When do informed traders arrive in foreign exchange markets?

Ramazan Gençay* Nikola Gradojevic† Faruk Selçuk‡

April 2008

Abstract

This article examines the implications of the existence of private information in the spot foreign exchange market. Our framework is a high-frequency version of a structural microstructure trade model that measures the market maker’s beliefs directly. We find that the underpinnings for the time-varying pattern of the probability of informed trading are rooted in the strategic arrival of informed traders on a particular hour-of-day, day-of-week, and geographic location (market). Specifically, we document that informed traders not only pick the low activity hours, but also attach the largest market weight to a particular market. The distributions of the estimated arrival rates confirm the commitment of the informed traders to strategic trading activities. In our framework, we acknowledge that an expected loss of informed trading to the market maker is a function of both the probability of informed trading and its likely impact on the price. The impact of the uninformed traders’ arrival on the daily foreign exchange price volatility is about twice the magnitude of the one for informed traders. These effects are in stark contrast to the findings from the hourly data that indicate dominance of informed traders. Finally, the results relate the informational content of trading to the trade size and suggest that the probability of the informed large trading is significantly higher than the probability of uninformed large trading.

Keywords: Foreign Exchange Markets; Volume; Informed Trading; Noise Trading

JEL No: G0, G1

*Corresponding author. Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada. Email: gencay@sfu.ca. Ramazan Gençay gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada. We are grateful to Olsen (www.olsen.ch), Switzerland for providing the data.

†Faculty of Business Administration, Lakehead University, 955 Oliver Road, Thunder Bay, ON P7B 5E1, Canada. Phone: 1 807 343-8419. Fax: 1 807 343-8443. Email: ngradoje@lakeheadu.ca

‡Department of Economics, Bilkent University, Bilkent 06800, Ankara, Turkey. Passed away during this research.
1. Introduction

Although information arrivals and the existence of informed and uninformed trades have received considerable attention in equity markets, the same cannot be said for the foreign exchange (FX) markets, mostly due to lack of transaction data for prices and trading volume.\(^1\) Easley and O’Hara (1992) have introduced a sequential trade model in which a market maker learns from both trades and 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, 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 in 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 the arrival rates of informed and uninformed trades to be time-varying. 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, but the actual trading is processed in the bookkeeping of particular markets with the major ones being 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 are different, 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. In contrast, the lowest market presence outside weekends can be found during the lunch-time break at the Tokyo Stock Exchange (TSE) when it is night in North America and Europe.

The intraday analysis of FX prices demonstrates that hourly returns exhibit seasonal fluctuations that are correlated with the hours of operation of the main worldwide market centers (Dacorogna et al. (2001)). The first contribution of this paper is to introduce a geographical (or the hour-of-day) component to the activity of informed and uninformed traders. Considering the low-transparency feature of the FX market, this exercise is of immense importance to our understanding of the market dynamics. Moreover, we also investigate the day-of-week effects of the arrival of informed and uninformed 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 measures the market maker’s beliefs directly, rather than searching for indirect evidence of informed trading. We find strong support for intraday geographic seasonality patterns in the arrival of informed traders, whereas the uninformed traders arrive uniformly. The intraday seasonality pinpoints two target markets for informed traders: the NYSE (during the last two hours of operation) and the TSE (during full trading hours). Noteworthy, the above-average activity of

---

\(^1\)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 about one year of daily data. To control for high-frequency noise effects and no-trade periods, we aggregate to hourly data in the present paper. The original data set contains tick-by-tick data over eight months of 2003-2004 or about 6.5 million data lines.
informed traders coincides with the low overall trading activity in both markets. This indicates to a certain extent strategic arrival timing of informed traders that not only pick the low activity hours, but also attach the largest market weight to a particular market - the TSE.\textsuperscript{2}

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, on a given day, a high (low) arrival rate of informed traders is observed, the estimated PIN is low (high). Also, informed traders strategically follow the arrival rates of the uninformed traders, i.e., tend to conceal their activity by transacting together with uninformed traders. This result is similar to the evidence by Kyle and Villa (1991) for the equity market 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 an expected loss of informed trading to the market maker 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 and the paper by Easley \textit{et al.} (1996b) only conjectures that the model explains the actual price dynamics. We extract the informational 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 significantly determine FX returns as predicted by Evans and Lyons (2002).

We also show that the arrival rates of informed and uninformed traders significantly determine hourly and daily FX rate volatility. The impact of the uninformed traders’ arrival on daily volatility is about twice the magnitude of the one for informed traders. On the other hand, the findings for the hourly data reveal persistent and dominant effects of the informed traders arrival on FX volatility. In addition, there is no evidence of the link between the PIN and volatility. The PIN can also be interpreted as a proportion of informed trading in the total trading activity (or the trade composition). The fact it is not informative for volatility means that, after sufficient trading activity takes place, information events are 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.\textsuperscript{3}

An important generalization of the model by Easley \textit{et al.} (1996b) that focuses on equity markets is the one that accounts for the trade size of a trader, which is commonly interpreted as an indicator

\textsuperscript{2}In contrast to our paper, Easley \textit{et al.} (2002) do not find any evidence of strategic behavior by informed traders. They document that uninformed traders seem to avoid informed traders by “herding”.

\textsuperscript{3}In a related study, Easley \textit{et al.} (2002) find that the trade composition does not forecast intraday volatility. Lei and Wu (2005) also recognize time series properties of the PIN on a panel of stocks and argue that the Easley \textit{et al.} (1996b) model should be extended with the time-varying PIN.
of informed trade. In traditional market microstructure theory, a larger trade size is typically viewed as being more informative since informed traders will try to trade a larger quantity to profit on their private information (see, for instance, Kyle (1985)). Consequently, one would expect to find a positive relationship between the trade size of a trader 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 the informed large trading is significantly higher than the probability of uninformed large trading. These findings are in accord with Easley et al. (1997a), but contrast Easley et al. (1997b) who document an insignificant difference in the probabilities of the informed and uninformed large trading.

In summary, our study contributes to the current market microstructure literature in several important ways. First, we use a unique high-frequency trading data for major FX rates which contains transaction prices and volume for one year and a total of ten million observations. Second, we extend the models by 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 model from Easley et al. (1996b). We not only utilize hourly data, but also estimate the model over 145 consecutive days and report both seasonal and average effects. Moreover, we are able to identify the arrival timing of both informed and uninformed traders with regard to the time-of-day, day of the week, and FX market. Fourth, we calculate the exact probability that an informed/uninformed trader trades a large amount and investigate its sensitivity to the definition of the large trade size. Fifth, we show that theoretical variables from the model can be informative for exchange rate determination.

The paper is organized as follows. Section 2 presents the model by Easley et al. (1996b) and analytically derives its log likelihood function. The next section describes the data and the estimates of the benchmark model. Section 4 extends the model to investigate the role of trade size. Section 5 examines the informativeness of the model estimates for the FX rate dynamics. Section 6 concludes.

2. Independent Arrival Model

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. The trades and the governing price process are generated from the quotes of the market maker over a trading day of twenty-four hours. Within any trading hour, the time is continuous, and the market maker is expected to buy and sell currencies from his posted bid and ask prices. The price process is the

---

4In this section, our framework follows Easley et al. (1996b).
expected value of the currency based on the market makers information set at the time of the trade.

The hourly arrival of news to the market occur with probability \( \alpha \). These are bad news with probability \( \delta \) and good news with \( 1 - \delta \) probability. Let \( \{s_i\} \) be the hourly price process over \( i = 1, 2, \ldots, 24 \) hours. \( s_i \) is assumed to be correlated across hours and will reveal the intraday seasonality and intraday persistence of the price behavior across these two classes of traders. The lower and upper bounds for the price process should satisfy \( s_b^i < s^n_i < s_g^i \) where \( s_b^i, s^n_i \) and \( s_g^i \) are the prices conditional on bad, no news and good news, respectively. Within each hour, the time is continuous and it is indexed by \( t \in [0, T] \).

On 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 \( \varepsilon \). Informed traders only trade when there is news and arrive at a rate of \( \mu \). All informed traders are assumed to be risk-neutral and competitive, and therefore they are expected to buy when there is good news and sell otherwise to maximize their profits.\(^5\) For good news hours, the arrival rates are \( \varepsilon + \mu \) for buy orders and \( \varepsilon \) for sell orders. For bad news hours, the arrival rates for buy orders are \( \varepsilon \), and \( \varepsilon + \mu \) for sell orders. When no news exists, the buy and sell orders arrive at a rate of \( \varepsilon \) 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 independent 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 at no news, bad news, and good news at time \( t \). Accordingly, his/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 at \( t \).

In Appendix A, we show that the probability that any trade occurring at time \( t \) is information-based by

\[
i(t) = \frac{\mu(1 - P_n(t))}{2\varepsilon + \mu(1 - P_n(t))}
\]

\(^5\)This assumption may seem inappropriate given that it rules out any strategic behavior. As will be shown in Section 3.2, informed traders have some tendency for strategic trading. Therefore, we concur that the assumption of risk-neutrality needs more defending, but for the sake of the model applicability, it will not be dropped.
2.1. The Likelihood

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

$$L_b((B, S)|\theta) = L_b(B|\theta)L_b(S|\theta) = e^{-(\mu + 2\varepsilon)T} \frac{\varepsilon^B(\mu + \varepsilon)^ST^{B+S}}{B!S!}$$  \hspace{1cm} (2)

where $\theta = (\alpha, \delta, \varepsilon, \mu)$. Similarly, on a no-event day, the likelihood of observing any sequence of orders that contains $B$ buys and $S$ sells is

$$L_n((B, S)|\theta) = L_n(B|\theta)L_n(S|\theta) = e^{-2\varepsilon T} \frac{\varepsilon^B(\mu + \varepsilon)^ST^{B+S}}{B!S!}$$  \hspace{1cm} (3)

On a good-event day, this likelihood is

$$L_g((B, S)|\theta) = L_g(B|\theta)L_g(S|\theta) = e^{-(\mu + 2\varepsilon)T} \frac{\varepsilon^S(\mu + \varepsilon)^ST^{B+S}}{B!S!}$$  \hspace{1cm} (4)

The likelihood of observing $B$ buys and $S$ sells on a day of unknown type is the weighted average of equations (2), (3), and (4) using the probabilities of each type of day occurring.

$$L((B, S)|\theta) = (1 - \alpha)L_n((B, S)|\theta) + \alpha\delta L_b((B, S)|\theta) + \alpha(1 - \delta)L_g((B, S)|\theta)$$

$$= (1 - \alpha)e^{-2\varepsilon T} \frac{\varepsilon^B(\mu + \varepsilon)^ST^{B+S}}{B!S!} + \alpha\delta e^{-(\mu + 2\varepsilon)T} \frac{\varepsilon^B(\mu + \varepsilon)^ST^{B+S}}{B!S!}$$

$$+ \alpha(1 - \delta)e^{-(\mu + 2\varepsilon)T} \frac{\varepsilon^S(\mu + \varepsilon)^ST^{B+S}}{B!S!}$$  \hspace{1cm} (5)

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,

$$L(M|\theta) = \prod_{i=1}^I L(\theta|B_i, S_i) = \prod_{i=1}^I \frac{e^{-2\varepsilon T} T^{B_i+S_i}}{B_i!S_i!} \times$$

$$\left[(1 - \alpha)e^{B_i+S_i} + \alpha\delta e^{-\mu T} \varepsilon^{B_i}(\mu + \varepsilon)^S_i + \alpha(1 - \delta)e^{-\mu T} \varepsilon^S_i(\mu + \varepsilon)^B_i\right]$$  \hspace{1cm} (6)
The log likelihood function is

\[ \ell(M|\theta) = \sum_{i=1}^{I} \ell(\theta|B_i, S_i) \]

\[ = \sum_{i=1}^{I} \left[ -2\epsilon T + (B_i + S_i) \ln T \right] \]

\[ + \sum_{i=1}^{I} \ln \left[ (1 - \alpha)\epsilon B_i + S_i + \alpha \delta e^{-\mu T} \epsilon B_i (\mu + \epsilon) S_i + \alpha (1 - \delta) e^{-\mu T} \epsilon S_i (\mu + \epsilon) B_i \right] \]

\[ - \sum_{i=1}^{I} (\ln B_i! + \ln S_i!) \]  (7)

As in Easley et al. (2002), the log likelihood function, after dropping the constant and rearranging\(^6\), is given by

\[ \ell(M|\theta) = \sum_{i=1}^{I} \left[ -2\epsilon + M_i \ln x + (B_i + S_i) \ln(\mu + \epsilon) \right] \]

\[ + \sum_{i=1}^{I} \ln \left[ \alpha (1 - \delta) e^{-\mu x^{S_i-M_i}} + \alpha \delta e^{-\mu x^{B_i-M_i}} + (1 + \alpha) x^{B_i+S_i-M_i} \right] \]  (8)

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

3. Data and Estimation

Our data set is from the OANDA FXTrade internet trading platform and consists of tick-by-tick foreign exchange transaction prices and 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 who mainly trade four major exchange rates. Since the bulk of all transactions (approximately 40 percent\(^7\)) involve only the U.S. Dollar - Euro (USD-EUR) trading, we focus our work on transactions involving only USD-EUR. Particularly, we analyze all USD-EUR buy and sell transactions (market, limit order executed, margin call executed, stop-loss, take-profit). In addition to price and volume, we also know the trader’s identity for each transaction. Table 1 summarizes the institutional

\(^6\)To derive Equation 8, the term \( \ln[x^{M_i}(\mu + \epsilon)^{B_i+S_i}] \) is simultaneously added to the first sum and subtracted from the second sum in Equation 7. This is done to increase computing efficiency and ensure convergence in the presence of large number of buys and sells as is the case with our data set.

\(^7\)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).
Figure 1: 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 selling arrivals at each hour are defined as the number of *unique* traders involved in buy or sell or both transactions in that hour. Sample period: October 5, 2003, 16:00 - May 14, 2004, 15:59 (3480 hours, 145 business days).

characteristics of the OANDA FXTrade trading platform. This platform is an electronic market making system (i.e., a market maker) that executes orders using the bid/ask prices that are realistic and prevalent in the marketplace. The prices are determined either by their private limit order book or by analyzing prices from the Interbank market. OANDA FXTrade policy is to offer the tightest possible bid/ask spreads (e.g., 0.0009% spread on the USD-EUR, regardless of the transaction size) and, like most market makers, earns profits from the spreads. Some of the other market features include continuous interest rate payments on a second by second basis, 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, one may conclude that OANDA FXTrade is designed to attract small, uninformed traders. However, as will be shown later, 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 some discussion. As opposed to the equity market, the FX market is a 24-hour decentralized market. Further, unlike the NYSE, it does not involve the so-called specialist who is responsible for maintaining a fair and orderly market, with an insight into the limit order book. The FX market exhibits a low level of transparency 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. Finally, the trading in the FX market is motivated by speculation, arbitrage and, importantly, currency inventory management. Dealers in the FX market generally eliminate inventory positions quickly (from below five minutes to half an hour). This process is sometimes referred to as the “hot-potato-trading” (Evans and Lyons (2002); Bjønnes and Rime (2005)). On the NYSE, however, the half-life of inventory averages over a week (Madhavan and Smidt (1993)). Thus, inventory management is an important component of an intraday FX trader activity.

Considering the features of OANDA FXTrade, it can be viewed as a “special case” of the FX market that can be approached with the Easley et al. (2002)’s model. First, OANDA FXTrade as a market maker 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 regard to the 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 the ones in equity markets. We will suggest in Section 5 how the intraday inventory management aspect of the FX market might be related to the findings.

Our preliminary analysis indicated that overall market activity was extremely low at certain days or weeks. Therefore, we eliminated all weekends, starting from every Friday 15:59:59 to Sunday 15:59:59, inclusive (all times are EST), the Christmas week (December 22-26), the first week of the year (December 29-January 2), and the Independence Day week (April 5-9) which 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. The final sample size is 3,480 hourly data, covering 145 business days from October 5, 2003, 16:00 to May 14, 2004, 16:00 EST. In the sample period, there are 667,030 sell and 666,133 buy transactions, with an average of approximately 6 transactions (3 buy and 3 sell) per minute. The volume of these transactions is a total of 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
Table 1: OANDA FXTrade institutional characteristics.

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

of the largest tick-by-tick FX data sets ever available for an academic study.\(^8\)

Since a trader’s identity for each transaction is known, we are able to identify the number of unique traders at each one-hour window. For estimation purposes, the number of buy arrivals at each hour \(B_t’\) is defined as the number of unique traders involved in buy transactions in that hour. The number of sell arrivals at each hour \(S_t’\) is defined similarly. Therefore, if an individual trader arrives in a given hour and conducts several buy (sell) transactions, this is counted as one buy (sell) arrival in that hour.

Figure 1 illustrates the number of hourly buy and sell arrivals \((B_t’, S_t’)\) and the sample autocorrelation functions. The strong daily seasonality and a time trend in both series are evident. Therefore, we first estimate the linear time trend, \(\hat{B}_t\) and \(\hat{S}_t\), from the trend regression which is free of seasonal and irregular fluctuations. Assuming a multiplicative separable seasonality, we divide the original series by the trend estimates \(\hat{B}_t\) and \(\hat{S}_t\) to obtain an estimate of the seasonal 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 seasonal index for each hour of the day, we averaged the values of \(\tilde{s}_t^B\) and

---

\(^8\)The electronic market for USD-EUR is largely dominated by EBS, followed by Reuters. However, unlike OANDA FXTrade, EBS does not provide high-frequency volume data nor the identity of the traders. Reuters does not report this information either. This makes OANDA FXTrade more suitable for our research goals.
\( \tilde{s}_i^S \) corresponding to the same hour of the day across the sample and obtained the final hour-of-day indices \( s_i^B \) and \( s_i^S \), \( i = 1, 2, \ldots, 24 \) for the 24-hour cycle. The diurnally\(^9\) adjusted number of buy and sell arrivals are obtained via

\[
B_i = \frac{B_i'}{s_i^B}, \quad S_i = \frac{S_i'}{s_i^S}, \quad i = 1, 2, \ldots, 24
\]

for each 145 days in the sample.

Figure 2 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 midnight 00:00 EST. The average number of unique buy and sell traders starts to increase after midnight and becomes above the hourly average before the LSE opens (03:00 EST). Similarly, a sharp increase is observed hours before the NYSE opens (9:30 EST). The number of traders start to decline after 10:00 and go below their hourly averages after 14:00. They remain relatively low and stable during the following hours until midnight. In the lower panel of Figure 2, the sample autocorrelation functions of the diurnally adjusted number of buy and sell arrivals are studied at hourly lags. The strong daily seasonality is removed and a strong persistence in both series is revealed.

### 3.1. Informed and uninformed trades: When do they arrive?

According to the Easley et al. (1996b) model, 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

\[
E(K) = \alpha \mu (2 \delta - 1)
\]

which provides information on the arrival of informed trades. When \( \mu \) 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 \( |K| \) contains information on the arrival of uninformed trades, \( \varepsilon \). Note that our measure of “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 \( \alpha \) is constant, the hour of the day average of the absolute trader imbalance \( |K| \) provides information on the intraday seasonality of

\(^9\)For 24-hour cycles, i.e., daily.
the orders from informed traders. In other words, it is possible to obtain a measure of the activity time of the informed traders as we know the number of unique individuals and corresponding trades at each hour of the day. Similarly, we can identify whether uninformed traders (liquidity traders) follow a distinct intraday pattern.

Figure 3 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
Figure 3: 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 deseasonalized data while the dashed line is for the raw data. For the deseasonalized data, the hour-of-day index of uninformed traders is relatively stable, therefore, indicating a non-strategic, uniform arrival. The same index of informed traders fluctuates according to the observed geographic seasonality pattern. For the raw data, the hour-of-day indices for both informed and uninformed traders exhibit similar patterns.

calculated from $B_t$ and $S_t$ so that we do not expect any hour-of-day effect a priori.\footnote{We also plot the hour-of-day indices for the raw data that use unadjusted $B_t$ and $S_t$ (dashed line). It can be observed that the arrival patterns are similar for both types of traders. Thus, we reveal “hidden” hour-of-day patterns that are present even after $B_t$ and $S_t$ are seasonally adjusted. Considering that these effects are strong enough to persist even after adjusting for intraday seasonalties, we will concentrate on the seasonally adjusted data for the} However,
the hourly activity of uninformed (liquidity) traders increases before the LSE, the NYSE, or the TSE open (03:00, 9:30 and 19:00). The activity is above the hourly average from 01:00 until 14:00. Notice that the variation of the hour-of-day index of uninformed traders is relatively small: it fluctuates between 0.95 and 1.04. Therefore, we may speculate that uninformed traders arrive any time of the day uniformly. However, the hour-of-day index of informed traders reveals a different picture. It is almost opposite to the hour-of-day index of uninformed traders, and the volatility of informed traders during the day is high: the index fluctuates 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 then picks up at around 03:00 when the LSE opens, reaches a high and dips down before the NYSE opens (9:30) when it is 14:30 in London. When the NYSE opens, the number of informed traders increases sharply until the market closes (16:00). This is followed by a decline until the TSE opens (19:00). The above-average Tokyo activity lasts until the market closes (01:00). The number of unique informed traders is well above the average after 14:59, which is one hour before the NYSE closes. This number remains above the average (except 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 lowest number of traders present in the market. The hour-of-day indices do not show any above-average arrival of informed traders during the LSE trading hours.

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

As we 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 \( \bar{TT} \) is 176.9 while \( \bar{K} = -1.4 \) and \( |\bar{K}| = 19.4 \). 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 (\( \bar{\delta} = (1 - 0.07)/2 = 0.47 \)). Buy and sell uninformed traders arrive at the market with an hourly arrival rate of \( \varepsilon \). The estimated sample statistics imply that this hourly rate is 79.7.

\[
E(TT) = \alpha \mu + 2\varepsilon = 176.9 = 19.4 + 2\varepsilon
\]

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:

\[
PIN = \frac{\alpha \mu}{\alpha \mu + 2\varepsilon} = 19.4/176.9 = 0.11
\]

remainder of the paper.

13
Table 2: Day-of-week indices of estimated parameters and the PIN.

<table>
<thead>
<tr>
<th>Index</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SI_\alpha$</td>
<td>1.14</td>
<td>0.89</td>
<td>0.91</td>
<td>1.1</td>
<td>0.96</td>
</tr>
<tr>
<td>$SI_\delta$</td>
<td>0.76</td>
<td>1.15</td>
<td>0.95</td>
<td>1.02</td>
<td>1.13</td>
</tr>
<tr>
<td>$SI_\varepsilon$</td>
<td>0.84</td>
<td>1.08</td>
<td>1.09</td>
<td>1.05</td>
<td>0.94</td>
</tr>
<tr>
<td>$SI_\mu$</td>
<td>0.85</td>
<td>1.09</td>
<td>1.14</td>
<td>0.94</td>
<td>0.98</td>
</tr>
<tr>
<td>$SI_{PIN}$</td>
<td>1.16</td>
<td>0.92</td>
<td>0.91</td>
<td>1.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

for the market maker’s initial belief. The estimated PIN in equity markets is usually between 0.15 and 0.25. The relatively low PIN in this case may indicate that the number of informed traders is low (small $\mu$) or the probability of the information event is low (small $\alpha$) or both. In any case, the risk of informed trade is relatively low for the market maker in this particular case.

In our estimations, we used above priors as our initial parameter set. That is, we set $\varepsilon = 78$, $\delta = 0.47$. Since we do not have a prior for $\alpha$ and $\mu$ separately, we assume $\alpha = 0.50$ so that $\mu = 38$. The log likelihood function in Equation 8 is maximized every day ($T = 24$) for the entire sample period (145 days). As a result, we have 145 different estimates of each parameter.\(^1\) The two probability parameters $\alpha$ and $\delta$ are restricted to $(0,1)$ and the two arrival rates to $(0,500)$ since the maximum observed number of 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.\(^2\) The probability of an event $\alpha$ is higher on Mondays and Thursdays. Given that an event occurs, the probability that it is a bad event $\delta$ is lower than the average on Mondays and Wednesdays. Therefore, we may speculate that Mondays were eventful days with good news for USD-EUR during the sample period. This may reflect the fact that there is a significant time period (48 hours) to “digest” the previous week’s information before Monday traders arrive.

Figure 4 plots the daily estimates of the probability of an event $\alpha$ (top left) and the probability that an event is bad news $\delta$ (top right).\(^3\) The estimated probability of an event $\hat{\alpha}$ fluctuates between 0.04 and 0.56 with an average of 0.30. This implies that there were no days without an event in an hour. The lowest estimate 0.04 shows that there was a day with only one hour with an event ($0.04 \times 24 \approx 1$ day). Similarly, the highest estimate 0.56 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

\(^{11}\)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 4, smaller log likelihood values indicate better fit.

\(^{12}\)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.

\(^{13}\)The estimates over 145 days in the sample are stable with regard to the reasonable choice of their starting values. The only case when the estimates start substantially changing is when $\mu_0 > 200$ and $\varepsilon_0 > 200$. 

\(^{14}\)
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 $\hat{\delta}$ is normally distributed with the $p$-value=0.418. This is an expected result as there was no significant trend in USD-EUR prices during the sample period.

Figure 4 also plots the daily estimates of the arrival rates. The estimated arrival rate of
uninformed traders \( \hat{\epsilon} \) exhibits a sharp increase at around the 60th day (in January 2004) from an average of approximately 50 to 80. The overall mean of this parameter is 77.8 (same as our 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. From Table 2, we see that both informed and uninformed traders arrive less than average on Mondays and Fridays. The highest arrival rates of both informed and uninformed traders are observed on Wednesdays. Noteworthy, the market maker attaches a non-normal component to the arrival of the informed traders which is not in line with the model assumption of the informed traders’ risk-neutrality. Another instance where the empirics diverges from the market assumptions (dominance of uninformed traders) is the fact that informed and uninformed traders have similar arrival rates. This can be interpreted in a way that although the institutional market characteristics encourage uninformed traders to trade, informed traders arrive equally likely.

Finally, the average estimated PIN is about 0.12 and is, thus, lower than 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 relatively high PIN, this is not reflected 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 fact that the distribution of \( \hat{\mu} \) is non-normal confirms the evidence of strategic activity by informed traders.

3.3. Independence of arrivals

The crucial underlying assumption of the model by Easley et al. (1996b) is the independence of information events across hours for each 24-hour sequence.\(^{14}\) Thus, while deriving the log likelihood function, we assume that the arrival of traders each hour, conditional on information events, is drawn from identical and independent distributions. Nevertheless, considering evidence of 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.

\(^{14}\)Easley et al. (1997b) test for independence assumption and find that information events are independent for their data set.
The estimated $\hat{\alpha}_i$'s ($i = 1, \ldots, 145$) help us to classify hours into ones when an event occurs and the ones when it does not. As denoted 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 and classify each event hour by one and nonevent hour by zero. The total number of event hours is denoted by $e_i$ and nonevent by $n_i$. The results indicate that the null hypothesis cannot be rejected at the 5% significance for 72 days whereas it is rejected for 73 days. This mixed evidence necessitates additional testing and we turn to the Ljung-Box portmanteau test (Ljung and Box (1978)) for white noise next.

We compute the Ljung-Box test statistic with up to the 10th 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 $\chi^2_{10}$. The null rejection frequencies at the 5% significance 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 assumption of independence is not plagued by this evidence as there are about 85% of the days where our model is in agreement with the assumption. Easley et al. (1997b) argue that the evidence of dependence does not affect the actual parameter estimates, but affects 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 would not have any major impact on our findings.

4. Informativeness of the trade size

In this section, we utilize a procedure similar to Easley et al. (1997b) that is theoretically outlined in Easley and O'Hara (1987). This approach involves allowing for both informed and uninformed traders to place large and small orders. The extended model now 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. This introduces two new parameters to the model: $\phi$ (probability that an uninformed trader trades a large amount) and $\omega$ (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 ($\alpha$, $\mu$, $\delta$ and $\varepsilon$) follow the notation from Section 2.

---

15 Under the null hypothesis of randomness of information events across hours, the total number of runs $r$ (sequences of ones or zeros) is normally distributed with $r = \frac{2e_i n_i}{e_i + n_i} + 1$ and $\sigma^2_r = \frac{(r-1)(r-2)}{(e_i + n_i)^2}$.  
16 For the null hypothesis of independence (or randomness) of information events over $I=24$ hours, this test is based on the following statistic: $Q_L = I(I+2) \sum_{\tau=1}^{L} \hat{\rho}^2_{\tau}$, where $L$ is typically chosen to be substantially smaller than $I$ and $\hat{\rho}^2_{\tau}$ is the sample autocorrelation coefficient at lag $\tau$.  
17 For simplicity, we ignore a no-trade outcome that is employed in Easley et al. (1997b) for a much smaller data set of stock prices. Also, using the procedure from Section 3, we deseasonalize each of the four new series to remove the observed daily seasonality.
Our goal is twofold: first to compare the estimated $\omega$ and $\phi$ 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 estimates. The testing procedure for trade size effects also 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, but we later show that this does not affect our main findings.\(^{18}\)

The derivation of the log likelihood function proceeds similarly to that of the restricted model and is presented in Appendix B. Table 3 lists the average estimates of $\alpha_i, \delta_i, \varepsilon_i, \mu_i, \omega_i, \text{ and } \phi_i \ (i = 1, \ldots, 145)$. 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.\(^ {19}\) However, the difference between the two sets of estimates of $\omega_i$ is statistically significant.\(^ {20}\) 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. It concludes that $\bar{\omega}$ is significantly greater than $\bar{\phi}$. There are about 68% of the days in the sample when $\omega_i > \phi_i$ 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 on average this is not the case.

It is essential to investigate whether the above findings 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 estimates.

We find that in the cutoff amount 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 the cutoff values above 8000 (e.g., 12000 in Table 4) distorts the results due to low frequency of such large trades. Similarly, it is unreasonable to consider trades above small cutoff values (e.g., 2000 in Table 4) “large” and the observed effects diminish. Such findings can also be found in 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)).\(^ {21}\) An interesting observation emerges from Table 4: the probability of both informed and uninformed large trading

\(^{18}\)Trade size is expressed in currency units of the base currency, i.e., the Euro.

\(^{19}\)The null hypothesis for this test is that the mean difference ($d$) 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 $\bar{d}$.

\(^{20}\)Also, the standard errors of $\hat{\omega}_i$ and $\hat{\phi}_i$ for the unconstrained model are consistently of order $10^{-4}$ and $10^{-5}$, respectively, thus indicating statistically significant difference in the probabilities.

\(^{21}\)Suppose the constrained model is found more appropriate. This would indicate a “pooling equilibrium” where informed traders submit both large and small orders roughly equal.
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 the third column represent the average estimates of the parameters in both the constrained \((\omega_i = \phi_i)\) and unconstrained \((\omega_i \neq \phi_i; i = 1, \ldots, 145)\) versions of the model. The last column contains differences in the mean value between the 145 parameters estimated from the constrained and unconstrained models. The \(p\)-value is received 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. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

5. Price dynamics

5.1. Model estimates, returns and volatility

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

\[
r_t = \alpha + \gamma M_t + v_t, \quad v_t \sim IID(0, \sigma^2)
\]

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

where \(r_t = \ln(P_t) - \ln(P_{t-1})\), \(M_t \in \{\alpha_t, \delta_t, \epsilon_t, \mu_t, \omega_t, \phi_t\}\) and \(t = 1, \ldots, 145\). By \(P_t\) we denote the daily close (USD/EUR) on day \(t\). In Table 5, we report the results for the \(M_t\) variables that were found significant in the first regression.\(^{22}\) Since the estimate for \(\delta_t\) is statistically significant with relatively high adjusted \(R^2\) value (0.2675), our results essentially confirm the ones from Easley et al. (1997b) that also find the probability of bad news informative for exchange rate determination. However, their estimated coefficient on the bad event probability variable is much larger than our

\(^{22}\)Although the coefficient on \(\omega_t\) is not significant for the whole sample, it is found 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.
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 that is reported in the brackets is received from the paired $t$-test for the null hypothesis of the difference being equal to zero. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2000 cutoff</th>
<th>8000 cutoff</th>
<th>12000 cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.31 (0.43)</td>
<td>0.30 (0.14)</td>
<td>0.30 (0.33)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.48 (0.53)</td>
<td>0.47 (0.66)</td>
<td>0.47 (0.38)</td>
</tr>
<tr>
<td>$\xi$</td>
<td>7.44 (0.59)</td>
<td>7.80 (0.15)</td>
<td>77.93 (0.19)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>74.54 (0.40)</td>
<td>83.11 (0.12)</td>
<td>82.65 (0.19)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.73 (0.00)**</td>
<td>0.36 -0.01 (0.00)**</td>
<td>0.24 -0.00 (0.02)**</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.70 (0.00)**</td>
<td>0.34 0.01 (0.00)**</td>
<td>0.23 0.00 (0.00)**</td>
</tr>
<tr>
<td>LLF</td>
<td>0.12 (0.15)</td>
<td>1.66 (0.09)*</td>
<td>0.97 (0.17)</td>
</tr>
</tbody>
</table>

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 ($\delta_t$). *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

<table>
<thead>
<tr>
<th>$r_t$</th>
<th>Coefficient</th>
<th>$M_t = \delta_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.007***</td>
<td>-0.015***</td>
</tr>
<tr>
<td>Standard error</td>
<td>0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td>$t$-statistic</td>
<td>6.26</td>
<td>-7.29</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.2675</td>
<td></td>
</tr>
</tbody>
</table>

$\hat{\gamma}$. More specifically, we find that a one-percent increase in $\delta_t$ 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 to the arrival of bad news than the FX market.

We next discuss the results of regressing 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 $\gamma$ for $\xi_t$ is about two times larger than the one for $\mu_t$. Considering the fact that the PIN is insignificant in the second regression, we can conclude that, although the 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
Table 6: Regression results - volatility.
This table reports the results of the OLS regression: \( r_t^2 = \alpha + \beta r_{t-1}^2 + \gamma M_t + \nu_t, \ t = 1, \ldots, 145 \). 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 \( \mu_t \), and the arrival rate of uninformed traders \( \varepsilon_t \). *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

<table>
<thead>
<tr>
<th>( r_t^2 )</th>
<th>Constant</th>
<th>( r_{t-1}^2 )</th>
<th>( M_t = \mu_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>0.000***</td>
<td>-0.205***</td>
<td>3.25e-07*</td>
</tr>
<tr>
<td>Standard error</td>
<td>0.000</td>
<td>0.081</td>
<td>1.72e-07</td>
</tr>
<tr>
<td>t-statistic</td>
<td>2.70</td>
<td>-2.51</td>
<td>1.89</td>
</tr>
<tr>
<td>p-value</td>
<td>0.008</td>
<td>0.013</td>
<td>0.061</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.0509</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( r_t^2 )</th>
<th>Constant</th>
<th>( r_{t-1}^2 )</th>
<th>( M_t = \varepsilon_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>0.000</td>
<td>-0.218***</td>
<td>7.03e-07***</td>
</tr>
<tr>
<td>Standard error</td>
<td>0.000</td>
<td>0.081</td>
<td>3.04e-07</td>
</tr>
<tr>
<td>t-statistic</td>
<td>0.73</td>
<td>-2.67</td>
<td>2.31</td>
</tr>
<tr>
<td>p-value</td>
<td>0.466</td>
<td>0.008</td>
<td>0.022</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.0626</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 the critical value \( \chi^2_\alpha = 5.991 \) for the 5% significance level. The null hypothesis is stated in the first column. \( \mu_t \) and \( \varepsilon_t \) \( (t = 1, \ldots, 145) \) 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.

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>( \chi^2 )</th>
<th>Prob &gt; ( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_t ) does not cause ( \varepsilon_t )</td>
<td>2.51</td>
<td>0.28</td>
</tr>
<tr>
<td>( \varepsilon_t ) does not cause ( \mu_t )</td>
<td>7.77</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Granger causality tests between the trader arrival rates. Essentially, the Granger causality test 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 \( \varepsilon_t \) and \( \mu_t \) is estimated using a standard bivariate framework. In Table 7, we report 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 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. Together with other findings from previous sections this suggests that uninformed traders arrive to the market first while informed traders follow by strategically timing their arrival. Therefore, the assumption of risk-neutrality

21
Table 8: Trade Imbalance Regression Results.
This table reports the results of the OLS regression: \( r_t = \alpha + \beta K_t + v_t, \quad t = 1, \ldots, 3479 \). 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. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

<table>
<thead>
<tr>
<th>( r_t )</th>
<th>Constant</th>
<th>( K_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>4.12e-05*</td>
<td>-2.37e-05***</td>
</tr>
<tr>
<td>Standard error</td>
<td>2.39e-05</td>
<td>9.90e-07</td>
</tr>
<tr>
<td>t-statistic</td>
<td>1.72</td>
<td>-23.98</td>
</tr>
<tr>
<td>p-value</td>
<td>0.085</td>
<td>0.000</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.1417</td>
<td></td>
</tr>
</tbody>
</table>

of informed traders is not appropriate in the FX market context as informed traders are clearly risk-averse.\(^{23}\)

5.2. Trading, returns and volatility

Section 3 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 the “trade imbalances” and can be interpreted as a variant of the market order flow (Evans and Lyons (2002)). We first test whether this variable can explain FX returns on an hourly basis by running the following regression:

\[
r_t = \alpha + \beta K_t + v_t, \quad v_t \sim IID(0, \sigma^2), \quad t = 1, \ldots, 3479
\]

where \( r_t \) is as defined before and \( K_t = S_t - B_t \) for each hour. Table 8 lists the estimates that essentially show that trade imbalances significantly determine hourly returns. Essentially, an increase in the trade imbalance by one unit (i.e., there exists one more 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 the evidence by 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 that is typically used in the FX microstructure literature (see e.g., Lyons (2001)), it would be interesting to understand the relationship between the two measures more clearly. For that purpose we construct hourly and daily market order flows based on the

\(^{23}\)One may expect that the arrival of informed traders is related only to the information flow, but, in this setting, they appear to strategically use their private information. Another possibility is that informed traders enter the market to not only establish speculative positions (information effects), but also to adjust their currency inventory (inventory effects), as mentioned previously.
Table 9: Order flow regression results.

This table reports the results of the OLS regression: \( r_t = \alpha + \beta X_t + \epsilon_t \). 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. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Hourly</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_t )</td>
<td>\begin{align*} &amp; 7.00 \times 10^{-06} - 5.67 \times 10^{-11}*** \end{align*}</td>
<td>\begin{align*} &amp; 0.0001396 - 1.33 \times 10^{-10}*** \end{align*}</td>
</tr>
<tr>
<td>Coefficient</td>
<td>0.0000255 &amp; 6.52 \times 10^{-12}</td>
<td>0.0005237 &amp; 3.20 \times 10^{-11}</td>
</tr>
<tr>
<td>Standard error</td>
<td>0.0000476 &amp; 0.0000578</td>
<td>0.00005237 &amp; 3.20 \times 10^{-11}</td>
</tr>
<tr>
<td>( t )-statistic</td>
<td>0.27 &amp; -8.70</td>
<td>0.27 &amp; -4.17</td>
</tr>
<tr>
<td>p-value</td>
<td>0.784 &amp; 0.000</td>
<td>0.790 &amp; 0.000</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.0210</td>
<td>0.0866</td>
</tr>
</tbody>
</table>

As expected, the 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 informational content of order flow is obscured by high-frequency noise and through aggregation this information becomes more pronounced. 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.

Considering that \( |K| \) and \( TT - |K| \) contain information on the hourly arrival of informed and uninformed traders, next, we examine their explanatory power with respect to hourly returns by using a regression similar to Equation (9). 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). That gives us 24 regressions with 144 observations each. In the top two panels, Figure 5 shows 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 with a cubic spline function.\(^{24}\) This transformation is convenient because it provides smooth transition in the trader behavior over the 24-hour cycle.

Clearly, the price impact is time-varying and, when significant, it is more pronounced for informed traders. Strong geographic seasonality patterns emerge again. The first region of activity commences at 03:00 when the LSE opens. This region ends after the NYSE opens (09:30). Sig-

\(^{24}\) 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)).
significant price impact is also observed before the NYSE closes and during most of the hours of operation of the TSE. These findings are in line with the ones for the hour-of-day index from Sec-
tion 3. Therefore, in general, when informed traders arrive above-average, they have 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 (10) 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 5 show the estimated \( \hat{\gamma}_i \) \((i = 0, \ldots, 23)\) for both types of traders.

It can be observed that informed traders dominate uninformed traders with regard to their influence on hourly FX volatility. The strongest geographic seasonality effects take place when both the NYSE and the LSE are open, i.e., roughly between 09:30 and 11:30. In general, during these particular hours informed traders appear to be capable of driving FX volatility and to a certain degree FX return, despite arriving below-average. Hence, this exercise extracts local information that was not obvious from the theoretical microstructure model.\(^{25}\)

It is also of interest to explore different patterns of FX volatility response to daily and hourly arrival rates of informed/uninformed traders. Recall that the impact of \( \varepsilon \) on volatility is more than twice as much as the impact of \( \mu \). First, we assess which percentile of the daily volatility that informed and uninformed traders are operating. This is carried out by comparing the absolute daily changes in volatility to the regression coefficients \( \gamma \) on \( \mu \) and \( \varepsilon \) in Equation (10) that are listed in Table 6. Based on the estimated volatility percentiles, we find that informed traders are operating roughly the 1st percentile while uninformed traders are operating between the 1st and the 5th percentile. We follow the same procedure for the hourly data and find that for informed traders the range of \( \gamma \) from Figure 5 (bottom panel) falls into the 25th percentile. Further, the range of \( \gamma \) for uninformed traders is narrower and it is found that they are operating roughly the 10th percentile of the 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.

6. Conclusions

Using a high-frequency version of the model by Easley et al. (1996b) for the FX market we address a number of important issues and provide new empirical findings. First, we estimate parameters that reflect 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, but also 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) seasonality to the day-of-week effects which

\(^{25}\)The Granger causality tests for the hourly arrivals yield similar findings to the ones for daily arrival rates: \( TT_t - |K_t| \) Granger-causes \(|K_t|\), but not vice-versa.
is substantiated by the Granger causality tests.

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 market 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 the trade composition (i.e., the PIN) has no effect on volatility.

One important advantage of our approach is the direct observability of the nature of currency orders and that attests to the accuracy of our results. As opposed to equity markets where data on trade classification are often unavailable, we are not forced to apply any trade classification algorithms 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 extention 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 that define a “large” trade, though some trade sizes that we find informative can also be interpreted as medium-sized. Therefore, the current paper adds to the literature that provides 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 independency of information events across hours. We, however, observe dependence for about 15% of the days in our sample and that may, to a certain extent, bias the standard errors of our estimates. Even though we conjecture that this bias would be minimal, it is noteworthy to emphasize that introducing dependency to the model presents a key future research direction. The impact of past transactions on the current transactions could be modeled by a latent parameter that would measure 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 overdispersion frequently found in transactions data that reduces the usefulness of the Poisson distribution. We hope that the evidence presented here will lead to a new structural microstructure model that would be applicable to both equity and FX markets.
Appendix A. Derivation of the PIN

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

$$P_n(t|S_t) = \frac{P_n(S_t|t)P_n(t)}{P(S_t)}$$

$$= \frac{P_n(S_t|t)P_n(t)}{P_n(S_t|t)P_n(t) + P_g(S_t|t)P_g(t) + P_b(S_t|t)P_b(t)}$$

$$= \frac{\varepsilon P_n(t)}{\varepsilon(1 - P_g(t) - P_b(t)) + \varepsilon P_g(t) + (\varepsilon + \mu)P_b(t)}$$

$$= \frac{\varepsilon P_n(t)}{\varepsilon + \mu P_b(t)}$$

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

$$P_b(t|S_t) = \frac{P_b(S_t|t)P_b(t)}{P(S_t)}$$

$$= \frac{(\varepsilon + \mu)P_b(t)}{\varepsilon + \mu P_b(t)}$$

and the posterior probability on good news is

$$P_g(t|S_t) = \frac{P_g(S_t|t)P_g(t)}{P(S_t)}$$

$$= \frac{\varepsilon P_g(t)}{\varepsilon + \mu P_b(t)}$$

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

$$b(t) = P_n(t|S_t)s^a_i + P_b(t|S_t)s^b_i + P_g(t|S_t)s^g_i = \frac{\varepsilon P_n(t)s^a_i + (\varepsilon + \mu)P_b(t)s^b_i + \varepsilon P_g(t)s^g_i}{\varepsilon + \mu P_b(t)}$$

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

$$a(t) = P_n(t|B_t)s^a_i + P_b(t|B_t)s^b_i + P_g(t|B_t)s^g_i = \frac{\varepsilon P_n(t)s^a_i + \varepsilon P_b(t)s^b_i + (\varepsilon + \mu)P_g(t)s^g_i}{\varepsilon + \mu P_g(t)}$$

27
The expected price conditional on $t$ is

$$E\left[ s_i | t \right] = P_n(t)s_i^n + P_b(t)s_i^b + P_g(t)s_i^g$$

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 conditional on time $t$.

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

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

$$= \frac{\varepsilon E\left[ s_i | t \right] + \mu P_b(t)s_i^b}{\varepsilon + \mu P_b(t)}$$

$$= \left[ 1 - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)} \right] E\left[ s_i | t \right] + \frac{\mu P_b(t)s_i^b}{\varepsilon + \mu P_b(t)}$$

$$= E\left[ s_i | t \right] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(E\left[ s_i | t \right] - s_i^b)$$

and

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

$$= \frac{\varepsilon E\left[ s_i | t \right] + \mu P_g(t)s_i^g}{\varepsilon + \mu P_g(t)}$$

$$= \left[ 1 - \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)} \right] E\left[ s_i | t \right] + \frac{\mu P_g(t)s_i^g}{\varepsilon + \mu P_g(t)}$$

$$= E\left[ s_i | t \right] + \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(s_i^g - E\left[ s_i | t \right])$$

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

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

$$= \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(s_i^g - E\left[ s_i | t \right]) + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(E\left[ s_i | t \right] - s_i^b)$$

The spread for the opening quotes is

$$d(0) = \frac{\mu P_g(0)}{\varepsilon + \mu P_g(0)}(s_i^g - E\left[ V_i \right]) + \frac{\mu P_b(0)}{\varepsilon + \mu P_b(0)}(E\left[ s_i - s_i^b \right])$$

$$= \frac{\mu \alpha (1 - \delta)}{\varepsilon + \mu \alpha (1 - \delta)}(s_i^g - E\left[ s_i | t \right]) + \frac{\mu \alpha \delta}{\varepsilon + \mu \alpha \delta}(E\left[ s_i | t \right] - s_i^b)$$

28
If good and bad events are equally likely, that is, if \( \delta = 1 - \delta, \delta = 0.5 \). Thus

\[
d(0) = \frac{\mu \alpha}{2 \varepsilon + \mu \alpha} (s_i^g - s_i^b)
\]

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

\[
i(t) = \frac{P_b(t) \mu + P_g(t) \mu}{P(B_t, S_t)} = \frac{\mu(1 - P_n(t))}{P_n(t) P_n(B_t, S_t|t) + P_g(t) P_g(B_t, S_t|t) + P_b(t) P_b(B_t, S_t|t)}
\]

\[
= \frac{\mu(1 - P_n(t))}{2 \varepsilon + \mu(1 - P_n(t))}
\]

**Appendix B. 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 on a bad news hour is

\[
L_b((LB, LS, SB, SS)|\theta) = L_b(LB|\theta) L_b(LS|\theta) L_b(SB|\theta) L_b(SS|\theta)
\]

\[
= e^{-(\mu + 2 \varepsilon) T} \varepsilon^{LB} (1 - \phi)^{SB} \varepsilon^{LS} (1 - \phi + \mu(1 - \omega))^{SS} T_{LB+LS+SB+SS}^{LB!LS!SB!SS!}
\]

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

\[
L_n((LB, LS, SB, SS)|\theta) = L_n(LB|\theta) L_n(LS|\theta) L_n(SB|\theta) L_n(SS|\theta)
\]

\[
= e^{-2 \varepsilon T} \varepsilon^{LB+LS} (1 - \phi)^{SB+SS} (\varepsilon T)^{LB+LS+SB+SS} T_{LB!LS!SB!SS!}
\]

On a good-event day, the likelihood is

\[
L_g((LB, LS, SB, SS)|\theta) = L_g(LB|\theta) L_g(LS|\theta) L_g(SB|\theta) L_g(SS|\theta)
\]

\[
= e^{-(\mu + 2 \varepsilon) T} \varepsilon^{LS} (1 - \phi)^{SS} (\varepsilon T)^{LB} (1 - \phi + \mu(1 - \omega))^{SB} T_{LB+LS+SB+SS}^{LB!LS!SB!SS!}
\]

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

\[
\ell(D|\theta) = \sum_{i=1}^{I} \ell(\theta|LB_i, LS_i, SB_i, SS_i)
\]

\[
= \sum_{i=1}^{I} [-2\varepsilon + M_i \ln x + N_i \ln y]
+ \sum_{i=1}^{I} [(LB_i + LS_i) \ln(\varepsilon \phi + \mu \omega) + (SB_i + SS_i) \ln(\varepsilon (1 - \phi) + \mu (1 - \omega))]
+ \sum_{i=1}^{I} \ln[(1 - \alpha) x^{LB_i+LS_i-M_i} y^{SB_i+SS_i-N_i} + \alpha \delta e^{-\mu x^{LB_i-M_i} y^{SB_i-N_i}}
+ \alpha (1 - \delta) e^{-\mu x^{LS_i-M_i} y^{SS_i-N_i}}]
\]

where \( M_i \equiv \min(LB_i, LS_i) + \max(LB_i, LS_i)/2, \) \( N_i \equiv \min(SB_i, SS_i) + \max(SB_i, SS_i)/2, \) \( y = \frac{\varepsilon (1 - \phi)}{\varepsilon (1 - \phi) + \mu (1 - \omega)} \in [0, 1], \) \( x = \frac{\varepsilon \phi}{\varepsilon \phi + \mu \omega} \in [0, 1]. \) 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


