The Rimini Centre for Economic Analysis

WP 25-09

Ramazan Gençay
Simon Fraser University, Canada
and
The Rimini Centre for Economic Analysis, Italy

Nikola Gradojevic
Lakehead University, Canada

Faruk Selçuk
Bilkent University, Turkey

PROFITABILITY IN AN ELECTRONIC FOREIGN EXCHANGE MARKET: INFORMED TRADING OR DIFFERENCES IN VALUATION?

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

1. Introduction

Structural macroeconomic spot exchange rate models ignore market microstructure effects in exchange rate determination. These models assume that markets are efficient in the sense that information is widely available to all market participants and that all relevant and ascertainable information is already reflected in exchange rates. In other words, from this point of view, exchange rates are not informed by microstructure variables. Even if price effects from currency order flows arise, they are quickly incorporated through the error term of an exchange rate equation. More specifically, as currency valuation depends primarily on macroeconomic information, the absence of firm-specific information implies a reduced potential for market maker losses to better informed traders (Bessembinder, 1994). In line with this, the existence of private information in the foreign exchange (FX) market implies that traders may be privately informed about macroeconomic fundamentals. Although such private information may exist, it would be challenging to characterize an appropriate exchange rate model. However, these difficulties do not arise with equities. In the equities case, in addition to public macroeconomic information, some traders may have information about firm characteristics (such as future asset prices) that other traders do not have (Glosten and Milgrom, 1985). Another explanation for the modest research activity on information arrival and the existence of informed and uninformed traders in FX markets could be the lack of transaction data for prices and trading volume.[1]

The pioneering work of Meese and Rogoff (1983) represented a turning point in explaining how asymmetric information can arise in FX markets. Meese and Rogoff (1983) found that macroeconomic factors were not sufficiently informative for exchange rate determination and forecasting. These findings resulted in extensive research on alternative model specifications and an increased focus on market microstructure analysis of exchange rates (Lyons, 2001). Generally speaking, market microstructure examines how the actual trading process and market structure affect price formation. In this view, information relevant for exchange rates is heterogeneous (asymmetric) and dispersed across FX traders. Using a general equilibrium framework, Evans and Lyons (2007) show that the information content in transaction activity (such as customer order flow) can predict exchange rates as well as macroeconomic fundamentals. Therefore, private information in the FX market appears to exist and is gradually revealed to the market via trading. Put differently, order flows convey heterogeneous expectations of future macroeconomic variables until they are aggregated and reflected in exchange rates. In a related work, Payne (2003) reveals substantial informed trading effects in an electronic FX market. Further empirical evidence on the existence of private information in spot FX markets can be found in Lyons (1995) and Yao (1998). A theoretical model involving asymmetric information in FX markets is provided in Vitale (2007). This

[1] The exceptions are Lyons (1995), Payne (2003), and Marsh and O'Rourke (2005). The first two papers use one week of trade-by-trade data, while Marsh and O'Rourke (2005) use approximately one year of daily data. To control for high-frequency noise effects and no-trade periods, we aggregate to hourly data in this paper. The original dataset contains tick-by-tick data over an eight month period between 2003 and 2004, or about 6.5 million data lines.

very recent paper shows how informed traders influence exchange rates by inventory management as well as through their private information. This work raises the important question of whether private information in the FX market is short-lived. A noteworthy finding is that long-lived private information enables traders to be more strategic in deciding how to optimally profit from their information advantage.

Given all the above, it may still be difficult to imagine circumstances in which FX market participants would have private information about macroeconomic variables. To clarify this issue, we first focus on the mapping from public information to prices. For example, a macroeconomic announcement is clearly observed by all market participants, but the participants may differ in their interpretations of how it will affect the exchange rate. Hence, FX traders' heterogeneous expectations about future exchange rate movements will be subsequently revealed through their order flows. The heterogeneity of expectations can also be understood as differences in valuation: some traders may place a higher value on a particular currency than others (Handa *et al.*, 2003). These differences may originate in taxes, liquidity shocks or behavioral considerations (e.g., overconfidence)(Daniel and Hirshleifer, 1998). Handa *et al.* (2003) show that the spread widens as valuation differences increase.

With respect to informed trading in equity markets, Easley and O'Hara (1992) have introduced a sequential trade model in which a market maker learns from both trades and from the lack of trades. That is, a market maker's beliefs are continuously updated with new information that may or may not be reflected in a transaction price. As a result, the timing of the trade plays an important role in price formation. In a series of papers, Easley *et al.* (1996a, 1997a,b) expand this work by modeling an equity market in which a competitive market maker trades a risky asset with informed and uninformed traders. Easley *et al.* (2002) further extend these models to allow for time-varying arrival rates of informed and uninformed trades. They show that both informed and uninformed traders are highly persistent in equity markets.

The FX market can generally be described as decentralized and worldwide. However, trading is processed in a few particular and major markets: London, New York and Tokyo. Thus, the total trading activity of informed and uninformed traders is comprised of the *geographic contributions* of individual market centers. The hours of operation of the market centers differ, but they jointly contribute to the aggregate market trading activity. For instance, the London Stock Exchange (LSE) and the New York Stock Exchange (NYSE) are both open from 09:30 to 11:30 EST. The lowest market activity except for weekends can be found during the lunch-time break of the Tokyo Stock Exchange (TSE), and during nighttime in North America and Europe.

The intraday analysis of FX prices demonstrates that hourly returns exhibit fluctuations correlated with the hours of operation of the major market centers (Dacorogna *et al.*, 2001). The first contribution of this paper is to introduce a geographical component to the activity of FX traders. Considering the low-transparency feature of the FX market, this exercise is of immense importance in understanding the market dynamics. Moreover, we also investigate the day-of-week effects of the arrival of FX traders. We draw our conclusions from the data, as well as from a continuous time sequential microstructure trade model in the spirit of Easley *et al.* (1996b). This model directly measures market maker's beliefs rather than searching for indirect evidence of informed trading. We find strong support for intraday geographic dependence in the arrival of informed traders, whereas the uninformed traders arrive uniformly.[2] The dependence on the time of day pinpoints two target markets for informed traders: the NYSE (during the last two hours of operation) and the TSE (during full trading hours). It is worth noting that the above-average activity of informed traders coincides with low overall trading activity in both markets. This indicates the strategic arrival timing of informed traders, who not only pick the low activity hours, but also attach the largest market weight to a particular market, the TSE.[3]

The model also reveals that the day-of-week effects represent a significant component of the trading by both types of traders. We show that the day-of-week arrival rates of informed traders are inversely related to the day-of-week probability of informed trading (PIN) values. In other words, when a high (low) arrival rate of informed traders is observed on a given day, the estimated PIN is low (high). Also, informed traders strategically follow the arrival rates of the uninformed traders, i.e., they tend to conceal their activity by transacting with uninformed traders. This result is similar to the equity market finding by Kyle and Villa (1991), where "noise trading" provides camouflage for a profitable takeover by a large corporate outsider. The distributions of the estimated arrival rates confirm the commitment of the informed traders to strategic trading activities.

The second main contribution of the present study is to quantify the price impact of informed and uninformed trading. More specifically, as in Odders-White and Ready (2007), we acknowledge that the market maker's expected loss from informed trading is a function of both the PIN and its likely impact on the price. It is important to stress that informed trading is not directly observable. The paper by Easley *et al.* (1996b) conjectures that the model only explains price dynamics. We extract the information content of the estimates by measuring the price impact over time. We find that the estimates of some of the model's probabilistic parameters can potentially be used to explain the fluctuations in daily FX returns. In addition, the hourly order imbalances have significant power in determining FX returns as predicted by Evans and Lyons (2002).

We also show that the arrival rates of informed and uninformed traders have significant power in determining hourly and daily FX rate volatilities. The impact of the uninformed traders' arrival on daily volatility is about twice the magnitude of that for informed traders. On the other hand, the findings for the hourly data reveal persistent and dominant effects of the arrival of informed traders on FX volatility. In addition, there is no evidence of the link between the PIN and volatility. The

[2] In the context of the paper, the term "informed" is used loosely: although we reveal trading patterns that are consistent with the informed trading hypothesis (please see Section 2), we do not preclude the possibility that the observed behavior is the result of differences in valuation.

[3] In contrast to our paper, Easley *et al.* (2002) do not find evidence of strategic behavior by informed traders. They document that uninformed traders seem to avoid informed traders by "herding".

PIN can also be interpreted as the proportion of total trading activity accounted for by informed trading (or the trade composition). The fact that it is uninformative for volatility means that after sufficient trading activity takes place, information is fully reflected in the price. More precisely, when the information is fully revealed, the trade composition (which by definition captures the degree of information discovery) does not matter for volatility.[4]

An important generalization of the equity market model by Easley *et al.* (1996b) accounts for a trader's trade size. In traditional market microstructure theory, a larger trade size is typically viewed as more informative, since informed traders will try to trade a larger quantity to profit from their private information (see, for instance, Kyle, 1985). Consequently, one would expect to find a positive relationship between the trade size and its impact on the price. This phenomenon is documented in Bjønnes and Rime (2005), who find that the quoted spread tends to increase with trade size in direct bilateral trading. We extend our framework to empirically test the potential difference in information content between large and small trades. The results relate the informational content of trading to the trade size and suggest that the probability of large, informed trading is significantly higher than the probability of large, uninformed trading. These findings agree with Easley *et al.* (1997a), but contrast with those of Easley *et al.* (1997b), which documents an insignificant difference in the probabilities of large, informed and large, uninformed trading.

In summary, our study contributes to the current market microstructure literature in several important ways. First, we use unique high-frequency trading data for major FX rates, which track individual transactions to reveal trading behavior consistent with the informed trading hypothesis. Second, we extend the models of Easley *et al.* (1996b) and Easley *et al.* (2002), while pointing to important theoretical and empirical considerations of adapting an equity market model to the FX market. This allows for a direct comparison of the observed effects in both markets. Third, our model represents a dynamic, high-frequency version of the Easley *et al.* (1996b) model. We use hourly data and estimate the model over 145 consecutive days, reporting both temporal and average effects. Moreover, we are able to identify the arrival of both informed and uninformed traders with regard to the time-of-day, day of the week, and the FX market. Fourth, we calculate the exact probability that an informed or uninformed trader trades a large amount and investigate the sensitivity to the definition of the large trade size. Fifth, we show that the model's theoretical variables can inform exchange rate determination.

The paper is organized as follows. Section 2 describes the data and the estimates of the benchmark model. Section 3 extends the model to investigate the role of trade size. Section 4 examines the informativeness of the model estimates for the FX rate dynamics, and Section 5 summarizes the study's conclusions.

[4] In a related study, Easley *et al.* (2002) find that the trade composition does not forecast intraday volatility. Lei and Wu (2005) also examine the time series properties of the PIN for a panel of stocks, arguing that the Easley *et al.* (1996b) model should be extended with the time-varying PIN.

2. Data and Estimation

In the early 1990s, the practice of switching from voice brokers to electronic trading systems rendered the FX market more transparent. However, early over-the-counter (OTC) FX market participants had no means of observing the market-wide order flow. The introduction of centralized electronic broking systems, such as Reuters and EBS, thus provided a new platform for research on the FX market microstructure. Although Reuters and EBS are dominant in electronic FX markets, they do not publicly report high-frequency volume data or the identity of the traders. In order to achieve our research goals, we turn to a smaller electronic FX market launched in 2001 by the OANDA Corporation, the OANDA FXTrade internet trading platform.

Our dataset consists of tick-by-tick foreign exchange transaction prices and the corresponding volumes for several exchange rates from October 1, 2003 to May 14, 2004. The number of active traders during this period is 4,983, and they mainly trade four major exchange rates.[5] The data show that the overall trading frequency increases from Monday to Wednesday (the peak) and falls from Thursday to Saturday. Using the trader's identity (trader ID), we next investigate the number of currency pairs traded by investors. We find that about 33% of the investors specialize in exactly one currency pair, about 11% in two currency pairs, and about 9% in three currency pairs. This decreasing trend leads to only 2-4% of active traders who deal in between 10 and 13 currency pairs. Hence, traders appear to specialize in a small number of currency pairs, in line with OANDA FXTrade's intention to attract small investors.

Since the bulk of all transactions (approximately 40 percent[6]) involve only U.S. Dollar - Euro (USD-EUR) trading, we focus on transactions involving only USD-EUR. In particular, we analyze all USD-EUR buy and sell transactions (market, limit order executed, margin call executed, stop-loss, and take-profit transactions). In addition to price and volume, we know the trader ID for each transaction, which ranges from 123 to 5904. The average number of USD-EUR transactions per trader is 512. Using the trader ID, we observe that a few traders transact very frequently in this currency pair (between 10,000 and 25,000 times) over the time period that spans the data (Figure 1). We also observe that day-of-the-week trading patterns for the USD-EUR transactions (trading frequency and volume) are similar to those of other currency pairs.

Further investigation shows that about 98% of the USD-EUR traders close their positions with a single trade and that about 90% of those "round-trip" transactions are closed intraday. The average duration of round-trip transactions ($\bar{\tau}$) is about 4 hours. To suggest potential informed trading activities, some of the traders would have to make consistent profits. We find excess round-trip profits in each month of our sample for 29 traders (and for 42 traders in the first four months of 2004). One interesting finding is that the trader with the highest excess return in the

[5]By "active", we refer to traders that did not simply receive interest on their positions, but placed orders during this period. The market share of these traders is approximately 86.4%.

[6]The next most active currency pairs are USD-CHF (7.88% share), GBP-USD (7.81% share), USD-JPY (6.42% share), and AUD-USD (5.98% share).

USD-EUR market is also the most successful one in the USD-CHF market. We also identified a single trader that dominates six different markets (AUD-JPY, AUD-USD, EUR-JPY, GBP-JPY, GBP-USD, USD-CAD) and another that dominates three currency pairs (EUR-CHF, EUR-GBP, GBP-CHF). Next, to verify the robustness of our findings, we apply a simple, model-free technique. We compute the price impact of the time t signed transaction on the change of mid-quote from time t to time $t + \tau$. The idea is to reveal whether (round-trip) trades "predict" mid-quote movement. On average, of the 29 consistently profitable traders in the EUR-USD market, we find that the top five most profitable traders[7] are on the correct side of the trade an astonishing 99% of the time. In comparison, an average trader in this market is correct in about 50-60% of his or her trades. Taken together, these results exhibit transaction behavior compatible with the informed trading hypothesis.

Table 1 summarizes the institutional characteristics of the OANDA FXTrade trading platform. This platform is an electronic market making system (i.e., a market maker) that executes orders using bid/ask prices that are realistic and prevalent in the marketplace. The prices are determined by their private limit order book or by analyzing prices from the Interbank market. The OANDA FXTrade policy is to offer the tightest possible bid/ask spread (e.g., 0.0009% spread on the USD-EUR, regardless of the transaction size). Like most market makers, they profit from the spreads. Some of the other market features include continuous, second-by-second interest rate payments, no limit on the transaction size, no requirement for minimum initial deposit, no charges for stop or limit orders, free quantitative research tools, and margin trading (maximum leverage of 50:1). Given these market characteristics, the OANDA FXTrade seems to be designed to attract small, uninformed traders. However, given the above findings, it is reasonable to assume that informed traders are also present in this market.

The theoretical model by Easley *et al.* (2002) is developed in the context of equity markets. Adapting it to the FX market requires care. As opposed to the equity market, the FX market is open 24 hours and is decentralized. Further, unlike the NYSE, it does not involve a so-called specialist responsible for maintaining fairness and order, with an insight into the limit order book. While the NYSE has recently introduced an open limit order book that provides a real-time view of the limit order book for all NYSE-traded securities, the FX market exhibits a low level of transparency. Finally, trading in the FX market is motivated by speculation, arbitrage and, importantly, inventory management of currencies. Dealers in the FX market are generally quick to eliminate inventory positions (from below five minutes to half an hour). This process is sometimes referred to as "hot-potato-trading" (Evans and Lyons, 2002; Bjønnes and Rime, 2005). On the NYSE, however, inventory has an average half-life of over a week (Madhavan and Smidt, 1993). Thus, inventory management is an important component of intraday FX trader activity.

The features of the OANDA FXTrade allow us to view it as a "special case" of the FX market that can be approached using the model by Easley *et al.* (2002). First, as a market maker, the

[7] Trader ID is withheld to preserve data confidentiality.

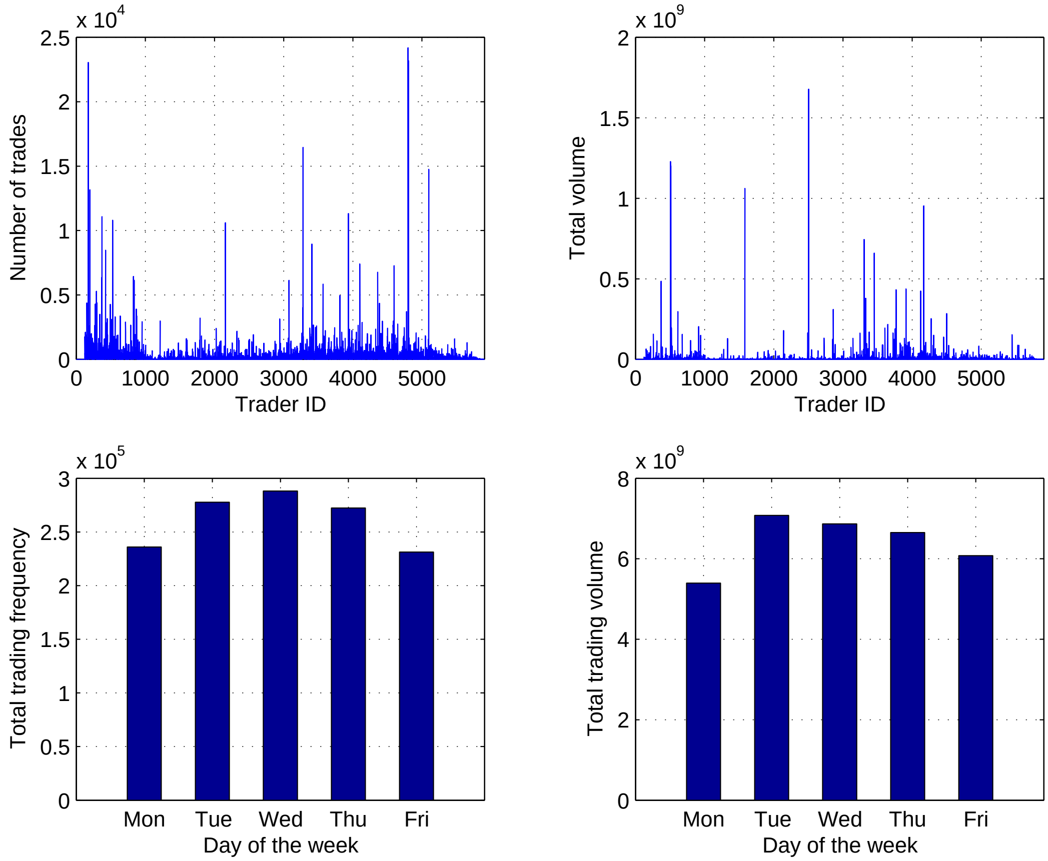

Figure 1: Top left: Total trading frequency (number of trades) per trader (trader ID). Top right: Total trading volume (in the units of base currency - EUR) per trader (trader ID). Bottom left: Total trading frequency (number of trades) for each day of the week. Bottom right: Total trading volume (in the units of base currency - EUR) for each day of the week.

OANDA FXTrade promotes transparency: spreads are clearly visible, past spreads are published for public view and current open orders on major pairs are visible to all market participants. In regards to trader behavior, as we only focus on the informational aspect (i.e., informed vs. uninformed), market participants in the FX market can be treated in a fashion similar to those in equity markets. Section 4 suggests how intraday inventory management may relate to the findings.

Our preliminary analysis indicates that overall market activity was extremely low on certain days or during certain weeks. Therefore, we eliminate weekends, starting from every Friday 15:59:59

Hours of operation	24 hours/7 days per week
Number of currency pairs	30
Number of active traders	4,983
Number of trades (USD-EUR)	667,030 sell transactions
	666,133 buy transactions
Average number of trades per day (USD-EUR)	192 sell transactions
	191 buy transactions
Total volume (USD-EUR)	32.6 billion USD
Average volume per day (USD-EUR)	224 million USD
Transaction types	Buy/Sell market (open or close)
	Limit order Buy/Sell
	Cancel order (reason: bound violation, insuff. funds, none)
	Change order
	Change stop loss (sl) or take profit (tp)
	Sell/Buy tp (close), Sell/Buy sl (close)
	Buy/Sell limit order executed (open or close)
	Order expired
	Sell/Buy margin called (close)
	Interest

Table 1: OANDA FXTrade institutional characteristics.

to Sunday 15:59:59 (all times are EST), including Christmas week (December 22-26), the first week of the year (December 29-January 2), and the week of Independence Day (April 5-9). This leaves us with 145 24-hour periods. In order to avoid extremely high-frequency noise and no-activity periods in small time windows, we aggregated the data over one-hour intervals. Aggregating over trading intervals smaller than one hour is not feasible, as this would not cover a sufficient number of buy and sell transactions for the model to be empirically applicable. On the other hand, longer trading intervals would "stretch" the assumptions of the model to a certain extent. For example, the news is assumed to arrive hourly (with probability α). It is unlikely that the flow of information would be less frequent, i.e. over longer time intervals. The final sample size is 3,480 hourly data points covering 145 business days, from October 5, 2003, 16:00 to May 14, 2004, 16:00 EST. There are 667,030 sell and 666,133 buy transactions in the sample period, with an average of approximately 6 transactions (3 buy and 3 sell) per minute. The transaction volume totals 32.6 billion USD-EUR contracts. According to the BIS Triennial Survey for 2004, the daily average turnover in the USD-EUR currency pair was 501 billion USD. Hence, on average, our data represent about 0.045% of the global daily USD-EUR trading volume. Nevertheless, it is one of the largest tick-by-tick FX datasets to be used in an academic study.

Since the trader's identity is known in each transaction, we are able to identify the number of unique traders in each one-hour window. For estimation purposes, the number of buy arrivals in each hour (B'_t) is defined as the number of unique traders involved in buy transactions in that hour. The number of sell arrivals in each hour (S'_t) is defined similarly. Therefore, the arrival of an individual trader who conducts several buy (sell) transactions in a given hour is counted as one buy (sell) arrival in that hour.

Figure 2 illustrates the number of hourly buy and sell arrivals (B', S') and the sample autocorrelation functions. We see strong daily time dependence and a time trend in both series. Therefore, we first estimate the linear time trend, $\hat{B}_t$ and $\hat{S}_t$, from the trend regression, which is free of temporal and irregular fluctuations. Assuming multiplicatively-separable time dependence, we divide the original series by the trend estimates $\hat{B}_t$ and $\hat{S}_t$ to obtain an estimate of the time component

$$\tilde{s}_t^B = \frac{B'_t}{\hat{B}_t}, \quad \tilde{s}_t^S = \frac{S'_t}{\hat{S}_t}.$$

In order to estimate a time index for each hour of the day, we average the values of $\tilde{s}_t^B$ and $\tilde{s}_t^S$ corresponding to the same hour of the day across the sample and obtain the final hour-of-day indices s_i^B and $s_i^S, i = 1, 2, \ldots 24$ for the 24-hour cycle. The 24-hour adjusted number of buy and sell arrivals are obtained via

$$B_i = \frac{B'_i}{s_i^B}, \qquad S_i = \frac{S'_i}{s_i^S}, \qquad i = 1, 2, \ldots, 24$$

for each 145 days in the sample.

Figure 3 studies the final hour-of-day indices s_i^B and $s_i^S, i = 1, 2, \ldots 24$, for the number of unique buy and sell traders starting at midnight 00:00 EST. The average number of unique buy and sell traders increases after midnight, rising above the hourly average before the opening of the LSE (at 03:00 EST). Similarly, we observe a sharp increase in the hours before the opening of the NYSE (at 9:30 EST). The number of traders declines after 10:00, falling below the hourly averages after 14:00. They remain relatively low and stable in the subsequent hours until midnight. In the lower panel of Figure 3, the sample autocorrelation functions of the diurnally-adjusted number of buy and sell arrivals are studied at hourly lags. The removal of the daily temporal component reveals strong persistence in both series.

2.1. Informed and uninformed trades: When do they arrive?

According to the Easley *et al.* (1996b) model (please see Appendix A), the expected value of the total number of trades per unit time, $E(TT) = E(S+B)$, is equal to the sum of the Poisson arrival rates of informed and uninformed trades:

$$E(TT) = \alpha(1-\delta)(\varepsilon + \mu + \varepsilon) + \alpha\delta(\mu + \varepsilon + \varepsilon) + (1-\alpha)(\varepsilon + \varepsilon) = \alpha\mu + 2\varepsilon$$

The expected value of the trade imbalance $E(K) = E(S - B)$ is given by

$$E(K) = \alpha\mu(2\delta - 1),$$

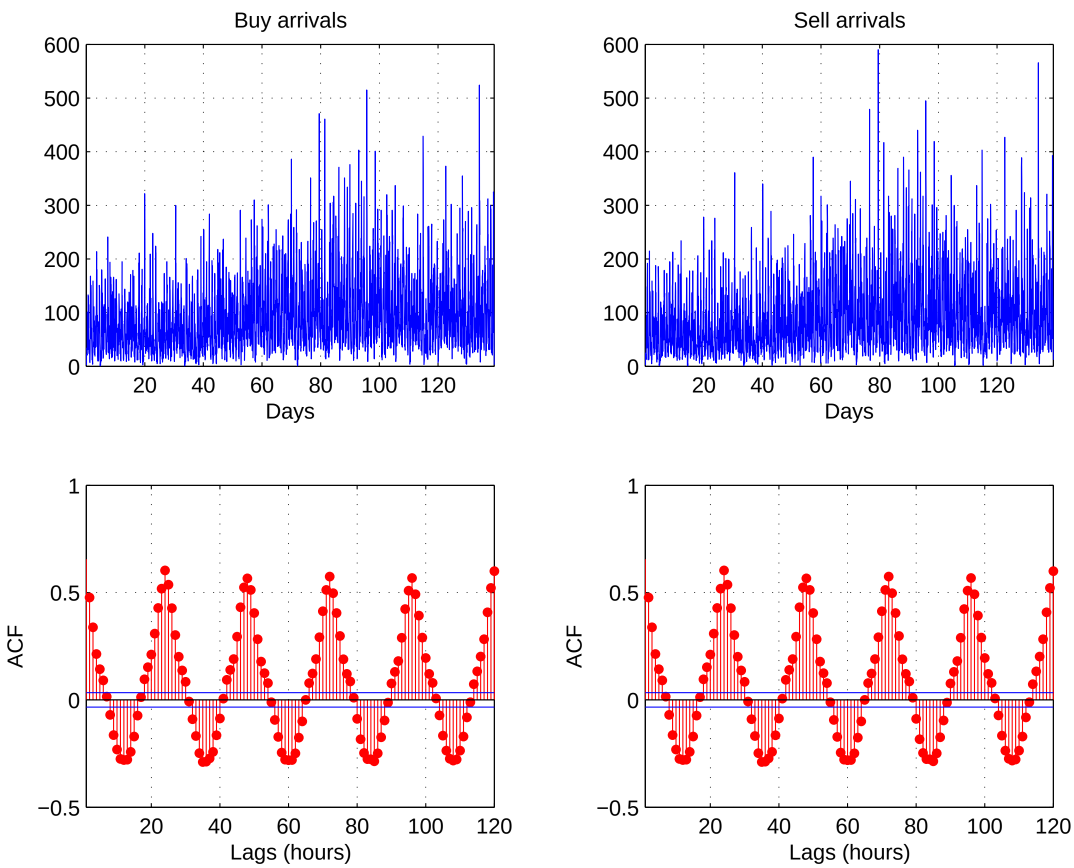

Figure 2: The number of hourly buy (top left) and sell (top right) arrivals (B', S') and the sample autocorrelation functions at 120 hourly lags (5 days). The buy and sell arrivals in each hour are defined as the number of *unique* traders involved in buy or sell or both types of transactions in that hour. Sample period: October 5, 2003, 16:00 - May 14, 2004, 15:59 (3480 hours, 145 business days).

which provides information on the arrival of informed trades. When μ is large, the following approximate relation holds:

$$E(|K|) \simeq \alpha\mu.$$

Accordingly, the absolute trade imbalance $|K|$ provides information on the arrival of informed trades, $\alpha\mu$, while the difference between the total trade TT and the absolute trade imbalance $|K|$

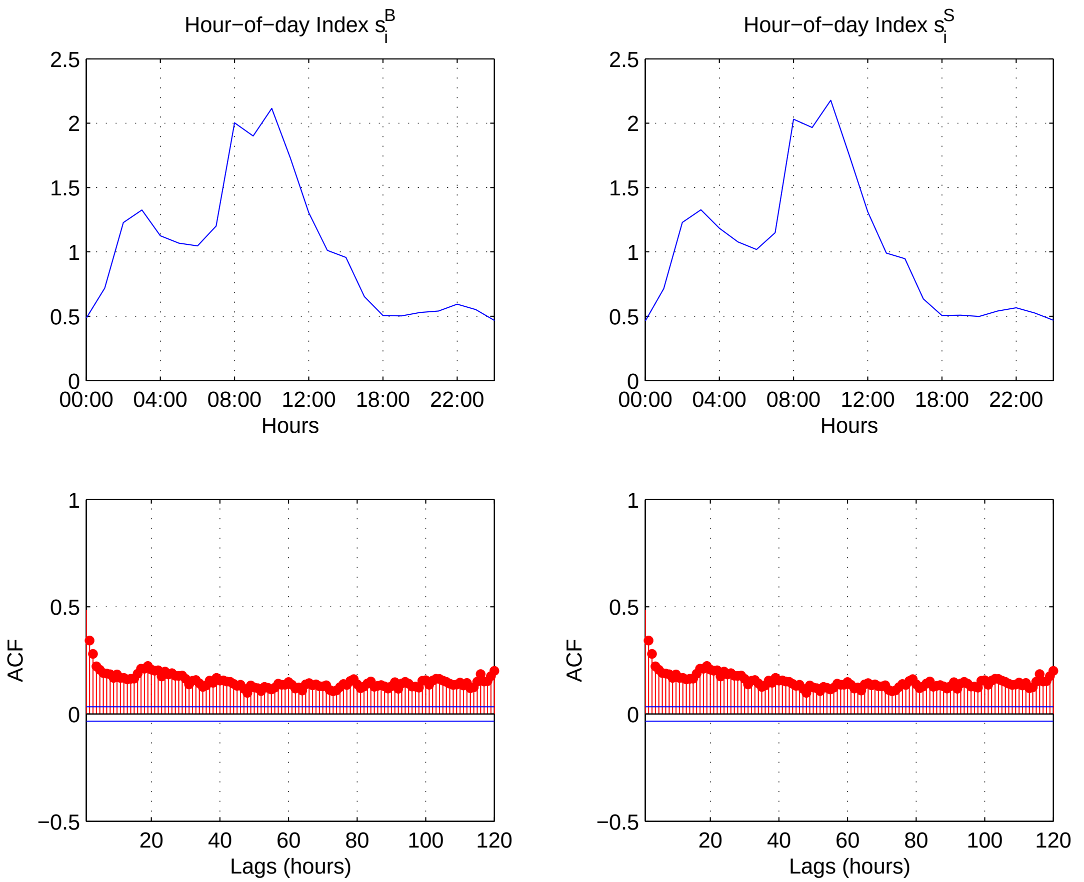

Figure 3: Top: Hour-of-day indices s_i^B and $s_i^S, i = 1, 2, \ldots 24$, for the number of unique buy (top left) and sell (top right) traders starting at midnight 00:00 EST. Bottom: Sample autocorrelation functions at 120 hourly lags (5 days) of the number of unique buy and sell traders.

contains information on the arrival of uninformed trades, ε. Note that our measure of the "number of buy and sell trades" in a given time period is the number of *unique* traders. Therefore, we can substitute the term "trader" for "trade" in the above expressions.

If we assume that the probability of information events α is constant, the hour-of-day average of the absolute trader imbalance $|K|$ provides information on the intraday time dependence of the orders from informed traders. In other words, since we know the number of unique individuals and corresponding trades at each hour of the day, we can obtain a measure of *the activity time* of the informed traders. We can similarly identify whether uninformed traders (liquidity traders) follow

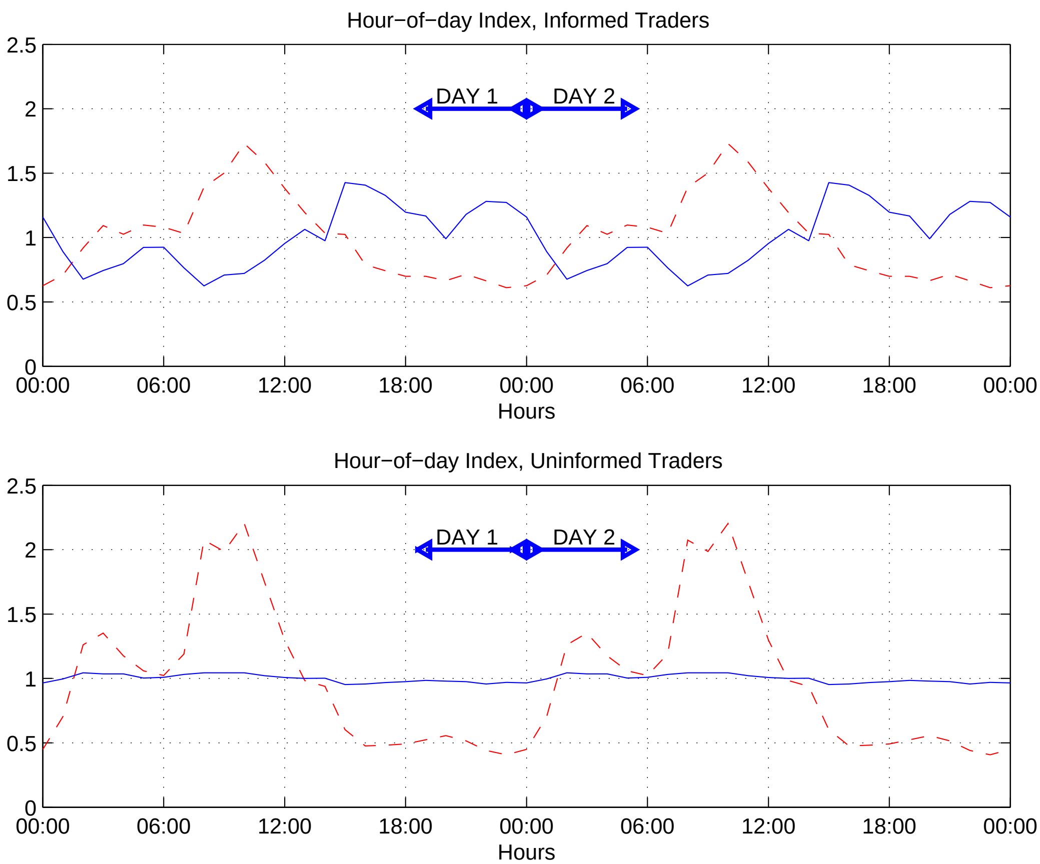

Figure 4: Hour-of-day indices of informed (top) and uninformed (bottom) traders over 48 hours, based on unbalanced traders ($|K|$) and balanced traders ($TT - |K|$). The solid line represents the data free of intraday fluctuations (the "de-seasonalized" data), while the dashed line represents the raw data. The hour-of-day index of uninformed traders is relatively stable for the de-seasonalized data, indicating a non-strategic, uniform arrival. The same index of informed traders fluctuates according to the observed dependence on market location. For the raw data, the hour-of-day indices of both informed and uninformed traders exhibit similar patterns.

a distinct intraday pattern.

Figure 4 plots the hour-of-day indices of informed (top) and uninformed (bottom) traders based on unbalanced traders ($|K|$) and balanced traders ($TT - |K|$). Note that both $|K|$ and TT are calculated from B_t and S_t so that we do not expect any hour-of-day effect *a priori*.[8] However, the hourly activity of uninformed (liquidity) traders increases before the openings of the LSE, NYSE, and TSE (at 03:00, 9:30 and 19:00, respectively). Activity exceeds the hourly average from 01:00 until 14:00. Notice that the variation in the hour-of-day index of uninformed traders is relatively small, fluctuating between 0.95 and 1.04. Therefore, we may speculate that uninformed traders arrive uniformly at any time of the day. However, the hour-of-day index of informed traders is almost the opposite of that of uninformed traders. The volatility of informed traders during the day is high, with the index fluctuating between 0.63 and 1.43. The number of informed traders falls sharply after 01:00, before all three major exchanges (the LSE, the NYSE, and the TSE) open. It picks up around 03:00 at the opening of the LSE, peaks and then dips down before the opening of the NYSE (at 9:30, or 14:30 in London). The number of informed traders increases sharply at the opening of the NYSE and until the market closes (at 16:00). This is followed by a decline until the opening of the TSE (at 19:00). Above-average activity in Tokyo lasts until the market closes (at 01:00). The number of unique informed traders is well above the average after 14:59, one hour before the NYSE closes. This number remains above the average (except at the opening of the TSE, at 19:00) until before the opening of the LSE (at 03:00). To conclude, informed traders appear to target the NYSE during the last two hours of operation and the TSE during full trading hours. Recall from Figure 2 that these are the hours with the fewest traders present in the market. The hour-of-day indices do not evidence the above-average arrival of informed traders during the LSE trading hours.

2.2. Estimation of the Easley *et al.* (1996b) model

As mentioned above, the sample estimates of $E(TT)$ and $E(|K|)$ provide prior information about the parameters of the Easley *et al.* (1996b) model. The sample mean of total unique trades $\overline{TT}$ is 176.9 while $\overline{K} = -1.4$ and $\overline{|K|} = 19.4$. From the equation above,

$$\frac{E(K)}{E(|K|)} = \frac{\alpha\mu(2\delta - 1)}{\alpha\mu} = \frac{-1.4}{19.4} = -0.07$$

Accordingly, the implied probability that an event is bad news is 0.47 ($\bar{\delta} = (1-0.07)/2 = 0.47$). Uninformed buy and sell traders arrive at an hourly rate of ε. The estimated sample statistics imply that this hourly rate is 79.7.

$$E(TT) = \alpha\mu + 2\varepsilon = 176.9 = 19.4 + 2\varepsilon,$$

[8] We also plot the hour-of-day indices for the raw data that use unadjusted B_t and S_t (dashed line). Similar arrival patterns are observed for the two types of traders. Thus, "hidden" hour-of-day patterns are present even after B_t and S_t are de-seasonalized. Since these effects are strong enough to persist even after adjusting for intraday time dependency, we will concentrate on the de-seasonalized data for the remainder of the paper.

Index	Monday	Tuesday	Wednesday	Thursday	Friday
SI_α	1.14	0.89	0.91	1.1	0.96
SI_δ	0.76	1.15	0.95	1.02	1.13
SI_ε	0.84	1.08	1.09	1.05	0.94
SI_μ	0.85	1.09	1.14	0.94	0.98
SI_{PIN}	1.16	0.92	0.91	1.01	1.00

Table 2: Day-of-week indices of estimated parameters and the PIN. Note: The day-of-week indices are found using the ratio-to-moving average method.

and $\hat{\varepsilon} = (176.9 - 19.4)/2 = 78.8$. Another measure based on the parameters of the Easley *et al.* (1996b) model is known as the PIN (see equation (1)):

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} = 19.4/176.9 = 0.11,$$

for the market maker's initial belief. The estimated PIN in equity markets is usually between 0.15 and 0.25. This is relatively low, which may indicate that the number of informed traders is low (small μ), that the probability of the information event is low (small α), or both. In this particular case, the market maker's risk of informed trading is relatively low.

This estimation uses the above priors as the initial parameter set. That is, we set $\varepsilon = 78, \delta = 0.47$. Since we do not have a prior for α and μ separately, we assume $\alpha = 0.50$ so that $\mu = 38$. The log likelihood function in equation (11) is maximized daily ($T = 24$) for the entire sample period (145 days). As a result, we have 145 different estimates of each parameter.[9] The two probability parameters α and δ are restricted to $(0, 1)$ and the two arrival rates to (0,500), since the maximum number of observed unique buy or sell traders in our sample is 474.

Table 2 reports the seasonal indices (day-of-week index) of the estimated parameters and the PIN.[10] The probability of an event α is higher on Mondays and Thursdays. Given that an event occurs, the probability that it is a bad event δ is lower than the average on Mondays and Wednesdays. Therefore, we speculate that Mondays were eventful, with good news for USD-EUR during the sample period. This may reflect that a significant time period (48 hours) is required to "digest" the previous week's information before Monday traders arrive.

Figure 5 plots the daily estimates of the probability of an event α (top left) and the probability that an event is bad news δ (top right).[11] The estimated probability of an event $\hat{\alpha}$ fluctuates

[9] The estimation is carried out in Matlab, using the optimization function "fmincon". As we minimize the negative log likelihood function for the results reported in Section 3, smaller log likelihood values indicate a better fit.

[10] The day-of-week indices, denoted by SI_i ($i \in \{\alpha, \delta, \varepsilon, \mu, PIN\}$), are found using the ratio-to-moving average method.

[11] The estimates over 145 days are stable with regard to the reasonable choice of their starting values. The only case in which the estimates begin to substantially change is when $\mu_0 > 200$ and $\varepsilon_0 > 200$.

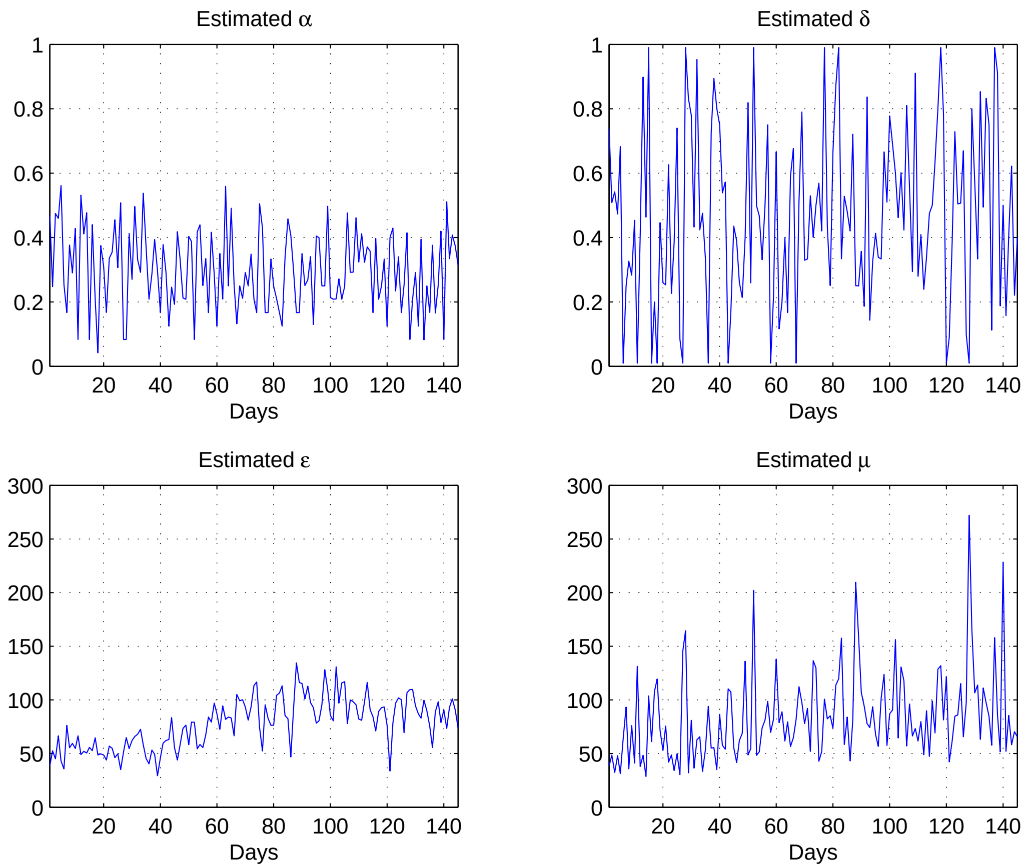

Figure 5: Daily estimates of the probability of an event α (top left), the probability that an event is bad news δ (top right), the arrival rate of uninformed traders ε (bottom left), and the arrival rate of informed traders μ (bottom right).

between 0.04 and 0.56, with an average of 0.30. This implies that there were no days without at least an event per hour. The lowest estimate is 0.04, which shows that there was a day with only one event in an hour ($0.04 \times 24 \approx 1$ day). Similarly, the highest estimate is 0.56, which shows that the most eventful day had 14 hours with an event. The Shapiro-Wilk test (Shapiro and Wilk, 1965) rejects the null hypothesis of normality at the 5% significance level, as the p-value is 0.035. Thus, for this sample, the market maker views the arrival of news as a non-normal process. The estimate that an event is bad news $\hat{\delta}$ lies in between 0 and 1, with an average of 0.47 (our initial parameter). Note that $(1-\delta)$ is the probability that an event is good news. For example, the 18th day of our sample covers the 24-hour period from October 28, Wednesday 16:00 to October 29, Thursday 15:59. On this particular day, $\hat{\alpha} = 0.0412$ and $\hat{\delta} = 0.01$. This means that there was only one hour with news (we do not know which hour) and the news was good ($1 - \hat{\delta} = 0.99$). According to the Shapiro-Wilk test, the estimate of δ is normally distributed, with p-value=0.418. This result is expected, as there was no significant trend in USD-EUR prices during the sample period.

Figure 5 plots the daily estimates of the arrival rates. The estimated arrival rate of uninformed traders $\hat{\varepsilon}$ exhibits a sharp increase around the 60th day (in January 2004), from an average of approximately 50 to 80. The overall mean of this parameter is 77.8 (the same as the initial parameter). The estimate follows a normal distribution as confirmed by the Shapiro-Wilk test. The estimated arrival rate of informed traders $\hat{\mu}$ seems to be stable with occasional jumps. The Shapiro-Wilk test strongly rejects the null hypothesis of normality. The overall average of this parameter is 83.5. Table 2 shows that both informed and uninformed traders arrive less often than average on Mondays and Fridays. The highest arrival rates of both informed and uninformed traders are observed on Wednesdays. It is worth noting that the market maker attaches a non-normal component to the arrival of the informed traders, which goes against the assumption that informed traders are risk neutral. Another instance of empirics diverging from the market assumptions (dominance of uninformed traders) is the fact that informed and uninformed traders have similar arrival rates. This can be interpreted as an equal likelihood for the arrival of informed and uninformed traders, despite the fact that the institutional market characteristics encourage the participation of uninformed traders in particular.

Finally, the average estimated PIN is about 0.12, lower than that observed in equity markets. Over the 145 days in the sample, the PIN ranges between 0.04 and 0.21. The seasonal day-of-week indices for the PIN point to Monday as the only above-average day. The PIN is below average on Tuesdays and Wednesdays, and SI_{PIN} is close to unity on both Thursdays and Fridays. Therefore, although Mondays are viewed as eventful days with a relatively high PIN, this does not result in extraordinarily high arrival rates of informed traders. Rather, their activity appears to be more subtle, with most of their trading taking place on days with lower-than-average news arrival and high arrival rates of uninformed traders. Hence, the PIN reveals that informed traders "conceal" their above-average activity on Tuesdays and Wednesdays as well as their below-average activity on Mondays and Fridays. In all, it appears that the arrival rates of informed traders are inversely related to the PIN values, i.e. high (low) arrival rates of informed traders imply low (high) PIN values. Moreover, informed traders strategically match the arrival rates of the uninformed traders, thereby camouflaging their trading activity. In addition, the non-normal distribution of $\hat{\mu}$ confirms the evidence of strategic activity by informed traders.

2.3. Independence of arrivals

The crucial underlying assumption in Easley *et al.* (1996b) is the hourly independence of information events in each 24-hour sequence.[12] Thus, while deriving the log likelihood function, we assume that the arrival of traders in each hour, conditional on information events, is drawn from identical and independent distributions. Nevertheless, considering evidence on the relationship between volatility clustering and trading volume (e.g. Gallant *et al.*, 1992), it would be useful to test whether the independence assumption holds.

As a first step, we follow Easley *et al.* (1997b) and use a runs test for each day in the sample. The estimated $\hat{\alpha}_i$'s ($i = 1, \ldots, 145$) help us to classify hours into one category in which an event occurs or another in which no event occurs. As noted previously, TT is the total number of trades. On each day, we order hourly TT from the smallest to the largest and classify the upper $100 \times \hat{\alpha}_i$ per cent as event hours. We then turn to the original TT sequence, classifying each event hour by one and each non-event hour by zero. The total number of event hours is denoted by e_i and non event hours by n_i.[13] The results indicate that the null hypothesis cannot be rejected at the 5% significance level for 72 days, although it is rejected for 73 days. This mixed evidence necessitates additional testing. We turn to the Ljung-Box portmanteau test (Ljung and Box, 1978) for white noise next.[14]

We compute the Ljung-Box test statistic with up to the 10^{th}-order serial correlation in levels of S, B, TT and K for each day. Hence, we compare 145 values for Q_L with the critical value χ^2_{10}. The null rejection frequencies at the 5% significance level are: for B (frequency = 24, or 17% of the days in the sample), for S (frequency = 21, or 14% of the days in the sample), for TT (frequency = 23, or 16% of the days in the sample), and for K (frequency = 18, or 12% of the days in the sample). We conclude that the independence assumption is not disproven by this evidence, as our model agrees with the assumption on about 85% of days. Easley *et al.* (1997b) argue that evidence of dependence does not affect the actual parameter estimates, but does affect their asymptotic standard errors. Since our inference is based on the mean values of the estimates and small standard errors (relative to the parameter values), we conjecture that not accounting for the dependence of information events does not have any major impact on our findings.

[12] Easley *et al.* (1997b) test for the independence assumption and find that information events in their dataset are independent.

[13] Under the null hypothesis of the randomness of information events across hours, the total number of runs r (sequences of ones or zeros) is normally distributed with $\bar{r} = \frac{2e_i n_i}{e_i + n_i} + 1$ and $\sigma_r^2 = \frac{(\bar{r}-1)(\bar{r}-2)}{(e_i+n_i)-1}$.

[14] For the null hypothesis of independence (or randomness) of information events over I=24 hours, this test is based on the following statistic: $Q_L = I(I+2)\sum_{\tau=1}^{L} \frac{\hat{\rho}_\tau^2}{I-\tau}$, where L is typically chosen to be substantially smaller than I and $\hat{\rho}_\tau^2$ is the sample autocorrelation coefficient at lag τ.

3. Informativeness of the trade size

In this section, we utilize a procedure similar to Easley *et al.* (1997b), theoretically outlined in Easley and O'Hara (1987). This approach allows for informed and uninformed traders to place both large and small orders. Our extended model relies on the number of unique large buy (LB), small buy (SB), large sell (LS) and small sell (SS) trades that represent the set of possible trade outcomes.[15] This introduces two new parameters: ϕ (the probability that an uninformed trader trades a large amount) and ω (the probability that an informed trader trades a large amount). Naturally, $(1-\phi)$ denotes the probability of a small uninformed trade and $(1-\omega)$ is the probability of a small informed trade. All other parameters (α, μ, δ and ε) follow the notation from Section 2.

Our goal is twofold: first, to compare the estimated ω and ϕ over 145 days ($\omega > \phi$ would mean that trade size conveys additional information to market participants); and second, to observe how changes in the cutoff trade size impact the estimates. The procedure of testing for trade size effects involves comparing the estimates of the restricted ($\omega = \phi$) and unrestricted ($\omega \neq \phi$) models. We initially set the cutoff amount that differentiates large from small trades to 5,000, and show later that this does not affect the major results.[16]

The derivation of the log likelihood function proceeds in a way similar to that of the restricted model, and is presented in Appendix C. Table 3 lists the average estimates of $\alpha_i, \delta_i, \varepsilon_i, \mu_i, \omega_i$, and ϕ_i ($i = 1, \ldots, 145$). The paired t-test of the equality of the means of the constrained and unconstrained models shows no significant difference for the first four parameters.[17] However, the difference between the two sets of estimates of ω_i is statistically significant.[18] Furthermore, including the trade size effects (unconstrained model) significantly increases the absolute value of the log likelihood function, thus indicating that the constraint is binding. The informativeness of the trade size is also confirmed by the unpaired t-test of the equality of $\bar{\omega}$ and $\bar{\phi}$ for the unconstrained model. The model concludes that $\bar{\omega}$ is significantly greater than $\bar{\phi}$. On about 68% of the days in the sample, $\omega_i > \phi_i$ and the difference in the probabilities ($\omega_i - \phi_i$) ranges from -0.12 to 0.14 ($i = 1, \ldots, 145$). Although there are 47 days when the probability of uninformed large trading exceeds the probability of informed large trading, we conclude that this is not the case on average.

It is essential to investigate whether the findings above are robust to the choice of the cutoff amount for a "large" trade. In Table 4, we report the results for 2000, 8000 and 12000 cutoff rates. We focus on the difference column from Table 3 and the mean values of the unconstrained

[15] For simplicity, we ignore the no-trade outcome considered in Easley *et al.* (1997b) for a much smaller dataset of stock prices. Also, using the procedure from Section 2, we de-seasonalize each of the four new series to remove the observed daily time dependence.

[16] Trade size is expressed in currency units of the base currency, i.e., the Euro.

[17] The null hypothesis for this test is that the mean difference ($\bar{d}$) between paired observations (constrained and unconstrained) of the estimated parameters is zero. The test statistic is calculated as $t = \frac{\bar{d}}{\sqrt{s_{\bar{d}}/145}}$, where $s_{\bar{d}}$ is the sample standard deviation for $\bar{d}$.

[18] Also, the standard errors of $\hat{\omega}_i$ and $\hat{\phi}_i$ for the unconstrained model are consistently on the order of 10^{-4} and 10^{-5}, respectively, thus indicating statistically significant differences in the probabilities.

Parameter	Benchmark model	Constrained	Unconstrained	Difference (p-value)
$\bar{\alpha}$	0.30	0.30	0.30	0.00 (0.17)
$\bar{\delta}$	0.47	0.47	0.47	0.00 (0.42)
$\bar{\varepsilon}$	77.8	77.16	77.12	0.04 (0.19)
$\bar{\mu}$	83.5	82.5	82.5	0.00 (0.99)
$\bar{\omega}$	-	0.43	0.45	-0.02 (0.00)***
$\bar{\phi}$	-	0.43	0.42	0.01 (0.00)***
$\overline{LLF}$	-15111	-12087.25	-12087.46	0.21 (0.08)*

Table 3: THE INFORMATION ROLE OF TRADE SIZE.
The first column lists the average estimates for the model, which do not account for the trade size. The second and third columns represent the average estimates of the parameters in the constrained ($\omega_i = \phi_i$) and unconstrained ($\omega_i \neq \phi_i$; $i = 1, \ldots, 145$) versions of the model, respectively. The last column contains the differences in mean value between the 145 parameters estimated from the constrained and unconstrained models. The p-value comes from the paired t-test for the null hypothesis of the difference being equal to zero. LLF denotes the value of the log likelihood function. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

estimates.

We find that in the cutoff range between 3000 and 8000, all the estimates are stable and the informed large trade size is more informative than the uninformed large trade size. The choice of cutoff values above 8000 (e.g., 12000 in Table 4) distorts the results due to the low frequency of such large trades. Similarly, it is unreasonable to consider trades above small cutoff values (e.g., 2000 in Table 4) to be "large," in which case the observed effects diminish. These findings confirm the work of Chakravarty (2001) and Anand and Chakravarty (2007), who find that medium-sized trades are the most informative. This can also be interpreted as a "separating equilibrium" outcome in which informed traders submit mainly large orders (Easley and O'Hara, 1987).[19] An interesting observation emerges from Table 4: the probability of both informed and uninformed large trading declines with the cutoff value. This can be explained by the fact that increasing the cutoff value eliminates the majority of the transactions that qualify as "large" trades.

4. Price dynamics

4.1. Model estimates, returns and volatility

As a final test of the model's usefulness, we regress exchange rate returns (and squared returns) on the theoretical variables that comprise the model. Our regressions are of the following form

$$r_t = \alpha + \gamma M_t + v_t, \quad v_t \sim IID(0, \sigma^2) \tag{1}$$

[19] Suppose that the constrained model is found to be more appropriate. This would indicate a "pooling equilibrium," where informed traders submit both large and small orders roughly equally.

Parameter	2000 cutoff	8000 cutoff	12000 cutoff
$\bar{\alpha}$	0.31 0.00 (0.43)	0.30 0.00 (0.14)	0.30 0.00 (0.33)
$\bar{\delta}$	0.48 0.00 (0.53)	0.47 0.00 (0.66)	0.47 0.02 (0.38)
$\bar{\varepsilon}$	74.43 0.02 (0.59)	78 0.00 (0.15)	77.93 0.05 (0.19)
$\bar{\mu}$	74.54 0.42 (0.40)	83.11 1.34 (0.12)	82.65 0.82 (0.19)
$\bar{\omega}$	0.73 -0.02 (0.00)***	0.36 -0.01 (0.00)***	0.24 -0.00 (0.02)**
$\bar{\phi}$	0.70 0.01 (0.00)***	0.34 0.01 (0.00)***	0.23 0.00 (0.00)***
$\overline{LLF}$	0.12 (0.15)	1.66 (0.09)*	0.97 (0.17)

Table 4: The robustness of the estimates with respect to "large" trade size. For each cutoff amount for a "large" trade (2000, 8000 and 12000), this table presents the average parameter estimates from an unconstrained model along with the average difference between the estimates from the two versions of the model: constrained ($\omega_i = \phi_i$) and unconstrained ($\omega_i \neq \phi_i$; $i = 1, \ldots, 145$). More precisely, each column represents the merged columns 3 and 4 from Table 3 for different cutoff amounts. LLF denotes the average value of difference between the log likelihood function for the two models. The p-value reported in the brackets comes from the paired t-test for the null hypothesis of the difference being equal to zero. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

$$r_t^2 = \alpha + \beta r_{t-1}^2 + \gamma M_t + v_t, \quad v_t \sim IID(0, \sigma^2) \tag{2}$$

where $r_t = \ln(P_t) - \ln(P_{t-1})$, $M_t \in \{\alpha_t, \delta_t, \varepsilon_t, \mu_t, \omega_t, \phi_t\}$ and $t = 1, \ldots, 145$. By P_t we denote the daily close (USD/EUR) on day t. Table 5 reports the results for the M_t variables found to be significant in the first regression.[20] Since the estimate for δ_t is statistically significant with a relatively high adjusted R^2 value (0.2675), these results essentially confirm those from Easley *et al.* (1997b), which also find the probability of bad news informative for price determination. However, their estimated coefficient on the bad event probability variable is much larger than ours $\hat{\gamma}$. More specifically, we find that a one-percent increase in δ_t decreases the exchange rate by about 1.5 cents, while the estimate in Easley *et al.* (1997b) is 0.61. This suggests that the equity market is more responsive than the FX market to the arrival of bad news.

We next discuss the results of regressing the squared FX returns, i.e., the measure of daily volatility, on the arrival rates (the only variables that are statistically significant in the second regression). We find that the impact of both arrival rates on volatility is positive and statistically significant. According to Table 6, the magnitude of the estimated γ for ε_t is about two times larger than that for μ_t. Since the PIN is insignificant in the second regression, we conclude that although volatility increases with the arrival rates of traders, it is independent of the trade composition.

Finally, to address the issue of the potentially strategic arrival of informed traders, we conduct Granger causality tests between the trader arrival rates. Essentially, the Granger causality test

[20] Although the coefficient on ω_t is not significant for the whole sample, it is significant for the first half of the sample. This means that the probability of informed traders submitting a large order affects the returns. These results are available by request from the authors.

r_t	Constant	$M_t = \delta_t$
Coefficient	0.007***	-0.015***
Standard error	0.001	-0.002
t-statistic	6.26	-7.29
p-value	0.000	0.000
Adjusted R^2	0.2675	

Table 5: REGRESSION RESULTS - RETURNS.
This table reports the results of the OLS regression: $r_t = \alpha + \gamma M_t + v_t,\ \ t = 1, \ldots, 145$. The dependent variable (r_t) is the return for the spot USD/EUR exchange rate at time t. The explanatory variable is the probability of a bad information event (δ_t). *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

assesses the ability of one series to forecast another. The idea is that if the informed traders move strategically to match the activity of uninformed traders, we may be able to forecast their arrival.

Granger causality between ε_t and μ_t is estimated using a standard bivariate framework. In Table 7, we report the results of regressions estimated with two lags. The results are similar if we change to 1, 3 or more lags. We cannot reject the hypothesis that the arrival rates of informed traders do not Granger-cause the arrival rates of uninformed traders. However, we can reject the hypothesis that the arrival rates of uninformed traders do not cause the arrival rates of informed traders. This indicates that our conjecture of the strategic arrival of informed traders is valid. Combined with the findings from previous sections, this suggests that uninformed traders arrive first, while informed traders follow after strategically timing their arrival. Therefore, the assumption that informed traders are risk neutral is not appropriate in the FX market context.[21]

4.2. Trading, returns and volatility

Section 2 presented the hour-of-day indices of informed and uninformed traders based on unbalanced traders ($|K|$) and balanced traders ($TT - |K|$). By definition, K represents "trade imbalances" and can be interpreted as a variant of the market order flow (Evans and Lyons, 2002). We test whether this variable can explain FX returns on an hourly basis by running the following regression:

$$r_t = \alpha + \beta K_t + v_t, \quad v_t \sim IID(0, \sigma^2), \quad t = 1, \ldots, 3479 \tag{3}$$

,
where r_t is defined as before and $K_t = S_t - B_t$ for each hour. Table 8 lists the estimates demonstrating that trade imbalances significantly determine hourly returns. Essentially, a one-unit increase

[21] One may expect the arrival of informed traders to be related only to the information flow. However, in this setting, they appear to use their private information strategically. Another possibility is that informed traders enter the market not only to establish speculative positions (information effects), but also to adjust their currency inventory (inventory effects), as previously mentioned.

r_t^2	Constant	r_{t-1}^2	$M_t = \mu_t$
Coefficient	0.000***	-0.205**	3.25e-07*
Standard error	0.000	0.081	1.72e-07
t-statistic	2.70	-2.51	1.89
p-value	0.008	0.013	0.061
Adjusted R^2	0.0509		
r_t^2	Constant	r_{t-1}^2	$M_t = \varepsilon_t$
Coefficient	0.000	-0.218***	7.03e-07**
Standard error	0.000	0.081	3.04e-07
t-statistic	0.73	-2.67	2.31
p-value	0.466	0.008	0.022
Adjusted R^2	0.0626		

Table 6: Regression results - volatility.
This table reports the results of the OLS regression: $r_t^2 = \alpha + \beta r_{t-1}^2 + \gamma M_t + v_t, \;\; t = 1, \ldots, 145$. The dependent variable (r_t^2) is the squared return for the spot USD/EUR exchange rate at time t. The explanatory variables are the lagged squared returns (r_{t-1}^2), the arrival rate of informed traders (μ_t), and the arrival rate of uninformed traders (ε_t). *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Null hypothesis	χ^2	Prob > χ^2
μ_t does not cause ε_t	2.51	0.28
ε_t does not cause μ_t	7.77	0.02

Table 7: Granger causality tests using daily observations.
This table lists the probabilities from Granger causality tests on the estimates of the arrival rates for 145 days in the sample. The test statistic is distributed as $\chi^2(df = 2)$, with critical value χ^2_{cr}=5.991 for the 5% significance level. The null hypothesis is stated in the first column. μ_t and ε_t ($t = 1, \ldots, 145$) denote the arrival rates of the informed and uninformed traders, respectively. The estimations are based on a standard bivariate framework. The figures in the third column are the probabilities of rejection.

in the trade imbalance (i.e., one additional unique seller relative to the buyers of the Euro over the one hour period) significantly decreases the USD/EUR exchange rate returns (by 2.37e-05). This confirms evidence from Evans and Lyons (2002) and many other authors who have documented (contemporaneous) microstructure effects in the FX market.

Since K_t is quite different from the order flow definition as the difference between buyer-initiated and seller-initiated transactions typically used in the FX microstructure literature (see, e.g., Lyons, 2001), it would be useful to understand the relationship between the two measures more clearly. For that purpose, we construct hourly and daily market order flows based on the total buying and selling volumes. The OANDA FXTrade lists the transaction type, and that allows us to aggregate buy/sell market orders. We follow this by regressing the exchange rate returns on order flow, as in equation (3). Table 9 shows the results of using both daily and hourly data. As expected, the

r_t	Constant	K_t
Coefficient	4.12e-05*	-2.37e-05***
Standard error	2.39e-05	9.90e-07
t-statistic	1.72	-23.98
p-value	0.085	0.000
Adjusted R^2	0.1417	

Table 8: Trade imbalance regression results.
This table reports the results of the OLS regression: $r_t = \alpha + \beta K_t + v_t, \ \ t = 1, \ldots, 3479$. The dependent variable (r_t) is the hourly return for the spot USD/EUR exchange rate at time t. The explanatory variable is the trade imbalance (K_t) aggregated over a one-hour period. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

	Hourly		Daily	
r_t	Constant	X_t	Constant	X_t
Coefficient	7.00e-06	-5.67e-11***	0.0001396	-1.33e-10***
Standard error	0.0000255	6.52e-12	0.0005237	3.20e-11
t-statistic	0.27	-8.70	0.27	-4.17
p-value	0.784	0.000	0.790	0.000
Adjusted R^2	0.0210		0.0866	

Table 9: Order flow regression results.
This table reports the results of the OLS regression: $r_t = \alpha + \beta X_t + v_t$. The dependent variable (r_t) is the hourly or daily return for the spot USD/EUR exchange rate at time t. The explanatory variable is market order flow (X_t) aggregated over one-hour (Hourly) or 24-hour (Daily) periods. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

impact of the order flow on FX returns is significant and negative in both cases. Surprisingly, the trade imbalance seems to be a more appropriate explanatory variable for the hourly data. Also, in line with other studies (Evans and Lyons, 2005, Gradojevic, 2007), the explanatory power of order flow increases with the aggregation to daily data where the adjusted R^2=0.09. It is possible that the information content of order flow is obscured by high-frequency noise and that this effect becomes more pronounced through aggregation. On the other hand, by counting the number of *unique* buyers and sellers, the trade imbalance variable reduces the impact of the returning, more frequent traders whose trades may be less informative.

Since $|K|$ and $TT - |K|$ contain information on the hourly arrival of informed and uninformed traders, we next examine their explanatory power with respect to hourly returns, using a regression similar to equation (1). Specifically, we run two regressions: one with $M_t = |K_t|$ and one with $M_t = TT_t - |K_t|$ ($t = 1, \ldots, 3479$). To pin down the hourly impact of informed and uninformed trading on returns, we aggregate the information for each hour over 145 days (144 FX returns).

This yields 24 regressions, each with 144 observations. The top two panels of Figure 6 show the absolute value of $\hat{\gamma}_i$ ($i = 0, \ldots, 23$) for both types of traders. The top right panel models the series of regression coefficients using a cubic spline function.[22] This transformation is convenient because it provides smooth transition in the trader behavior over the 24-hour cycle.

The price impact is time-varying, and when significant, is more pronounced among informed traders. Strong geographic dependence emerges again. The first region of activity commences at 03:00, the opening of the LSE. This region ends after the opening of the NYSE (at 09:30). We also observe a significant price impact before the closing of the NYSE and during most of the hours of operation of the TSE. These findings are in line with those for the hour-of-day index from Section 2. Therefore, in general, the above-average arrival of informed traders has a substantial effect on the exchange rate movements.

Our final objective is to assess the impact of the hourly arrival of informed and uninformed traders on FX volatility. We estimate equation (2) with hourly squared returns and, as before, use $M_t = |K_t|$ for informed traders and $M_t = TT_t - |K_t|$ for uninformed traders ($t = 1, \ldots, 3479$). Here, we follow the same approach as for returns. The bottom two panels of Figure 6 show the estimated $\hat{\gamma}_i$ ($i = 0, \ldots, 23$) for both types of traders.

We observe that informed traders dominate uninformed traders with regard to their influence on hourly FX volatility. The strongest geographic dependence effects take place when both the NYSE and the LSE are open, i.e., between roughly 09:30 and 11:30. During these particular hours, informed traders appear capable of driving FX volatility and, to a certain degree, FX returns, despite below-average arrival. Hence, this exercise extracts local information that was not obvious from the theoretical microstructure model.[23]

It is also of interest to explore the different patterns of FX volatility responses to daily and hourly arrival rates of informed/uninformed traders. Recall that the impact of ε on volatility is more than twice as much as the impact of μ. First, we compute the percentile of daily volatility operated by informed and uninformed traders. This is done by comparing the absolute daily changes in volatility to the regression coefficients (γ) on μ and ε in equation (2), listed in Table 6. Based on the estimated volatility percentiles, we find that informed traders operate roughly the 1^{st} percentile, while uninformed traders operate between the 1^{st} and the 5^{th} percentiles. We follow the same procedure for hourly data, finding that the range of γ from Figure 6 (bottom panel) falls into the 25^{th} percentile for informed traders. Further, the range of γ for uninformed traders is more narrow. We find that they operate in roughly the 10^{th} percentile of hourly volatility changes. Hence, it appears that the impact of informed traders on hourly volatility becomes *averaged out* (more than it does for the uninformed traders) when it is translated into the daily data.

[22] Cubic spline is an interpolation method that fits a curve by constructing piecewise third-order polynomials that pass through original data points (Burden and Faires, 2004).

[23] The Granger causality tests for the hourly arrivals yield findings similar to those for the daily arrival rates: $TT_t - |K_t|$ Granger-causes $|K_t|$, but not vice-versa.

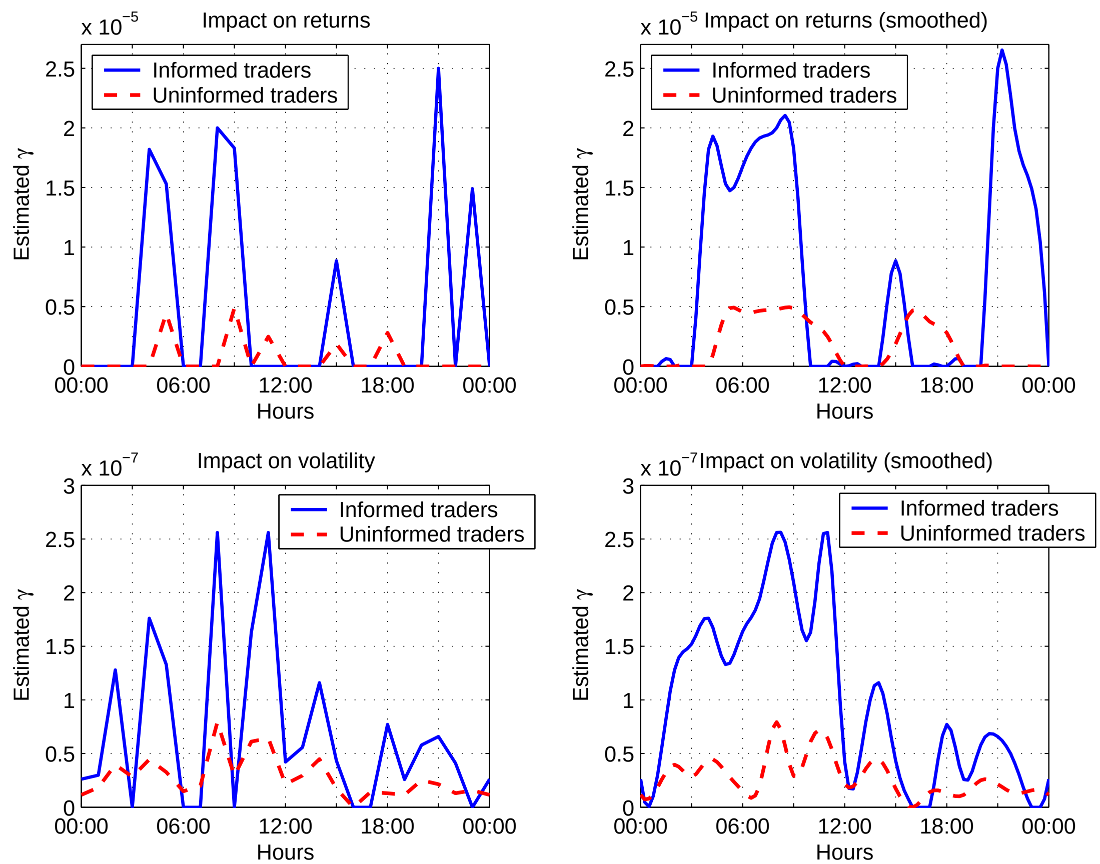

Figure 6: Top left: Geographic dependence of the impact of the traders on hourly FX returns. Top right: Geographic dependence of the impact of the traders on hourly FX returns smoothed using the cubic spline interpolation method. Bottom left: Geographic dependence of the impact of the traders on hourly FX volatility. Bottom right: Geographic dependence of the impact of the traders on hourly FX volatility smoothed using the cubic spline interpolation method. The absolute value of $\hat{\gamma}_i$ (the impact of informed traders on FX returns or volatility) for each hour ($i = 0, \ldots, 23$) related to informed traders is given by a solid line. $\hat{\gamma}_i$'s (the impact of uninformed traders on FX returns or volatility) for uninformed traders are given by a dashed line. Sample period: October 5, 2003, 16:00 - May 14, 2004, 15:59 (3480 hours, 145 business days).

5. Conclusions

Since the seminal work of Meese and Rogoff (1983) and their findings of a weak relationship between exchange rates and market fundamentals, significant attention has been devoted to the market microstructure of the FX market. Market microstructure proponents argue that order flows (or signed transaction flows) convey dispersed private information and incorporate it into exchange rates. On one hand, asymmetric information may exist in the FX market due to, for instance, differences in valuation (Handa *et al.*, 2003). These effects can arise due to behavioral considerations (e.g., overconfidence, overreaction to news, etc.). Other research contributions that corroborate the private information hypothesis are summarized in Lyons (2001). On the other hand, it is difficult to imagine that traders are privately informed about macroeconomic factors (Bessembinder, 1994). Some very recent research evidence also supports the latter notion (Sager and Taylor, 2008; Gradojevic and Neely, 2008). These papers document the poor predictive ability of order flows in relation to exchange rate movements.

This paper utilizes a unique high-frequency dataset from OANDA to search for direct evidence of private information in an electronic spot FX market. Our model-free analysis in Section 2 reveals that some traders may possess private information in this market. We uncover two lines of evidence that support the informed trading hypothesis: consistent trading profitability and the ability to predict mid-quote movements. These intriguing findings motivate us to develop a high-frequency version of the model by Easley *et al.* (1996b) for the FX market and to address a number of important issues. First, we estimate parameters reflecting the market maker's beliefs about the arrival of informed traders to the market and the risk of informed trading. We establish the exact timing of the arrival of not only informed traders, but also that of uninformed traders. The findings indicate a strong strategic component in the activity of the informed traders that is not observed for the uninformed traders. This phenomenon operates at different levels, starting from the geographic (intraday) dependency to the day-of-week effects, and substantiated by Granger causality tests. The model by Easley *et al.* (1996b) does not disentangle the differences in valuation aspect of the trading behavior from informed trading (like in Handa *et al.* 2003). Thus, we leave this potential extension to future research. Our results may also be explained by differences in valuation. This phenomenon does not represent private information *per se*, as in equity markets, but implies the existence of asymmetric information effects. However, it is important to stress that these findings challenge the traditional view of exchange rate determination.

In addition to examining the behavior of informed traders, our evidence on the impact of the model parameters on returns and volatility sheds new light on the microstructure of FX markets. The movements in FX returns appear to be driven by the probability of an information event and the probability of large informed trading. In particular, the market maker views both of these probabilities as having a negative impact on the price. Our study also indicates that FX rate volatility increases with the arrival rates of uninformed and informed traders, but that the trade composition (i.e., the PIN) has no effect on volatility.

An important advantage of our approach is that the nature of currency orders is directly observable, which attests to the accuracy of our results. This differs from equity markets, where trade classification data is often unavailable, and this meant that trade classification algorithms were not required to differentiate between buyer- and seller-initiated trades (see, e.g., Lee and Ready, 1991). For example, Boehmer *et al.* (2007) find that trade misclassification results in a downward bias in the estimate of the PIN for the Easley *et al.* (1996b) framework.

The extension of the model to account for the role of volume reveals that the transactions of informed traders are related to larger trade sizes. These findings are robust with regard to reasonable choices for cutoff points defining a "large" trade, although medium-sized trades can be informative. Therefore, the current paper adds to the literature providing evidence on the link between informed trading and larger trade sizes (e.g., Easley *et al.*, 1997a, Menkhoff and Schmeling, 2007, Chakravarty, 2001 and Anand and Chakravarty, 2007).

The model assumes independence of information events across hours. However, we observe dependence for about 15% of the days in our sample. This may bias the standard errors of our estimates. Though we conjecture this bias to be minimal, it is worth emphasizing that introducing dependency to the model is a key direction of future research. The impact of past transactions on current transactions could be modeled by a latent parameter measuring the degree of serial correlation. In this context, testing for inter-day dependency may offer broader insight into FX trading patterns and strategies. Another future research avenue we envisage concerns the over-dispersion frequently found in transactions data, which reduces the usefulness of the Poisson distribution.[24] We hope that the evidence presented here will lead to a new structural microstructure model applicable to both equity and FX markets.

[24]This point is also made in Wuensche (2007), who proposes mixed Poisson distributions to capture the characteristics of the trade data. Other potential weaknesses of the Easley *et al.* (1996b) model can be found in Venter and De Jongh (2004).

Appendix A. Independent Arrival Model (Easley *et al.*, 1996b)

The model consists of informed and uninformed traders and a risk neutral competitive market maker. The traded asset is a foreign currency for the domestic currency. Similar to the portfolio shifts model (Evans and Lyons, 2002), the trades and the governing price process are generated by the quotes of the market maker over a twenty-four hour trading day. Within any trading hour, the market maker is expected to buy and sell currencies from his posted bid and ask prices. The price process is the expected value of the currency based on the market maker's information set at the time of the trade.

The hourly arrival of news occurs with the probability α. This represents bad news with probability δ and good news with $1-\delta$ probability. Let $\{p_i\}$ be the hourly price process over $i = 1, 2, \ldots, 24$ hours. p_i is assumed to be correlated across hours and will reveal the intraday time dependence and intraday persistence of the price behavior across these two classes of traders. The lower and upper bounds for the price process should satisfy $p_i^b < p_i^n < p_i^g$ where p_i^b, p_i^n and p_i^g are the prices conditional on bad, no news and good news, respectively. Within each hour, time is continuous and indexed by $t \in [0, T]$.

In any trading hour, the arrivals of informed and uninformed traders are determined by independent Poisson processes. At each instant within an hour, uninformed buyers and sellers each arrive at a rate of ε. Informed traders only trade when there is news, and arrive at a rate of μ. All informed traders are assumed to be risk neutral and competitive, and are therefore expected to maximize profits by buying when there is good news and selling otherwise.[25] For good news hours, the arrival rates are $\varepsilon + \mu$ for buy orders and ε for sell orders. For bad news hours, the arrival rates are ε for buy orders and $\varepsilon + \mu$ for sell orders. When no news exists, the buy and sell orders arrive at a rate of ε per hour.

The market maker is assumed to be a Bayesian who uses the arrival of trades and their intensity to determine whether a particular trading hour belongs to a no news, good news or bad news category. Since the arrival of hourly news is assumed to be independent, the market maker's hourly decisions are analyzed independently from one hour to the next. Let $P(t) = (P_n(t), P_b(t), P_g(t))$ be the market maker's prior beliefs with no news, bad news, and good news at time t. Accordingly, his or her prior beliefs before trading starts each day are $P(0) = (1-\alpha, \alpha\delta, \alpha(1-\delta))$.

Let S_t and B_t denote sell and buy orders at time t. The market maker updates the prior conditional on the arrival of an order of the relevant type. Let $P(t|S_t)$ be the market maker's updated belief conditional on a sell order arriving at t. $P_n(t|S_t)$ is the market maker's belief about no news conditional on a sell order arriving at t. Similarly, $P_b(t|S_t)$ is the market maker's belief about the occurrence of bad news events conditional on a sell order arriving at t, and $P_g(t|S_t)$ is the market maker's belief about the occurrence of good news conditional on a sell order arriving

[25] This assumption may seem inappropriate, given that it rules out any strategic behavior. As shown in Section 2.2, informed traders have some tendency to trade strategically. Therefore, we concur that the assumption of risk neutrality needs defending, but we retain it for the sake of the model applicability.

at t.

The probability that any trade occurring at time t is information-based is (please see Appendix B)

$$i(t) = \frac{\mu(1 - P_n(t))}{2\varepsilon + \mu(1 - P_n(t))} \tag{4}$$

Since each buy and sell order follows a Poisson Process at each trading hour and orders are independent, the likelihood of observing a sequence of orders containing B buys and S sells in a bad news hour of total time T is given by

$$L_b((B,S)|\theta) = L_b(B|\theta)L_b(S|\theta) = e^{-(\mu+2\varepsilon)T}\frac{\varepsilon^B(\mu+\varepsilon)^S T^{B+S}}{B!S!}, \tag{5}$$

where $\theta = (\alpha, \delta, \varepsilon, \mu)$.

Similarly, in a no-event hour, the likelihood of observing any sequence of orders that contains B buys and S sells is

$$L_n((B,S)|\theta) = L_n(B|\theta)L_n(S|\theta) = e^{-2\varepsilon T}\frac{\varepsilon^{B+S}T^{B+S}}{B!S!} \tag{6}$$

In a good-event hour, this likelihood is

$$L_g((B,S)|\theta) = L_g(B|\theta)L_g(S|\theta) = e^{-(\mu+2\varepsilon)T}\frac{\varepsilon^S(\mu+\varepsilon)^B T^{B+S}}{B!S!} \tag{7}$$

The likelihood of observing B buys and S sells in an hour of unknown type is the weighted average of equations (2), (3), and (4) using the probabilities of each type of hour occurring.

$$\begin{aligned} L((B,S)|\theta) &= (1-\alpha)L_n((B,S)|\theta) + \alpha\delta L_b((B,S)|\theta) + \alpha(1-\delta)L_g((B,S)|\theta) \\ &= (1-\alpha)e^{-2\varepsilon T}\frac{\varepsilon^{B+S}T^{B+S}}{B!S!} + \alpha\delta e^{-(\mu+2\varepsilon)T}\frac{\varepsilon^B(\mu+\varepsilon)^S T^{B+S}}{B!S!} \\ &\quad + \alpha(1-\delta)e^{-(\mu+2\varepsilon)T}\frac{\varepsilon^S(\mu+\varepsilon)^B T^{B+S}}{B!S!} \end{aligned} \tag{8}$$

Because hours are independent, the likelihood of observing the data $M = (B_i, S_i)_{i=1}^{I}$ over twenty-four hours ($I = 24$) is the product of the hourly likelihoods,

$$\begin{aligned} L(M|\theta) &= \prod_{i=1}^{I} L(\theta|B_i, S_i) = \prod_{i=1}^{I} \frac{e^{-2\varepsilon T}T^{B_i+S_i}}{B_i!S_i!} \times \\ &\quad \left[(1-\alpha)\varepsilon^{B_i+S_i} + \alpha\delta e^{-\mu T}\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + \alpha(1-\delta)e^{-\mu T}\varepsilon^{S_i}(\mu+\varepsilon)^{B_i}\right] \end{aligned} \tag{9}$$

The log likelihood function is

$$\begin{aligned}
\ell(M|\theta) &= \sum_{i=1}^{I} \ell(\theta|B_i, S_i) \\
&= \sum_{i=1}^{I} [-2\varepsilon T + (B_i + S_i)\ln T] \\
&\quad + \sum_{i=1}^{I} \ln\left[(1-\alpha)\varepsilon^{B_i+S_i} + \alpha\delta e^{-\mu T}\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + \alpha(1-\delta)e^{-\mu T}\varepsilon^{S_i}(\mu+\varepsilon)^{B_i}\right] \\
&\quad - \sum_{i=1}^{I} (\ln B_i! + \ln S_i!) \qquad (10)
\end{aligned}$$

As in Easley *et al.* (2002), the log likelihood function, after dropping the constant and rearranging[26], is given by

$$\begin{aligned}
\ell(M|\theta) &= \sum_{i=1}^{I} [-2\varepsilon + M_i \ln x + (B_i + S_i)\ln(\mu+\varepsilon)] \qquad (11) \\
&\quad + \sum_{i=1}^{I} \ln\left[\alpha(1-\delta)e^{-\mu}x^{S_i - M_i} + \alpha\delta e^{-\mu}x^{B_i - M_i} + (1+\alpha)x^{B_i+S_i-M_i}\right],
\end{aligned}$$

where $M_i \equiv \min(B_i, S_i) + \max(B_i, S_i)/2$, and $x = \frac{\varepsilon}{\varepsilon+\mu} \in [0, 1]$.

[26] To derive equation (11), the term $\ln[x^{M_i}(\mu+\varepsilon)^{B_i+S_i}]$ is simultaneously added to the first sum and subtracted from the second sum in equation (10). This is done to increase computational efficiency and to ensure convergence in the presence of a large numbers of buys and sells, as is the case in our dataset.

Appendix B. Derivation of the PIN

By Bayes' rule, the market maker's posterior probability with no news at time t, if an order to sell arrives at t, is

$$\begin{aligned} P_n(t|S_t) &= \frac{P_n(S_t|t)P_n(t)}{P(S_t)} \\ &= \frac{P_n(S_t|t)P_n(t)}{P_n(S_t|t)P_n(t) + P_g(S_t|t)P_g(t) + P_b(S_t|t)P_b(t)} \\ &= \frac{\varepsilon P_n(t)}{\varepsilon(1 - P_g(t) - P_b(t)) + \varepsilon P_g(t) + (\varepsilon + \mu)P_b(t)} \\ &= \frac{\varepsilon P_n(t)}{\varepsilon + \mu P_b(t)}, \end{aligned}$$

where $P_n(S_t|t)$ is the probability of the arrival of a sell order conditional on no news at time t, $P_g(S_t|t)$ is the probability of the arrival of a sell order conditional on good news at time t, and $P_b(S_t|t)$ is the probability of the arrival of a sell order conditional on bad news at time t.

Similarly, the posterior probability on bad news is

$$\begin{aligned} P_b(t|S_t) &= \frac{P_b(S_t|t)P_b(t)}{P(S_t)} \\ &= \frac{(\varepsilon + \mu)P_b(t)}{\varepsilon + \mu P_b(t)}, \end{aligned}$$

and the posterior probability on good news is

$$\begin{aligned} P_g(t|S_t) &= \frac{P_g(S_t|t)P_g(t)}{P(S_t)} \\ &= \frac{\varepsilon P_g(t)}{\varepsilon + \mu P_b(t)} \end{aligned}$$

The bid price, $b(t)$, conditional on S_t at time t at hour i is

$$b(t) = P_n(t|S_t)p_i^n + P_b(t|S_t)p_i^b + P_g(t|S_t)p_i^g = \frac{\varepsilon P_n(t)p_i^n + (\varepsilon + \mu)P_b(t)p_i^b + \varepsilon P_g(t)p_i^g}{\varepsilon + \mu P_b(t)}$$

Similarly, the ask price $a(t)$ is the market maker's expected value of the asset conditional on the history prior to t and on B_t.

Thus, the ask at time t at hour i is

$$a(t) = P_n(t|B_t)p_i^n + P_b(t|B_t)p_i^b + P_g(t|B_t)p_i^g = \frac{\varepsilon P_n(t)p_i^n + \varepsilon P_b(t)p_i^b + (\varepsilon + \mu)P_g(t)p_i^g}{\varepsilon + \mu P_g(t)}$$

The expected price conditional on t is

$$E[p_i|t] = P_n(t)p_i^n + P_b(t)p_i^b + P_g(t)p_i^g,$$

where $P_n(t)$, $P_b(t)$ and $P_g(t)$ are the prior beliefs of the market maker for no news, bad news and good news at time t.

Substituting the expected price equation into the equations for bid and ask prices yields

$$\begin{aligned} b(t) &= \frac{\varepsilon P_n(t)p_i^n + \varepsilon P_b(t)p_i^b + \varepsilon P_g(t)p_i^g + \mu P_b(t)p_i^b}{\varepsilon + \mu P_b(t)} \\ &= \frac{\varepsilon E[p_i|t] + \mu P_b(t)p_i^b}{\varepsilon + \mu P_b(t)} \\ &= \left[1 - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}\right] E[p_i|t] + \frac{\mu P_b(t)p_i^b}{\varepsilon + \mu P_b(t)} \\ &= E[p_i|t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(E[p_i|t] - p_i^b), \end{aligned}$$

and

$$\begin{aligned} a(t) &= \frac{\varepsilon P_n(t)p_i^n + \varepsilon P_b(t)p_i^b + \varepsilon P_g(t)p_i^g + \mu P_g(t)p_i^g}{\varepsilon + \mu P_g(t)} \\ &= \frac{\varepsilon E[p_i|t] + \mu P_g(t)p_i^g}{\varepsilon + \mu P_g(t)} \\ &= \left[1 - \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}\right] E[p_i|t] + \frac{\mu P_g(t)p_i^g}{\varepsilon + \mu P_g(t)} \\ &= E[p_i|t] + \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(p_i^g - E[p_i|t]) \end{aligned}$$

Let $d(t) = a(t) - b(t)$ be the spread at time t.

$$\begin{aligned} d(t) &= E[p_i|t] + \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(p_i^g - E[p_i|t]) - \left[E[p_i|t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(E[p_i|t] - p_i^b)\right] \\ &= \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(p_i^g - E[p_i|t]) + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(E[p_i|t] - p_i^b) \end{aligned}$$

The spread for the opening quotes is

$$\begin{aligned} d(0) &= \frac{\mu P_g(0)}{\varepsilon + \mu P_g(0)}(p_i^g - E[V_i]) + \frac{\mu P_b(0)}{\varepsilon + \mu P_b(0)}(E[p_i] - p_i^b) \\ &= \frac{\mu\alpha(1-\delta)}{\varepsilon + \mu\alpha(1-\delta)}(p_i^g - E[p_i|t]) + \frac{\mu\alpha\delta}{\varepsilon + \mu\alpha\delta}(E[p_i|t] - p_i^b) \end{aligned}$$

If good and bad events are equally likely, that is, if $\delta = 1 - \delta$, $\delta = 0.5$. Thus

$$d(0) = \frac{\mu\alpha}{2\varepsilon + \mu\alpha}(p_i^g - p_i^b)$$

The probability that any trade occurring at time t is information-based is

$$\begin{aligned} i(t) &= \frac{P_b(t)\mu + P_g(t)\mu}{P(B_t, S_t)} \\ &= \frac{\mu(1 - P_n(t))}{P_n(t)P_n(B_t, S_t|t) + P_g(t)P_g(B_t, S_t|t) + P_b(t)P_b(B_t, S_t|t)} \\ &= \frac{\mu(1 - P_n(t))}{2\varepsilon + \mu(1 - P_n(t))} \end{aligned}$$

Appendix C. Derivation of the log likelihood function for the extended model

The likelihood of observing a sequence of orders with LB large buys, SB small buys, LS large sells and SS small sells in a bad news hour is

$$L_b((LB,LS,SB,SS)|\theta) = L_b(LB|\theta)L_b(LS|\theta)L_b(SB|\theta)L_b(SS|\theta)$$
$$= e^{-(\mu+2\varepsilon)T}\frac{(\varepsilon\phi)^{LB}[\varepsilon(1-\phi)]^{SB}(\varepsilon\phi+\mu\omega)^{LS}[\varepsilon(1-\phi)+\mu(1-\omega)]^{SS}T^{LB+LS+SB+SS}}{LB!LS!SB!SS!},$$

where $\theta = (\alpha,\delta,\varepsilon,\mu,\omega,\phi)$. On a no-event day, the likelihood of observing a sequence of LB large buys, SB small buys, LS large sells and SS small sells is

$$L_n((LB,LS,SB,SS)|\theta) = L_n(LB|\theta)L_n(LS|\theta)L_n(SB|\theta)L_n(SS|\theta)$$
$$= e^{-2\varepsilon T}\frac{\phi^{LB+LS}(1-\phi)^{SB+SS}(\varepsilon T)^{LB+LS+SB+SS}}{LB!LS!SB!SS!}$$

On a good-event day, the likelihood is

$$L_g((LB,LS,SB,SS)|\theta) = L_g(LB|\theta)L_g(LS|\theta)L_g(SB|\theta)L_g(SS|\theta)$$
$$= e^{-(\mu+2\varepsilon)T}\frac{(\varepsilon\phi)^{LS}[\varepsilon(1-\phi)]^{SS}(\varepsilon\phi+\mu\omega)^{LB}[\varepsilon(1-\phi)+\mu(1-\omega)]^{SB}T^{LB+LS+SB+SS}}{LB!LS!SB!SS!}$$

As before, the likelihood of observing LB large buys, SB small buys, LS large sells and SS small sells is the weighted average of the above equations:

$$L((LB,LS,SB,SS)|\theta) = (1-\alpha)L_n(.|\theta) + \alpha\delta L_b(.|\theta) + \alpha(1-\delta)L_g(.|\theta)$$

Since we use hourly data, the likelihood of observing the data $D = (LB_i, LS_i, SB_i, SS_i)_{i=1}^{I}$ over twenty-four hours ($I = 24$) is the product of the hourly likelihoods as follows

$$L(D|\theta) \;=\; \prod_{i=1}^{I} L(\theta|LB_i, LS_i, SB_i, SS_i)$$

The log likelihood function is

$$\begin{aligned}
\ell(D|\theta) &= \sum_{i=1}^{I} \ell(\theta|LB_i, LS_i, SB_i, SS_i) \\
&= \sum_{i=1}^{I} [-2\varepsilon + M_i \ln x + N_i \ln y] \\
&\quad + \sum_{i=1}^{I} [(LB_i + LS_i) \ln(\varepsilon\phi + \mu\omega) + (SB_i + SS_i) \ln(\varepsilon(1-\phi) + \mu(1-\omega))] \\
&\quad + \sum_{i=1}^{I} \ln[(1-\alpha) x^{LB_i + LS_i - M_i} y^{SB_i + SS_i - N_i} + \alpha\delta e^{-\mu} x^{LB_i - M_i} y^{SB_i - N_i} \\
&\quad + \alpha(1-\delta) e^{-\mu} x^{LS_i - M_i} y^{SS_i - N_i}],
\end{aligned}$$

where $M_i \equiv \min(LB_i, LS_i) + \max(LB_i, LS_i)/2$, $N_i \equiv \min(SB_i, SS_i) + \max(SB_i, SS_i)/2$, $y = \frac{\varepsilon(1-\phi)}{\varepsilon(1-\phi)+\mu(1-\omega)} \in [0,1]$ and $x = \frac{\varepsilon\phi}{\varepsilon\phi+\mu\omega} \in [0,1]$. Here, to receive the final expression, the terms $\ln[x^{M_i}(\mu\omega + \varepsilon\phi)^{LB_i + LS_i}]$ and $\ln[y^{N_i}(\mu(1-\omega) + \varepsilon(1-\phi))^{SB_i + SS_i}]$ are added to and subtracted from the right-hand side of the likelihood equation.

References

Anand, A. and Chakravarty, S. (2007). Stealth trading in options markets. *Journal of Financial and Quantitative Analysis*, **42**, 167–188.

Bessembinder, H. (1994). Bid-ask spreads in the interbank foreign exchange market. *Journal of Financial Economics*, **35**, 317–348.

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

Boehmer, E., Grammig, J., and Theissen, E. (2007). Estimating the probability of informed trading-does trade misclassification matter? *Journal of Financial Markets*, **10**(1), 26–47.

Burden, R. and Faires, J. D. (2004). *Numerical Analysis*. Brooks Cole, Florence.

Chakravarty, S. (2001). Stealth-trading: Which traders trades move stock prices? *Journal of Financial Economics*, **61**(2), 289–307.

Dacorogna, M., Gençay, R., Muller, U., Olsen, R., and Pictet, O. (2001). *An Introduction to High-Frequency Finance*. Academic Press, San Diego.

Daniel, K. and Hirshleifer, D.and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, **53**(6), 1839–1885.

Easley, D. and O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, **19**, 69–90.

Easley, D. and O'Hara, M. (1992). Time and the process of security prices adjustment. *Journal of Finance*, **47**, 577–605.

Easley, D., Kiefer, N., and O'Hara, M. (1996a). Cream-skimming or profit-sharing? The curious role of purchased order flow. *Journal of Finance*, **51**, 811–833.

Easley, D., Kiefer, N., O'Hara, M., and Paperman, J. (1996b). Liquidity, information and infrequently traded stocks. *Journal of Finance*, **51**, 1405–1436.

Easley, D., Kiefer, N., and O'Hara, M. (1997a). The information content of the trading process. *Journal of Empirical Finance*, **4**, 159–185.

Easley, D., Kiefer, N., and O'Hara, M. (1997b). One day in the life of a very common stock. *Review of Financial Studies*, **10**, 805–835.

Easley, D., Engle, R., O'Hara, M., and Wu, L. (2002). Time-varying arrival rates of informed and uninformed trades. Working paper. hhttp://www.bnet.fordham.edu/lwu.

Evans, M. and Lyons, R. (2002). Order flow and exchange rate dynamics. *Journal of Political Economy*, **110**, 170–180.

Evans, M. and Lyons, R. (2005). Meese-rogoff redux: Micro-based exchange rate forecasting. *American Economic Review*, **95**(2), 405–414.

Evans, M. and Lyons, R. (2007). Exchange rate fundamentals and order flow. Working Paper.

Gallant, A., Rossi, P., and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, **5**(2), 199–242.

Glosten, L. and Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, **14**, 71–100.

Gradojevic, N. (2007). The microstructure of the Canada/U.S. dollar exchange rate: A robustness test. *Economics Letters*, **94**(3), 426–432.

Gradojevic, N. and Neely, C. (2008). The dynamic interaction of order flows and the CAD/USD exchange rate. Working Paper.

Handa, P., Schwartz, R., and Tiwari, A. (2003). Quote setting and price formation in an order driven market. *Journal of Financial Markets*, **6**, 461–489.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica*, **53**(6), 1315–1335.

Kyle, A. and Villa, J.-L. (1991). Noise trading and takeovers. *RAND Journal of Economics*, **22**(2), 54–71.

Lee, C. and Ready, M. (1991). Inferring trade direction from intraday data. *Journal of Finance*, **46**, 733–746.

Lei, Q. and Wu, G. (2005). Time-varying informed and uninformed trading activities. *Journal of Financial Markets*, **8**(2), 153–181.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

Lyons, R. (1995). Test of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics*, **39**, 321–351.

Lyons, R. (2001). *The Microstructure Approach to Exchange Rates.* The MIT Press, Cambridge, USA.

Madhavan, A. and Smidt, S. (1993). An analysis of changes in specialist inventories and quotations. *Journal of Finance*, **48**, 1595–1628.

Marsh, I. W. and O'Rourke, C. (2005). Customer order flow and exchange rate movements: Is there really information content? Working Paper.

Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, **14**, 3–24.

Menkhoff, L. and Schmeling, M. (2007). Whose trades convey information? Evidence from a cross-section of traders. Working Paper.

Odders-White, E. R. and Ready, M. J. (2007). The probability and magnitude of information events. *Journal of Financial Economics*. Forthcoming.

Payne, R. (2003). Informed trade in spot foreign exchange markets: An empirical investigation. *Journal of International Economics*, **61**, 307–329.

Sager, M. and Taylor, M. (2008). Commercially available order flow data and exchange rate movements: Caveat emptor. *Journal of Money, Credit and Banking*, **40**(4), 585–625.

Shapiro, S. and Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrica*, **52**(3-4), 591–611.

Venter, J. and De Jongh, D. (2004). Extending the ekop model to estimate the probability of informed trading. Working Paper.

Vitale, P. (2007). Optimal informed trading in the foreign exchange market. CEPR Discussion Paper no. 6553.

Wuensche, O. (2007). Using mixed Poisson distributions in sequential trade models. Working Paper.

Yao, J. (1998). Market making in the interbank foreign exchange market. Working Paper No. S-98-3.