Real Trading Patterns and Prices

in Spot Foreign Exchange Markets

By

Jon Danielsson
and
Richard Payne

DISCUSSION PAPER 320

March 1999

FINANCIAL MARKETS GROUP
AN ESRC RESEARCH CENTRE

LONDON SCHOOL OF ECONOMICS

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

ISSN 0956-8549-320

Real Trading Patterns and Prices in Spot Foreign Exchange Markets

Jón Daníelsson and Richard Payne*
Financial Markets Group
London School of Economics

This version March 26, 1999

Abstract

Most of the existing empirical literature on FX market microstructure uses indicative quote data derived from Reuters EFX screens. This paper examines the adequacy of such data as proxies for firm, tradeable quotes. We present a comparison of prices (and volumes) derived from Reuters D2000–2 electronic inter-dealer broking system with contemporaneous data from EFX. Tick-by-tick data is available from both sources, covering October 6-10, 1997. Our main comparative results are as follows. EFX midquote returns are consistently more volatile than their D2000–2 counterparts and display strong moving average effects which are not present in the D2000–2 returns. EFX spreads bear little or no relation to the inside spreads derived from D2000–2. In terms of information flows, D2000–2 returns lead those on EFX by up to 3 minutes and, further, contribute around 90% of all information impounded in quotes. A bivariate GARCH analysis also indicates a dominant role for D2000–2 in price discovery. On the positive side, however, EFX quotation frequency correlates well with D2000–2 transaction frequency.

*Preliminary, comments welcome. The authors research papers can be downloaded from their homepages, `cep.lse.ac.uk/~jond` for Danielsson and `cep.lse.ac.uk/~payne` for Payne. Their respective e-mail addresses are `j.danielsson@lse.ac.uk` and `r.payne@lse.ac.uk`. We would like to thank Casper De Vries and Charles Goodhart for helpful comments and Reuters Group PLC and Olsen and Associates for providing the data.

1 Introduction

In recent years, there has been a large increase in the amount of research on the behaviour of high–frequency exchange rate quotations. This increase has been due to both theoretical advances and improvements in FX data availability. Intra–day exchange rate data provides a natural testing ground for market microstructure hypotheses and models of time–varying volatility and the increased availability of such data, largely due to the willingness of certain corporations to make data available to academics at low cost, has been a strong impetus to research. Finally, as foreign exchange markets have the highest transaction volumes of any markets, foreign exchange data has unique potential to answer outstanding questions in market microstructure.

The data used in previous academic work has primarily been derived from the pages of Reuter's information systems, specifically the EFX page.[1]. EFX data, however, has a number of shortcomings. First, and most importantly from a microstructure perspective, it contains no measure of traded currency volumes. This renders many interesting microstructure hypotheses untestable. Second, the bid and ask quotes derived from EFX screens are *indicative* rather than *firm*. This means that such quotes are not binding commitments to trade from the originator and hence they may not be accurate measures of tradeable exchange rates. Also, whilst EFX system gives a timestamp for the entry of a quote pair, no such timing is given for the exit of quotes. Hence there is no information on the effective lifetime of EFX quotes. Lastly, each EFX bid and ask quote pair is input by a single dealer. As such, these quotes are likely to reflect dealer specific characteristics (e.g. inventories or beliefs) and may be a poor representation of 'market quotes'.

Despite these shortcomings, EFX data has been used in numerous studies. EFX midquotes are widely used as proxies for traded exchange rate prices in the study of intra-day exchange rate volatility (Baillie and Bollerslev (1991), Dacorogna, Müller, Nagler, Olsen, and Pictet (1993), Andersen and Bollerslev (1997) and Payne (1996)), triangular arbitrage relationships (de Jong, Mahieu, and Schotman (1998)) and intraday technical trading rule performance (Curcio, Goodhart, Guillaume, and Payne (1997)). EFX spreads have been used as measures of FX market liquidity in studies such as Bollerslev and Melvin (1994) and Hartmann (1996). Finally, EFX quote frequency is used as a proxy for traded currency volumes in Bollerslev and Domowitz

[1] EFX is the current equivalent of the older data feed called FXFX.

(1993) and Melvin and Yin (1996).

The objective of this study is to assess the validity of these assumptions by comparing the properties of the EFX quote data with those of tradeable, at the market quotes. This is done via a comparison of series derived from EFX with corresponding measures derived from Reuters D2000–2 electronic broking system for one week in October 1997.[2] D2000–2 is an automated inter-dealer trading system widely used by European and North American currency traders. As such, the bid and ask quotes derived from the system are *firm*. Second, the best bid and ask quote series from D2000–2 are calculated from those limit orders input by a large number of dealers and are hence likely to represent inside, market quotes much better than the EFX data. Finally, in conjunction with the D2000–2 quote series, we have a tick–by–tick indication of traded USD/DEM volumes.

Our analysis occurs in four stages. In a first step we compare the intra–day market activity patterns (intra-day seasonals) derived from D2000–2 and EFX data. We then go on to give a statistical comparison of the midquote data from both sources, examining return variance and measures of return dependence amongst other things. Our third exercise examines the information content of the D2000–2 and EFX midquote series. This is done via the cointegrating VAR model introduced in Hasbrouck (1995). Finally, we analize information flows between D2000–2 and EFX from a volatility perspective, employing the bivariate GARCH parameterization due to Engle and Kroner (1995).

Our main results are as follows. A graphical analysis of activity patterns from EFX and D2000–2 demonstrates that the intra–day pattern in EFX quote frequency correlates very strongly with that in traded D2000–2 volume. In contrast, the intra–day pattern in EFX spreads shows them to be poor measures of the inside, market spread. This is likely to be due to the single dealer nature of the EFX data. Our basic statistical analysis demonstrates EFX returns to be excessively volatile when compared to D2000–2 quote returns. Further, D2000–2 returns are essentially uncorrelated (as efficient market theory would predict) whilst EFX returns display significant negative first order autocorrelation. Hence the negative moving average component which is a prominent (and widely discussed) feature of much prior work on EFX data is shown

[2]The data covers the week of October 6 to October 10, 1997. The Reuters data is available for academic research from the Financial Markets Group at the London School of Economics. See `http://cep.lse.ac.uk/fmg`. The EFX data was obtained from Olsen and Associates. See `http://www.olsen.ch`.

to be a facet of their indicative and single dealer nature.

We believe that many of the results reported above (and subsequently) derive from the manner in which dealers set EFX quotes. In particular, our hypothesis is that, in general, a dealer inputting an EFX quote only wishes to trade on one side of the market, for example the bid. The trader therefore enters a competitive bid quote. By the nature of EFX, however, he is required to enter an offer also and this is formed by adding a standard (relatively large) spread to the bid. Such behaviour will imply insensitivity of spreads to market liquidity and 'excess volatility' in EFX quotes precisely as described above.

Through a cross-correlation analysis, we go on to demonstrate that EFX returns tend to lag those on D2000–2 by around 2 to 3 minutes on average. Refining our lead–lag analysis via the use of a cointegrating VAR technique, we find that around 90% of all USD/DEM relevant information is impounded into the exchange rate via D2000–2 and from there enters the EFX data. Hence, the extent to which the EFX system contributes to price discovery can be seen to be negligible, but relevant information is absorbed with a very short lag. Finally, our bivariate GARCH analysis shows that, while there are volatility spillovers in both directions between D2000–2 and EFX, spillovers from D2000–2 to EFX are stronger than those with the reverse causality.

The structure of the paper is as follows. The next Section contains a brief description of the nature of the spot FX markets and how the D2000–2 and EFX systems fit in to the market. Sections 2.2 and 2.3 give an overview of the two data sources and the structure of each data set. Section 3.1 provides our analysis of the activity patterns on EFX and D2000–2, mainly through a set of graphics. In Section 3.2 we describe the basic statistical nature of the two sets of midquote returns. Section 4 contains an analysis of the information content of each of the two return series and our bivariate GARCH results. Section 5 provides conclusions and presents ideas for further work.

2 The Spot FX Market

In this section we give a brief overview of the structure of the markets for spot foreign exchange, and the origins of our data. We then give a more detailed explanation of the D2000–2 and EFX data structures. Appendix A contains detailed information on the processing of the D2000–2 data.

2.1 Spot FX Market Structure

The spot FX market has grown tremendously in recent years. According to the BIS surveys of FX activity in 1995 and 1998 (Bank for International Settlements (1998)), total spot volume in all exchange rates has risen annually by 4% over the last three years to a daily turnover figure of just over $590bn. This represents a slowdown in the growth of the spot markets from that experienced in the period covering 1989 to 1995. Futures and swap markets, however, have grown by 10% annually over the same period. Of the global turnover in FX markets the most commonly traded currency was the US dollar which was party to 44% of transactions, and the second largest was the German mark with 15%. The largest trading center was London, with 32% of global activity.

Spot turnover is not generated by a single coherent market however. Transaction activity can be divided into a number of segments, classified by the participants to a given FX transaction and the means by which trade occurs. Activity can be subdivided into inter–dealer and customer–dealer trade.[3] Of these, inter–dealer trade is by far the largest, accounting for around 75% of all spot activity. The importance of customer–dealer trades should not be discounted however. They are completely opaque and strong anecdotal and empirical evidence suggests that dealers derive important information regarding the state of the market and likely evolution of exchange rates from their customer contacts (see Lyons (1995), Yao (1997) and Payne (1999) for empirical analysis of asymmetric information in FX markets based on the information contained in customer order flow.)

Inter–dealer activity itself is heavily segmented. Trades can occur in three basic ways. The first of these is direct inter–dealer trade. Here, dealers interact bilaterally, primarily via electronic communications systems such as the Reuters' D2000–1 system. The initiator of the communication contacts the second dealer and requests two–way prices which are recognized as good for a given normal trade size. The initiator then decides whether to buy or sell and for what quantity. If a transaction results, the occurrence of the deal and its characteristics are known only to the counterparties involved. Historically, another important method of inter–dealer trade was via voice brokers. These are intermediaries whose sole activity is the matching of buy and sell requests submitted by individual dealers. Information on trades consummated through voice brokers is available through intercom systems located on the desk of

[3] A 'customer' is defined as any non-dealer transaction participant.

each spot dealer. The final segment of activity, electronically brokered trade, has grown enormously in the last few years, mainly at the expense of voice brokered trade. The two purveyors of electronic brokerage services are EBS and Reuters, the latter through their D2000–2 system from which our data is derived. Both function as essentially closed order driven systems with liquidity supply via limit order and liquidity drained through market orders, and the direct crossing of bid and offer limit orders. Activity on these systems is observable to any subscriber to the service via the EBS/D2000–2 screen, aside from information on the identity of those involved in trading and quoting activity.

From the preceding discussion it is clear that spot FX order flow is heavily fragmented with the different market segments having varying levels of transparency. One source of FX price information which is available to all market participants, however, is that provided by the Reuters' EFX system. This screen provides a continuously updated sequence of exchange rate quotations, also advertising the institution which quoted. EFX does not contain any data on traded volumes but gives an evolving picture of the quotes available from other dealers. These quotes, however, are not 'firm' but 'indicative' i.e. they do not present a binding commitment from the advertising institution to trade at these prices. It has been argued in prior studies of FX market microstructure that reputation considerations would almost force those submitting EFX quotes to treat them as firm, but this is an assertion which has not been empirically validated. The main objective of the current study is to analyze how the indicative and single–dealer nature of EFX quotes affect their accuracy as measures of true market prices. Further we seek to examine whether activity patterns from the EFX system give a fair representation of actual trading activity on the USD/DEM market.

2.2 The Reuters' D2000–2 Dealing System and Data

The Reuters D2000–2 data set consists of all entries onto the D2000–2 system for the week of October 6 to October 10, 1997. The system display and its basic trading mechanism are as follows.

A subscriber to D2000–2 sees the following items on the trading screen, for up to 6 exchange rates;

- Best bid and offer limit order prices

- The quantity available for trade at the best bid and offer
- An indicator of the characteristics of the last trade

These data items are available to us on a tick–by–tick basis. Furthermore, the data made available contains information not available to market participants, specifically every subsidiary limit order on the D2000–2 order book at every point in time.[4] The entry and exit time of all limit and market orders are supplied to the one hundredth of a second. Hence we can examine variations in liquidity supply which are unknown to those actually participating in trade.

The basic trading mechanism of the system is as follows. Limit orders are queued via price and then time priority. In general, market orders will hit the best outstanding limit order on a given side of the market. There are exceptions to this rule however. In a small number of cases in the data set a market order failed to complete as the submitter of the order was credit constrained. Second, for any transaction to occur between D2000–2 participants, the participants must have opened bilateral credit lines. Hence, at some points in time, the submitter of a market order may find the best limit order unavailable to him as no such credit channel has been opened with the initiator of the limit order. This implies that market orders may occur outside the touch.

Finally, trade also occurs when bid and offer limit orders cross i.e. the book contains a bid limit order with price greater than or equal to the best outstanding limit offer.[5] These crosses occur automatically on the system and are straightforward to retrieve from the data set supplied. However, the fact that all participants on D2000-2 may not have agreed credit also leads to a number of situations when the bid-ask spread on D2000-2 is negative as (bid and offer) orders that should cross do not do so. We delete such observations before constructing the data used in this study.

For the five trading day period included in the data set, there were 130535 system entries, with most occurring between 6 and 16 GMT. There are four main types of system entry, these being;

1. Bid limit order entry

[4]By a subsidiary limit order we mean an offer order with price above the current best or a bid order with price below the best.

[5]The price of such a transaction is that of the limit order entered earliest i.e. the system treats the order entering latest similar to a market order.

2. Offer limit order entry

3. Take: a market buy order

4. Hit: a market sell order

In addition there are a few other entry types, most of which can be reclassified within one of the four preceding categories. The most common of these other entries are IDEALs. These are the simultaneous input of bid and offer limit orders, each for the same quantity, by a given participant.

In our empirical analysis, the following basic D2000–2 variables were employed. First, the best bid and offer in the system at every observation point were used as our basic buy and sell prices. From these, the midquote and bid–ask spread were constructed. Further, as measures of FX market activity we constructed time series of the number and aggregate quantity of limit orders outstanding as well as transaction volume. The structure of the raw data set and construction of these variables are discussed in Appendix 1.

2.3 The EFX Data

A widely available source of price information in the FX market is given by the quote stream appearing on the Reuters EFX page. These data have also been extensively employed in the extant academic literature on FX market microstructure. See Dacorogna, Müller, Nagler, Olsen, and Pictet (1993), Bollerslev and Melvin (1994), Andersen and Bollerslev (1997) and Payne (1996) among others.

Over the sample period there were 32121 quotes entries on the EFX system. As detailed in Section 1, these data are less rich than the D2000–2 data. The quotes given are not 'firm' and there is no indication of traded volume. Furthermore, rather than representing the best outstanding bid and offer prices, the EFX quotes are from a single dealer and, as such, are likely to reflect idiosyncrasies in his or her position and information relative to the state of the market as a whole.

As the EFX data have been widely used and their properties are well understood, we present only a cursory description of the data structure. Each line of the data file contains the following items;

- Quote entry timestamp

- Bid and offer quotes
- Codes detailing country, city and institution of submitter

In our analysis, information on the location of the submitting institution is ignored. The stream of single–dealer quote pairs is treated as a homogeneous event time series which is converted to a fixed calendar time sampling for many of our empirical exercises.[6]

3 Statistical Features of D2000–2 and EFX data

In order to assess the quality of the EFX data as proxy for prices, we present a series of statistical results designed to demonstrate the key features of the data. Raw (tick–by–tick) data does not lend itself readily to this comparison. While the D2000–2 data is time stamped to 1/100th of a second, the EFX data has a minimum time between quotations of 2 seconds. In addition, the noise in ultra high frequency data is considerable. We therefore converted the tick–by–tick D2000–2 and EFX data into calendar time–series with a 20 second sampling frequency. The final observations in each 20 second interval were taken as valid at the end of the interval and all D2000–2 trade activity within each interval was aggregated.

It has been suggested to us that our sampling convention of taking the final EFX observation in each interval biases our results towards making the EFX data look bad due to their single dealer nature. The suggested alternative was to attempt to combine information from current and past EFX quotes in order to gain a set of bids and offers which better approximate at the market quotes.[7] Our response to this is two-fold. First, more or less all prior papers using EFX have used the quotes precisely as we do currently.[8] Second, implementation of this suggestion implies that one needs to use a necessarily ad hoc procedure for estimating the lifetimes of EFX

[6] In order to convert both EFX and D2000-2 data from event to calendar time, the following procedure was used. For EFX quotes, the final observation pair in each calendar time interval was recorded. The D2000-2 quotes used are the best limit bid and offer prices outstanding at the end of each interval. For EFX quote frequency and D2000-2 transaction frequency/volume, the number/quantity of such events occurring in each interval was calculated. Finally, the number and quantity of limit orders outstanding at the end of each calendar time interval was recorded.

[7] A very simple example would be to take the highest bid and lowest offer from the last 10 EFX quote pairs as the 'market' bid and offer at every observation time.

[8] The only exception to this of which we are aware is Bollerslev and Domowitz (1993).

quotes.[9] Hence, we continue with the sampling scheme mentioned above.

Our analysis in this section breaks down into two segments. First we present the intra–day seasonal patterns in the EFX and D2000–2 data. We then go on to analyze the statistical characteristics of quote returns from the two systems before presenting raw correlations between series of interest.

3.1 Seasonal Patterns

Figure 1 presents a time plot of liquidity supply on D2000–2, measured by both the aggregate quantity in $million and the number of limit orders outstanding in a given 20 second interval. These figures clearly show the existence of strong intra-day seasonal patterns in D2000–2 activity. The system is very quiet in the GMT evening and overnight period with very light liquidity supply whilst from 6 to 18 GMT outstanding liquidity on the system is very high. Note that the hours during which D2000–2 is busy correspond broadly to European and North American trading hours, reflecting both the lack of impact of the D2000–2 system in Asia as well as the pre–eminence of London as FX trading center.[10]

Our first comparison of the D2000–2 data and statistics derived from the EFX data examines the correspondence between EFX quote frequency and D2000–2 transaction frequency. This is presented in Figure 2. In general, Figure 2 shows that EFX quote frequency correlates very strongly with D2000–2 transaction frequency. Both are high relative to their unconditional means in the periods covering 6 to 10 GMT and 12 to 16 GMT, while both measures are close to zero on average from 18 GMT to 6 GMT. The only real difference comes in the period from 10 to 12 GMT when EFX quote frequency stays relatively high while transaction frequency dips strongly on D2000–2. Nonetheless, the overall impression from Figure 2 is that the patterns in EFX quote frequency are likely to represent those in transacted volumes very well. The conclusion is confirmed by the evidence in Figure 3, which demonstrates that (as one might expect) the seasonal patterns in D2000–2 transaction frequency and unsigned transaction volume are almost identical.

The comparison between EFX spreads and their D2000–2 counterparts is given in Figure 4. This figure paints a far less impressive picture of the extent to which EFX

[9]Further, there is no guarantee that such an ad hoc scheme will generate sensible output in terms of non-negative spreads, for example.

[10]Electronic broking in Asia is dominated by the EBS consortium through the Minex system.

tracks true market conditions as proxied by those on D2000–2. The first fact apparent from Figure 4 is that there is essentially no intra–day seasonal pattern in the EFX bid–offer spread. D2000–2 spreads, on the other hand, vary widely across the GMT day. During European and North American trading hours D2000–2 spreads are an order of magnitude lower than those observed in the GMT evening and overnight period. As such, D2000-2 spreads seem to follow a similar intra-day pattern to the U-shape found for spreads observed on many major stock markets (see Foster and Viswanathan (1990) and Biais, Hillion, and Spatt (1995) for results from the NYSE and Paris Bourse.) A closer look at D2000–2 spreads, however, reveals a small increase in average spreads around midday such that the intra-day pattern is more of a W rather than U-shape.

Hence, in comparison we see that EFX spreads are much greater than those on D2000–2 during peak D2000–2 trading hours and much lower in the complementary period of the day. This immediately implies that the EFX data is likely to understate true market liquidity throughout the hours from 6 to 18 GMT and overstate liquidity from 18 to 6 GMT.

We believe that these features of EFX spreads are caused by the fact that 40T000 [illegible] midquote returns are distorted also. EFX quote frequency on the other hand, shares a very similar seasonal pattern to aggregate liquidity demand measures on D2000–2 and, as such, may be considered a good indicator of market trading activity.

3.2 Characterizing D2000–2 and EFX Returns

A complementary picture to that derived from activity patterns emerges from a statistical analysis of the characteristics of the EFX mid–quote return, as well as D2000–2 transaction price and midquote returns.

Tables 1 and 2 present the first four moments of the two midquote return series along with their first autocorrelation and a fifth order Box–Ljung statistic. Table 3 contains identical statistics for D2000-2 transaction price returns. Given the evidence of strong seasonal patterns in activity from Figures 1 to 5, statistics are computed for seven non–overlapping subsamples of the entire data sample. The first six of these subsamples represent data from two hour sections of the trading day, starting with the period from 6 to 8 GMT and ending with 16 to 18 GMT, aggregating across all days in the sample. For example, the third subsample contains observations from the 10 to 12 GMT time period of all 5 trading days, ordered by time of entry.[11] The final subsample consists of all other returns i.e. those observed between 18 and 6 GMT.

It is immediately clear that EFX midquote returns are around 50% more volatile than those on D2000–2. This links nicely with the analysis of patterns in absolute returns from Section 3.1. Interestingly, all volatility series are inversely related to measures of market activity and liquidity i.e. at peak trading times, volatility is at its lowest level. This is likely to be due to the effect of illiquidity on 20 second midquotes in thin trading periods. Figure 6 gives a graphical depiction of this result using the estimated return variances. Return skewness is higher for both D2000–2 return series than for EFX, and interestingly, the sign changes in the skew between subsamples are the same in all three series. Analysis of fourth moments shows D2000–2 returns to have consistently greater excess kurtosis than returns on the EFX system. However, tests for the existence of the fourth moment of returns, using a procedure proposed by Danielsson and de Vries (1997) indicate that the fourth moment, and hence the kurtosis, is unbounded in all cases.

[11] Clearly, then, each subsample contains 4 breaks (between the last observation in that period on day k and the first on day $k+1$. When appropriate, these breaks are modelled in estimation using dummy variables.

Much prior work using EFX data has noted that returns contain a negative moving average component. Various explanations have been put forward for this phenomenon, e.g. the effect of idiosyncratic inventory positions of individual dealers, differing dealers working on different information sets and noise in the EFX data. Analysis of our data demonstrates that the EFX data used here displays negative first order autocorrelation also. The Box–Ljung statistics in Table 2 indicates that the hypothesis of a lack of up to fifth order autocorrelation can be rejected at 5% in all of the subsamples. A very different picture emerges when examining the D2000–2 data. Table 1 shows that in active periods (i.e. 6 to 16 GMT) autocorrelation in the midquote return is statistically and economically insignificant. It is only in periods of illiquidity that returns demonstrate any dependence. Analysis of transaction price returns in Table 3 confirms this result.

Again, we could form an explanation of the magnitudes of first order autocorrelations using our hypothesis for EFX quote setting behaviour. Assume an agent who wishes to trade at the bid. He submits a competitive bid to EFX and computes the ask by adding a large, standard spread. Another agent who, instead, wishes to trade at the offer will enter a competitive offer and subtract the standard spread to gain his bid quote. Note that, even if both of these agents agree on the competitive bid and offer, the midquote submitted by the first will exceed that submitted by the second. Then, if agents wanting to trade at the bid and ask arrive randomly to EFX, quote returns will contain negative first order autocorrelation. This is exactly the intuition used in the spread estimator of Roll (1984).

Table 4 presents mean levels of market activity measures derived from the D2000–2 and EFX data. These numbers confirm the evidence from seasonal patterns in Figures 1 to 5. The measures of liquidity supply (number of orders outstanding and their aggregate size) are at peak levels between 8 and 10 GMT, the number of orders dips slightly between 10 and 12 GMT before rising again and tailing off towards 18 GMT. As one would expect, D2000–2 bid–offer spreads move inversely to liquidity supply while the lack of variability in EFX spreads is clearly identifiable. Transaction activity follows a very similar pattern to orders. At peak times each 20 second interval contains 3 or 4 transactions on average, with a maximum recorded intensity of close to 40 deals in one interval.

A preliminary look at the relationships between those market activity measures derived from the D2000–2 and EFX data is given in Table 5 in the form of contemporaneous cross–correlations. Again, these results tend to confirm the insights from the seasonal patterns. D2000–2 spreads are strongly negatively correlated with D2000–2 liquidity supply measures. D2000–2 transaction intensity correlates far less well with spreads however. EFX spreads are extremely poorly correlated with all other series while EFX quote frequency does a fairly good job of predicting both D2000–2 liquidity and transaction activity. Hence, prior insights are corroborated in that EFX spreads seem to convey little information regarding underlying spot market activity while EFX quote frequency can be regarded as a fair proxy for volumes and liquidity supply.

4 The Inter–Relationship Between D2000–2 and EFX Quotes

Much prior FX market microstructure research has used the indicative EFX quotes as proxies for firm quotes. However, certain authors have questioned the accuracy of the EFX quote process in this context. A standard concern about the EFX data is that in busy market hours, dealers may not submit information to EFX as they are concentrating on dealing over the phone and their trading screens. Alternatively, it may be the case that bounds on the processing capability of the EFX system imply that the EFX screen does not represent the information submitted to the system precisely.

Below we examine the relationship between firm D2000–2 midquotes and their indicative EFX counterparts using three approaches. First a set of Q–statistics for cross–correlations between EFX and D2000–2 at various leads and lags are presented. These statistics are presented separately for those time–of–day subsamples defined in Section 3.2. Second, a cointegrating VAR for the returns is estimated and the information share of each midquote process is derived via the methodology in Hasbrouck (1995). Again, separate estimations are undertaken for each of the seven intra–day subsamples. Finally, a bivariate GARCH model is estimated for the residuals from the conditional mean VAR structure. This enables us to analyse the extent to which D2000–2 and EFX volatility are linked.

4.1 Cross-correlation Analysis

Table 6 contains Box-Ljung Q-statistics for cross correlations between D2000-2 and EFX midquote returns for the 7 previously defined subsamples of the GMT day.[12] There are four Q–statistics associated with each subsample and these are calculated as follows. Q_5 is the statistic relevant to cross–correlation between D2000–2 returns and lags 1 to 5 of EFX returns. Similarly, Q_{10} is relevant to correlation between D2000–2 returns and lags 6 to 10 of the EFX return. Q_{-5} and Q_{-10} are calculated analogously correlating EFX returns with lagged D2000–2 returns.

Figure 7 plots the cross correlations between D2000–2 and EFX returns calculated using data from the period 6 to 18 GMT. It is immediately clear that there is a strong asymmetry in correlation with those for lags -1 to -7 (approximately) being significantly positive while those from 1 to 7 are not. The implication is that D2000–2 returns tend to lead EFX returns i.e. EFX returns are predictable with D2000–2 returns while the converse is not true. Given the 20 second sampling of the data, the estimated cross-correlations imply that EFX returns are predictable between 2 and 3 minutes ahead using D2000–2 returns. As the D2000–2 midquote is formed from firm quotes and the EFX midquote is indicative this is in line with intuition.

The results in Table 6 give a similar picture to that on Figure 7. However, for most periods of the day, Table 6 shows that the first 5 lags of EFX returns have some predictive power for current D2000–2 returns. In general though, for peak trading hours the statistics Q_{-5} and Q_{-10} are an order of magintude greater than Q_5 and Q_{10} implying D2000–2 returns affect subsequent EFX returns to a far greater degree than the converse.

4.2 Cointegrating VAR analysis

In Section 3 we demonstrated that the midquote returns derived from D2000–2 and EFX differ substantially along certain dimensions. EFX returns are more volatile than their D2000–2 counterparts and also have stronger temporal dependence. Further, Section 4.1 showed that D2000–2 returns predict their EFX counterparts far better than in the converse direction. These observations, however, do not imply that the EFX data is inferior in terms of the speed at which USD/DEM relevant

[12]Tables 1 and 2 indicate that both D2000–2 and EFX returns contain MA(1) components. We filter the MA(1) structure from each sample before constructing the cross-correlations.

information is assimilated into prices. To make such as assertion, a different analysis is needed.

To get at this issue we employ the cointegrating VAR framework developed in Hasbrouck (1995). This structure permits us to study the speed of information assimilation into D2000-2 and EFX quotes and to compute the contribution of each series to overall price discovery. The basic empirical model is as follows;

$$r_t = \mu + \alpha z_{t-1} + \beta(L) r_{t-1} + \epsilon_t, \quad E(\epsilon_t \epsilon_t') = \Omega \tag{1}$$

where $r_t = (r_t^{D2}, r_t^{EFX})'$, is a vector containing the D2000–2 and EFX midquote returns, $\mu = (\mu_1, \mu_2)'$, $\alpha = (\alpha_1, \alpha_2)'$, $\epsilon = (\epsilon_{1t}, \epsilon_{2t})'$ and $\beta(L)$ is a conformable polynomial in the lag operator. Equation (1) is just an error correction representation for the pair of return series where we have assumed that the difference between the two midquote variables is $I(0)$ i.e. $z_t = q_t^{D2} - q_t^{EFX}$. This assumption guarantees that the two price series cannot diverge and is shown to be valid through a series of unit root tests on the midquotes and the difference between them.[13] We refer to this difference as the pricing error from now on. In equilibrium the EFX and D2000–2 midquotes are identical (i.e. $z_t = 0$.) This implies that one measure of each system's contribution to price discovery can be gained through comparison of the α coefficients. Given the way in which we have constructed z_t one would expect α_1 to be negative and α_2 positive, but an asymmetry in their sizes would indicate one system reacting more to deviations from equilibrium than the other and hence an asymmetry in price dis-

Imeasure

con4

thee

The long run impacts of both innovations on both return series are summarized in the value of $\Phi(1)$. Note that, due to the cointegration between the pair of return series, $\Phi(1)$ must contain two identical rows. Denote a row of $\Phi(1)$ by ϕ.

If we now assume that the innovations to the VAR are uncorrelated, the calculation of the information share of each market is straightforward. In particular, the information share for market j is calculated as;

$$S_j = \frac{\phi_j^2 \Omega_{jj}}{\phi \Omega \phi} \tag{3}$$

However, this assumption is unlikely to be satisfied. Hence we must modify the preceding estimator. In order to bound the information share we use a Choleski decomposition of the variance–covariance (VCV) matrix. An upper bound on the information share of system i is obtained from this decomposition when the innovation to equation i is represented in the first row of the VCV matrix. A lower bound on i's information share is obtained when system i is represented in the second row of the VCV. Denoting the Choleski factor of Ω by F, the information share of series j is then;

$$S_j = \frac{([\phi F]_j)^2}{\phi \Omega \phi} \tag{4}$$

where $[\phi F]_j$ is the j^{th} element of the given row vector. In the empirical results which follow, we present only the lower bound on the D2000–2 information share.

The properties of the pricing error (z_t) are presented in Table 7, broken down across intervals of the trading day. Several facts are immediately apparent. First, the difference between the two midquote series is small on average. Second, the variance of the pricing error covaries negatively with D2000–2 trading volume. In times of heavy D2000–2 activity D2000–2 and EFX midquotes stay closer together on average. Lastly, the dependence of the pricing error, measured by the first order autocorrelation and a fifth order Box–Ljung statistic, is also inversely related to D2000–2 volume. Hence, when trading volume is large the pricing error is less persistent, implying that deviations from equilibrium are removed more speedily. Also, this final result suggests that the ECM structure in equation (1) will not be stable across the trading day and we therefore estimate the ECM separately for each of our trading day subsamples.

Table 8 contains a selection of the estimated ECM parameters for all seven trading day subsamples plus the estimated D2000–2 information share for each. The crucial ECM parameters are α_1 and α_2 and we expect the former to be negative while the latter should be positive. For all of the trading day subsamples, the sign of α_2 is as expected i.e. when the D2000–2 midquote exceeds the EFX midquote, the EFX midquote adjusts upwards. In only 4 of the 7 subsamples does α_1 take the expected sign.

With regard to the relative size and significance of the parameters on the pricing error, a clear asymmetry is visible. For the subsamples covering 6 to 18 GMT, α_2 is always an order of magnitude larger than α_1 and is always more significant. The lagged equilibrium error is always significant in the EFX equation but is only significant in the D2000–2 equation in 3 of 6 cases. Further, the R^2 for the EFX return equation is always an order of magnitude greater than that for the D2000–2 equation. These results imply that the majority of any dis-equilibrium in the system is removed through the adjustment of EFX quotes. D2000–2 quotes react very weakly to dis-equilibrium.

The exception is for the overnight subsample (6pm to 6am.) In this case the parameters on the lagged equilibrium error have similar size and significance level. This is likely due to the lack of activity on both D2000–2 and EFX during this interval.

A final point regarding the ECM results relates to those parameters which have been omitted from Table 8, specifically the coefficients on lagged EFX and D2000–2 returns in each equation.[14] For the D2000–2 return equation, neither lagged own returns nor lagged EFX returns are significant. On the other hand, both sets of lagged returns are significant in the EFX equation with lagged D2000–2 returns generally having a positive effect and lagged EFX returns a negative impact on current EFX returns.[15] Note that as all equilibrium variation in the EFX quote occurs through the coefficient on z_{t-1} this predictability reflects inefficiency in the EFX quote process.

Finally, the last column of Table 8 presents the estimated lower bound on the D2000–2 information share for each subsample. It is clear that, for those hours of the day when D2000–2 is active (i.e. 6am to 6pm), it is the dominant location for price

[14] Full results are available upon request from the authors.

[15] These results are also consistent with our autocorrelation analysis from Section 3.2.

discovery with minimal information shares in excess of 85%. For the overnight period, however, the lack of D2000–2 activity implies its share drops to around 35%. Within the effective trading day, the information share on D2000–2 follows an inverted U–shape, such that the information share is minimized in the early GMT morning and early GMT evening.

4.3 Bivariate GARCH Analysis

A standard interpretation of time–varying asset price volatility asserts that variation is a reaction to the arrival of new information. In our setting, information flows relevant to USD/DEM quotes on EFX and D2000–2 should be identical. If we further assume that the quote streams on the two systems incorporate information with equal speed then one would not expect volatility on one system to lead that on another. We examine this notion using a bivariate GARCH representation for the D2000–2 and EFX data. Rather than using the returns series as the basis for these estimations we have used the residuals from the VAR specifications of the previous subsection. The residuals are used so as to remove any covariation between the series which [illegible]

Estimated parameters from the BEKK specification for residual returns are given in Table 9.[16] From these parameters it is clear that both volatility series are strongly autocorrelated, a result which is in line with those from univariate GARCH specifications on intra-day FX data. The diagonal elements of A and B are all positive and strongly significant. An interesting feature of the results appears in the off diagonal elements. The upper right coefficients of A and B (i.e. a_{12} and b_{12}) are greater in magnitude and more significant than a_{21} and b_{21}. This would tend to imply that D2000-2 volatility affects EFX return volatility to a greater degree than in the converse direction. A further result which agrees with those from univariate volatility models (and the prior kurtosis measures for returns) is that the coefficients in Table 9 imply that the unconditional variance-covariance matrix does not exist, a multivariate equivalent of IGARCH.[17]

A clearer picture of such asymmetries may be apparent if we explicitly calculate the representation for each volatility series from the results. These are given below.[18]

$$\begin{aligned}
\sigma^2_{D2,t} &= 2.04 \times 10^{-10} + 0.09 r^2_{D2,t-1} + 0.012 r_{D2,t-1} r_{EFX,t-1} + 0.0004 r^2_{EFX,t-1} + \\
&\quad 0.92 \sigma^2_{D2,t-1} - 0.038 \sigma_{D2-EFX,t-1} + 0.0004 \sigma^2_{EFX,t-1} \\
\sigma^2_{EFX,t} &= 2.82 \times 10^{-09} + 0.020 r^2_{D2,t-1} - 0.092 r_{D2,t-1} r_{EFX,t-1} + 0.11 r^2_{EFX,t-1} + \\
&\quad 0.0064 \sigma^2_{D2,t-1} + 0.14 \sigma_{D2-EFX,t-1} + 0.81 \sigma^2_{EFX,t-1}
\end{aligned}$$

Observations from these representations are as follows. First, in line with results from univariate GARCH models of intra-day FX volatility, coefficients on lagged own conditional variance dominate coefficients on lagged own returns in size. Looking at the cross effects, our previous comments are borne out but the size of these effects is small on average. Coefficients on lagged D2000-2 variables in the conditional variance equation for EFX returns are an order of magnitude greater than coefficients on lagged EFX variables in the D2000-2 equation. Hence volatility spillovers, although small, seem more pronounced from firm to indicative quotes. Note also that the

[16]Previous research on intra-day FX volatility has shown that the intra-day seasonal patterns in volatility can bias estimated volatility process coefficients. Hence in Table 10 we present estimated coefficients from the bivariate GARCH using residual returns with deseasonalised volatility. Results on volatility spillovers from the two tables are qualitatively similar.

[17]The unconditional covariance matrix is given by $\Omega = (I - [A \otimes A]' - [B \otimes B]')^{-1} \ ec(V'V)$.

[18]We have not provided standard errors for these numbers are they are products of the estimated parameters.

conditional covariance between returns has much greater effects on EFX volatility than on D2000-2 volatility.

5 Conclusion

In recent years, a large literature on empirical FX microstructure has emerged. Most of this work has been conducted using data derived from indicative EFX quotes. This paper analyses the adequacy of such indicative quotes via a comparison of their properties with those of firm quotes drawn from the electronic FX broking system D2000–2.

Our results will probably come as little surprise to researchers who have worked with the EFX data before. First, EFX midquote returns are excessively volatile when compared to returns in the midquote derived from the best D2000–2 bid and offer i.e. the EFX indicative quotes are more volatile than their firm quote counterparts. Second, price relevant information is shown to be impounded in D2000–2 quotes first and such information only subsequently enters the EFX price. Across the entire trading day, this implies that EFX returns are predictable with D2000–2 returns observed two to three minutes previously. We also demonstrate similar causality patterns in the volatility series where lagged D2000–2 volatility affects current EFX volatility, but the converse is not true. It should be noted, however, the the volatility spillovers appear to be small in magnitude.

In prior research on EFX data, many researchers have noted that returns contain a strong, negative, first-order moving average component. Several explanations have been proposed for this phenomenon (including heterogeneous expectations, inventory discrepancies between dealers and noise,) most of which build on the fact that the EFX quote pairs are single-dealer quotations. Both D2000-2 midquote returns and transaction price returns contain no such moving average effects during peak trading hours (6 to 18 GMT). Hence, the dependence in EFX quotes is not representative of that in firm, market data and can be thought of as due to their indicative, single dealer nature.

With regard to the patterns in market activity statistics obtained from these data, the following results emerge. The bid-ask spread on D2000-2 broadly follows the U-shaped intra-day pattern observed in spreads on many other asset markets (e.g the NYSE and the Paris Bourse. See Foster and Viswanathan (1990) and Biais, Hillion, and Spatt (1995).) EFX spreads, however, are flat on average across the trading day and seem to contain zero information on the state of the market or the nature/position of the inputting dealer. The EFX data fares much better when quotation frequency is analysed. Specifically, we demonstrate that EFX quote frequency has a strong positive correlation with market order frequency on D2000-2. As such it may be considered a valid proxy for FX transactions frequency.

Most of the preceding results can be explained if our hypothesis regarding the EFX quote submission behaviour of dealers is valid. We hypothesise that most agents submitting to D2000-2 only really wish to deal on one side of the market, say the ask, and hence place a competitive ask quote. The bid quote is obtained by subtracting a standard, large spread from this ask. This implies that spreads will contain little or no information and quote returns will be excessively volatile and have non-zero first order autocorrelations. We are at present constructing a direct test of this hypothesis.

Further work in this area is likely to prove fruitful, not only for those attempting to understand FX markets in particular, but also those studying market microstructure in general. The D2000–2 data set is very rich, and we have only used a small portion of it in this work. There are many interesting questions that can be addressed the data, and we are currently working on a few. For example, it is possible to construct the entire excess supply and demand curves for foreign exchange from the D2000-2 limit orders, and we hope to analyze the dynamic evolution of these curves, and how they interact with price changes, volume, and the indicative quotes.

A Processing the D2000–2 data

The D2000–2 dataset has 130,535 observations, or entries, where each entry has a series of fields:

n index number of observation (only for limit orders)

t_1 entry time

t_2 exit time

E_1 Entry type

E_2 Entry result

p price

q quantity requested

qt quantity traded

S StatusPositioned field

M MsgnamePositoned field

The new entry type field E_1 almost always takes the values [Enter**Offer**, Enter**Bid**, Enter**Take**, Enter**Hit**]. Other values indicate errors. Offer in the terminology of the D2000–2 system is the same as ask. Hit and Take are market orders (hit for bid and take for ask)

The result field E_2 takes the values [Entry**Cancelled4**, Entry**Removed**, Entry**Hit**, Entry**Taken**].

The field S and M indicate what happens to entries, errors, etc.

A.1 Limit Orders

If E_1 = [Enter**Offer**, Enter**Bid**] it indicates a new limit order. This limit order can either enter into the limit order book, or if it matches or exceeds the best countertype, result in an instant trade, called cross. The remaining quantity, if any enters the limit order book. When a cross occurs we treat it as a new transaction with the transaction time as the later entry time. Note that a cross may not always happen in practice, see discussion below. The limit orders are sorted in price–age order, providing an entire supply and demand curve for the currency. The difference between the best bid and ask is denoted as inside spread.

A.2 Market orders

If E_1 = [Enter**Take**, Enter**Hit**] it indicates a market order. The price may equal or exceed the best price. In the latter case we record transaction time as market order exit time and transaction price as the price of the limit order with which the market order matched. Note that a transaction may not always take place as expected. See discussion below.

A.3 Events

Every time a new entry enters the system it triggers one or more events. It may simply enter the order book (one event) or as a market order be a party to one transaction and trigger 3 events (entry or market order, transaction, exit or limit order). As each limit order can be a party to many transaction, and each market order can hit/take many limit orders the number of events is large. A subset of the events is transaction events. In that case we only record one event for the result of the entire transaction, regardless of whether one or more limit orders were party to the transaction.

A.4 Aggregation

We can either use the data in event time or aggregate the date. When we aggregate, we sample the markets at fixed intervals, e.g. every 5 minutes, and use the current value of the order book, or the latest recorded transaction.

A.5 Processing

On order to get the necessary data we created a computational model of the market, where we mimic the market in a specialized computer program. This involves building and maintaining the entire limit order book at any given time point. The benefit is that one has a complete set of information about the markets at any time and can extract any dataset of interest. In addition, the integration of EFX quotes into the model is straightforward. Only a small sample of available information is used in this paper, but we use the data in other research papers.

If each limit and market order was processed according to the guidelines above, building the computational model would be relatively straight forward. However we encountered significant difficulties. In a number of cases where a trade should occur it does not. For example, say \$1m is available for sale at 1.76, and we observe a market order for the purchase of \$1m at 1.76, $(p = 1.76, q = 1)$ yet the transaction does not take place. In some cases the market order has $q_t = 0$, clearly indicating this, but in many cases the market order may take a lower priority offer (same price, later entry time). Many variations of this type exist. The reasons for this is typically that the counterparties may not have bi–lateral credit, or possibly a delay to processing time (entries are recorded and processed to the 1/100 of a second). However we do not observe this directly. If even one entry is processed incorrectly in the computational model, it causes serious problems in the processing of subsequent entries. In effect, one has to process each of these anomalous entries separately, complicating the modelling process[19]. We were able to validate our computational model by processing the entire dataset without an error, where the error checking was very strict. For example, if a market order does not hit the best limit order, but the second best, but one processes the data as if it hit the best, then the qt (quantity traded) field will be incorrect for both limit orders at exit time. By checking that the computational model does not result in such errors, a strong validation for the model is achieved.

[19]We received very helpful information from Reuters regarding the processing of anomalous events.

References

ANDERSEN, T., AND T. BOLLERSLEV (1997): "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High-Frequency Returns," *Journal Of Finance*, 52, 975–1005.

BAILLIE, R., AND T. BOLLERSLEV (1991): "Intra-Day and Inter-Market Volatility in Foreign Exchange Markets," *Review of Economic Studies*, 58, 565–586.

BANK FOR INTERNATIONAL SETTLEMENTS (1998): *Central Bank Survey of Foreign Exchange and Derivatives Market Activity in April 1998: Preliminary Global Data.* B.I.S., Basle.

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.

BOLLERSLEV, T., AND I. DOMOWITZ (1993): "Trading Patterns and Prices in the Interbank Foreign Exchange Market," *Journal of Finance*, 48, 1421–1444.

BOLLERSLEV, T., AND M. MELVIN (1994): "Bid-ask Spreads and Volatility in the Foreign Exchange Market: An Empirical Analysis," *Journal of International Economics*, 36, 355–372.

CURCIO, R., C. GOODHART, D. GUILLAUME, AND R. PAYNE (1997): "Do Technical Trading Rules Generate Profits? Conclusions from the Intra-Day Foreign Exchange Market," *International Journal of Finance and Economics*, forthcoming.

DACOROGNA, M., U. MÜLLER, R. NAGLER, R. OLSEN, AND O. PICTET (1993): "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance*, 12, 413–438.

DANIELSSON, J., AND C. G. DE VRIES (1997): "Beyond the sample: Extreme quantile and probability estimation," London School of Economics, FMG DP 298.

DE JONG, F., R. MAHIEU, AND P. SCHOTMAN (1998): "Price Discovery in the Foreign Exchange Market," *Journal of International Money and Finance*, forthcoming.

Engle, R., and K. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11(1), 122–150.

Foster, D., and S. Viswanathan (1990): "A Theory of the Interday Variations in Volume, Variance and Trading Costs in Securities Markets," *Review of Financial Studies*, 3, 593–624.

Hartmann, P. (1996): "Trading Volumes and Transaction Costs in the Foreign Exchange Market: Evidence from Daily Dollar-Yen Spot Data," Discussion Paper 232, Financial Markets Group, London School of Economics.

Hasbrouck, J. (1995): "One Security, Many Markets: Determining the Contributions to Price Discovery," *Journal of Finance*, 50, 1175–1199.

Lyons, R. (1995): "Tests of Microstructural Hypotheses in the Foreign Exchange Market," *Journal of Financial Economics*, 39, 321–351.

Melvin, M., and X. Yin (1996): "Public Information Arrival, Exchange Rate Volatility and Quote Frequency," Working Paper, Arizona State University.

Payne, R. (1996): "Announcement effects and seasonality in the intra-day foreign exchange market," Discussion Paper 238, Financial Markets Group, London School of Economics.

——— (1999): "Informed Trade in Inter-Dealer FX Markets: an Empirical Investigation," Working Paper, Financial Markets Group, London School of Economics.

Roll, R. (1984): "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139.

Yao, J. (1997): "Market Making in the Interbank Foreign Exchange Market," Working Paper, New York University.

Table 1: Summary Statistics for 20 second D2000-2 Midquote Return Subsamples

Data	Mean	Var.	Skew	Kurt.	ρ_1	$Q(5)$
6am to 8am	-0.000523	0.000615	-14.12	367.66	-0.03	9.91
8am to 10am	-0.000587	0.000479	-18.89	535.17	-0.02	4.41
10am to 12pm	-0.000401	0.000500	9.22	400.78	0.02	5.01
12pm to 2pm	-0.000293	0.001170	-12.77	418.83	-0.03	3.20
2pm to 4pm	-0.000329	0.000951	-18.17	510.75	-0.01	1.25
4pm to 6pm	-0.000413	0.000550	1.21	237.95	-0.13*	34.00*
6pm to 6am	-0.000017	0.000818	-0.79	128.87	-0.24*	633.80*

Notes: the first our columns of the table give the sample mean, variance, skewness and kurtosis of returns. The next column gives the first order return autocorrelation and the final column a fifth order Box-Ljung statistic for return dependence.

Table 2: Summary Statistics for 20 second EFX Midquote Return Subsamples

Data	Mean	Var.	Skew	Kurt.	ρ_1	$Q(5)$
6am to 8am	-0.000523	0.000902	-6.51	140.10	-0.18*	62.07*
8am to 10am	-0.000563	0.000702	-8.54	192.29	-0.19*	65.31*
10am to 12pm	-0.000440	0.000897	3.41	157.45	-0.17*	56.48*
12pm to 2pm	-0.000279	0.001549	-8.24	249.22	-0.15*	39.71*
2pm to 4pm	-0.000326	0.001501	-12.86	325.48	-0.21*	85.35*
4pm to 6pm	-0.000403	0.000834	1.76	88.71	-0.16*	46.33*
6pm to 6am	-0.000028	0.000384	-2.56	136.07	-0.13*	295.73*

Notes: the first our columns of the table give the sample mean, variance, skewness and kurtosis of returns. The next column gives the first order return autocorrelation and the final column a fifth order Box-Ljung statistic for return dependence.

Table 3: Summary Statistics for 20 second D2000-2 Transaction Return Subsamples

Data	Mean	Var.	Skew	Kurt.	ρ_1	$Q(5)$
6am to 8am	-0.000528	0.000616	-14.09	368.37	-0.03	3.81
8am to 10am	-0.000593	0.000498	-18.25	510.04	-0.02	3.41
10am to 12pm	-0.000400	0.000551	7.66	327.12	0.02	6.81
12pm to 2pm	-0.000285	0.001218	-11.51	370.68	-0.05	5.18
2pm to 4pm	-0.000329	0.000969	-18.13	504.41	0.00	3.03
4pm to 6pm	-0.000399	0.000424	4.23	384.24	-0.04	6.42
6pm to 6am	-0.000021	0.000443	-0.52	600.47	-0.13*	270.51*

Notes: the first our columns of the table give the sample mean, variance, skewness and kurtosis of returns. The next column gives the first order return autocorrelation and the final column a fifth order Box-Ljung statistic for return dependence.

Table 4: Summary Statistics for 20 second D2000-2 Order book and Transaction data

Subsample	D2 *sp*	EFX *sp*	Orders	Depth	Deals	Vol.
6am to 8am	0.016	0.040	54.62	99.03	2.96	5.26
8am to 10am	0.010	0.037	95.57	185.56	3.56	6.81
10am to 12pm	0.013	0.039	92.84	185.26	2.92	5.40
12pm to 2pm	0.013	0.038	98.06	176.77	4.77	8.81
2pm to 4pm	0.019	0.040	69.45	132.70	2.53	4.49
4pm to 6pm	0.040	0.037	32.39	68.59	0.38	0.61
6pm to 6am	0.165	0.036	11.10	27.33	0.08	0.11

Notes: *sp* is the percentage bid-offer spread. Orders and depth refer to the number of outstanding D2000-2 limit orders and their aggregate size respectively. Deals is a count of the number of transactions in each interval and Vol is the aggregate transacted volume in a given interval.

Table 5: Cross-correlations of D2000-2 and EFX activity variables

Data	Correlation						
D2 *sp*	1	-	-	-	-	-	-
Orders	-0.50	1	-	-	-	-	-
Depth	-0.47	0.95	1	-	-	-	-
EFX *sp*	0.04	-0.04	-0.05	1	-	-	-
EFX Quotes	-0.26	0.43	0.37	-0.01	1	-	-
Deals	-0.13	0.29	0.26	-0.00	0.38	1	-
Volume	-0.11	0.27	0.24	-0.01	0.34	0.94	1

Notes: *sp* is the percentage bid-offer spread. Orders and depth refer to the number of outstanding D2000-2 limit orders and their aggregate size respectively. EFX quotes refers to the number of EFX quotes posted in each interval. Deals is a count of the number of transactions in each interval and Volume is the aggregate transacted volume in a given interval.

Table 6: Q-statistics for cross-correlations between D2000–2 and EFX returns

Subsample	Q_{-10}	Q_{-5}	Q_5	Q_{10}
6-18 GMT	75.2	1194.0	149.5	19.2
6-8 GMT	13.9	145.2	16.7	7.4
8-10 GMT	20.1	167.0	58.0	6.3
10-12 GMT	98.9	528.3	42.0	21.8
12-14 GMT	8.2	317.3	36.2	7.7
14-16 GMT	14.1	165.3	37.5	6.3
16-18 GMT	3.9	31.9	31.7	4.4
18-6 GMT	2.7	12.5	4.9	5.3

Notes: The table presents the Q-statistics for cross-correlations between D2000-2 and EFX returns. Q_5 is the statistic relevant to cross–correlation between D2000–2 returns and lags 1 to 5 of EFX returns. Similarly, Q_{10} is relevant to correlation between D2000–2 returns and lags 6 to 10 of the EFX return. Q_{-5} and Q_{-10} are calculated analogously correlating EFX returns with lagged D2000–2 returns.

Table 7: Properties of Price error between D2000-2 and EFX

Subsample	Median	$\bar{z}$	σ_z^2	ρ_{1z}	$Q(10)$
6am to 8am	0	-1.5×10^{-5}	1.23×10^{-7}	0.36^*	416.3^*
8am to 10am	0	5.2×10^{-6}	9.11×10^{-8}	0.34^*	494.7^*
10am to 12pm	5×10^{-5}	2.3×10^{-5}	2.25×10^{-7}	0.57^*	1276.3^*
12pm to 2pm	5×10^{-5}	3.7×10^{-5}	1.22×10^{-7}	0.23^*	202.2^*
2pm to 4pm	0	-2.5×10^{-5}	1.34×10^{-7}	0.31^*	406.7^*
4pm to 6pm	5×10^{-5}	-8.4×10^{-5}	2.86×10^{-7}	0.59^*	1695.5^*

Notes: z_t is defined as the price error at t. ρ_{1z} is the first autocorrelation of the pricing error. $Q(10)$ is the 10th order Box-Ljung statistic for the pricing error.

Table 8: Cointegrating VAR results for D2000-2 and EFX returns

Subsample	Lags	α_1	t(α_1)	R^2_{D2}	α_2	t(α_2)	R^2_{EFX}	D2 share
6am to 8am	2	-0.05	-2.10	0.03	0.55	9.35	0.34	90.5
8am to 10am	1	-0.03	-1.77	0.02	0.56	16.26	0.33	94.7
10am to 12pm	1	0.09	1.65	0.03	0.48	3.93	0.34	93.0
12pm to 2pm	1	0.06	0.20	0.02	0.72	22.48	0.41	92.9
2pm to 4pm	1	-0.05	-3.05	0.04	0.51	11.82	0.34	89.9
4pm to 6pm	2	-0.05	-2.43	0.14	0.27	8.40	0.17	88.8
6pm to 6am	7	-0.03	-4.74	0.10	0.01	5.34	0.08	35.4

Notes: α_1 and α_2 are the coefficients on the lagged pricing error in the D2000-2 and EFX return equations respectively. The columns immedaitely following give the t-values for these coefficients. The R^2 for the D2000-2 and EFX return equations are also presented and, finally, we present the lower bound on the D2000-2 information share.

Table 9: Bivariate GARCH results: raw 20 second data

Parameter	Coeff	t-value
v_{11}	1.43×10^{-5}	38.37
v_{12}	1.33×10^{-5}	6.02
v_{22}	5.14×10^{-5}	52.95
a_{11}	0.96	1053.18
a_{12}	0.08	37.70
a_{21}	-0.02	-15.64
a_{22}	0.90	367.68
b_{11}	0.30	88.28
b_{12}	-0.14	-22.51
b_{21}	0.02	6.39
b_{22}	0.33	66.59

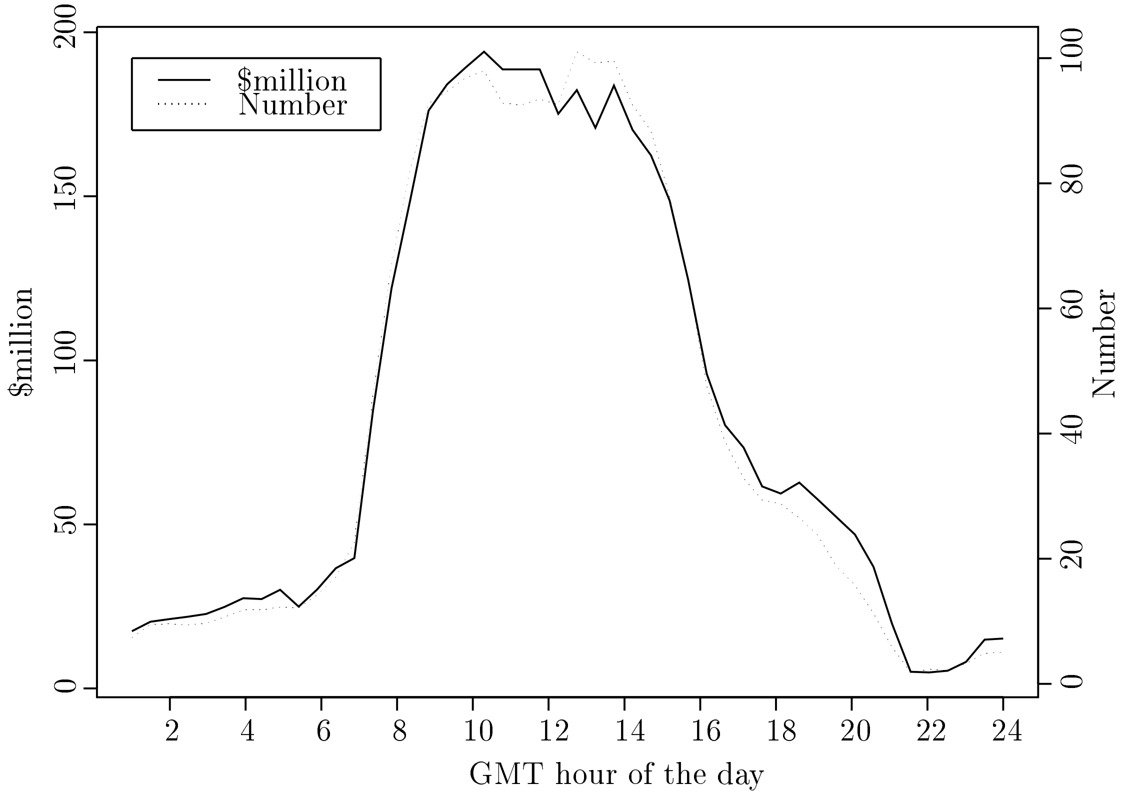
Notes: estimated coefficients from the bivariate GARCH specification from residual return data.

Table 10: Bivariate GARCH results: deasonalised volatility

Parameter	Coeff	t-value
v_{11}	0.18	19.87
v_{12}	-0.05	-0.49
v_{22}	0.15	2.34
a_{11}	0.89	112.39
a_{12}	0.54	33.16
a_{21}	0.17	6.26
a_{22}	-0.89	-103.58
b_{11}	0.34	95.19
b_{12}	-0.05	-12.68
b_{21}	-0.01	-2.18
b_{22}	0.27	54.05

Notes: estimates of the bivariate GARCH parameters from the residual return data with deseasonalised volatility.

Figure 1: Intra-day Seasonal Patterns in Aggregate Quantity ($m) and Aggregate Number of D2000-2 Limit Orders Outstanding



Notes: The basic data was constructed using a 20 second sampling. The limit order quantity data were then aggregated across 30 minute intervals of the trading day and averaged to give the mean 20 second limit order quantity within each 30 minute segment of the GMT day.

Figure 2: Intra-day Seasonal Patterns in EFX Quotation Frequency and D2000–2 Market Order Frequency

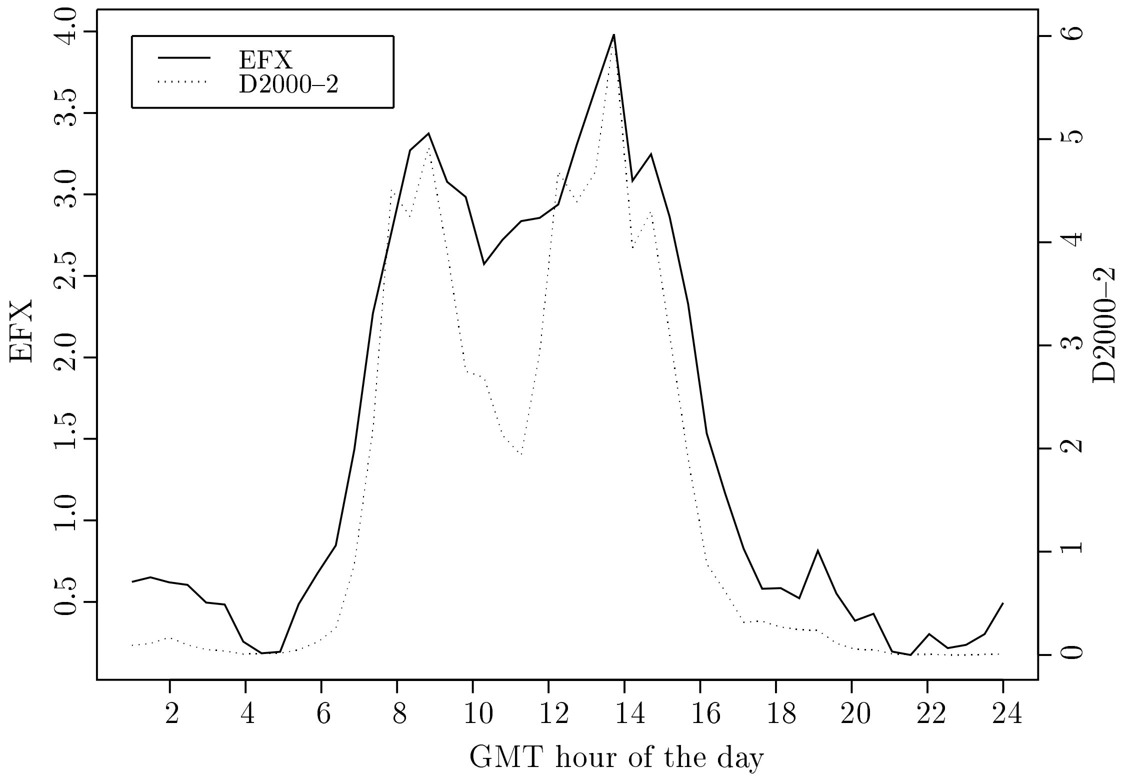


Figure 3: Intra-day Seasonal Patterns in D2000-2 Transaction frequency and volume

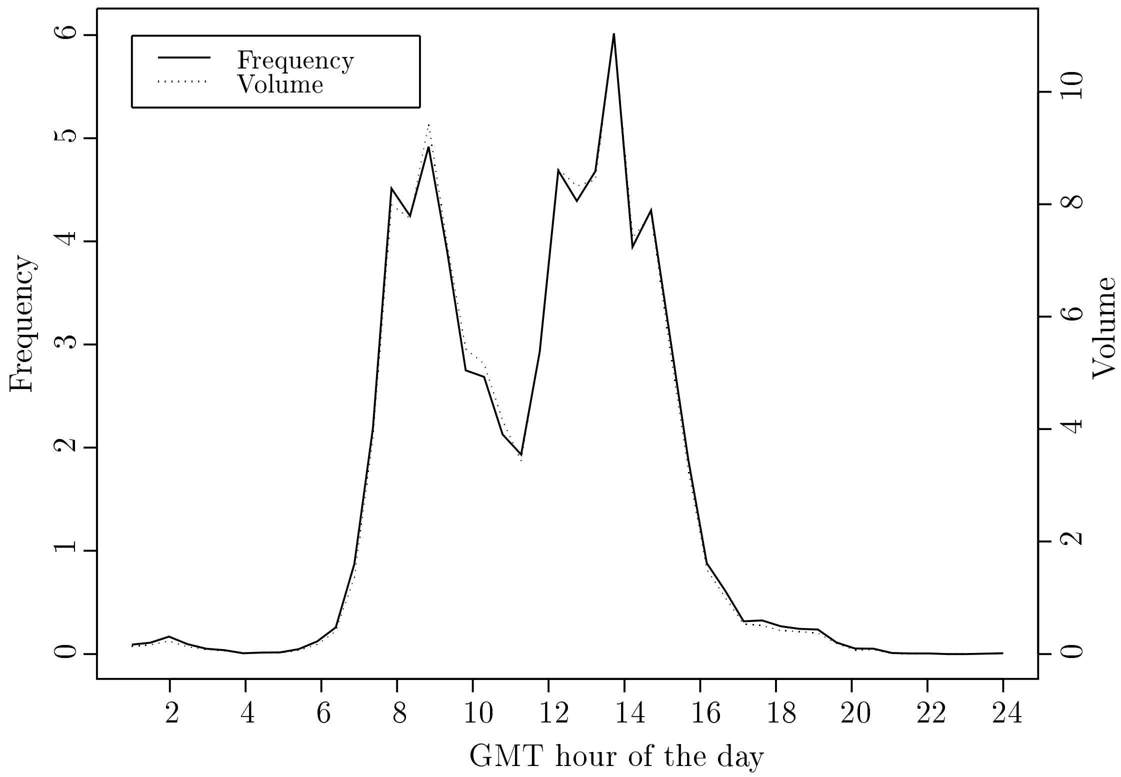


Figure 4: Intra-day Seasonal Patterns in D2000-2 and EFX Spread

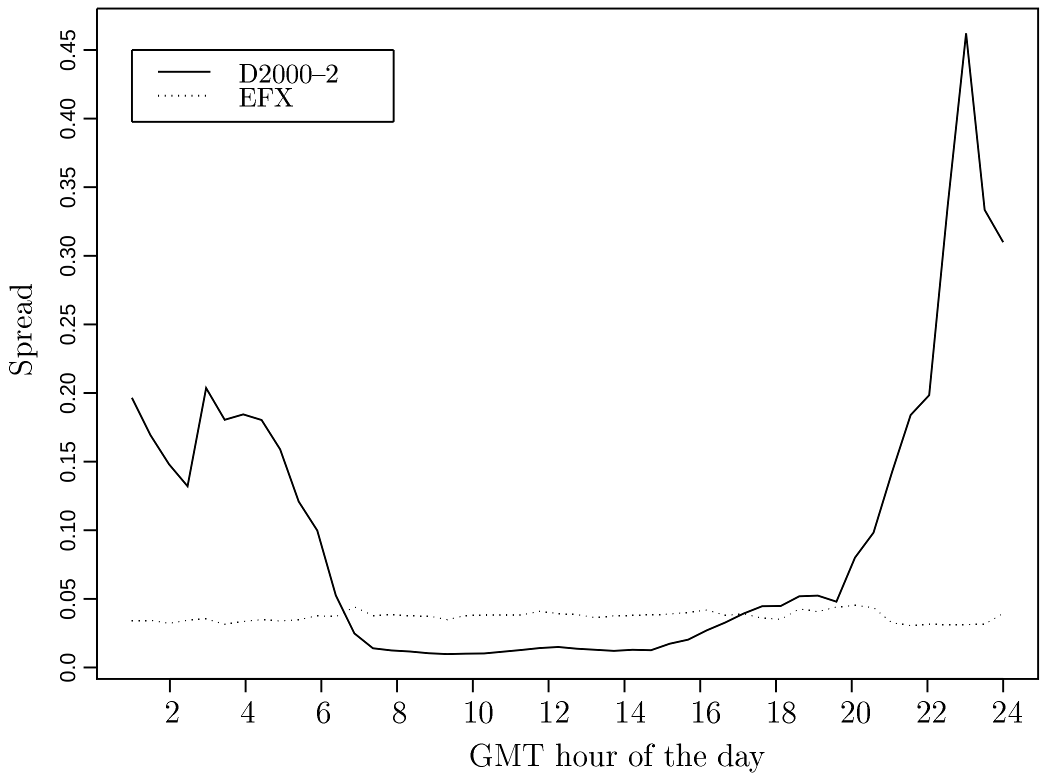


Figure 5: Intra-day Seasonal Patterns in D2000–2 and EFX Absolute Return

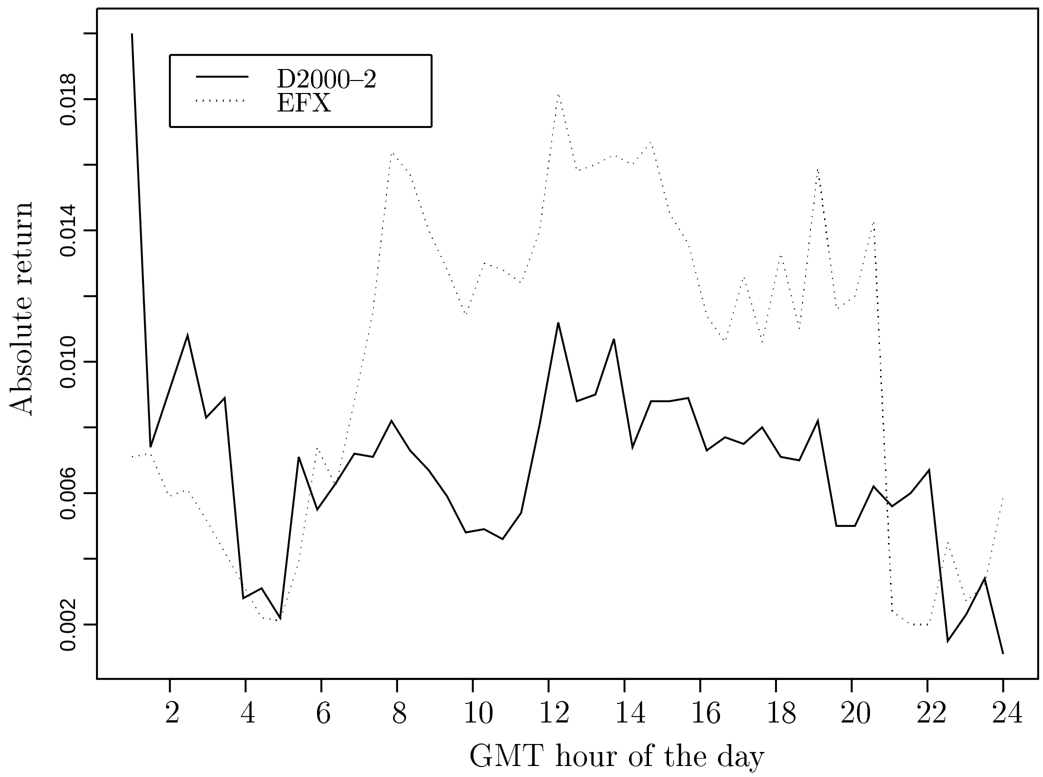


Figure 6: Intra-day Pattern in Return Variances

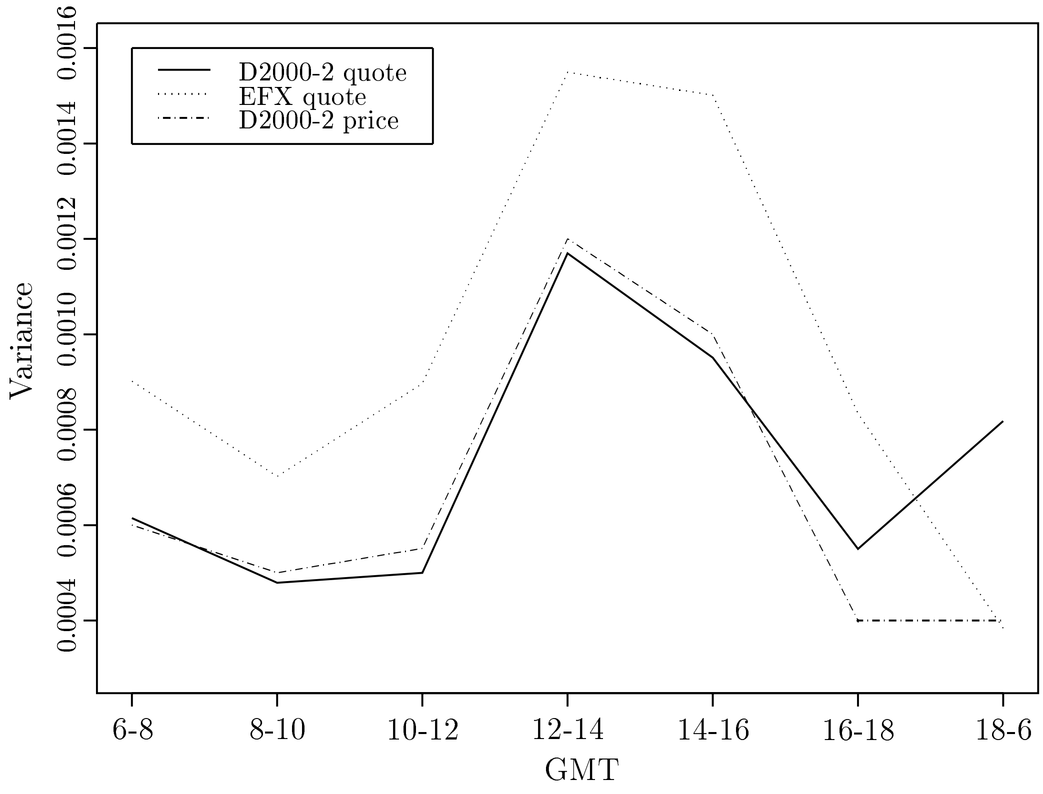


Figure 7: Cross Correlations Between D2000–2 and EFX (with 95% Confidence Band)