

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

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, while 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 which appeals to 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 rationalize using the pro-rata algorithm by arguing that it encourages liquidity supply on markets with low volatility, such as short-term interest

rate futures (STIRs).

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 simple theoretical model of equilibrium behaviour to explain the latter two of these stylized facts (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 this model will formalize. The model is specified in the next section, and then equilibrium behaviour 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 400 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 available counterparty limit orders at the best price 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 1/400 of its volume with each market order (and adverse rounding rules detailed in the Appendix may make that proportion even lower). By contrast, a limit order of much larger size may execute a more satisfactory quantity for its submitter.

Consequently, traders may have incentives to submit limit orders of larger quantity than their true trading intention, when market depths are large. This can create escalating feedback in offered depths. Furthermore, because traders are offering more than they intend to trade, they may have a higher propensity to cancel their limit orders once they have achieved their trading objectives.

To assess this line of reasoning, we now study equilibrium depths in a theoretical model.

4.2 Model specification

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 is also large, so these traders submit asks, rather than selling cheaply with a market sale. Furthermore the best ask, which is the upper limit of the bid-ask spread, is effectively fixed for the duration of the game.

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. Because it is clear and simple, we will consider a simultaneous game where all quantity decisions are taken at the same moment.

Then one market purchase arrives and trades with one or more of the submitted asks. Its size, M, is determined by nature as the realization of a random variable of 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 of that amount that actually executed (i.e. filled).

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^*. 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

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

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, $Q_{-i} = (n-1)Q_i$.

4.3 Model analysis and equilibrium characterization

If $Q_{-i} = 0$, then, from (4), $f_i = \min(M, Q_i)$. 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^*. Matters are however otherwise if competing traders are present:

Proposition 4.2 *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 4.2 formalizes the principle that when facing great depth at the ask (bid), an asker (bidder) has an incentive to *increase* the size of her limit order. Because of Lemma 4.1, this entails **over-offering**, by offering a larger limit order than the quantity, S^*, which is really desired. A typical best response function of this form is illustrated in Figure 2.

Definition 3 (substantial overtrading risk) *The market has **substantial overtrading risk** if for $x > 1$,*

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

with $\alpha \in (0, 1)$, so that M has a fat 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 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 4.4 uses to deduce equilibrium existence.

Lemma 4.3 *Suppose that the market has substantial overtrading risk. Then the elasticity of $\hat{Q}$ with respect to Q_{-i} is less than unity:*

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

Proof. See Appendix. ■

Proposition 4.4 *Suppose that the market has substantial overtrading risk. Then a unique equilibrium exists which is symmetric.*

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 4.2 and Lemma 4.3,

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

It follows from this that such intersection always exists and is unique. This point of intersection describes the unique and symmetric equilibrium. ■

Proposition 4.5 *As the number of players, n, increases, the equilibrium degree by which* each *player over-offers rises, as does the amount of overtrading.*

Proof. Note that the function $\hat{Q}(\cdot)$ is invariant to n. From Lemma 4.3 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, and risks a greater overtrade. ■

So strategic interaction among competing limit order suppliers may create escalating feedback in limit order size and depths, which can explain the very high offered depths that we observe. The next Proposition outlines a further consequence: large cancellation rates.

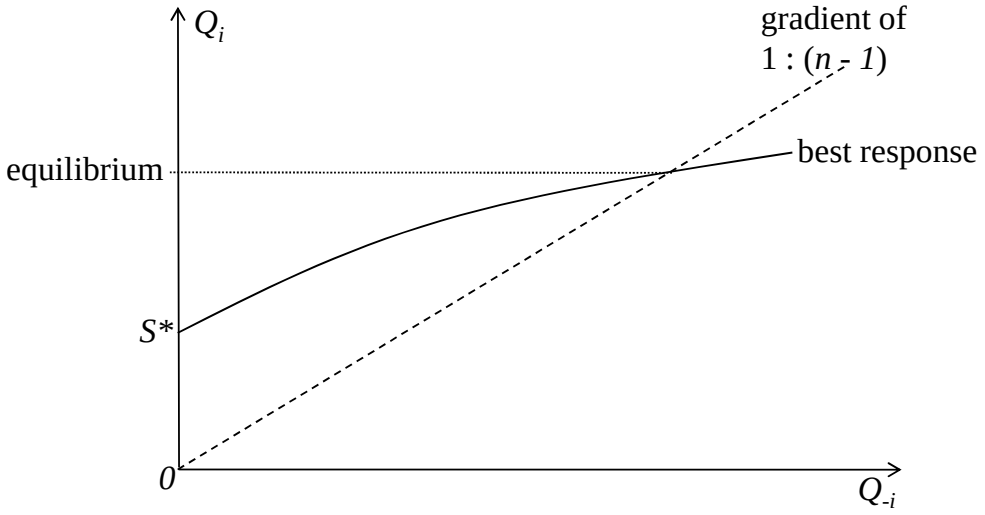

Figure 2: Illustration of equilibrium in the pro-rata market game.

Proposition 4.6 *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 4.3 $\left(Q_i - \frac{M}{n}, 0\right)$ 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.

4.4 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. This can be explicated in the following way: as Copeland and Galai (1983) notes, submitting a limit order gives a free trading option to other market participants. Whenever this order is of great size, then the option granted to the market is great. If we observe in the data that very large market orders

are placed, exercising this option, then offering such an option has a cost which can deter the submission of very large limit orders: bounding depth. It is therefore helpful to check that in our data we see evidence of overtrading risk.

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

Some important fixed income futures have a distinctive *one-tick pro-rata* market microstructure. Using rich tick-level data from a majority (by traded volume) of futures markets worldwide, we show that these markets are distinctive with regard to four features: a bid-ask spread which very seldom exceeds the price tick size; a pro-rata matching rule; very high offered depths at the quotes; and a high propensity to withdraw offers to trade, i.e. to cancel limit orders.

In a simple and intuitive theoretical model of a one-tick pro-rata market, high offered depths arise due to the submission of oversized limit orders, brought about by a strategic interaction among liquidity suppliers. 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. This helps to explain to the very high cancellation rates observed on our sample markets, where overtrading risk apparently constrains offered depths quite weakly, while nevertheless ensuring a finite equilibrium inside depth at the quotes.

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 be an important exercise for future research, and would be of interest to policymakers and the designers of exchanges.

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.

Copeland, T. and D. Galai (1983). Information effects on the bid-ask spread. *Journal of Finance 38*, 1457–1469.

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.

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.

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.

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

Large, J. (2007). A market-clearing role for inefficiency on a limit order book. University of Oxford, Nuffield College, Economics Group Working Paper.

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 4.2

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 4.3

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

$$rE\left[2(f_i - S^*)\frac{\partial f_i}{\partial Q_i}\right], \tag{14}$$

which is zero in equilibrium. 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 $M > D$,

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

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

$$2(2Q_i - S^*)Pr(M > D) - Df_M(D)2(Q_i - S^*)\left(1 - \frac{Q_{-i}}{D}\right) \tag{15}$$

which is

$$2(2Q_i - S^*)Pr(M > D) - 2Q_i(Q_i - S^*)f_M(D) \tag{16}$$

This is positive if

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

We now try to prove the central inequality, (19), given the distribution of M. From Proposition 4.2, $Q_i > S^*$, so that

$$\frac{Q_i(Q_i - S^*)}{2Q_i - S^*} < \frac{Q_i^2}{Q_i} = Q_i < D. \tag{18}$$

So, the inequality follows if for any x (including $x = D$),

$$\frac{Pr(M > x)}{f_M(x)} > x, \tag{19}$$

which is easily proven if $Pr(M > x) = \frac{1}{x^\alpha}$.

Therefore, when r rises,

$$\frac{\partial}{\partial Q_i} E\left[(f_i - S^*)^2\right] > 0, \tag{20}$$

and it would be better for player i to reduce Q_i. Thus $\hat{Q}(rQ_{-i}) < r\hat{Q}(Q_{-i})$ for small r and the lemma follows.

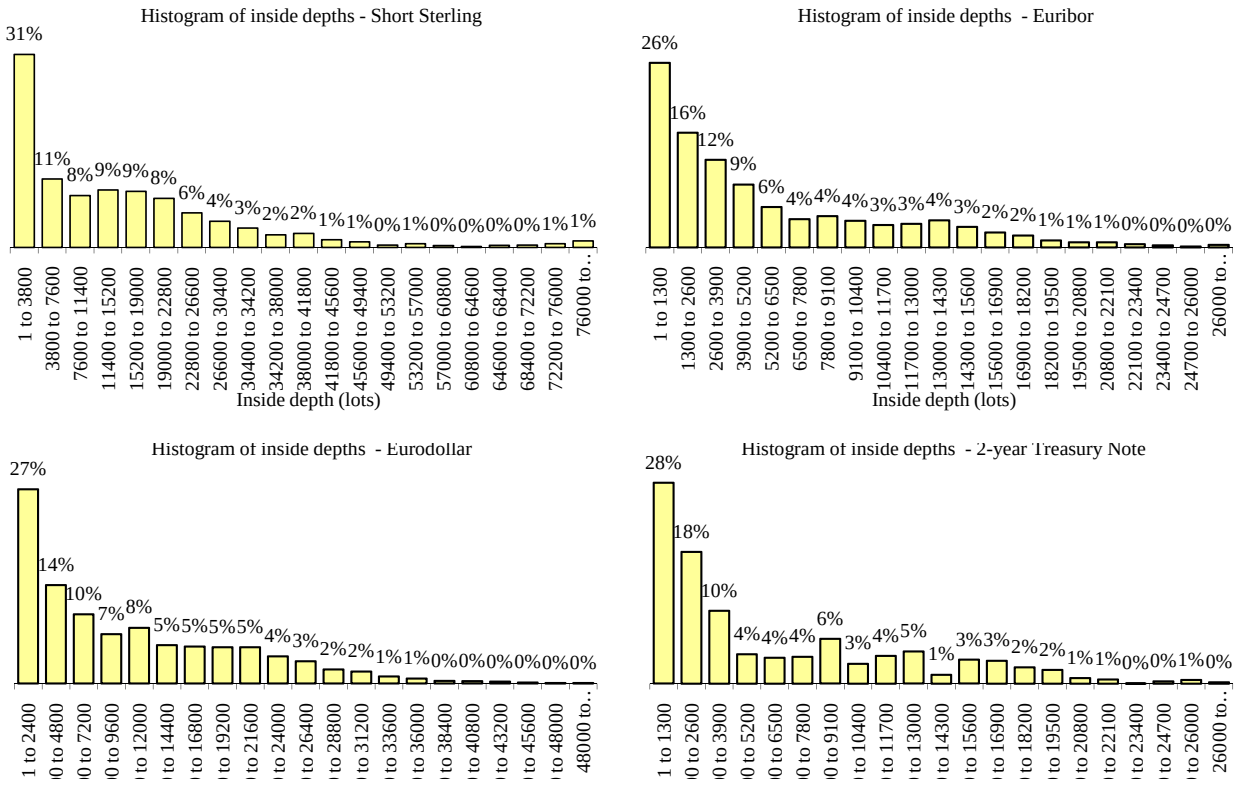

Figure 3: These histograms are derived by averaging the histogram of inside depths at the best bid, and the histogram of inside depths at the best ask. Except for the right-most bin in each histogram, bins are equally sized.

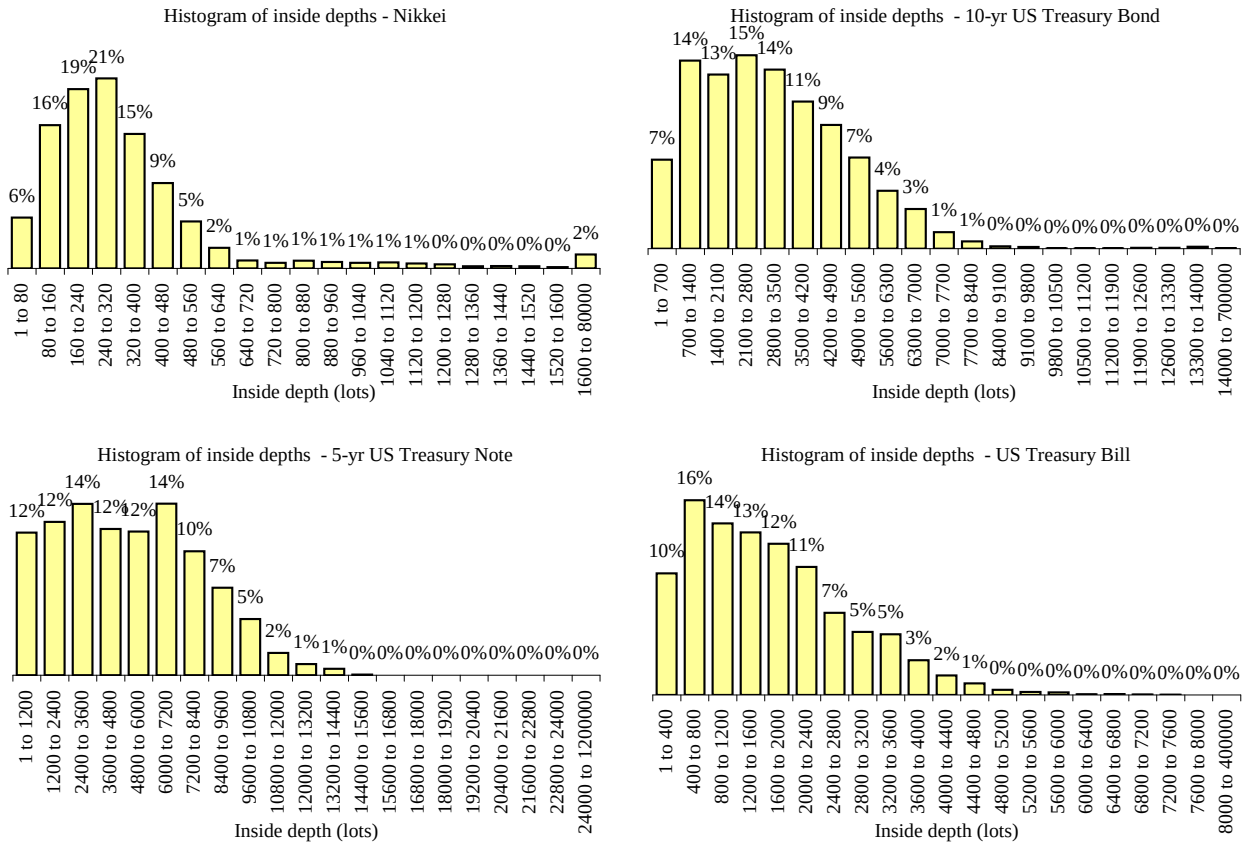

Figure 4: These histograms are derived by averaging the histogram of inside depths at the best bid, and the histogram of inside depths at the best ask. Except for the right-most bin in each histogram, bins are equally sized.

CFS Working Paper Series:

No.	Author(s)	Title
2008/39	**Giovanni Cespa** **Thierry Foucault**	Insiders-Outsiders, Transparency and the Value of the Ticker
2008/38	**Rachel J. Huang** **Alexander Muermann** **Larry Y. Tzeng**	Hidden Regret in Insurance Markets: Adverse and Advantageous Selection
2008/37	**Alexander Muermann** **Stephen H. Shore**	Monopoly Power Limits Hedging
2008/36	**Roman Inderst**	Retail Finance: Bringing in the Supply Side
2008/35	**Dimitris Christelis** **Tullio Jappelli** **Mario Padula**	Cognitive Abilities and Portfolio Choice
2008/34	**Tullio Jappelli** **Marco Pagano**	Information Sharing and Credit: Firm-Level Evidence from Transition Countries
2008/33	**Tullio Jappelli** **Marco Pagano**	Financial Market Integration under EMU
2008/32	**Fancis X. Diebold** **Georg H. Strasser**	On the Correlation Structure of Microstructure Noise in Theory and Practice
2008/31	**Günter Franke** **Jan Pieter Krahnen**	The Future of Securitization
2008/30	**Karin Assenmacher-Wesche** **Stefan Gerlach**	Financial Structure and the Impact of Monetary Policy on Asset Prices

Copies of working papers can be downloaded at http://www.ifk-cfs.de