Quasi-Centralized Limit Order Books

Martin D. Gould*†‡, Mason A. Porter§¶, and Sam D. Howison§‖

‡ CFM–Imperial Institute of Quantitative Finance, Imperial College, London SW7 2AZ
§ Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX2 6GG
¶ CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP
‖ Oxford–Man Institute of Quantitative Finance, University of Oxford, Oxford, OX2 6ED

February 4, 2015

Abstract

A quasi-centralized limit order book (QCLOB) is a limit order book (LOB) in which financial institutions can only access the trading opportunities offered by counterparties with whom they possess sufficient bilateral credit. We perform an empirical analysis of a recent, high-quality data set from a large electronic trading platform that utilizes QCLOBs to facilitate trade. We find many significant differences between our results and those widely reported for other LOBs. We also uncover a remarkable empirical universality: although the distributions describing order flow and market state vary considerably across days, a simple, linear rescaling causes them to collapse onto a single curve. Motivated by this finding, we propose a semi-parametric model of order flow and market state in a QCLOB on a single trading day. Our model provides similar performance to that of parametric curve-fitting techniques, while being simpler to compute and faster to implement.

Keywords: Limit order books; quasi-centralized liquidity; market microstructure; foreign exchange; curve collapse.

I. Introduction

More than half of the world's financial markets utilize electronic limit order books (LOBs) to facilitate trade [53]. In contrast to quote-driven systems, in

*Corresponding author. Email: m.gould@imperial.ac.uk.
†MDG also completed part of this work while at the University of Oxford.

which prices are set by designated market makers, trade in an LOB occurs via a continuous double-auction mechanism whereby institutions submit orders that state their desire to buy or sell a specified quantity of an asset at a specified price. Active orders reside in a queue until they are either cancelled by their owner or execute against an order of opposite type. Upon execution, the owners of the relevant orders trade the agreed quantity of the asset at the agreed price.

During the past 20 years, a large body of empirical and theoretical work has addressed a specific type of LOB, in which all institutions are able to trade with all others (see [30] for our recent review). We call this market organization a *centralized LOB*. Although several large platforms (including the London Stock Exchange (LSE) Electronic Trading Service [61], NASDAQ [48], and the Euronext Universal Trading Platform [23]) employ centralized LOBs, many others utilize alternative LOB configurations with different rules. In contrast to the wealth of publications on centralized LOBs, discussion of alternative LOB configurations has remained limited to a handful of technical descriptions of matching mechanisms on specific platforms [5, 26, 52, 55]. Given their widespread use, detailed study of alternative LOB mechanisms is an important and timely task.

A prominent example of an alternative LOB mechanism is the one utilized by multi-institution trading platforms in the foreign exchange (FX) spot market. On such platforms, institutions specify *credit limits* for their trading counterparties. Each institution can only access the trading opportunities offered by counterparties with whom they possess sufficient bilateral credit. We call this market organization a *quasi-centralized limit order book (QCLOB)* because different institutions have access to different subsets of a centralized liquidity pool. Examples of platforms that utilize QCLOBs include Reuters [62], EBS [22], and Hotspot FX [40], which together facilitate a mean turnover in excess of US $0.5 trillion each day.[1]

Despite this enormous volume of trade, a lack of adequate data has hindered discussion of many important questions regarding QCLOBs. Do the statistical properties of QCLOBs differ from those of centralized LOBs? Do arbitrage opportunities arise? How do institutions assess market state when deciding how to act? In this paper, we present an empirical study of a recent, high-quality data set from Hotspot FX, which enables us to address these issues.

Our results highlight many important differences between centralized LOBs and QCLOBs. For example, we observe much lower levels of order flow at the prevailing quotes and a much higher ratio of active liquidity to market order flow than has been reported by studies of centralized LOBs. We also identify sustained periods during which bid-ask spreads are negative. Due to the extremely high levels of market activity on Hotspot FX, we are able to perform both cross-sectional (i.e., between different currency pairs) and longitu-

[1] According to the 2013 Triennial Central Bank Survey [4], the mean daily turnover of the global FX spot market exceeds US $2 trillion, which surpasses the mean daily turnover of the New York Stock Exchange (NYSE) by a factor of more than 50 [49] and outstrips the daily global gross domestic product by a factor of more than 5 [25]. The market consists of several constituent parts, including the spot, forwards, options, and swaps markets. The spot market accounts for approximately 38% of the market's total volume.

dinal (i.e., across different time periods) comparisons of our findings. We find several longitudinal differences in market activity, and thereby argue that using long-run statistical averages to formulate short-run forecasts may produce misleading results. We also uncover a remarkable empirical universality: applying a simple, linear rescaling to the distributions describing order flow and market state causes the data to collapse onto a single curve. Motivated by this finding, we propose a semi-parametric model of these distributions that provides similar performance to parametric curve-fitting techniques, while being simpler to compute and faster to implement.

Our findings are important for several reasons. First, they present a detailed overview of recent trading activity on a large electronic trading platform. Second, they illustrate similarities and differences between market activity on different trading days. Third, they highlight how the statistical properties of QCLOBs differ from those of centralized LOBs. Fourth, they motivate a semi-parametric model for the distributions describing order flow and market state in a QCLOB. Together, these results help to illuminate the delicate interplay between order flow, liquidity, and price formation for a widely used but hitherto unexplored market organization.

The remainder of this paper is organized as follows. In Section II, we provide a detailed description of centralized LOBs and QCLOBs, highlight the important differences between these mechanisms, and present several definitions that we use throughout the paper. In Section III, we describe the data that forms the basis for our empirical study and discuss the Hotspot FX platform. In Section IV, we describe the methodology that we adopt for our empirical study. We present our main results in Section V. In Section VI, we discuss our results and compare our findings to those of several empirical studies of centralized LOBs. We conclude in Section VII. In Appendix A, we provide a detailed description of the structure and format of the Hotspot FX data. In Appendix B, we describe our method of performing parametric fits to the daily data. In Appendix C, we describe how we quantify the strength of curve collapse when rescaling each day's data.

II. Centralized and Quasi-Centralized Limit Order Books

Let $\Theta = \{\theta_1, \theta_2, \ldots\}$ denote the set of institutions that trade a given asset on a given platform. In an LOB, these institutions interact by submitting orders. An *order* $x = (p_x, \omega_x, t_x)$ submitted at time t_x with price p_x and size $\omega_x > 0$ (respectively, $\omega_x < 0$) is a commitment by its owner to sell (respectively, buy) up to $|\omega_x|$ units of the asset at a price no less than (respectively, no greater than) p_x.

Whenever an institution submits a buy (respectively, sell) order x, an LOB's trade-matching algorithm checks whether it is possible for x to *match to* an active sell (respectively, buy) order y such that $p_y \leq p_x$ (respectively, $p_y \geq p_x$).

If so, the matching occurs immediately and the owners of the relevant orders agree a trade for the specified amount at the specified price. If $|\omega_x| > |\omega_y|$, any residue size of x is then considered for matching to other active sell (respectively, buy) orders, until either x becomes fully matched or there are no further active sell (respectively, buy) orders eligible for matching to x. Any portion of x that does not match becomes *active* at the price p_x, and it remains active until it either matches to an incoming sell (respectively, buy) order or is *cancelled*.

Orders that result in an immediate matching upon arrival are called *market orders*. Orders that do not — instead becoming active orders — are called *limit orders*. Some platforms allow other order types (such as fill-or-kill, stop-loss, or peg orders [41]), but it is always possible to decompose the resulting order flow into limit and/or market orders. Therefore, we study LOBs in terms of these simple building blocks.

The *global*[2] *LOB* $\mathcal{L}(t)$ is the set of all active orders for the given asset on the given platform at time t. The *global bid price* $b(t)$ is the highest price among active buy orders in $\mathcal{L}(t)$. The *global ask price* $a(t)$ is the lowest price among active sell orders in $\mathcal{L}(t)$. The *global bid-ask spread* is $s(t) = a(t) - b(t)$. The *global mid price* is $m(t) = [b(t) + a(t)]/2$.

A. *Centralized LOBs*

In a *centralized LOB*, all institutions face the same trading opportunities because all institutions can trade with all others. Whenever an institution θ_i submits a buy (respectively, sell) market order, the order matches to the highest-priority active sell (respectively, buy) order that is owned by another institution $\theta_j \neq \theta_i$, irrespective of the identities of θ_i and θ_j. For a detailed discussion of centralized LOBs, see [30].

B. *Quasi-Centralized LOBs*

In a QCLOB, each institution θ_i privately assigns a *credit limit* $c_{(i,j)} \geq 0$ to each other institution $\theta_j \in \Theta$. Institutions can only access the trading opportunities offered by counterparties with whom they possess sufficient bilateral credit. More precisely, if the total value v of trades agreed between θ_i and θ_j but not yet completed[3] satisfies both $v < c_{(i,j)}$ and $v < c_{(j,i)}$, then we say that θ_i and θ_j are *trading partners* and we write $\theta_i \leftrightarrow \theta_j$. Otherwise, we write $\theta_i \not\leftrightarrow \theta_j$. An institution θ_i can ensure that $\theta_i \not\leftrightarrow \theta_j$ (i.e., ensure that they never trade with θ_j) by setting $c_{(i,j)} = 0$.

When accessing a QCLOB trading platform, institutions cannot see the state of the global LOB $\mathcal{L}(t)$. Instead, each institution sees only the active orders owned its trading partners. This yields local versions of several key concepts (see Figure 1). Institution θ_i's *local LOB* $\mathcal{L}_i(t)$ is the subset of active orders in $\mathcal{L}(t)$ that are owned by θ_i's trading partners. Institution θ_i's *local bid price*

[2] We use the term *global* to highlight the differences between these definitions and the local definitions in Section II.B.

[3] In the FX spot market, trades agreed on day d are completed on day $d + 2$.

$b_i(t)$ is the highest stated price among active buy orders in $\mathcal{L}_i(t)$. Institution θ_i's *local ask price* $a_i(t)$ is the lowest stated price among active sell orders in $\mathcal{L}_i(t)$. Institution θ_i's *local bid-ask spread* is $s_i(t) = a_i(t) - b_i(t)$. Institution θ_i's *local mid price* is $m_i(t) = [b_i(t) + a_i(t)]/2$. When an institution θ_i submits a buy (respectively, sell) market order, the order matches to the highest-priority active sell (respectively, buy) order in $\mathcal{L}_i(t)$ (i.e., the highest-priority active sell (respectively, buy) order owned by an institution θ_j such that $\theta_i \leftrightarrow \theta_j$).

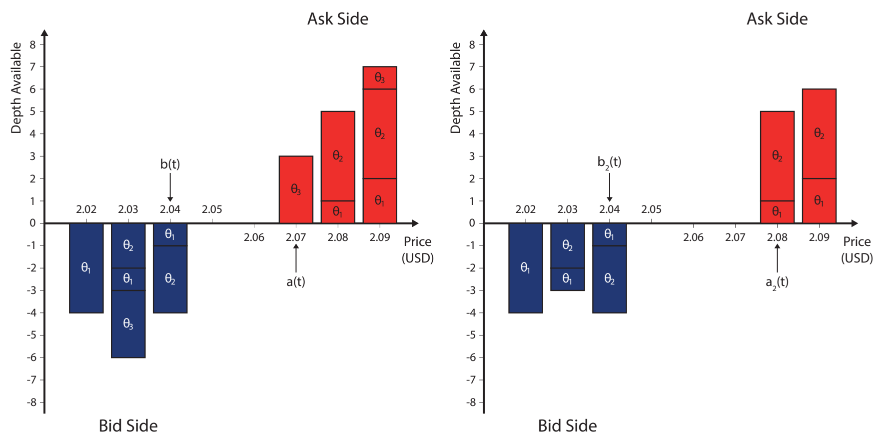

Figure 1: Schematic of (left) a global LOB $\mathcal{L}(t)$ and (right) institution θ_2's corresponding local LOB $\mathcal{L}_2(t)$, for a QCLOB populated by institutions $\Theta = \{\theta_1, \theta_2, \theta_3\}$ with $\theta_1 \leftrightarrow \theta_2$ and $\theta_1 \leftrightarrow \theta_3$ but $\theta_2 \not\leftrightarrow \theta_3$. To illustrate the role of trading partnerships, we label each active order in the figure according to its owner, but trading platforms do not disseminate this information.

C. *The Trade-Data Stream*

In addition to their local LOB $\mathcal{L}_i(t)$, each institution in a QCLOB can access a *trade-data stream* that lists the price, time, and direction[4] of every trade that occurs. All institutions can see all entries in the trade-data stream in real time, irrespective of whether they are trading partners with the institutions involved in a given trade. Therefore, although institutions in a QCLOB do not have access to information regarding which trading opportunities are available to others, they have access to a detailed historical record of previous trades.

[4]The *direction* of a trade describes whether it is buyer-initiated or seller-initiated. A *buyer-initiated* (respectively, *seller-initiated*) trade occurs when a buy market order matches to an active sell order (respectively, sell market order matches to an active buy order).

III. The Hotspot FX Data and Platform

We have been granted access to a recent, high-quality data set from Hotspot FX [40, 41, 42], which is one of the largest multi-institution trading platforms in the FX spot market [50]. The platform serves a broad range of trading professionals, such as banks, financial institutions, hedge funds, high-frequency traders, corporations, and commodity trading advisers [40].

The data describes all limit order arrivals, cancellations, and trades between the hours of 08:00:00–17:00:00 GMT for the EUR/USD (Euro/US dollar), GBP/USD (Pounds sterling/US dollar), and EUR/GBP (Euro/Pounds sterling) currency pairs[5] on 30 trading days during May–June 2010. According to the 2013 Triennial Central Bank Survey [4], global trade for EUR/USD, GBP/USD, and EUR/GBP constitutes about 24%, 9%, and 2%, respectively, of the FX market's total turnover We provide a detailed description of the Hotspot FX data in Appendix A.

Trade for each currency pair occurs in a separate QCLOB with price-time priority — that is, priority is first given to the active orders with the best (i.e., highest buy or lowest sell) price, and ties are broken by selecting the active order with the earliest submission time t_x. For each of the three currency pairs, the minimum order size is 0.01 units of the base currency and the tick size[6] is 0.00001 units of the counter currency [41].

IV. Methodology

A. Time Scales

We perform all of our calculations in *event time,* whereby we advance the clock by 1 unit whenever a limit order arrives. Measuring time in this way helps to remove the nonstationarities that occur in calendar time due to irregular bursts of trading activity [14, 31, 45, 58, 63]. The number of market order arrivals and active order cancellations varies in each time unit. We reset the clock at the start of each trading day so that the first limit order arriving after 08:00:00 GMT has $t_x = 1$.

B. Coordinate Frames

Most existing studies of LOBs use a coordinate frame that we call *quote-relative coordinates,* in which prices are measured relative to $b(t)$ and $a(t)$ (see, e.g., [7, 10, 14, 15, 18, 32, 33, 46, 51, 67]). The *quote-relative price of an order x at time t* is

$$\phi(p_x, t) := \begin{cases} b(t) - p_x, & \text{if } x \text{ is a buy order,} \\ p_x - a(t), & \text{if } x \text{ is a sell order.} \end{cases} \tag{1}$$

[5] A price for the currency pair XXX/YYY denotes how many units of the *counter currency* YYY are exchanged per unit of the *base currency* XXX.

[6] The *tick size* is the smallest permissible price interval between different orders. All orders must arrive with a price specified to the accuracy of the tick size.

In a centralized LOB, the use of quote-relative coordinates is motivated by the notion that institutions monitor $b(t)$ and $a(t)$ when deciding how to act. In a QCLOB, by contrast, institutions do not know the values of $b(t)$ and $a(t)$ (see Section II.C). It is, therefore, unclear whether quote-relative coordinates provide a useful framework for studying QCLOBs.

Recall from Section II.C that institutions trading in a QCLOB have access to a trade-data stream that lists the prices of all previous trades. As an alternative to quote-relative coordinates, we introduce a coordinate frame that we call *trade-relative coordinates*, in which we measure prices relative to those of the most recent trades. In contrast to quote-relative prices, all institutions are able to calculate trade-relative prices in real time. Let $B(t)$ and $A(t)$ denote, respectively, the price of the most recent seller-initiated and buyer-initiated trades that occurred at or before time t. The *trade-relative price of an order x at time t* is

$$\Phi(p_x, t) := \begin{cases} B(t) - p_x, & \text{if } x \text{ is a buy order,} \\ p_x - A(t), & \text{if } x \text{ is a sell order.} \end{cases} \tag{2}$$

C. *Trading Days*

Due to the extremely high levels of activity on Hotspot FX, we are able to estimate several properties of order flow and LOB state on each trading day separately. By contrast, most existing empirical studies of LOBs aggregate market activity from multiple trading days or multiple different assets to obtain sufficiently many data points to perform statistically stable estimation [7, 10, 14, 18, 24, 32, 33, 46, 51, 67].

We choose a single trading day as our longitudinal unit for three reasons. First, a single trading day represents a structural cycle on Hotspot FX because the platform automatically cancels all active orders at the end of each day [41]. Second, a single trading day provides a compromise between including enough data points to ensure statistical stability and including enough units to perform useful longitudinal comparisons. Third, several empirical studies have reported that most institutions implement their investment decisions and trading strategies over a single trading day [2, 8, 54]. To such institutions, statistics that describe market behaviour over this time horizon are likely to be the most useful.

V. Results

A. *LOB Activity*

Table I lists summary statistics describing aggregate LOB activity across all 30 trading days. In terms of both limit order and market order arrivals, EUR/USD is the most active and EUR/GBP is the least active of the three currency pairs. Similarly, the mean total size of active orders in $\mathcal{L}(t)$ is largest for EUR/USD and smallest for EUR/GBP. In all cases, the aggregate level of activity for buy orders is approximately equal to the corresponding level of activity for sell orders.

Limit order arrivals outstrip market order arrivals by approximately 3 orders of magnitude for EUR/USD and GBP/USD and by almost 4 orders of magnitude for EUR/GBP. In all cases, market orders constitute less than 0.05% of the total arriving order flow, which indicates that the vast majority of limit orders end in cancellation rather than matching.

For each of the three currency pairs, the mean size of arriving limit orders is more than double the corresponding mean size of arriving market orders. The modal size of all order arrivals and departures is exactly 1 million units of the base currency, which indicates that many institutions prefer to use large, round-number volumes for their orders, even though the minimum order size on Hotspot FX is just 0.01 units of the base currency (see Section III). Only a small percentage of market orders match at more than one price (this behaviour is sometimes called "walking up the book").

Each of the three currency pairs undergoes periods during which the global bid-ask spread is negative. Therefore, we are able to deduce that not all institutions are trading partners with all others. Otherwise, any arriving buy order x with $p_x \geq a(t)$ (respectively, sell order x with $p_x \leq b(t)$) would match and would not become active, so the spread would never become negative.

B. Daily Activity Levels

Figure 2 shows the total size of arriving limit orders and market orders on each of the 30 days in our sample. For each of the three currency pairs each day, the difference between the total size of buy and sell limit orders is below 1.5% of the total size of limit order arrivals on that day. By contrast, the corresponding statistic for market orders frequently exceeds 15%. Therefore, it is common to observe a substantial imbalance between buy and sell market order volumes on a single day.

Although aggregate market activity levels vary considerably across trading days, particularly active or particularly quiet days tend to coincide for each of the three currency pairs. This suggests that common, exogenous factors play an important role in institutions' trading decisions. During May 2010, the European Central Bank announced and implemented a series of measures to combat financial instability within the Eurozone [60], such as providing loans to countries in financial difficulties, recapitalizing financial institutions, and purchasing bonds from member states [59]. The large changes in daily aggregate activity levels during this period (compared to the relatively calm period of June 2010) suggest that the implementation of such measures and the uncertainty surrounding their announcements strongly influenced aggregate activity levels in the FX spot market.

		EUR/USD		GBP/USD		EUR/GBP	
		Buy	Sell	Buy	Sell	Buy	Sell
Total Size (units of base currency $\times 10^9$)	Limit Orders	150430	150707	117928	118006	92379	92218
	Market Orders	67	71	23	23	5	7
	Cancellations	150339	150620	117896	117973	92370	92210
Total Number (orders $\times 10^3$)	Limit Orders	67981	68027	65410	65678	44085	43897
	Market Orders	83	85	43	44	7	8
	Cancellations	67878	67927	65359	65628	44076	43888
Modal Size (units of base currency $\times 10^6$)	Limit Orders	1.00	1.00	1.00	1.00	1.00	1.00
	Market Orders	1.00	1.00	1.00	1.00	1.00	1.00
	Cancellations	1.00	1.00	1.00	1.00	1.00	1.00
Mean Size (units of base currency $\times 10^6$)	Limit Orders	2.21	2.22	1.80	1.80	2.10	2.10
	Market Orders	0.803	0.833	0.539	0.509	0.761	0.791
	Cancellations	2.21	2.22	1.80	1.80	2.10	2.10
Percentage of Market Orders that Match at Several Different Prices		8.27%	8.54%	6.70%	5.90%	3.93%	4.52%
Mean Total Size of Active Orders (units of base currency $\times 10^6$)		283	296	164	166	94	95
Mean Bid-Ask Spread (ticks)		3.62		9.54		10.1	
Percentage of Time for which the Bid-Ask Spread is Negative		9.99%		4.08%		0.23%	

Table I: Summary statistics for aggregate activity on all 30 trading days that we study.

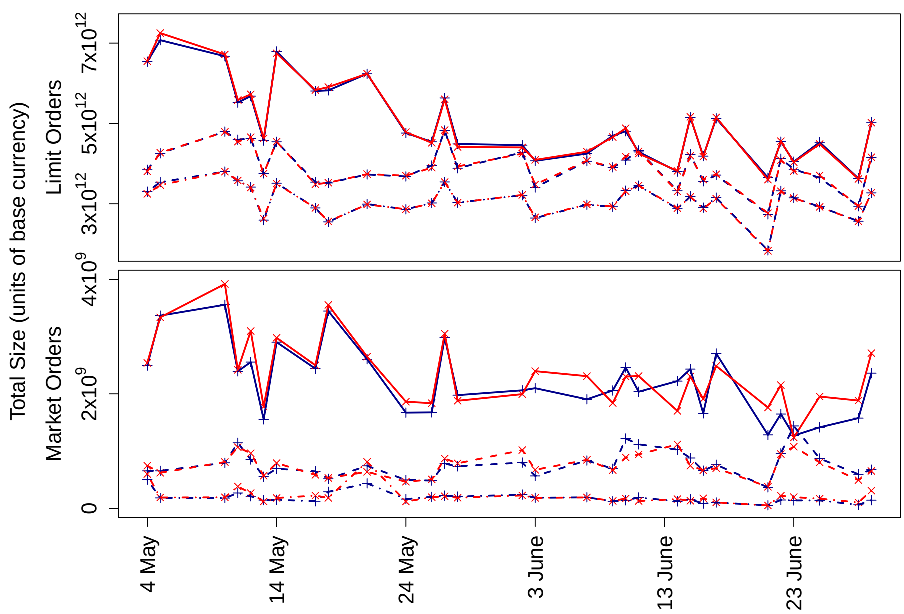

Figure 2: Total size of arriving (top panel) limit orders and (bottom panel) market orders for (solid curves) EUR/USD, (dashed curves) GBP/USD, and (dotted-dashed curves) EUR/GBP. The blue curves denote buy orders and the red curves denote sell orders.

C. Activity on a Single Trading Day

In this section, we calculate several statistical properties for a single trading day. We arbitrarily choose to present the results for 4 May 2010, which is the first day in our sample. In Section V.D, we investigate how these statistical properties vary across trading days.

Figure 3 shows the quote-relative and trade-relative price distributions of limit order arrivals on 4 May 2010. For each of the three currency pairs, and in both quote-relative and trade-relative coordinates, the maximum limit order arrival rate occurs at a strictly positive relative price. Limit orders for EUR/USD and EUR/GBP tend to arrive closer to the best quotes than do those for GBP/USD, and limit orders for EUR/USD tend to arrive closer to the most recently traded price than do those for GBP/USD and EUR/GBP. Limit orders tend to arrive more often at trade-relative prices that are integer multiples of 10 than they do at neighbouring prices.

In their study of the LSE, [46] used a generalized t distribution to model the distributions of quote-relative prices of arriving orders. We also find that this distribution provides a good fit to limit order arrivals on Hotspot FX, in both quote-relative and trade-relative coordinates. We describe our method of fitting the distribution in Appendix B. Several other parametric distributions with more than four parameters (most notably, the five-parameter logistic distribution [29]) also fit the data well, but the inclusion of additional parameters increases the computational complexity of the optimization required for fitting.

Figure 4 shows our fit of the generalized t distribution to the quote-relative and trade-relative distributions of limit order arrivals for EUR/USD. The results for the other currency pairs are qualitatively similar. In quote-relative coordinates, the fits perform well over the majority of the domain, although they fail to capture the strong kurtosis of the data and therefore perform less well in the upper tail. In trade-relative coordinates, the fits performs very well over the whole domain.

Figure 5 shows the quote-relative and trade-relative distributions of cancellations for each of the three currency pairs. In contrast to limit order arrivals, cancellations can only occur at non-negative quote-relative prices, because the lowest possible quote-relative price of an active order is 0 (which occurs for orders at $b(t)$ or $a(t)$). Each of the three currency pairs' quote-relative cancellation distributions have a local maximum at 0, but cancellations for EUR/USD tend to occur closer to the most recently traded price than do those for the other two currency pairs. For strictly positive quote-relative prices, the cancellation distributions have qualitatively similar shapes to the corresponding distributions for limit order arrivals. In trade-relative coordinates, the cancellation distributions are extremely similar to the limit order arrival distributions at all prices.

In trade-relative coordinates, we again find that a generalized t distribution provides a good fit to the distribution of active order cancellations (see Figure 6 for EUR/USD; the results for the other currency pairs are similar). In quote-relative coordinates, the local maximum in active order cancellations at a quote-relative price of 0 hinders this approach because the shape of the generalized t

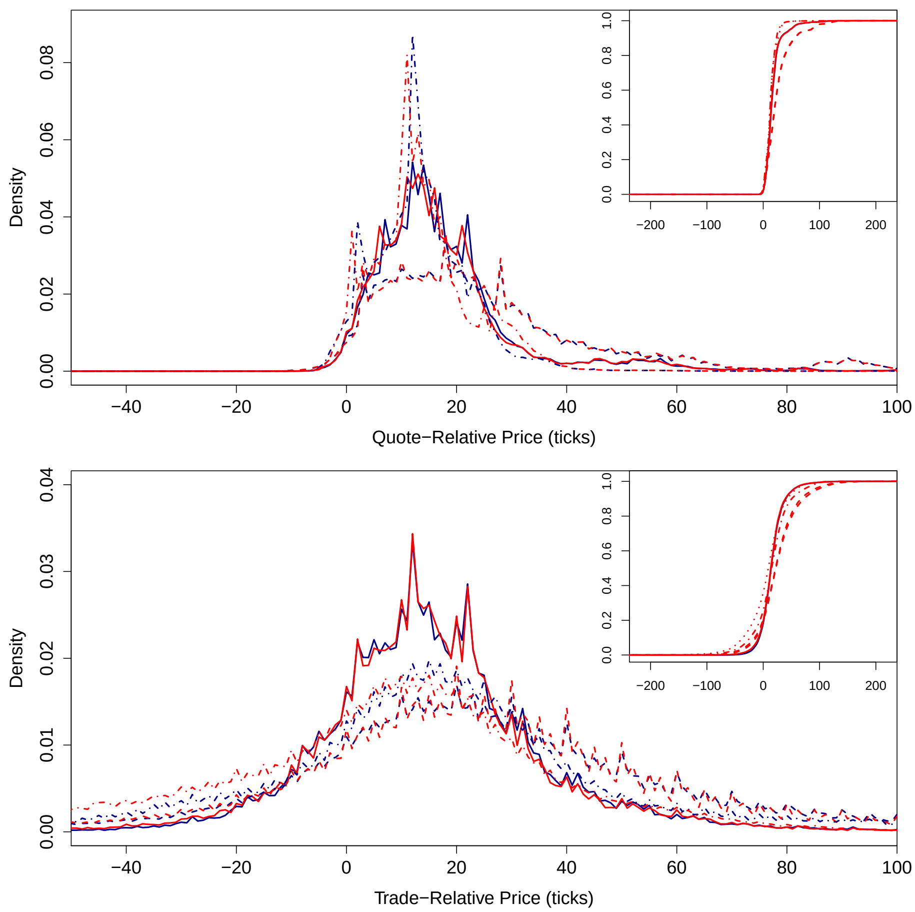

Figure 3: Distributions of limit order arrivals for (solid curves) EUR/USD, (dashed curves) GBP/USD, and (dotted-dashed curves) EUR/GBP on 4 May 2010 in (top) quote-relative and (bottom) trade-relative coordinates. The blue curves denote buy orders and the red curves denote sell orders. The main plots show the empirical density functions and the inset plots show the empirical cumulative density functions (ECDFs).

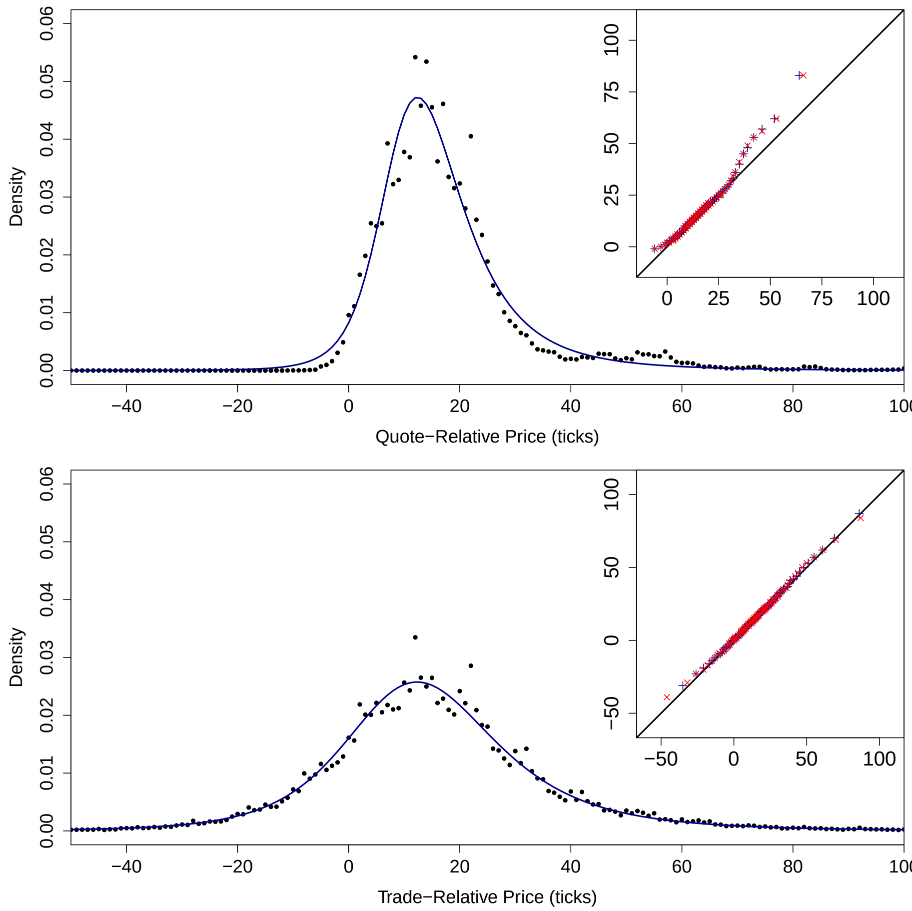

Figure 4: Fits of the generalized t distribution to the distribution of limit order arrivals for EUR/USD on 4 May 2010 in (top) quote-relative and (bottom) trade-relative coordinates. The main plots show (black circles) the empirical density functions for buy limit orders and (blue curves) the corresponding fits of the generalized t distribution. The results for sell orders and for the other currency pairs are qualitatively similar. Inset are quantile-quantile (Q-Q) plots of (vertical axis) the ECDFs versus (horizontal axis) our fits of the generalized t distribution. The points denote the $1^{\text{st}}, 2^{\text{nd}}, \ldots, 99^{\text{th}}$ percentiles of the distributions for (blue crosses) buy orders and (red crosses) sell orders. The solid black lines denote the diagonal. The results for the other currency pairs are qualitatively similar.

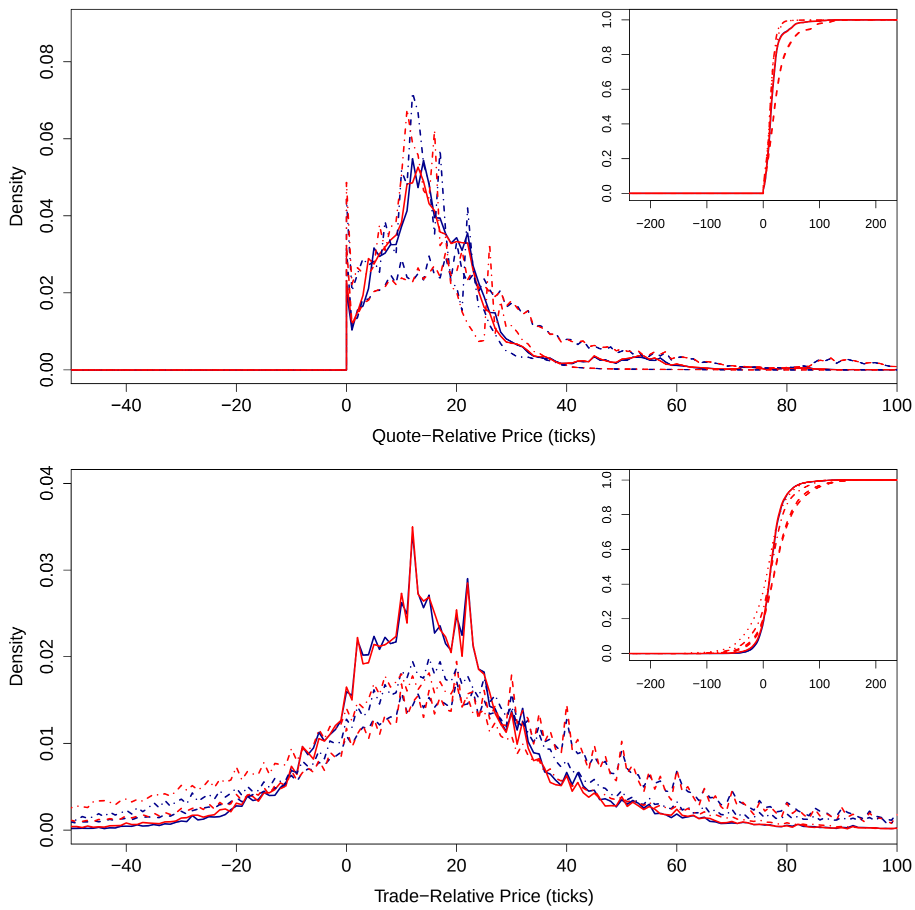

Figure 5: Distributions of cancellations by (top) quote-relative and (bottom) trade-relative price for (solid curves) EUR/USD, (dashed curves) GBP/USD, and (dotted-dashed curves) EUR/GBP on 4 May 2010. The blue curves denote buy orders and the red curves denote sell orders. The main plots show the empirical density functions and the inset plots show the ECDFs.

distribution does not capture this feature of the data. Moreover, similarly to the fit for limit order arrivals, the fitted distribution fails to capture the empirical behaviour in the upper tail.

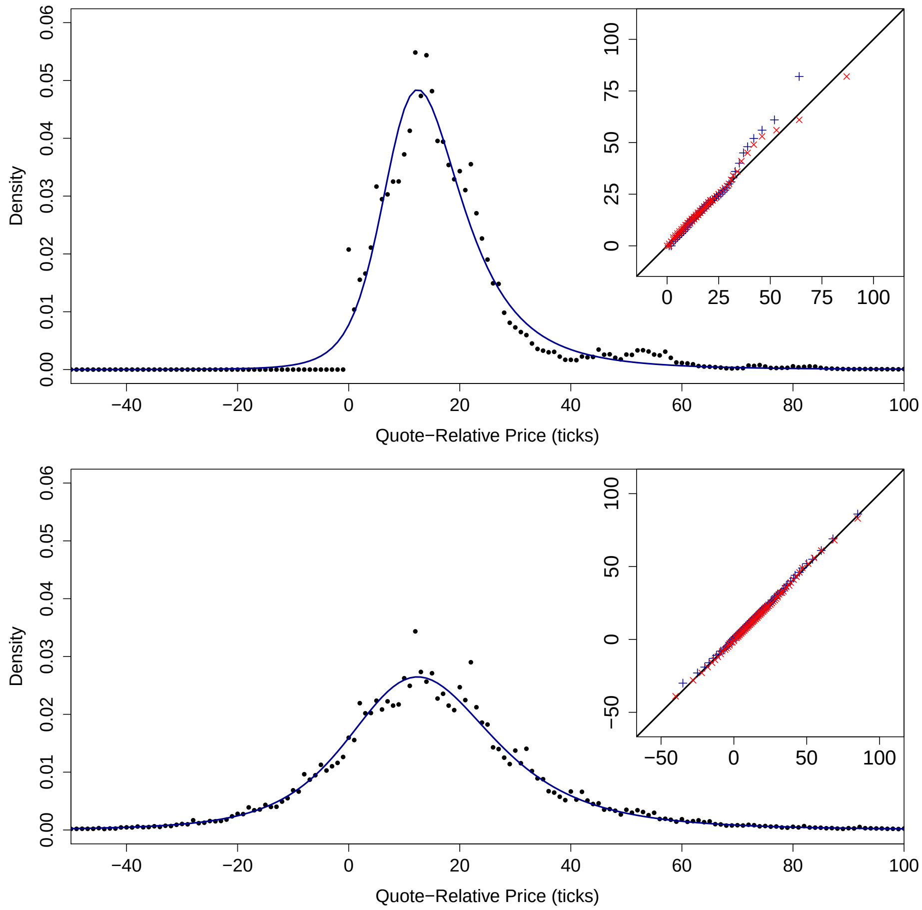

Figure 6: Fits of the generalized t distribution to the distribution of cancellations for EUR/USD on 4 May 2010 in (top) quote-relative and (bottom) trade-relative coordinates. The main plots show (black circles) the empirical density functions for active buy order cancellations and (blue curves) the corresponding fit of the generalized t distribution. The results for sell orders and for the other currency pairs are qualitatively similar. Inset are Q-Q plots of (vertical axis) the ECDFs versus (horizontal axis) our fits of the generalized t distribution. The points denote the $1^{\text{st}}, 2^{\text{nd}}, \ldots, 99^{\text{th}}$ percentiles of the distributions for (blue crosses) buy orders and (red crosses) sell orders. The solid black lines denote the diagonal. The results for the other currency pairs are qualitatively similar.

Figure 7 shows the mean depths (i.e., the mean total size of active orders in the global LOB) at the given quote-relative and trade-relative prices. By definition, the mean depth is 0 for all negative quote-relative prices. Although all three currency pairs have a local maximum in mean depth at the best quotes, in each case it is much smaller than the corresponding local maximum in the cancellation distributions. In both quote-relative and trade-relative coordinates, the mean depth at small quote-relative prices is substantially larger for EUR/USD than it is for GBP/USD and EUR/GBP.

Similarly to our results for limit order arrivals and cancellations, we find that the generalized t distribution provides a good fit to the distributions of normalized mean depths (i.e., the mean depths expressed as a fraction of the mean total size of all active orders of the given type at all prices) over the whole domain in trade-relative coordinates, but that the fits perform less well for the upper tail of the distributions in quote-relative coordinates (see Figure 8 for EUR/USD; the results for the other currency pairs are qualitatively similar).

Because the mean depth varies substantially across relative prices, we also calculate the *mean cancellation ratio,* which we measure by rescaling the total size of cancelled active orders at a given relative price by the corresponding mean depth (see Figure 9). In quote-relative coordinates, the mean cancellation ratio varies considerably with relative price, but this is unsurprising given that institutions in a QCLOB are not able to calculate quote-relative prices in real time and therefore cannot use this information when deciding whether to cancel an order. In trade-relative coordinates, by contrast, two interesting insights emerge. First, each of the three currency pairs' mean cancellation ratios exhibit a strong round-number periodicity. This indicates that the average lifetime of an active order at a trade-relative price that is an integer multiple of 5 is longer than that of an active order at a neighbouring trade-relative price. Second, aside from this round-number effect, the mean cancellation ratios for EUR/USD and GBP/USD are approximately constant for negative trade-relative prices, then decrease for positive trade-relative prices. At all trade-relative prices, the cancellation ratio for EUR/GBP is higher than it is for the other two currency pairs. However, the round-number effect is particularly strong for this currency pair, so it is more difficult to discern the variation in mean cancellation ratio across trade-relative prices.

D. Comparisons Across Trading Days

In this section, we investigate how the statistics and distributions in Section V.C vary across trading days.

Figure 10 shows the ECDFs of limit order arrivals, active order cancellations, and normalized mean depths at given quote-relative and trade-relative prices for EUR/USD on each of the 30 days in our sample. The results for GBP/USD and EUR/GBP are qualitatively similar. Despite the similarities between their shapes, the ECDFs exhibit substantial quantitative differences across different days. On some days, the majority of order arrivals and cancellations occur over a narrow range of small relative prices; on other days, the range of relative prices over which such activity occurs is wider and a larger fraction of order arrivals

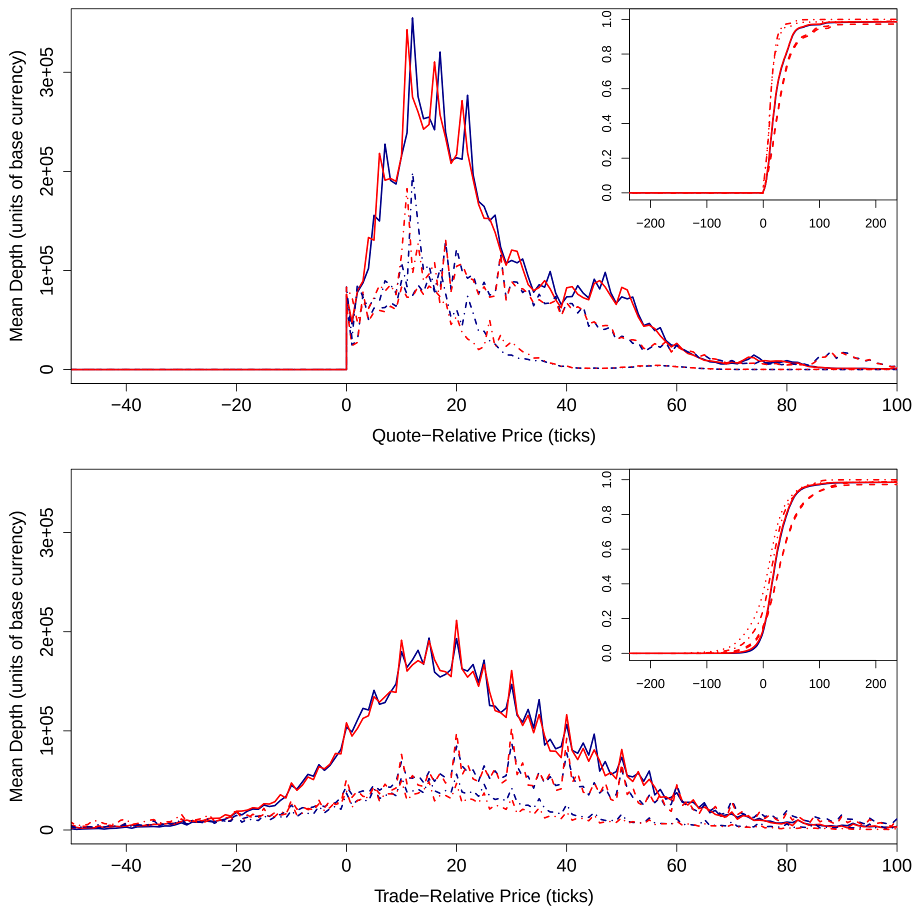

Figure 7: Mean depths at given (top) quote-relative and (bottom) trade-relative prices for (solid curves) EUR/USD, (dashed curves) GBP/USD, and (dotted-dashed curves) EUR/GBP on 4 May 2010. The blue curves denote the mean total size of active buy orders and the red curves denote the mean total size of active sell orders. The main plots show the empirical mean depths (in units of the base currency) and the inset plot shows the normalized empirical cumulative mean depths (i.e., the empirical cumulative mean depths expressed as a fraction of the mean total size of all active orders of the given type at all prices).

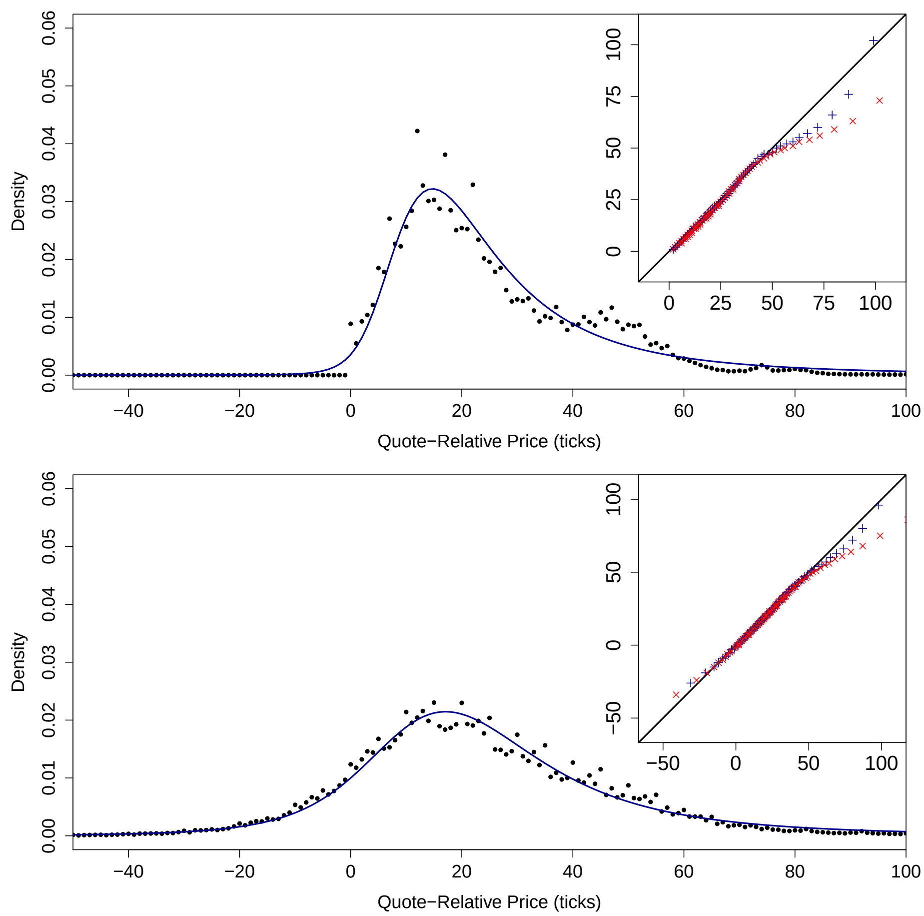

Figure 8: Fits of the generalized t distribution to the normalized distribution of mean buy order depths for EUR/USD on 4 May 2010 in (top) quote-relative and (bottom) trade-relative coordinates. The main plots show (black circles) the empirical density functions and (blue curves) the fits of the generalized t distribution. The results for sell orders and for the other currency pairs are qualitatively similar. Inset are Q-Q plots of (vertical axis) the ECDFs versus (horizontal axis) the fits of the generalized t distribution. The points denote the $1^{\text{st}}, 2^{\text{nd}}, \ldots, 99^{\text{th}}$ percentiles of the distributions for (blue crosses) buy orders and (red crosses) sell orders. The solid black lines denote the diagonal. The results for the other currency pairs are qualitatively similar.

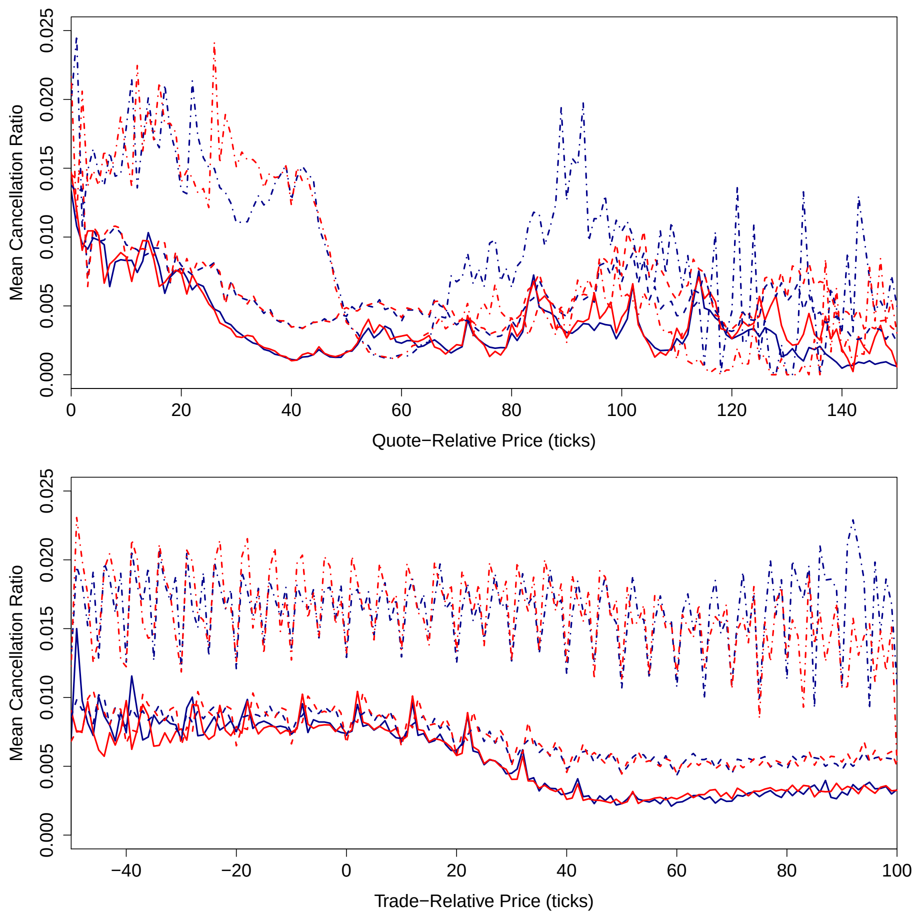

Figure 9: Mean cancellation ratio at given (top) quote-relative and (bottom) trade-relative prices for (solid curves) EUR/USD, (dashed curves) GBP/USD, and (dotted-dashed curves) EUR/GBP on 4 May 2010. The blue curves denote buy orders and the red curves denote sell orders.

and cancellations occur deep into the LOB. These observations suggest that the distributions' first two moments vary considerably across trading days.

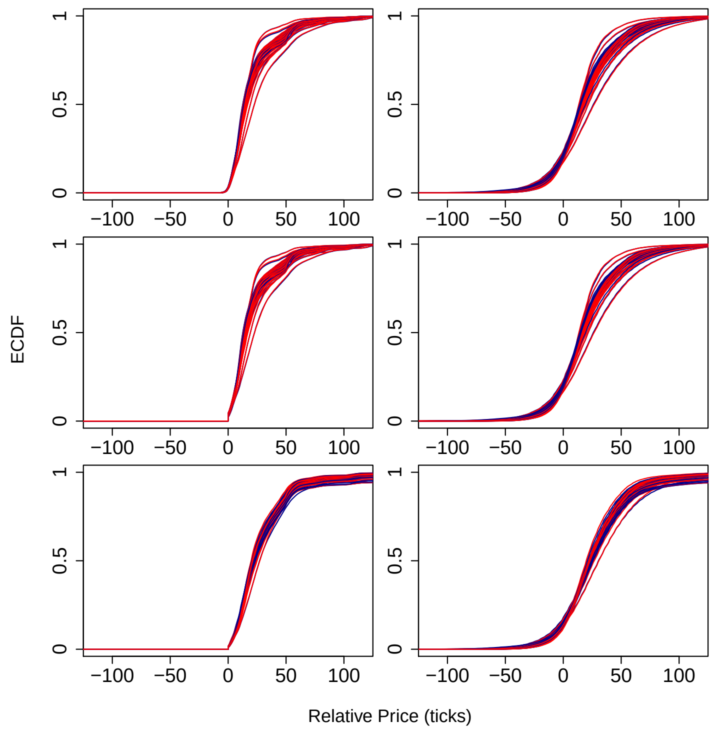

Figure 10: ECDFs for EUR/USD (top row) limit order arrivals, (middle row) cancellations, and (bottom row) normalized mean depths at given (left column) quote-relative and (right column) trade-relative prices. Each blue (respectively, red) curve denotes buy- (respectively, sell-)order activity on a single day. The corresponding plots for GBP/USD and EUR/GBP are qualitatively similar.

To investigate the extent to which differences in the first two moments account for the observed differences between the daily distributions, we rescale each day's data by subtracting its sample mean and dividing by its sample standard deviation.[7] Figure 11 shows the same ECDFs after rescaling the data in this way. The results for GBP/USD and EUR/GBP are qualitatively similar. In quote-relative coordinates, the rescaling causes a strong collapse over the first three quartiles of the distributions of limit order arrivals and cancellations, but daily differences in the distributions' upper quartiles prevents a strong collapse in the upper tail. In trade-relative coordinates, the rescaling causes a strong collapse over the whole domain onto a single, universal curve. In both quote-relative and trade-relative coordinates, the collapse for the distributions

[7]We use a trimmed sample mean and trimmed sample standard deviation to exclude a very small number of orders with extremely large relative prices (see Appendix A).

of normalized mean depths is weaker than for the order-flow distributions due to a handful of orders with extremely large relative prices that remain active for long periods on some days.[8]

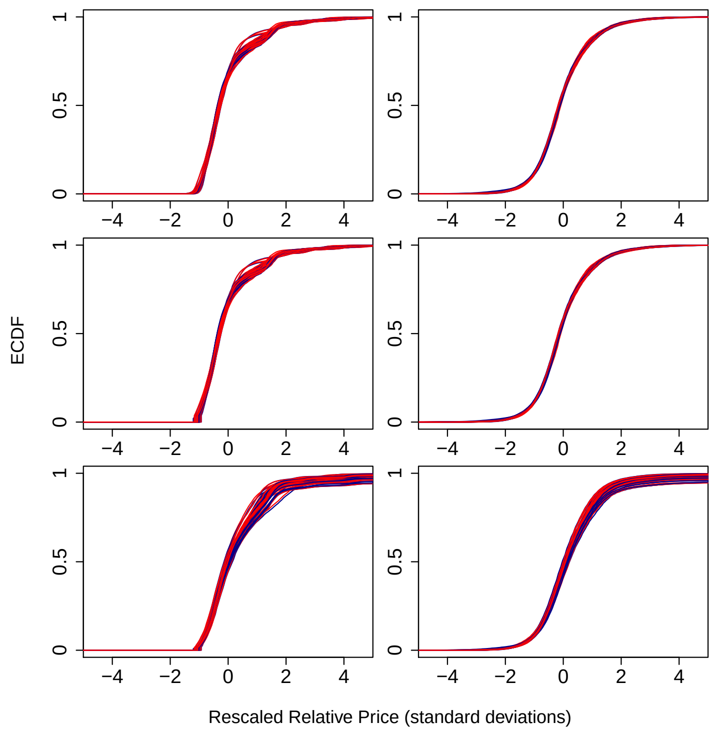

Figure 11: ECDFs for EUR/USD (top row) limit order arrivals, (middle row) cancellations, and (bottom row) normalized mean depths at given (left column) quote-relative and (right column) trade-relative prices, after rescaling each day's data by subtracting its sample mean and dividing by its sample standard deviation. Each blue (respectively, red) curve denotes buy- (respectively, sell-)order activity on a single day. The corresponding plots for GBP/USD and EUR/GBP are qualitatively similar.

To quantify the strength of this curve collapse, we compute the mean ratio $\overline{C}$ of the Cramér–von Mises (CvM) distances [19] between the distributions before and after applying the rescaling (see Table II).[9] We provide a detailed discussion

[8]To verify that such extreme-priced orders were indeed the primary reason for the weaker collapse of these distributions, we repeated our calculations after excluding all active orders with a relative price of more than 5 standard deviations from the mean. We found that the resulting curve collapse was similar to that for limit order arrivals and cancellations.

[9]We also find qualitatively similar results when using the Kolmogorov–Smirnov (KS) distance [56]. There are many other possible distance measures [21] that we could use; we choose the CvM and KS distances because they are widely used, easy to interpret, and fast to compute.

of our methodology in Appendix C.

Coordinates	Order Flow	EUR/USD		GBP/USD		EUR/GBP	
		Buy	Sell	Buy	Sell	Buy	Sell
Quote Relative	Limit Orders	4.12	3.77	3.84	3.86	4.34	5.13
	Cancellations	3.75	3.46	3.64	3.55	4.08	4.97
	Mean Depths	1.39	1.28	2.23	1.77	2.50	2.54
Trade Relative	Limit Orders	16.5	16.1	20.0	25.5	17.8	11.4
	Cancellations	16.3	16.1	19.1	24.3	17.7	11.2
	Mean Depths	2.78	2.83	12.4	9.89	10.3	6.82

Table II: Mean CvM ratios $\overline{C}$ (see the description in the main text and Appendix C) for limit order arrivals, cancellations, and mean depths. Values greater than 1 indicate that rescaling each day's data to account for differences in its first two moments reduces the mean distance between the daily distributions. Larger values correspond to stronger curve collapse.

In quote-relative coordinates, the reductions in CvM distance for limit orders and cancellations range from a factor of about 4 to a factor of about 5. This indicates a moderately strong curve collapse. The corresponding reductions for normalized mean depths are weaker, due to a small number of extreme-priced orders that remain active for long periods of time and thereby prevent stronger collapse in the upper tails of these distributions.

In trade-relative coordinates, the reductions in CvM distance for limit order arrivals and cancellations range from a factor of about 11 to a factor of more than 25. This indicates very strong curve collapse. Again, the corresponding reductions for normalized mean depths are weaker (particularly for EUR/USD), but they still indicate a moderate curve collapse for EUR/USD and a strong curve collapse for GBP/USD and EUR/GBP.

E. A Semi-Parametric Model of Order Flow and LOB State

In Section V.C, we noted that the generalized t distribution provides a good fit to the distributions of limit order arrivals, cancellations, and normalized mean depths on a single trading day. The curve collapse that we observe in Figure 11 and Table II motivates the following alternative, semi-parametric approach to modelling these distributions.

For a single trading day d, let μ_d and σ_d denote, respectively, the mean and standard deviation of a specified property (e.g., EUR/USD buy limit order arrivals in trade-relative coordinates). Given data from a set D of trading days, rescale the data on each day d by subtracting μ_d then dividing by σ_d, then aggregate the rescaled data for all days into a single data set. To obtain the model for the distribution on another trading day $d' \notin D$, apply the inverse rescaling for day d' to the aggregated data set (i.e., multiply each entry in the aggregated data set by $\sigma_{d'}$ then add $\mu_{d'}$).

Figure 12 shows the result of applying this process to use the other 29 days in our sample to model the trade-relative distribution of limit order arrivals for EUR/USD. The results for cancellations and for the other currency pairs are all qualitatively similar. The results for normalized mean depths and for the distributions in quote-relative coordinates are slightly weaker in the upper tail due to the weaker curve collapse in this region (see Figure 11), but given that such activity corresponds to limit orders with very low fill probabilities, we do not regard a close fit in this region to be as important as it is for the main body of the distribution, where the fits are strong.

In all cases, the performance of our semi-parametric method is similar to that of fitting the generalized t distribution directly to the data (see Figure 4). However, our semi-parametric approach offers a considerable computational advantage: after computing the aggregated data set (which, given a historical database of trading days, can be performed offline and in advance of fitting the single trading day), performing the semi-parametric fit requires only a simple multiplication and addition. By contrast, fitting the generalized t distribution requires numerical optimization of a nonlinear objective function (see Appendix B), which is much slower to perform.

VI. Discussion

Our results in Section V raise many interesting points for discussion. In this section, we address these points and compare our findings to those reported by empirical studies of centralized LOBs.

In their study of centralized LOBs on the LSE and Paris Stock Exchange, [65] proposed that the ratio of the mean total size of active orders to the mean total size of market orders on a single day could be used as a simple measure of liquidity. The authors reported ratios in the range 100 to 1000 for the stocks that they studied. On Hotspot FX, the same ratios vary between roughly 10 for EUR/GBP and 100 for EUR/USD (see Table I). This suggests that global liquidity in $\mathcal{L}(t)$ is much more plentiful on Hotspot FX than was the case for the stocks that Wyart *et al.* studied. However, each institution on Hotspot FX can only access a subset of this global liquidity pool, so it remains unclear whether the liquidity levels experienced by individual institutions are really as plentiful as this statistic would suggest.

The Hotspot FX data provides some evidence that several institutions are trading partners with relatively few others. For example, we observe periods when $s(t) < 0$. During such periods, $a(t) < b(t)$, so an arbitrage opportunity exists for any institution θ_k that is trading partners with both the owner θ_i of an active order at $b(t)$ and the owner θ_j of an active order at $a(t)$. Negative spreads often persist for several seconds, which suggests that there are no institutions $\theta_k \in \Theta$ with the necessary trading partnerships to capitalize on these arbitrage opportunities. Therefore, among institutions that place limit orders close to the best quotes, there appear to be many pairs $\theta_i \nleftrightarrow \theta_j$ with no mutual trading partners.

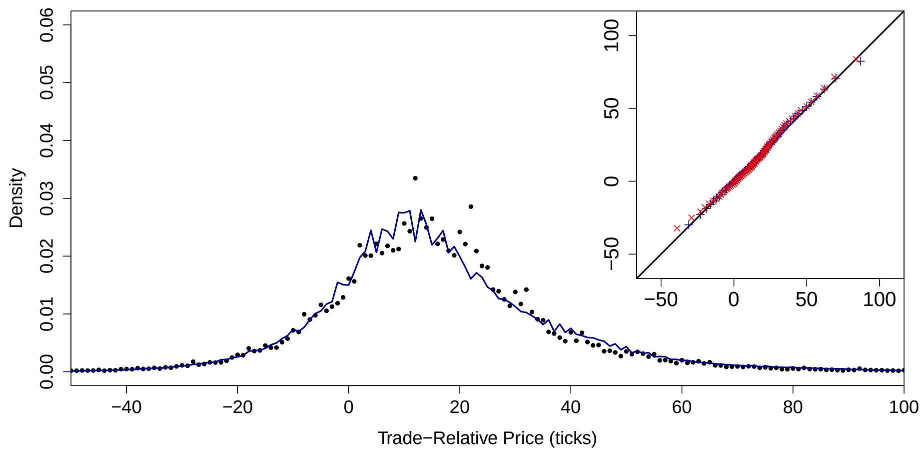

Figure 12: Semi-parametric fit of the distribution of limit order arrivals for EUR/USD on 4 May 2010 in trade-relative coordinates. The main plot shows (black circles) the empirical density function for buy limit orders and (blue curve) the corresponding semi-parametric fit obtained by rescaling and aggregating the data from all other days in our sample then inverting the rescaling according to the mean and standard deviation on 4 May 2010. The results for sell orders, for cancellations, and for the other currency pairs are qualitatively similar. Inset is a Q-Q plot of (vertical axis) the ECDFs versus (horizontal axis) our semi-parametric fits of the distribution. The points denote the $1^{\mathrm{st}}, 2^{\mathrm{nd}}, \ldots, 99^{\mathrm{th}}$ percentiles of the distributions for (blue crosses) buy orders and (red crosses) sell orders. The solid black line denotes the diagonal. The results for cancellations and for the other currency pairs are qualitatively similar.

Although the data does not enable us to calculate the local LOBs for specific institutions, our results suggest that institutions monitor $\mathcal{L}_i(t)$ carefully when deciding how to act. For example, we observe few market orders that match at several different prices, which suggests that many institutions implement selective liquidity-taking strategies [64] by monitoring $\mathcal{L}_i(t)$ and only submitting market orders with a size smaller than the depth at $b_i(t)$ or $a_i(t)$. Correspondingly, we find that the mean size of market orders is less than half of the mean size of limit orders (see Table I). Together, these observations are also consistent with a strategy called "order splitting" [1, 9, 43], by which institutions that wish to conduct large trades submit several small market orders to minimize their market impact.

In addition to monitoring $\mathcal{L}_i(t)$, our results also suggest that many institutions calculate trade-relative prices when deciding how to act. In centralized LOBs, quote-relative distributions often contain strong round-number periodicities at integer multiples of 5 or 10 ticks [15, 33, 47, 66]. We find no evidence for this behaviour on Hotspot FX (which is unsurprising given that institutions in a QCLOB cannot calculate quote-relative prices), but we instead find similar round-number periodicities in the trade-relative distributions (see Figure 3 for limit order arrivals, Figure 5 for cancellations, Figure 7 for normalized mean depths, and Figure 9 for cancellation ratios). Such periodicities are extremely unlikely to emerge by chance, so it seems that institutions regard $B(t)$ and $A(t)$ as important sources of information.

Several empirical studies of centralized LOBs have reported that the maximum limit order arrival rate occurs at a quote-relative price of 0 [7, 10, 32, 37, 46, 51]. By contrast, the maximum limit order arrival rate on Hotspot FX occurs at a strictly positive quote-relative price (see Figure 3). We propose the following explanation for this observation. In a QCLOB, each institution θ_i sees the values of $b_i(t)$ and $a_i(t)$, but cannot see the values of $b(t)$ and $a(t)$. By definition, $b_i(t) \leq b(t)$ and $a_i(t) \geq a(t)$, so if each institution bases its trading decisions on $b_i(t)$ and $a_i(t)$, and if $b_i(t)$ and $a_i(t)$ both typically reside at strictly positive quote-relative prices, then the maximum arrival rate of the aggregate limit order flow generated by all institutions will occur at a strictly positive quote-relative price.

Empirical studies of centralized LOBs have similarly reported that cancellations occur most often among active orders at $b(t)$ and $a(t)$ and less often among active orders deeper into the LOB [18, 51]. Several authors have conjectured that the high number of cancellations at these prices indicate that many institutions compete for priority at the best quotes, and that the lower cancellation rates among other orders indicate that their owners aim to profit from price movements on longer time horizons [15, 51, 67]. We also observe a local maximum in the distribution of cancellations at a quote-relative price of 0 (see Figure 5), but we find the distribution's global maximum to be strictly positive. After rescaling to account for differences in the mean depths, we find that the quote-relative cancellation ratios vary considerably, with no clear trend (see Figure 9). In trade-relative coordinates, we find that the cancellation distributions closely resemble those of limit order arrivals, with a slightly lower cancellation ratio among orders with larger trade-relative prices.

Many centralized LOBs have been reported to exhibit a "hump" shape that first increases then subsequently decreases away from the spread [10, 33, 37, 51]. [53] conjectured that such a hump represents a trade-off between an optimism that limit orders placed far from the spread may eventually result in a significant profit and a pessimism that such orders may never match. We also observe a hump shape in the mean state of $\mathcal{L}(t)$ in both quote-relative and trade-relative coordinates (see Figure 7). For the LOBs examined by the other empirical studies, market orders accounted for about 10%–30% of the total arriving order flow, and therefore played an important role in maintaining the hump shape of $\mathcal{L}(t)$ [15, 27, 35, 44, 51]. On Hotspot FX, by contrast, market orders constitute less than 0.05% of the total arriving order flow (see Table I). Therefore, the hump shapes that we observe are primarily a consequence of similar shapes in the distributions of limit order arrivals and cancellations.

Why do institutions submit so many limit orders, given that so few result in trades? We propose two possible explanations. First, institutions may place limit orders on several different trading platforms simultaneously to increase their chance of receiving a matching. If one such order matches, an institution can simply cancel the duplicates on other platforms. [17] recently noted that this strategy, which they call "overbooking", becomes more prominent as the number of venues for a given asset increases. The large number of different platforms in the FX spot market [12] could result in a large volume of cancellations from institutions that adopt this strategy. Second, many high-frequency and algorithmic trading techniques involve the submission and cancellation of large numbers of limit orders [6, 13, 36, 39]. The recent surge in popularity of trading strategies that utilize these techniques could account for a high percentage of the cancellations that we observe [3].

In both quote-relative and trade-relative coordinates, the distributions of limit order arrivals, cancellations, and normalized mean depths on Hotspot FX exhibit considerable variation across different trading days (see Figure 10). In all cases, however, rescaling the data to account for differences in the first two moments significantly reduces the mean pairwise CvM distance between daily distributions (see Table II). In trade-relative coordinates, the resulting curve collapse for limit order arrivals and cancellations is particularly strong. Given the turbulent macroeconomic activity that occurred during this period, we regard such strong curve collapse to be surprising, because it indicates that the first two moments provide significant explanatory power for daily order flow and highlights that the vast majority of daily variations in order flow manifest as simple, linear transformations to a single, universal curve.

VII. Conclusions

During the past decade, a rich and diverse literature has helped to illuminate many important aspects of trading via LOBs. To date, however, almost all work in this area has addressed only centralized LOBs, in which all institutions can trade with all others. In this paper, we have provided a detailed description of an alternative LOB configuration, which we called a QCLOB, and have performed an empirical analysis of a recent, high-quality data set from a large electronic trading platform that utilizes this mechanism to facilitate trade. Our results reveal many important differences between QCLOBs and centralized LOBs, and thereby underline the urgent need for detailed investigations of other widely used market organizations to complement the sizeable literature on centralized LOBs.

We introduced a coordinate frame that we called trade-relative coordinates, in which prices are measured relative to those of the most recent trades. We found that trade-relative coordinates illuminate several interesting properties of order flow and LOB state that are not apparent when measuring prices relative to the prevailing quotes, as is common when studying centralized LOBs. Although our use of trade-relative coordinates was motivated by the structure of a QCLOB, we conjecture that this coordinate frame may also provide useful insight into centralized LOBs. In many markets, the rise in popularity of electronic trading has led to a sharp increase in the frequency of order arrivals near the best quotes [6, 13, 16, 36, 39]. Such arrivals cause the values of $b(t)$ and $a(t)$ — and, therefore, the quote-relative prices of all orders — to fluctuate rapidly. By contrast, trade-relative prices change only when a trade occurs, and therefore avoid the difficulties caused by the extremely high update frequency of the best quotes. It would be interesting to perform an empirical analysis of a centralized LOB in trade-relative coordinates, to compare the results with our findings.

In a recent study of the LSE, [2] conjectured that the statistical properties of financial markets change every day. At present, however, many of the most widely discussed LOB models operate under the assumption that order flow is governed by stochastic processes with fixed rate parameters [15, 18, 24, 46, 57, 64]. The empirical verification of such models has typically consisted of comparing their output to long-run statistical averages from large data sets. Our results, together with those of Axioglou and Skouras, bring into question the usefulness of using long-run statistical averages to forecast activity on a specific day. It would be interesting to study the performance of several existing LOB models to assess their performance on shorter timescales. Given that regulators require many institutions to make risk calculations on a daily basis, we regard this to be a task of primary importance for future research.

Finally, we note that our statistical analysis in this paper has mainly examined the aggregate order flow and global LOB $\mathcal{L}(t)$. An extremely important topic for future research is to asses how the organization of a QCLOB impacts price formation. We aim to address this, and many other related questions, in our forthcoming work.

Appendix A. Structure of the Hotspot FX Data

For each currency pair each day, the Hotspot FX data consists of two files. The first file contains a transcription of the trade-data stream that the platform disseminates to all institutions in real time (see Section II.C), along with the size of each trade. The second file begins with a *snapshot* that describes the state of the global LOB $\mathcal{L}(t)$ at 08:00:00 GMT, then contains a list of all subsequent limit order arrivals and active order departures until 17:00:00 GMT. By first constructing the initial LOB state (according to the snapshot) then processing each subsequent order arrival or departure, we are able to reconstruct the global LOB $\mathcal{L}(t)$ at any time t during 08:00:00–17:00:00 GMT. However, Hotspot FX do not disclose any information regarding trade partnerships on the platform, and the data does not contain any information about institutions' identities, so it is not possible to reconstruct local LOBs from the data.

Active Order Departures

In the Hotspot FX data, an active order departure denotes that an active order was either cancelled by its owner or fully matched to an incoming market order. The data provides no way to deduce with certainty whether such an event relates to a cancellation or a complete matching. When studying order-flow distributions, we treat all active order departures as cancellations. The percentage of active order departures that are actually due to complete matching is extremely low, because market orders constitute about 0.05%, about 0.02%, and less than 0.01% of arriving order flow for EUR/USD, GBP/USD, and EUR/GBP, respectively (see Table I). Therefore, incorrectly classifying a tiny fraction of departures in this way should have negligible impact on our results.

The Trade-Data Stream

To prevent institutions from exploiting trade information in real time, Hotspot FX lags dissemination of the trade-data stream by 1 second. For example, if a trade occurs at 08:34:26.346 GMT, the time stamp in the trade-data stream is 08:34:26.346 but the information is not disseminated to other institutions until 08:34:27.346. When calculating the $B(t)$ and $A(t)$ series, we therefore add 1 second to the times reported in the trade-data stream to align them with the time that the relevant information appeared on the Hotspot FX trading screen.

Extreme-Priced Orders

For each of the three currency pairs, a very small number of order arrivals and cancellations occur at extremely large relative prices. For example, all EUR/USD trades described in the data occur within the price interval \$1.10 to \$1.40, but some sell limit orders arrive with a price of more than \$500.00.

We do not judge such orders to represent a serious intention to trade. Instead, we believe that institutions submit such orders either in error or to check their connectivity to the Hotspot FX platform.

Although only a very small number of such orders appear in the data, their extreme prices can greatly affect the sample moments. Therefore, when we calculate estimates of the first two moments in Section V.D, we calculate trimmed means and trimmed standard deviations [38] by removing all order arrivals and cancellations that occur with a relative price of more than 1000 ticks. For each of the three currency pairs, for both buy and sell orders, and in both quote-relative and trade-relative coordinates, trimming the data in this way removes less than 0.05% of the total order flow. We also obtain qualitatively similar results if we instead trim the top 1 percentile of the ECDFs of relative prices.

Inferring Market Orders

If a market order matches to several different active orders, then the Hotspot FX data reports each partial matching as a separate line in the trade-data stream, with a time stamp that differs from the previous line by at most 1 millisecond. In the absence of explicit details regarding order ownership, we regard all entries that correspond to a trade of the same direction and that arrive within 1 millisecond of each other as originating from the same market order. For each of the three currency pairs, the mean inter-arrival time between trades is of the order of several seconds, so it is unlikely that two separate market orders would arrive within 1 millisecond. We regard any incorrectly grouped market orders as a source of noise in the data.

The data does not provide a reliable way to perform inference about incoming orders that partially match and partially become active. For such orders, we treat the matched part as a market order and the unmatched part as a separate limit order.

Appendix B. Fitting the Generalized t Distribution

If Z is a random variable from the standard normal distribution and V is an independent random variable from the chi-squared distribution with ν degrees of freedom, then the random variable

$$T = \sigma \frac{Z + \xi}{\sqrt{V/\nu}} + \mu \qquad \text{(B1)}$$

follows a *generalized t distribution*. The parameters μ, σ, and ξ extend the classical Student's t distribution by providing explicit control over the distribution's mean, variance, and skewness, respectively [28].

For each day $d = 1, 2, \ldots, 30$, we perform our fit of the generalized t distribution to a given property of the Hotspot FX data (e.g., EUR/USD buy limit order arrivals in trade-relative coordinates) by minimizing the Cramér–von Mises (CvM) distance [19]

$$C = \int_p \left(F_d(p) - F(p; \mu, \sigma, \xi, \nu)\right)^2 dF(p; \mu, \sigma, \xi, \nu) \tag{B2}$$

between the ECDF F_d of the given property on day d and the cumulative density function F of the generalized t distribution with parameters μ, σ, ξ, and ν. We use Newton's method [20] to optimize the objective function in Equation (B2) over these parameters. On a standard desktop computer, this process requires approximately 1 to 2 minutes of computation to fit the distribution of a given property for a given currency pair on a given day.

Fitting the distribution by minimizing the CvM distance is equivalent to minimizing a least-squares objective function that assigns greater weight to the regions of the distribution with greater density. It is also possible to fit the generalized t distribution via moment-matching [34] or maximum-likelihood [11] techniques, but the resulting estimates do not perform as well due to the existence of a handful of orders with extremely large relative prices that have a large impact on the sample moments and maximum-likelihood estimates (see Appendix A).

Appendix C. Quantifying the Strength of Curve Collapse

To quantify the strength of curve collapse caused by rescaling each day's data, we calculate the mean of the ratio of CvM distances (see Equation (B2)) between the ECDFs of a chosen property on a given pair of trading days, before and after applying the rescaling. More precisely, we calculate

$$\overline{C} = \frac{1}{30 \times 29} \sum_{\substack{d_1, d_2 \\ d_1 \neq d_2}} \frac{C^1_{d_1,d_2}}{C^2_{d_1,d_2}}, \tag{C1}$$

where $C^1_{d_1,d_2}$ denotes the CvM distance between the ECDFs of a chosen property (e.g., EUR/USD buy limit order arrivals in quote-relative coordinates) on days d_1 and d_2, and $C^2_{d_1,d_2}$ denotes the CvM distance between the same ECDFs after rescaling the data on day d_2 by subtracting the mean for day d_2 and dividing by the standard deviation for day d_2, then multiplying the result by the standard deviation for day d_1 and finally adding the mean for day d_1. Larger values of $\overline{C}$ correspond to stronger collapse of the ECDFs. Observe that we do not rescale the data from both days and measure the distance between the rescaled distributions directly, but instead apply the inverse rescaling from day d_1 to the rescaled data from day d_2. This ensures that we measure our results in the units of price for both $C^1_{d_1,d_2}$ and $C^2_{d_1,d_2}$, rather than using the units of rescaled price in $C^2_{d_1,d_2}$.

Acknowledgements

We thank Bruno Biais, Jean-Philippe Bouchaud, Rama Cont, J. Doyne Farmer, Austin Gerig, Ben Hambly, Nikolaus Hautsch, Gabriele La Spada, Sergei Maslov, Stephen Roberts, Cosma Shalizi, and Thaleia Zariphopoulou for useful discussions. We also thank Jonas Haase, Terry Lyons, Rich Plummer-Powell, and Justin Sharp for technical support. MDG and SDH thank the Oxford-Man Institute of Quantitative Finance, and MDG thanks EPSRC and the James S. McDonnell Foundation for supporting this research.

REFERENCES

[1] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

[2] C. Axioglou and S. Skouras. Markets change every day: evidence from the memory of trade direction. *Journal of Empirical Finance*, 18(3):423–446, 2011.

[3] Bank for International Settlements. Triennial central bank survey: report on global foreign exchange market activity in 2010. Technical report, Bank for International Settlements, available at `http://www.bis.org/publ/rpfxf10t.pdf`, 2010.

[4] Bank for International Settlements. Foreign exchange turnover in April 2013: preliminary global results. Technical report, Bank for International Settlements, available at `http://www.bis.org/publ/rpfx13fx.pdf`, 2013.

[5] W. Barker. The global foreign exchange market: growth and transformation. *Bank of Canada Review*, 4:3–12, 2007.

[6] B. Biais, T. Foucault, and S. Moinas. Equilibrium high-frequency trading. Technical report, Toulouse School of Economics, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1834344`, 2011.

[7] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *The Journal of Finance*, 50(5):1655–1689, 1995.

[8] G. H. Bjønnes and D. Rime. Dealer behavior and trading systems in foreign exchange markets. *Journal of Financial Economics*, 75(3):571–605, 2005.

[9] J. P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In T. Hens and K. R. Schenk-Hoppé, editors, *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–160. North–Holland, Amsterdam, The Netherlands, 2009.

[10] J. P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.

[11] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, Pacific Grove, CA, USA, 2001.

[12] Celent. Electronic platforms in foreign exchange trading. Technical report, Available at `http://www.e-forex.net/Files/surveyreportsPDFs/Celent%20FX%20report.pdf`, 2007.

[13] A. Chaboud, B. Chiquoine, E. Hjalmarsson, and C. Vega. Rise of the machines: Algorithmic trading in the foreign exchange market. Technical report, Federal Reserve, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1501135`, 2011.

[14] A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review I: empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[15] D. Challet and R. Stinchcombe. Analyzing and modeling $1 + 1d$ markets. *Physica A*, 300(1–2):285–299, 2001.

[16] R. Cont. Statistical modeling of high-frequency financial data. *Signal Processing Magazine, IEEE*, 28(5):16–25, 2011.

[17] R. Cont and A. Kukanov. Optimal order placement in limit order markets. *arXiv:1210.1625*, 2014.

[18] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563, 2010.

[19] H. Cramér. On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928.

[20] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice–Hall, Englewood Cliffs, NJ, USA, 1983.

[21] M. M. Deza and E. Deza. *Dictionary of Distances*. Elsevier, Amsterdam, The Netherlands, 2006.

[22] EBS. `https://emea.ebsspot.com/SHARED/HELP/userguide.pdf`, 2011.

[23] Euronext. Retrieved 23 November 2013 from `https://europeanequities.nyx.com/connecting/universal-trading-platform`, 2013.

[24] J. D. Farmer, P. Patelli, and I. I. Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences*, 102(6):2254–2259, 2005.

[25] International Monetary Fund. World economic outlook. Technical report, International Monetary Fund, available at `http://www.imf.org/external/pubs/ft/weo/2013/01/pdf/text.pdf`, 2013.

[26] P. Gallardo and A. Heath. Execution methods in foreign exchange markets. *Bank for International Settlements Quarterly Review*, 1:83–91, 2009.

[27] Á. Gereben and N. Kiss. A brief overview of the characteristics of interbank Forint/Euro trading. *Magyar Nemzeti Bank Bulletin*, 1(2):21–26, 2010.

[28] W. S. Gosset. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

[29] P. G. Gottschalk and J. R. Dunn. The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Analytical Biochemistry*, 343(1):54–65, 2005.

[30] M. D. Gould, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

[31] C. Gouriéroux, J. Jasiak, and G. Le Fol. Intra-day market activity. *Journal of Financial Markets*, 2(3):193–226, 1999.

[32] G. F. Gu, W. Chen, and W. X. Zhou. Empirical regularities of order placement in the Chinese stock market. *Physica A*, 387(13):3173–3182, 2008.

[33] G. F. Gu, W. Chen, and W. X. Zhou. Empirical shape function of limit-order books in the Chinese stock market. *Physica A*, 387(21):5182–5188, 2008.

[34] A. R. Hall. *Generalized Method of Moments*. Oxford University Press, Oxford, UK, 2005.

[35] J. Hasbrouck and G. Saar. Limit orders and volatility in a hybrid market: The Island ECN. Technical report, NYU, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1294561`, 2002.

[36] T. Hendershott, C. M. Jones, and A. J. Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33, 2011.

[37] B. Hollifield, R. A. Miller, and P. Sandås. Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4):1027–1063, 2004.

[38] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. Wiley, New York, NY, USA, 2009.

[39] A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun. The flash crash: the impact of high-frequency trading on an electronic market. *Working Paper, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1686004`*, 2011.

[40] Knight Capital Group. http://www.hotspotfx.com/overview/index.jsp, 2015.

[41] Knight Capital Group. http://www.hotspotfx.com/download/userguide/HSFX/HSFX_UserGuide_wrapper.html, 2015.

[42] Knight Capital Group. http://www.hotspotfx.com/products/hotspot_volumes.jsp, 2015.

[43] F. Lillo, S. Mike, and J. D. Farmer. Theory for long memory in supply and demand. *Physical Review E*, 71(6):066122, 2005.

[44] I. Lo and S. G. Sapp. Order aggressiveness and quantity: how are they determined in a limit order market? *Journal of International Financial Markets, Institutions and Money*, 20(3):213–237, 2010.

[45] R. N. Mantegna and H. E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge, UK, 1999.

[46] S. Mike and J. D. Farmer. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1):200–234, 2008.

[47] G. H. Mu, W. Chen, J. Kertész, and W. X. Zhou. Preferred numbers and the distributions of trade sizes and trading volumes in the Chinese stock market. *The European Physical Journal B*, 68(1):145–152, 2009.

[48] NADSAQ. http://business.nasdaq.com/trade/markets/index.html, 2015.

[49] New York Stock Exchange. http://www.nyxdata.com/nysedata/asp/factbook/main.asp, 2015.

[50] E. Pan. FX trading and technology in 2012. Technical report, StreamBase Systems, available at http://www.tradersmagazine.com/news/fxsurveysb.php, 2012.

[51] M. Potters and J. P. Bouchaud. More statistical properties of order books and price impact. *Physica A*, 324:133–140, 2003.

[52] D. Rime. New electronic trading systems in foreign exchange markets. In D. C. Jones, editor, *New Economy Handbook*, pages 469–504. Academic Press, San Diego, CA, USA, 2003.

[53] I. Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, 2009.

[54] M. J. Sager and M. P. Taylor. Under the microscope: the structure of the foreign exchange market. *International Journal of Finance and Economics*, 11(1):81–95, 2006.

[55] L. Sarno and M. P. Taylor. The microstructure of the foreign-exchange market: a selective survey of the literature. *Princeton Studies in International Economics*, 89, 2001.

[56] N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin*, 2(2):3–26, 1939.

[57] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.

[58] J. A. Stephan and R. E. Whaley. Intraday price change and trading volume relations in the stock and stock option markets. *The Journal of Finance*, 45(1):191–220, 1990.

[59] The European Financial Stability Facility. `http://www.efsf.europa.eu/attachments/EFSF%20FAQ%202014-07-28.pdf`, 2014.

[60] The European Financial Stability Facility. `http://www.efsf.europa.eu`, 2015.

[61] The London Stock Exchange. `www.londonstockexchange.com/products-and-services/trading-services/sets/sets.htm`, 2015.

[62] Thomson–Reuters. `https://dxtrapub.markets.reuters.com/docs/Matching_Rule_Book.pdf`, 2011.

[63] I. M. Toke. "Market making" in an order book model and its impact on the spread. In F. Abergel, Chakraborti A., B. K. Chakrabarti, and Mitra M., editors, *Econophys-Kolkata V*, pages 49–64, Milan, Italy, 2011. Springer.

[64] B. Tóth, Y. Lempérière, C. Deremble, J. De Lataillade, J. Kockelkoren, and J. P. Bouchaud. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):021006, 2011.

[65] M. Wyart, J. P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid-ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.

[66] L. Zhao. *A Model of Limit Order Book Dynamics and a Consistent Estimation Procedure*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2010.

[67] I. Zovko and J. D. Farmer. The power of patience: a behavioral regularity in limit order placement. *Quantitative Finance*, 2(5):387–392, 2002.