Empirical Market Microstructure: An Analysis Of The Brl/Us$ Exchange Rate Market Using High-Frequency Data

Márcio Poletti Laurini
Luiz Gustavo Cassilatti Furlani
Marcelo Savino Portugal
Copyright Insper. Todos os direitos reservados.

É proibida a reprodução parcial ou integral do conteúdo deste documento por qualquer meio de distribuição, digital ou impresso, sem a expressa autorização do Insper ou de seu autor.

A reprodução para fins didáticos é permitida observando-se a citação completa do documento.
EMPIRICAL MARKET MICROSTRUCTURE: AN ANALYSIS OF THE BRL/US$ EXCHANGE RATE MARKET USING HIGH-FREQUENCY DATA

MÁRCIO POLETTI LAURINI
IBMEC SÃO PAULO AND IMECC-UNICAMP
LUIZ GUSTAVO CASSILATTI FURLANI
PPGE-UFRGS AND SICREDI
MARCELO SAVINO PORTUGAL
PPGE-UFRGS

ABSTRACT. This article provides an analysis of empirical microstructure for the BRL/US$ exchange rate market using high-frequency bid and ask quote data. The aims of the article are to verify the importance of the presence of asymmetric information in price dynamics, to build a model for the price discovery process and to analyze the empirical determinants of the spread between bid and ask through a conditional model that captures an asymmetric response to the spread regarding past information.

The asymmetric information hypothesis is tested through a nonparametric test of conditional independence for the Markov property. A model for price discovery is built using a vector error correction between bid and ask, controlling for duration and volatility. As a result of this vector, we build an equilibrium spread deviation series, and we show that the conditional distribution of equilibrium spread deviations responds asymmetrically to the spread changes and expected conditional volatilities and durations. This is made by using the quantilogram and a quantile autoregression as tools for modeling the asymmetry effects. We relate the findings to some facts presented in the theoretical literature on market microstructure.

keywords: market microstructure, emerging market, spread, Markov property, asymmetric response, quantile regression
JEL Codes - G14, C22, C14.

1. INTRODUCTION

The analysis of empirical microstructure effects on exchange rate markets has gained great momentum in recent years. It is well recognized that in short-run asset pricing may be more closely related to market structures than the factors related to asset fundamentals, as pointed out by Flood & Taylor (1996). The literature on market microstructure indicates that factors such as transaction costs, stock balance and liquidity premia may play a more crucial role in prices in the short run than factors associated with macroeconomic fundamentals.

The literature on exchange rates encourages the analysis of market microstructure effects by providing evidence that the conventional macroeconomic approach to exchange rate determination can only explain long-run movements and extreme situations, such as in hyperinflation events and exchange rate crises. In normal exchange rate market situations, exchange rate movements are defined by the market microstructure (e.g. Flood & Taylor (1996), Taylor (1995) and Frankel & Rose (1995)).

Another factor that allows assessing exchange rate market microstructure effects is the large availability of information about intraday exchange rate operations, provided by proprietary trading systems, such as Reuters 2000-2 Dealing System, the Electronic Broking System (EBS), Spot Dealing System and, in Brazil, the Sisbex system. Information is collected systematically and made

1Corresponding author Márcio Poletti Laurini - Ibmec São Paulo and IMECC-Unicamp. Rua Quatá 300, CEP 01546-042. São Paulo, SP - Brazil. Phone 55 11 4504 24 26. Fax 55 11 4504 2350. email - marcipl@isp.edu.br. Luiz Gustavo Cassilatti Furlani - PPGE-UFRGS. Marcelo Savino Portugal - PPGE-UFRGS and CNPq Associated Researcher.

Reference works in the literature of microstructure of the market exchange are Frenkel et al. (1996) and Sarno & Taylor (2002). The monograph of Lyons (2001) is an extensive study of the microstructure of the exchange market based on the order flow.
publicly available through systems such as Reuters and Bloomberg Data License (used herein), allowing for comprehensive studies on market microstructure using high frequency bid and ask quote data (tick by tick operations).

The availability of this information allows assessing some exchange rate market characteristics that cannot be systematically explained by usual macroeconomic models. Among unexplained effects we have the persistence of exchange rate returns in intraday data, related to deviations of the martingale property from the returns, which translates into violations of market efficiency and into the principle of no arbitrage. Other effects that are not accounted for by macroeconomic analysis are the determinants of the spread between bid and ask; the importance of the information captured by order flows and its predictive power over future rates; the impacts of chartist analysis on the exchange rate market; influence of trading volume, spatial location of agents and volatility in price setting; and the importance of private information for the determination of prices and spreads.

A remarkable difference between market microstructure models and macroeconomic models concerns assumed theoretical restrictions. Macroeconomic models are often based on representative agent structures, symmetric information, rational expectations and absence of transaction costs. Market microstructure models, on the other hand, are often characterized by asymmetric and heterogeneous structures\(^2\). There are several types of agents in this market, such as traders, market makers and customers with distinct strategic goals and information sets.

Since the exchange rate market is decentralized\(^3\) and its operators are physically distant, the information sets are distinct among agents, rendering private information relevant to the price setting process. These different sets of information can allow arbitrage situations, which is indeed quite common and could affect the degree of market efficiency, as reported in Flood (1994).

The wealth of information obtained from intraday data allows for the assessment of issues that would not be accounted for by lower frequency data analysis. In addition to prices, intraday transactions include other interesting sources of information, such as the time elapsed between two operations in the market (order durations). The time elapsed between two orders is linked to the arrival of new information in the market and is also an inherent liquidity measure\(^4\). This information is relevant in market microstructure models since prices are likely to be affected by recent transactions (Hasbrouck (1991) and Dufour & Engle (2000)), that is, prices and the spread in the subsequent transaction will be affected by previous prices and also by trading volume, spread, and time of previous transactions.

The aim of the present article is to assess the empirical effects of market microstructure based on intraday bid and ask quotes in the R$/US$ exchange rate market. We evaluate the importance of private information in the market by testing the Markov property (Section 5). The result of this test encourages the development of an empirical price discovery model using a vector error correction model using bid and ask prices and the previous price durations and quote volatility as explanatory variables, allowing to check the impact of recent operations on prices.

Using the results of this model, we assess the determinants of equilibrium spread deviations between bid and ask by developing a model that enables the asymmetric response of the spread to the previous information set, by means of the quantilogram (Linton & Whang (2007)) and quantile autoregressions (Koenker & Xiao (2006)).

The paper is organized as follows: Sections 2 and 4 describe the data used, show some characteristics of these series and comment on the relationship with previous studies using exchange market data; Section 5 checks for the presence of asymmetric information by testing the Markov

\(^{2}\)See O'Hara (1995) for a review of the theoretical models of asymmetry of information in the context of market microstructure and Hasbrouck (2007) to empirical implications.


\(^{4}\)For a review of informational content in durations and econometric models for conditional durations see Engle & Russell (1998) and Engle (2000). Fernandes & Grammig (2005a) and Fernandes & Grammig (2005b) and references for specification and testing for conditional duration models.
property; Section 7 describes the vector error correction model used for price discovery; and Section 7 analyzes the effects of asymmetry on the conditional distribution of the spread. Section 8 concludes.

2. DATABASES AND PREVIOUS STUDIES.

Despite the extensive literature on exchange rate market microstructure, there has been a scarcity of research into the BRL/US$ exchange rate microstructure, one of the most significant emerging markets. The most important studies on the BRL/US$ exchange rate market microstructure are those by García & Urban (2004) and Wu (2007). The former takes an in-depth look into institutions and the operation of the interbank currency exchange market in Brazil, describing the agents, institutions and the existing trading mechanisms. In addition, the paper provides econometric evidence of a shift of Granger causality from the futures market to the spot market, using daily data.

The article by Wu (2007) is a comprehensive study of Brazilian exchange market microstructures based on daily data collected from the Sisbaex system, the Central Bank database containing all the consolidated currency transactions in Brazil. The complete order flow database used by Wu (2007) is the only study in the international literature that includes 100% of a country’s official currency operations and enables sorting out the effects of exchange rate movements related to trading and financial operations, Central Bank interventions in the currency exchange market, and the consequences of these movements on the exchange rate.

There are, however, criticisms against the database used by Wu (2007). The first one concerns the fact that the data set does not correspond to the publicly available information at the time of the agents’ decision. Some of the information used is not disclosed by the Central Bank and some is made known but with a large time delay, and therefore it is not the same as the data set used by agents in the decision-making process in the intraday market.

Our analysis differs from previous ones because we use high-frequency bid and ask quotes. This method is analogous to those used by Goodhart (1989) and Bollerslev & Domowitz (1993) and allows assessing the effects of microstructure on the operation of the currency exchange market observed in intraday quotes.

Our data are based on the spot market quotes provided by the Bloomberg Data License database. This database format is known as FXFX DATA. This system collects information about the operations carried out in several markets, including Sisbex (Trading system of the Brazilian Mercantile & Futures Exchange) and over-the-counter operations, based on information gathered from several market participants. The sample used in this study contains all order fed into the system, starting on May 28, 2006 and ending on November 30, 2006. The data set format is shown in Table 1.

---

Lyons (2001) contains a full description of the various databases used in microstructure of foreign exchange markets.
The data file contains eight columns. The first column specifies the traded asset (currency). The second column contains the time at which the operation was fed into the system, with accuracy in seconds. The third and fourth columns show the bid identification and the bid price; the fifth column presents the agent who provided the bid value. The sixth, seventh and eighth columns show the ask, ask price, and the operator who provided the ask value. Note that the data do not always specify the operator in charge of the bid and ask, since anonymous order are allowed.

The high-frequency exchange rate market data have a lot of limitations compared to other databases used in empirical microstructure. The Transaction and Quotes (TAQ) for stocks traded on the New York Stock Exchange (NYSE), for instance, contains additional information such as the price and volume of transactions, instead of only indicative quotes. Exchange rate data show only the behavior of quotes, but do not reveal the traded value. Other limitations include the absence of information on the trading volume and the impossibility to find out whether the order was initiated by a buyer or seller, which is provided by the Transaction Orders and Quotes (TORQ) database compiled by Hasbrouck (1992). The lack of these informations (transaction prices, volumes and whether the transaction was initiated by a buyer or seller) about effective transactions results from the absence of a disclosure clause for exchange rate operations.

Although the scope of market microstructure analysis regarding the dataset is limited by the no disclosure, one should note that this is the dataset publicly available to spot market operators, and analyzed in other studies as Goodhart (1989) and Bollerslev & Domowitz (1993).

There is some evidence that the omission of transactions does not affect the estimation results, as pointed out by Goodhart (1989), but the literature has demonstrated some problems with the use of quotes because they are just indications and not transactions. Lyons (1996) shows that interdealer spreads are lower than those of indicative quotes. To dismiss such criticisms one may say that these quotes correspond to the information made publicly available in the spot market; studies as the one by Lyons (1996), which uses data from a private dealer, have two shortcomings: the time interval is too short (weeks at most) and they capture a private dealer’s trading behavior and might not necessarily represent the behavior of the market, due to the large heterogeneity of agents in the exchange rate market.

Note that the high-frequency data used in this paper contain two important pieces of information that are not provided by other studies involving the exchange rate market, such as that conducted by Wu (2007). The information on trading time allows us to build a variable for order duration, which is given by the time elapsed between the arrival of two orders to the market. This variable provides information about market liquidity and volatility, and its behavior represents the arrival of new information to the market (e.g. Engle & Russell (1998), Engle (2000) and Fernandes & Grammig (2005a)) Another variable is the conditional volatility derived from a GARCH model for the mid-quote between the bid and ask, used as a proxy for the traded price. The use of a mid-quote as trading price is justified in the microstructure literature, since in some models, the mid-quote is related to the fundamental asset price.

3. DATA CONSTRUCTION AND FILTERING

Time series containing information on bid and ask prices are used in this paper, as shown in Table 1. Two additional variables are built: the duration variable, given by the number of seconds between the arrival of two orders, and the mid-quote variable, given by the average between the bid and ask prices, which will be used to build the volatility proxy using a GARCH model.

Real-time storage of data yields a relatively large number of operations with incorrect information. The correction was made by eliminating clearly discrepant observations due to mistypings (e.g.: a bid recorded as 0.219 instead of 2.19), observations with negative spreads or spreads that are not compatible with the local spread behavior. These outliers were filtered out by the use of a filtering rule that regards as outliers operations whose spread between bid and ask is larger than 10

---

6 For this conclusion Lyons (1996) uses the order flow from a particular dealer.

7 See Hasbrouck (2007) for a review of procedures and models used in empirical modeling of market microstructure.

8 See Falkenberry (2002) for details on the problems in the processing of high-frequency data
standard deviations from the spread. This rule captures mainly mistypings in the database. After filtering out this information, our database comprises 279,737 tick-by-tick observations between May 28, 2006 and November 30, 2006.

Another consistency check regards the time of the transaction, which measures whether the time was recorded correctly and whether it follows the correct sequence. Note that, in line with the microstructure literature, we did not restrict trading hours because exchange rate market quotes are negotiated around the clock, since this market operates nonstop in the three major trading locations (United States, Europe, and Japan). Restricted information would lead to the exclusion of data that could serve as a benchmark for pricing other financial instruments, as information from other markets outside the Brazilian trading hours may affect the exchange rate determination in the domestic market. Around 4% of the quotes is negotiated outside the normal business hours of Sisbex system in the BM&F, the most important market for BRL/US$ exchange rate.

The construction of the volatility variable occurred through a GARCH(1,1) model for the mid-quote price. This variable is built in such a way that it can be used as a proxy for the volatility of the traded price, which is not observed directly. This variable is obtained through a GARCH(1,1)$^{9}$ model estimated by maximum likelihood with estimated parameters ($\omega = 1.51E09, \alpha = .210275, \beta = .7773732$). The duration series is obtained by the number of seconds between two orders.

4. Descriptive Statistics and Periodic Patterns

4.1. Descriptive Statistics and Unit Root Tests. Figures 1 and 2 show the graphs for the bid-ask series, spread, duration and volatility variables. The figures reveal that the periods with an increase in spreads correspond to the highest values obtained for the bid-ask series. Table 2 presents the descriptive statistics for these variables, as well as Phillips-Perron unit root tests for the bid-ask series and bid-ask log returns. The test results indicate that we should not reject the unit root null for the bid-ask spread and that the log return series are stationary at 1% significance level. As usual in financial time series, the Gaussian distribution is rejected for the bid and ask series.

4.2. Periodic Patterns. A stylized fact in high frequency financial series is the presence of intraday periodic patterns, e.g. Zivot & Yen (2003). To analyze periodic patterns, we executed a nonparametric fitting procedure that captures intraday periodic patterns, using the smoothing spline model for the spreads, durations and volatilities. A smoothing spline (Green & Silverman (1994)) can be defined as the solution to the minimization of the following function:

\[ S_\lambda(g) = \sum_{i=1}^{n} (Y_i - g(X_i))^2 + \lambda \int (g''(x))^2 dx \]

$^{9}$The complete model is not placed in the text by restrictions of space, but can be obtained by request to the authors.
Figure 2. Durations and Volatilities

Table 2. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Asks</th>
<th>Bids</th>
<th>Asks Log Returns</th>
<th>Bids Log Returns</th>
<th>Durations</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.163719</td>
<td>2.162170</td>
<td>-2.67E-07</td>
<td>-2.07E-07</td>
<td>12.75741</td>
<td>4.17E-08</td>
</tr>
<tr>
<td>Median</td>
<td>2.162000</td>
<td>2.160500</td>
<td>0.000000</td>
<td>0.000000</td>
<td>1.999999</td>
<td>2.47E-08</td>
</tr>
<tr>
<td>Maximum</td>
<td>2.236000</td>
<td>2.355400</td>
<td>0.010915</td>
<td>0.0100925</td>
<td>3563.100</td>
<td>5.00E-09</td>
</tr>
<tr>
<td>Minimum</td>
<td>2.122000</td>
<td>2.120000</td>
<td>-0.009199</td>
<td>-0.008751</td>
<td>0.000000</td>
<td>1.54E-08</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.022946</td>
<td>0.02887</td>
<td>0.000187</td>
<td>0.000203</td>
<td>72.92300</td>
<td>1.76E-06</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.515426</td>
<td>2.490172</td>
<td>247.0023</td>
<td>176.0372</td>
<td>814.3413</td>
<td>4827.558</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>14518.0</td>
<td>14433.53</td>
<td>6.93E+08</td>
<td>8.20E+11</td>
<td>7.96e+09</td>
<td>7.46E+14</td>
</tr>
<tr>
<td>Prob.</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Figure 3 shows that the spread is often higher outside the Brazilian market trading hours (interval of 2-6 a.m. hours). The figure shows that the spread tends to increase at opening and closing hours in the Brazilian currency exchange market, which is analogous to the “U” pattern obtained by Bollerslev & Domowitz (1993) and consistent with the theoretical model for spread determination by Bollerslev & Domowitz (1993). A relevant effect is that the spread tends to increase around 5 p.m., which is the time limit for currency operations at Sisbacen. This effect can be rationalized by the fact that unrecorded operations must be canceled, producing adjustment costs, inventory imbalance, and problems in risk margins.

The periodic pattern for the duration series shows that the time length between two quotes is quite long outside trading hours in Brazil, but shorter, with a slight increase around 5 p.m. The periodic patterns for the volatility series show large volatility outside trading hours (due to the smaller number of quotes, price jumps are higher), and that volatility tends to increase during the opening and the closure of trading hours in Brazilian market, which is possibly associated with the adjustment of trading positions. Note that these figures indicate a possible correlation between spread, duration and volatility. This association will be tested in sections 6 and 7.

An additional hypothesis about the distribution of spreads concerns the existence of clusters in the spread, which may be consistent with price collusion (Christie et al. (1994), Hasbrouck (1999)). The hypothesis of price collusion can be summarized as the tendency of spread quotes
towards yielding multiple values for the minimally allowed variation. In the case of exchange rate series, for instance, the minimum spread is 0.001, but spread distributions tend to concentrate in a certain multiple value for the minimum value. An easy way to test this effect is by checking the distribution of the last digit of the spread. Under the null of no price clustering, the last digit of the spread should be uniformly distributed and the proportion of values in each digit should be statistically equal. This is the approach followed by McGloarty et al. (2006), who provide a comprehensive analysis of the clustering effect on exchange rates.

Table 3 shows the distribution of the last digit of spread. Note that the last digit concentrates around values 1, 2 and 3, indicating that spread values tend to range from 0.001 to 0.003. The test for the equality of proportions is carried out by using the $\chi^2$ for equality of proportions, which rejects the null hypothesis of equal proportions. Even though the test rejects the null of no clustering, it does not indicate the possible cause for price clusters. The studies by Hasbrouck (1999) and Hasbrouck (1999) discuss some possible causes for this effect, and Hasbrouck (1999) complements this analysis with a dynamic model. In these models, price clustering may be related to a stochastic cost for liquidity provision incurred by the market maker. In section 6 we show that the equilibrium spread has a value around 0.0025, and that the concentration of values immediately below 0.001 and 0.003 units is related to short-run deviations from the equilibrium spread.

5. Markovian Property

The property of market efficiency, obtained by the assumption of agents’ rationality and efficient processing of the available data, implies that asset prices should be compatible with a first-order Markov process. Thus, the price at time $t$ should only depend on the most recent information at $t-1$ plus an innovation process:

$$P_t = P_{t-1} + \varepsilon_t$$

If the information is efficiently processed, all the information available up to period $t-1$ should be contained in price $P_{t-1}$, and thus price variation should correspond to a non systematic error.

\footnote{Formally, the information process is a filtration $\mathcal{F}_t$ given by a crescent sequence of sub-$\sigma$-algebras $B_t \subset B_u \subset B$ for $0 \leq t \leq u$, defined in an space of probability $(\Omega,B,P)$. We assume the usual filtration in this article, and suppress notation of the filtration.}
Table 3. Distribution of the spread last digit

<table>
<thead>
<tr>
<th>Last Digit</th>
<th>Size</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>36833</td>
<td>6.017438%</td>
</tr>
<tr>
<td>1</td>
<td>135928</td>
<td>48.59136%</td>
</tr>
<tr>
<td>2</td>
<td>111837</td>
<td>39.97934%</td>
</tr>
<tr>
<td>3</td>
<td>11260</td>
<td>4.025409%</td>
</tr>
<tr>
<td>4</td>
<td>1259</td>
<td>0.4500656%</td>
</tr>
<tr>
<td>5</td>
<td>434</td>
<td>0.1551457%</td>
</tr>
<tr>
<td>6</td>
<td>445</td>
<td>0.159078%</td>
</tr>
<tr>
<td>7</td>
<td>893</td>
<td>0.3192284%</td>
</tr>
<tr>
<td>8</td>
<td>657</td>
<td>0.2348635%</td>
</tr>
<tr>
<td>9</td>
<td>191</td>
<td>0.06827842%</td>
</tr>
</tbody>
</table>

\[ \chi^2 \text{Test Stat.} \quad p-value \]

\[ H_0: \text{equal proportions} \quad 842649.8 < 2.2e-16 \]

process\textsuperscript{11}. Another characteristic related to no-arbitrage conditions (e.g. Harrison & Kreps (1979) and Harrison & Pliska (1981)) is that conditional expectation \( E_Q[P_{t+k} | F_t] = P_t \), i.e., in the risk-neutral measure the price should be a martingale process, which leads to the concept of equivalent martingale measure.

However, microstructure models based on asymmetric information, such as those by Glosten & Milgrom (1985) and Easley & O'Hara (1987), predict that the existence of different information sets between agents affects the Markov property of bid and ask prices. In these models, asymmetric information causes prices at \( t \) to depend upon the whole trading history and not only upon the most recent information, invalidating the Markov property for prices and indirectly characterizing some type of market inefficiency. A discussion about this issue can be found in Flood (1994), who shows that the decentralization of agents in the currency exchange market represents a less efficient information-based form than in stock markets. Decentralization slows down the dissemination of information, and therefore the prices are correlated not only with the most recent price, but also with a long set of prices in the past.

Thus, we can check for the presence of asymmetric information using Markov property tests. The test proposed by Fernandes & Amaro de Matos (2007) takes into account the irregular pattern of price quotes over time in high-frequency financial data. This is a nonparametric test for conditional independence, based on the null hypothesis that if the Markov property holds, the length of time between both operations should be independent of the realization of the variable related to asset prices, that is, the spread between bid and ask.

The null hypothesis of the test derived by Fernandes & Amaro de Matos (2007) is given by:

\[ H_0 : \quad f_{iX_j}(d_i, x, d_j) = f_{i|X}(d_i) f_{x|j}(x, d_j) \]

where \( f_{iX_j}(d_i, x, d_j) \) represents the joint density of duration \( d_i \), of spread \( x \) and duration \( d_j \), \( f_{i|X}(d_i) \) the conditional density of duration \( d_j \) and \( f_{x|j}(x, d_j) \) the joint distribution between spread \( d_j \) and duration \( j \), for \( i > j \). If the conditional independence property holds, the null hypothesis given by equation 3 is equivalent to the validity of the Markov property.

The test derived by Fernandes & Amaro de Matos (2007) is based on the weighted quadratic distance between \( f_{iX_j}(d_i, x, d_j) - f_{i|X}(d_i) f_{x|j}(x, d_j) \), when densities are replaced with nonparametric density estimators. The test statistic is given by:

\[ 11\] Different types of market efficiency are related to distinct assumptions about the process \( \epsilon_t \). Efficiency type III would be associated with a process uncorrelated \( \epsilon_t \) correlated, whereas type II efficiency would be given more restrictive process \( \epsilon_t \) assuming independence. The more restrictive Efficiency type I is obtained assuming that the process \( \epsilon_t \) is independent and identically distributed. See Campbell et al. (1997).
Table 4. Nonparametric test of Markov Property

<table>
<thead>
<tr>
<th>Series</th>
<th>Test Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spread/Duration</td>
<td>135.10</td>
<td>0.000</td>
</tr>
<tr>
<td>Adjusted Spread/Duration</td>
<td>102.74</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ \Lambda \hat{f} = \frac{1}{n} \sum_{i=1}^{n} w_i(d_{k+j}, X_k, d_k) \left[ \hat{f}_i(d_{k+j}, X_k, d_k) - \frac{\hat{f}_i X_i(d_{k+j}, X_k)}{f_X(X_k)} \right] \]

Where estimators \( \hat{f}_i \) are kernel density estimators for joint, marginal and conditional densities and \( w_i \) is a weighting function. Fernandes & Amaro de Matos (2007) show that the statistic given by equation 5 has a standard normal asymptotic distribution:

\[ \hat{\lambda}_n = \frac{nb_n^{3/2} \Lambda \hat{f} - b_n^{-3/2} \delta \hat{\lambda}}{\sigma \hat{\lambda}} d N(0,1). \]

where \( n \) stands for the sample size, \( b_n \) is the bandwidth value, \( \delta \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} w_i X_i(d_{k+j}, X_k, d_k) \hat{f}_i(d_{k+j}, X_k, d_k) \)

\[ \sigma \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} w_i^2 X_i(d_{k+j}, X_k, d_k) \hat{f}_i^2(d_{k+j}, X_k, d_k) \]

and \( c_k \) and \( v_k \) are constants that depends on the selected kernel function.

To test the Markov property on prices, we followed the method proposed by Fernandes & Amaro de Matos (2007) and we used the spread series at time \( t \), and duration series at times \( t+1 \) and \( t \), and we use the log-spread and log-duration series, in compliance with the same methodology. We calculated the marginal, conditional and joint density estimators using a quartic kernel and the same rules applied by Fernandes & Amaro de Matos (2007) for the selection of the bandwidth \( c_k \) and \( v_k \). Table 4 shows the Markov property test results for gross spreads and durations and also for adjusted spreads and durations extracting the periodic pattern found in Section 4. The results indicate rejection of the null hypothesis that the first-order Markov property holds for any significance level in both data sets. This evidence supports the conclusion achieved by Richard Lyons - "Contrary to the asset approach - exchange rate determinations is not wholly a function of public news.\(^{12}\), where we assume that the public information is given by the past information on spreads and durations.

The rejection of the null hypothesis of the Markov property confirms the presumed existence of asymmetric information effects pointed out by Glosten & Milgrom (1985) and Easley & O'Hara (1987) for the exchange rate market. Our results are analogous to those obtained by Flood (1994), showing that the differential information between agents is key to the determination of spreads and that the existence of different information affects the agents’ price discovery process. Note that it is possible to explain the violation of the Markov property by the operational structure of the currency exchange market based on a framework that includes several dealers with different locations\(^{13}\).

The results of rejection of Markov property indicates that the price discovery process will not be based only on the price established in the immediately preceding time period, but on the whole set of past information. Note that this property directly affects the model used for price discovery, which will have a larger number of lags for the adjustment of the short-run dynamics.

\(^{12}\) Lyons (2001), pg 9.
\(^{13}\) See Lyons (2001) for a discussion on the effects of a structure of multiple dealers on the foreign exchange market.
6. Vector Error Correction Models for Price Discovery

One of the predictions of the market microstructure literature in the presence of asymmetric information is that agents should discover the actual equilibrium price of the asset in a process known as price discovery. In this process, agents seek to determine the fundamental asset price, which is contained in current quotes but contaminated with microstructure noise. When the price process is based on the existence of several prices for the same asset, e.g., stocks traded on several stock exchanges or existence of bid and ask prices, the discovery of the fundamental asset price is related to a mechanism of search for an equilibrium price between several quotes for the same asset. This is equivalent to the existence of correction mechanisms for deviations of prices in each quote from equilibrium prices.

This multivariate price discovery process can be represented by a vector error correction model in event time. By assuming a bivariate vector of prices \( P_t = [p_{1t}, p_{2t}] \), the vector error correction model is represented as follows:

\[
\Delta P_t = \mu_0 + A_1 \Delta P_{t-1} + A_2 \Delta P_{t-2} + ... + A_k \Delta P_{t-k} + \gamma (Z_{t-1} - \mu_1) + \lambda_1 X_{t-1} + ... + \lambda_j X_{t-k}
\]

\[
Z_{t-1} = [p_{1t-1} - B_1 p_{2t-1}]
\]

In this model, vector \( Z_{t-1} \) represents the deviations of the long-run equilibrium values between \( p_1 \) and \( p_2 \) and \( B_1 \) are the coefficient vectors in the equilibrium relationship; \( \gamma \) is the coefficient vector that controls the error correction mechanism; coefficients \( A_i \) represent the short-run coefficients; \( X_{t-k} \) is a vector of explanatory variables that are not cointegrated with \( p_1 \) and \( p_2 \) and \( \lambda \) is a coefficient vector that captures the influence of \( X_{t-k} \) on the short-run dynamics of \( \Delta P_t \).

The VECM (Vector Error Correction Model) seeks to decompose the price adjustment dynamics into two components - one that is linked to the short-run dynamics, given by components \( A_i \Delta P_{t-k} \) and \( \lambda_i X_{t-k} \) and the other one linked to the dynamics of the long-run equilibrium deviations given by cointegration vector \( Z_{t-1} \). In the context of the price discovery model, the cointegration vector represents the equilibrium between two asset price measures, and this equilibrium includes a measure of the fundamental asset price, given by \( Z_{t-1} \) in the VECM.

In our study, the price vector is given by the bid and ask prices \( P_t = [bid_t, ask_t] \) and thus the cointegration vector captures the relationship of the equilibrium spread value given by \( ask - B \cdot bid \). The aim of the VECM proposed in this study is to decompose bid and ask variations at time \( t \) into two components: a short-run component given by past bid and ask variations and explanatory variables and another component linked to the long-run equilibrium, given by the correction of the deviations of the long-run relationship between bid and ask.

The violation of the Markov property shows that the price discovery process cannot be based only on the set of information about the immediately previous observation \( \Delta P_{t-1}, X_{t-1} \), since quotes at time \( t-1 \) do not contain all the information available in the market (the private information shown in previous transactions is not instantly incorporated into the prices, causing the violation of the Markov property presented in Section 5). The price discovery process does not depend only on the immediately previous price, but on the whole trading history, a property that is compatible with some asymmetric information models such as Glosten & Milgrom (1985) and Easley & O'Hara (1987).

In our VECM model, we also included the values of past durations and past conditional volatilities of the mid-quote price as explanatory variables. These two variables are included as a way to check whether the bids and asks respond to the effects of liquidity and uncertainty, using durations and volatilities as proxies for these effects.

---


15 For a review of multivariate models of price discovery see Hasbrouck (2007), and also Engle (2000) for models in event time. Event time is the use of order of operations as time index, replacing the calendar time as index of the stochastic process, and generating irregular spaced time series. Hasbrouck (2007) discusses the advantages of the use of the event time in the microstructure studies, which are related to time deformation process.
The specification of the VECM is valid in the presence of an equilibrium vector between the bid and ask, a hypothesis that can be verified using a cointegration test. To test for the existence of an equilibrium relationship, we used Johansen’s cointegration test, whose results for the bid and ask log series are shown in Table 5. The specification of the cointegration test is an error correction model using 120 lags for variations in the bid and ask logs, 24 lags for past durations and 20 lags for past volatilities, where the specification was determined by Schwartz Information Criteria. As discussed in Section 5, this large number of lags is related to the violation of the Markov property. The test results show that we rejected the null hypothesis of nonexistence of a cointegration vector with p-value of 0.001 by both tests (Rank and Trace) obtained by Johansen’s procedure, indicating the existence of a long-term equilibrium mechanism between bid and ask logs and the validity of vector error correction model.

Note that Johansen’s cointegration test is based on the assumptions of normal distribution in the residuals and absence of structural breaks. The hypothesis of normal distributions is not valid for our dataset, since kurtosis and asymmetry indicate non-Gaussian distributions. Therefore, the critical values used by the test must be affected by this violation. Note, however, that there is an a priori economic reason for the existence of a cointegration vector between the bid and ask, since an imbalance between the bid and ask, representing a non-stationary spread, leads to systematic arbitrage opportunities. Thus, we did not reject the evidence in favor of a cointegration hypothesis, even with possible distortions in the power of the test.

The vector error correction model is partially shown in Table 6, where we present the estimated cointegration vector (cointegration equation), the loading matrix for the correction of long-run deviations (error correction) and the first two lags of the short-run mechanism. The cointegrating equation shows that the normalized cointegration vector for the bid-ask log is $[1 -1.000949]$, which represents an equilibrium spread of .00259.

This equilibrium spread value can be explained by three basic factors: costs related to market dealers’ functions, costs of handling currency inventory and a factor related to the asymmetric information given by adverse selection. The dealers’ costs are linked to the provision of immediate liquidity. The inventory costs are related to the provision of liquidity and to the possibility that the dealers may be operating with agents who have privileged information (insiders). The adverse selection problem is given by the fact that the dealers are not able to distinguish agents with demands for liquidity and hedge from insiders, thus increasing the spread for both classes of agents.

The estimated loading matrix $\gamma$ is given by the value of -0.015364 for the ask log variation and 0.014691 for the bid log variation. We can interpret these signs as follows: positive deviations from the equilibrium spread are adjusted by reducing the ask price and increasing the bid price, but the mechanism is the opposite for spreads below the equilibrium value, being characterized by an increase in the ask price and a decrease in the bid price. The short-run mechanisms are harder to analyze, due to the large number of lags and the change in the sign of coefficients. There is

---

16 See Frenkel et al. (1996) and Sarno & Taylor (2002) for detailed references on this explanation of determinants of the spread.
evidence that 120 lags correspond on average to 20 minutes at the calendar time, showing that the mean time for the incorporation of information can be approximated by this value.

Figure 4 shows the generalized response impulse functions obtained by the VECM estimation in Table 6. The figure indicates that the shocks converge around 50 observations to their permanent value, and show stable convergence to the long-run values.

7. Spread Determination

The VECM estimated in Section 6 allows determining an empirical models for the equilibrium spread value and the correction mechanism for equilibrium spread deviations, but it does not allow the direct identification of the factors that influence and impact the spread. To assess spread determinants, we begin by investigating the empirical characteristics of the spread deviation series, created by the cointegrating equation in the VECM. Based on the observed characteristics, we formulated an asymmetric response model for the conditional distribution of the spread based on quantile autoregression (Koenker & Xiao (2006)) and the results of the quantilogram estimates (Linton & Whang (2007)).

Note that the existence of cointegrating vector between the log bid and the log ask is equivalent to the existence of a stationary process for spread deviations. To describe this process, we first formulated a linear model for spread deviations. This method is based on regressions on the spread as those described in Jorion (1996). Following the methodology of Jorion (1996), we developed an

### Table 6. Error Correction Model

<table>
<thead>
<tr>
<th></th>
<th>Long Run</th>
<th>Dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cointegrating Equation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOG(ASK(-1))</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>LOG(BID(-1))</td>
<td>-1.000949</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.7E-05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-575570.]</td>
<td></td>
</tr>
<tr>
<td><strong>Error Correction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D(LOG(ASK))</td>
<td>-0.015364</td>
<td>0.014691</td>
</tr>
<tr>
<td></td>
<td>(0.00089)</td>
<td>(0.00094)</td>
</tr>
<tr>
<td></td>
<td>[-17.2781]</td>
<td>[15.9325]</td>
</tr>
<tr>
<td>D(LOG(BID))</td>
<td>0.087691</td>
<td>-0.238502</td>
</tr>
<tr>
<td></td>
<td>(0.00209)</td>
<td>(0.00214)</td>
</tr>
<tr>
<td></td>
<td>[32.5363]</td>
<td>[-70.4570]</td>
</tr>
<tr>
<td>DUR(-1)</td>
<td>-2.08E-06</td>
<td>-4.17E-06</td>
</tr>
<tr>
<td></td>
<td>(1.4E-07)</td>
<td>(1.5E-07)</td>
</tr>
<tr>
<td></td>
<td>[-16.5732]</td>
<td>[-27.4075]</td>
</tr>
<tr>
<td>DUR(-2)</td>
<td>-1.38E-07</td>
<td>-9.12E-07</td>
</tr>
<tr>
<td></td>
<td>(7.2E-08)</td>
<td>(1.5E-07)</td>
</tr>
<tr>
<td></td>
<td>[-16.5732]</td>
<td>[-5.61120]</td>
</tr>
<tr>
<td>GARCH(-1)</td>
<td>7.768290</td>
<td>80.278222</td>
</tr>
<tr>
<td></td>
<td>(3.20714)</td>
<td>(3.48201)</td>
</tr>
<tr>
<td></td>
<td>[23.5601]</td>
<td>[25.63098]</td>
</tr>
<tr>
<td>GARCH(-2)</td>
<td>-69.03694</td>
<td>-14.68382</td>
</tr>
<tr>
<td></td>
<td>(4.73599)</td>
<td>(5.00135)</td>
</tr>
<tr>
<td></td>
<td>[-14.7038]</td>
<td>[-34.2260]</td>
</tr>
</tbody>
</table>
autoregressive linear model for the spread by adding variables that represent the expected values of volatility and duration, measuring the expected effects of uncertainty and liquidity.

The formulation of this model seeks to control the stochastic impacts of the dealer’s costs and the impact of expected risk and liquidity values on the determination of equilibrium spread deviations (which can be seen as deviations on the average costs embedded in the equilibrium spread) discussed in Section 6.

The model is based on a third-order autoregressive process for equilibrium spread deviations, with the addition of one-step ahead predictions for volatility and duration. Volatility forecasts are obtained by the same GARCH model used for the construction of the volatility variable. For duration forecasts, we estimated and create forecasts by mean of an Autoregressive Conditional Duration (ACD) model 18 (Engle & Russel (1998)). The aim of incorporating the predictions is to include the effect of agent’s expectations about the volatilities and durations expected for the next transaction in spread determination.

The results obtained in this spread decomposition (7) show that there is a high dependence of the spread on past spreads (the persistence measured by the sum of the autoregressive coefficients

---

17 An alternative way of looking at the determination of bids and asks is the use of permanent-transitory decompositions for time series (Hasbrouck (2007)). The permanent component would be linked to fundamentals of the asset and the transitory components to microstructure effects.

18 The estimated model is a Autoregressive Conditional Duration (ACD) model, with estimated parameters:

\[ \psi_t = 1.37e - 05 + 0.190683x_2 + 0.88451\psi_t \] where \( x_t \) are price durations and \( \psi_t \) the conditional durations. More general ACD models could be used; see Bauwens & Giot (2000), Bauwens et al. (2002) and Fernandes & Grammig (2005b) for generalizations for the ACD model.
Other important effects are the signs obtained for the coefficients related to volatility and duration forecasts. We obtained a positive sign for volatility and a negative but non significant sign for duration. This may be interpreted as an additional premium in the spread for uncertainty. The non significance of duration can be interpreted as a measure of high liquidity in this market, where the agents do not have to pay a premium for urgent transactions. These effects can be interpreted as the dealers protection against uncertainty (volatility can be related to the arrival of insiders with privileged information and protection against a higher loading cost for an increase in volatility). Similar effects are obtained in Glassman (1987), who found positive correlation between spread and volatility.

A possible shortcoming of this model is the symmetric treatment of spread deviations: spreads below the equilibrium value are treated just like those values above the equilibrium spread. Note that these situations are intuitively associated with different market situations. Therefore, the imposition of the same response in both situations can be an invalid restriction.

7.1. Quantilogram. To check for a possible asymmetric response in spread deviations, we used a tool known as quantilogram, derived by Linton & Whang (2007). The quantilogram is a generalization of the correlogram to the modeling of the dependence in conditional quantiles of the time series distribution. The quantilogram is also a measure of directional predictability, as discussed in Linton & Whang (2007) and is part of the general literature on tests for market efficiency.

Let $y_{1}, y_{2}, \ldots$ be a stationary process whose marginal distribution with quantiles $\mu_{\alpha}$ for $\alpha \in (0, 1)$. In the null hypothesis of no directional predictability conditional on quantile $\alpha$:

$$E[\psi_\alpha(y_t - \mu_\alpha)|\mathcal{F}_{t-1}] = 0$$

where $\psi_\alpha(x) = \alpha - 1(x < 0)$ is an indicator function that measures if the variable $t$ hit the quantile $\alpha$ and $\mathcal{F}_{t-1} = y_{t-1}, y_{t-2}, \ldots$ is the usual filtration. Under the null hypothesis of no directional predictability, if the variable at time $t-1$ is below quantile $\alpha$, the chance is no more than $\alpha$ for series $y$ to achieve this quantile again at time $t$. Violations of this hypothesis are evidence for asymmetric behavior.

---

\footnote{The analyzed series $y$ can be a directly observed process or residual of a model estimated in a first stage, as in our case. Linton & Whang (2007) derived asymptotic distributions valid in the two situations. The inference in quantilogram also remains valid in the presence of general heteroscedastic components, as stationary GARCH process.}
EMPIRICAL MARKET MICROSTRUCTURE: AN ANALYSIS OF THE BRL/US$ EXCHANGE RATE MARKET USING HIGH-FREQUENCY DATA

The quantilogram has two advantages over the usual directional predictability measures: the estimation of conditional quantiles is robust to the presence of outliers and some quantiles of the distribution of asset returns have a straightforward interpretation in risk management, being related to measures such as Value At Risk and Expected Shortfall. Note that there is a similar interpretation in the analysis of the spread - if the higher quantiles of the spread distribution show persistence, this effect affects the dealers incomes and the transaction costs in this market.

To measure the dependence in conditional quantiles, the quantilogram derived by Linton & Whang (2007) is given by the following expression:

\[ \rho_{\alpha k} = \frac{E[\psi_{\alpha}(y_{t+k} - \hat{\mu}_{\alpha})]\psi_{\alpha}(y_{t+k} - \mu_{\alpha})]}{E[\psi_{\alpha}^2(y_{t} - \hat{\mu}_{\alpha})]} \]

The sample estimator is given by:

\[ \hat{\rho}_{\alpha k} = \frac{\sum_{t=1}^{T-k} \psi_{\alpha}(y_{t} - \hat{\mu}_{\alpha})\psi_{\alpha}(y_{t+k} - \hat{\mu}_{\alpha})}{\sqrt{\sum_{t=1}^{T-k} \psi_{\alpha}^2(y_{t} - \hat{\mu}_{\alpha})} \sqrt{\sum_{t=1}^{T-k} \psi_{\alpha}^2(y_{t} - \mu_{\alpha})}} \]

where the estimator for \( \hat{\mu}_{\alpha} \) is given by the sample quantile, which is an estimator of the process:

\[ \hat{\mu}_{\alpha} = \arg \min_{\mu} \sum_{t=1}^{T} \rho_{\alpha}(y_{t} - \mu) \]

The null hypothesis of no directional predictability is given by

\[ H_0 \ E[\psi_{\alpha}(y_{t} - \mu_{\alpha})\psi_{\alpha}(y_{t+k} - \mu_{\alpha})] = 0 \]

for the each \( \alpha \) quantile. Figure 5 shows the quantilogram estimated for quantiles (0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99). The quantilogram estimated for these quantiles shows that the same shape of autoregressive dependence found in conditional mean occurs in the quantiles, but note that the dependence intensity is different in each quantile, indicating a correlation in the first lags close to .4 for quantile .01 and an upward trend in the persistence of higher quantiles; in quantile 0.99 (highest spreads), persistence is close to .95.

This asymmetric effect demonstrates that the lowest percentiles are characterized by low persistence and fast reversion to the unconditional quantile, whereas for the points where the spread is way above the equilibrium values (percentiles greater than .90), large persistence exists. Note that this asymmetry effect is of major financial importance, since it shows that high spreads tend to be more persistent than lower ones. Again, we can interpret this effect as a response of dealers to unanticipated shocks, such as increase in uncertainty and higher currency inventory maintenance costs.

7.2 Quantile Autoregression. The quantilogram reveals different time dependence patterns for each conditional quantile, but it does not represent a complete parametric model. This indicates the necessity to build a model for the conditional distribution of the spread for each quantile using an autoregressive structure, but also controlling for volatility effects and expected durations, analogous to the linear model estimated for the spread.

A possible tool for this type of analysis is the quantile autoregression model (Koenker & Xiao (2006)), which consists in formulating a quantile regression model using the lags of the dependent variable as explanatory variables. In a quantile regression (Koenker & Basset (1978)), the objective function is directly formulated as a function of the conditional quantile, minimizing the function:

\[ E[(y_{t} - \mu_{\alpha})|F_{t-1}] = 0 \]
\[(14) \quad \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_{\alpha}(y_i - x_i'\beta(\alpha)) \]

which corresponds to a loss function \(\rho_{\alpha}\) conditional on quantile \(\alpha\), where \(\alpha \in (0,1)\). We define the loss function as \(\rho_{\alpha}(u) = u(\alpha - I(u < 0))\), where \(I(\cdot)\) is an indicator function, and \(u\) is the difference between the observed value \(y_i\) and the value predicted by \(x_i'\beta(\alpha)\). Estimators for \(\hat{\beta}(\alpha)\) are obtained by minimizing the loss function given by 14, obtained by the expected value of \(\rho_{\alpha}(y_i - x_i'\beta(\alpha))\) relative to each \(\beta(\alpha)\).

Note that the quantile regression model can be extended to autoregressive structures, using a quantile regression for the autoregressive process\(^{20}\).

\[(15) \quad Q_{y_t}(\alpha|y_{t-1}, ..., y_{t-p}) = \beta_0(\tau) + \beta_1(\alpha)y_{t-1} + ...\beta_p(\alpha)y_{t-p} \]

Thus a first-order autoregressive quantile model (QAR(1)) can be written as:

\[(16) \quad Q_{y_t}(\alpha|y_{t-1}) = \beta_0(\alpha) + \beta_1(\alpha)y_{t-1} \]

Note that we can represent this model as \(y_t = \beta_0(U_t) + \beta_1(U_t)y_{t-1}\) where \(U_t\) is uniformly distributed between \((0,1)\). In this formulation, the traditional autoregressive AR(1) model is obtained when \(\beta_0(u) = \sigma \Phi^{-1}\) and \(\beta_1(u) = \beta_1\). To model the possible asymmetry structure in the response of spread deviations, we built the following model:

\[(17) \quad Q_{y_t}(\alpha|y_{t-1}, ..., y_{t-p}) = \beta_0(\tau) + \nu(\alpha)E[Vol|\mathcal{F}_{t-1}] + \delta(\alpha)E[Dur|\mathcal{F}_{t-1}] + \beta_1(\alpha)y_{t-1} + ...\beta_p(\alpha)y_{t-p} \]

which is a version of the conditional quantile of the model estimated in Table 7.

We estimated\(^{21}\) these models for the same quantiles \((0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99)\) used in the quantilogram (8). The quantile regression model (Equation 17) reveals the asymmetric effect already shown by the quantilogram - the persistence of shocks is lower for the quantiles below the conditional median and increases for quantiles above the conditional median, being close to 1 in these quantiles, and confirming the asymmetric response of spread deviations.

The estimated coefficients related to volatility and duration show another interesting effect. The coefficients of volatility always have negative signs for quantiles below the median and positive signs for percentiles above the median. The effect of duration is clearer for percentiles above the median, where the coefficients are always positive and statistically significant.

This asymmetry effect indicates that volatility and duration increase the spreads for spread deviations above the conditional median, and this effect is enhanced in higher quantiles. We may interpret this asymmetry effect as asymmetric relationship of the spread with volatility and duration - spreads above the equilibrium spread show high persistence and are positively influenced by volatility and by expected durations, whereas spreads below the equilibrium spread are poorly persistent and negatively correlated with expected volatility and durations, indicating an asymmetric mechanism of reversion to the equilibrium spread.

8. Conclusions

In this paper, we assessed some empirical properties related to market microstructure using high-frequency bid and ask quote data for the BRL/US$ exchange rate market. The paper shows that some stylized facts observed in the international literature on exchange rate market microstructure are valid for the BRL/US$ series and introduces new tools for the analysis of empirical microstructure effects.


\(^{21}\)For a complete reference about quantile regression see Koenker (2005). We use the method of rank inversion for calculating the variance-covariance matrix of the parameters.
Among the effects analyzed herein, we observed that the violation of the Markov property implies that there is no immediate incorporation of new information into prices in this market, resulting in a structure with long-range dependence in terms of bid and ask returns. To capture this long-range dependence structure, we built a price discovery model using a vector error correction model, parameterizing this process of information incorporation and obtaining an estimate for the equilibrium spread, which is interpreted in the microstructure literature as a measure of the average costs of liquidity provision and stock loading by dealers who operate in this market.

The modeling of the spread shows that there is a mechanism of asymmetric response for this variable, where spread values above and below the equilibrium value react differently to the previous information about the spread, volatility and durations. Spread values above the equilibrium value show high persistence and react positively and proportionally to the quantile in relation to the expected volatility and conditional duration, whereas we found an inverse relationship for quantiles below the spread distribution, with a negative correlation of the spreads with the expected volatilities and durations and low persistence, which characterizes a nonlinear mean reversion in the spreads.

This analysis of asymmetry using tools such as quantilogram for the identification of asymmetry in conditional quantiles and the modeling of this structure using quantile regression models is original in the literature on currency exchange market microstructure. Such empirical evidence points to new stylized facts that should be added to theoretical models of market microstructure.
References


