Price Discovery in Order Books: Evidence from FX in Brazil

Paulo Ramos* Alex Ferreira †

April 5, 2025

Abstract

Abstract: We investigate price discovery in the Brazilian FX futures market using over 14 million bid and ask quotes from a volatile trading session marked by central bank interventions. Unlike previous studies on executed trades, we examine quote-level data to detect leadership dynamics. Applying wavelet-based and traditional econometric methods, we find frequent shifts between quoting sides, suggesting that informed and uninformed traders intermittently shape price formation.

Keywords: Price Discovery; Wavelet Analysis; Foreign Exchange Market.

JEL classification: C14, G14, G15

*University of São Paulo, paulo-ramos@usp.br

†University of São Paulo, alexferreira@usp.br

1 Introduction

Agents with heterogeneous information sets populate financial markets. In the market microstructure literature, this asymmetry is central to the price discovery process: trading—not merely the arrival of public news—serves as the channel through which informed agents convey private signals to the market. This is how dispersed information is aggregated and incorporated into prices.

This study contributes to a growing literature on price discovery, which has developed a variety of measures to identify the informational leadership of venues, instruments, or agents. Seminal contributions include (Hasbrouck, 1995), (Baillie et al., 2002), and (de Jong, 2002), whose methods have been applied to equity markets ((Booth et al., 1999); (Chakravarty et al., 2004)) and commodities ((Garbade and Silber, 1983); (Figuerola-Ferretti and Gonzalo, 2010)). In the context of FX markets, studies such as (Cabrera et al., 2009) and (Osler et al., 2011) analyze whether spot or futures markets dominate in processing new information, with results often pointing to spot market leadership in developed currency pairs.

Our analysis focuses on the Brazilian FX futures market—a highly liquid and relatively deregulated environment where agents with varying levels of information interact. Prior literature on emerging markets, such as (De Boyrie et al., 2012) and (Kumar, 2018), finds mixed evidence of informational dominance between spot and futures prices. However, these studies rely on executed trades and parametric models in the time domain, typically using frameworks like vector error correction models (VECM), and have not examined the limit order book directly.

In contrast, we investigate whether the best bid and ask quotes—before execution—carry directional information about price formation. This question remains largely unexplored in the FX literature, where price discovery is typically examined through executed trades. To our knowledge, no previous study has analyzed order book quotes in this context. By shifting the focus from transaction to quoted prices, we ask whether one side of the book consistently leads the other, whether leadership fluctuates over time, or whether no systematic pattern emerges. The empirical analysis relies on a high-frequency dataset from the Brazilian FX futures market, consisting of over 14 million observations from November 27$^{\text{th}}$, 2019—a trading day marked by heightened volatility and central bank interventions. Our main findings challenge the notion of a consistent asymmetry between bid and ask quotes. Instead, we observe a diffuse and dynamic price discovery process where leadership alternates frequently, reflecting the interplay of informed and uninformed traders on both sides of the order book.

This question answered in this paper is relevant to researchers, policymakers, and market participants. For academics, it broadens the scope of price discovery analysis by shifting attention from executed trades to the informational content of quoted prices. Policymakers—particularly in floating exchange rate regimes like Brazil's—must understand quote-level dynamics in FX markets. This is essential for interpreting liquidity conditions and anticipating warranted intervention. Persistent bid-ask asymmetries can signal strategic behavior and inform quoting practices and risk management decisions for market makers.

We approach this question using high-frequency data from November 27$^{\text{th}}$, 2019—a trading day marked by heightened volatility and two central bank spot interventions.[1] The dataset comprises over 14 million observations from the mini futures contract ($10,000 USD), including 5,096,477 sell orders and 9,422,183 buy orders. Rather than relying on trade data, we examine the best bid and ask quotes using a bivariate wavelet framework based on the Cross Wavelet Transform (XWT), which allows us to track lead-lag relationships across time and frequency. We also construct simulated series to benchmark phase-based metrics and isolate structural dynamics.

A recent strand of the literature has incorporated spectral methods into price discovery analysis. Wavelet-based techniques have been applied to commodity and exchange rate markets to examine short-term fluctuations and co-movement patterns. (Joseph et al., 2015) and (Nigatu and Adjemian, 2020) use Wavelet Coherence (WTC) to study price discovery and market integration. Other studies, such as (Kumar et al., 2016) and (Ghosh and Chaudhuri, 2017), explore forecasting exchange rates with wavelet tools, but none

[1] A pre-announced spot sale at 09:30 and an unanticipated one at 12:39, see (Ferreira et al., 2025).

apply to price discovery in FX futures. Our work contributes to this frontier by analyzing the dynamics of quoting behavior through a time-frequency lens.

We also build on the empirical literature on the Brazilian FX market, including (Sung, 2021), (Garcia et al., 2015), (Laurini et al., 2008), and (Ferreira et al., 2025), which address interventions, liquidity, and the relationship between spot and futures. What distinguishes our study is the use of quote-level data and the construction of a continuous metric of price discovery, as opposed to interval-based or aggregated measures.

Our results indicate no robust evidence of persistent asymmetry between the bid and ask sides on the selected trading day. Parametric measures such as Information Share and Component Share produce conflicting outcomes depending on the temporal aggregation level, while the wavelet-based analysis reveals frequent, short-lived shifts in leadership. Rather than pointing to a systematic advantage on either side of the book, the findings suggest a more diffuse price formation process, potentially consistent with informed and uninformed traders operating intermittently on both sides in response to evolving market conditions.

The remainder of the paper is organized as follows. Section 2 presents the methodological framework, including econometric benchmarks and the wavelet-based approach. Section 3 describes the dataset and simulation design. Section 4 reports the results of both simulated and empirical analyses. Section 5 concludes with implications for price discovery and suggestions for future research.

2 Methodology

Model. Before moving to the econometric aspects of the model, some conceptual considerations need to be made. The first concerns interpreting the lead-lag relationship between quotes: we assume that the lead/anticipation phenomena can be due to asymmetric information. We suppose that private information is converted into orders and, therefore, conducts the innovation process efficiently. The second one involves timing: they vanish quickly. There is no reason to consider that the opportunities to profit from private information would take too long because of market efficiency.

We are interested in understanding how a wavelet-based metric behaves in front of different scenarios and comparing it to usual price discovery metrics. Because analytical solutions are hard to obtain, we simulate series that closely resemble our data in such a way that they capture the most usual dynamics. We use assumptions that are reasonable and consistent with the observed patterns. Therefore, assume three hypotheses:

- **H1** - Bid and ask quotes follow a vector auto-regressive structure (VAR);
- **H2** - Bid and ask quotes are non-stationary;
- **H3** - There is a long-run equilibrium between these quotes, i.e they are $CI(1,1)$

H1 comes from the intuitive idea that if a good has different prices, they must relate. **H2** is a consequence of the efficient price being a martingale and, consequently, non-stationary. **H3** comes as a consequence of (Roll, 1984) definition that the bid-ask spread corresponds to a small region where the efficient price varies - thus the quotes should not drift way in the long-run.

Suppose f_t^b and f_t^a represent, respectively, the best bid and ask quotes in the FX futures market. Assume that the data generating process follows this dynamics $\forall t \in \{1, 2, ..., n\}$:

$$f_t^b = \rho_b f_{t-1}^b + \delta_b f_{t-1}^a + \gamma^b + \epsilon_t^b, \tag{1}$$

$$f_t^a = \rho_a f_{t-1}^a + \delta_a f_{t-1}^b + \gamma^a + \epsilon_t^a. \tag{2}$$

where $\epsilon_i \sim N(\sigma_i^2, 0)$ are iid for $i = \{a, b\}$.

Given **H3**, from the (Engle and Granger, 1987) representation theorem, we can write our system as a Vector Error Correction Model (VECM), where $\beta = (\beta_a, \beta_b)$ is the cointegration vector and $\alpha = (\alpha_a, \alpha_b)$ represents the speed of adjustment. In other words, quotes react to a deviation from the long-run equilibrium converging to the mid-quote:

$$\Delta f_t^b = \alpha_b(\beta_a f_{t-1}^a - \beta_b f_{t-1}^b) + \epsilon_t^b, \tag{3}$$

$$\Delta f_t^a = \alpha_a(\beta_a f_{t-1}^a - \beta_b f_{t-1}^b) + \epsilon_t^a. \tag{4}$$

Rewriting and normalizing β and α, we have $(\beta_a, \beta_b) = \left(1, \frac{(1-\rho_a)}{\delta_a}\right)$ and $(\alpha_a, \alpha_b) = \left(\delta_a, \frac{-(\delta_a\delta_b)}{(1-\rho_a)}\right)$. In order to $f_t^a - \beta_b f_t^b \sim CI(1,1)$ it must be the case that $\rho_a > -1$ and $\delta_a\delta_b + \rho_a^2 < 1$. These restrictions follow from the root conditions of our primitive system's characteristic equation, which also imposes that one of our auto-regressive parameters must be endogenously defined based on our previous normalisation:

$$\rho_b = \frac{(1-\rho_a) - \delta_a\delta_b}{(1-\rho_a)}. \tag{5}$$

Given the VECM, we can map the econometric measures of price discovery from the primitive system. Denote Ω as the variance-covariance matrix of the innovative process (ϵ_t) of both quotes. Additionally, from our hypotheses of the quotes being $I(1)$, consider Ξ_i from the vector moving average (VMA) representation, $\Delta f^i = \Xi_i(L)\epsilon_t^i$, where L represents the lag operator. Then, the information share (IS) and component share (CS) to the i-th quote can be written as:

$$IS_i = \frac{\Xi_i^2 \Omega_i}{\Xi\Omega\Xi'}, \tag{6}$$

$$CS_i = \frac{\alpha_{\perp i}}{\alpha_{\perp i} + \beta_{\perp i}}. \tag{7}$$

The intuition of both metrics relies on efficient price. The efficient price is the true value of an asset and all the available information is immediately incorporated into it (Fama, 1970). The IS proposes that the variance of this unobservable efficient price is the source of price discovery. In this fashion, each quote's percentage of this variance is precisely its IS. On the other side, the CS decomposes this efficient price into 2 components: transitory (short-term and $I(0)$) and permanent (long-term and $I(1)$). The CS of a quote corresponds to its percentage of the permanent component from the efficient price. These measures will be applied to both the actual and the simulated series - for comparison with the wavelet approach.

Continuous Wavelet Transform. A wavelet is a function that can be compressed or stretched by its scale parameter a and moved along the timeline by its location or translation parameter τ (Daubechies, 1994). For a time series x_t, we can denote its CWT as $W_x(a,\tau)$, which can be written as:

$$W_x(a,\tau) = \int_{-\infty}^{\infty} x_t \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-\tau}{a} \right) dt, \tag{8}$$

where $\psi(t)$ stands for the mother wavelet and * denotes the complex conjugate.

To be considered a mother wavelet, a function must satisfy (Farge, 1992) two admissibility conditions : (i) it must have zero mean, $\int_{-\infty}^{\infty} \psi(t)dt = 0$; (ii) and unit energy, $\int_{-\infty}^{\infty} ||\psi(t)||^2 dt = 1$.

Naturally, as we are dealing with discrete time series, we are interested in the case where $t \in \{1, 2, ..., n\}$. Considering an uniform time step δt, the discretized version of (1) is (Torrence and Compo, 1998):

$$W_{n,x}(a) = \frac{\delta t}{\sqrt{s}} \sum_{t=0}^{n-1} x_t \psi^* \left(\frac{\delta t(t-n)}{a} \right). \tag{9}$$

As it can be seen in (8), the WT coefficients form a $m \times n$ matrix where m stands for the number of scales. In addition, considering that the data arrive at a unitary time interval, which yields $\delta_t = 1$, and that our smallest scale would precisely this unit, what gives us $a_0 = 1$, the j-th scale is defined as (Torrence and Compo, 1998):

$$a_j = a_0 2^{j\delta_s}, \delta_s = \log_2 \left(\frac{n\delta_t}{a_0} \right) m^{-1}. \tag{10}$$

Once our $\psi(t)$ are complex-valued, our $W_{n,x}$ may also be divided into a real part and an imaginary part. Let's denote them, respectively, as $\Re\{W_{n,x}(a)\}$ and $\Im\{W_{n,x}(a)\}$. Then, the power spectrum of $W_{n,x}(a)$ can be defined as $||W_{n,x}(a)||^2$, which shows how the energy of the signal x_t is distributed among the different frequencies and at different moments.

After understanding the idea behind the power spectrum, we can go further and present the cone of influence (COI). The COI, as in (Torrence and Compo, 1998), is the region in the WT power spectrum where its values are less affected by the edge effects - the consequence of approximating a finite signal with a truncated infinite integral. For a WT translated in such a way that t_0 is its center, the boundaries of COI are given by $|t - t_0| = g(a)$, where $g(.)$ is the e-folding time of the given mother wavelet. Naturally, we limit our conclusions to the area within the COI

Mother Wavelet Selection. As previously shown, a WT depends on the selected mother wavelet (wavelet function). In this sense, (Li et al., 2009) the selection of the mother wavelet interferes with the accuracy and efficiency of the transformation and this selection is case-dependent. Surveying the literature, most the applications of WTs use mother wavelets based on previous works. So far, we have not been able to find WTs' applications in economics or finance that test their parametrization.

Some metrics can be used to assess the fit of a wavelet function to the original signal: (i) signal-to-noise ratio (SNR) (Matz et al., 2009); (Agarwal et al., 2017); (ii) mean squared error (MSE) (Agarwal et al., 2017); (iii) Shannon entropy of wavelet coefficients (EWC) (Li et al., 2009). While the first two are usually employed to analyze discrete WTs, the EWC can be used for both CWTs and DWTs and is defined as:

$$EWC(a) = \sum_{i=1}^{n} \frac{w_{a,i}^2}{\left(\sum_{a=1}^{m} \sum_{i=1}^{n} w_{a,i}^2\right)} ln \left(\frac{w_{a,i}^2}{\sum_{a=1}^{m} \sum_{i=1}^{n} w_{a,i}^2)} \right). \tag{11}$$

In this framework (Li et al., 2009), the higher the energy concentrated at scale a, the lower the EWC. Therefore, for a fixed m, we select the wavelet function which achieves the minimum EWC compared to the others. For a m grid, we choose the pair $(m, \psi(t))$ that presents the global minimum EWC.

In our particular case, we define our m grid as $\{80, 81, ..., 100\}$ for $n > 80$ and $\{0.08n, 0.081n, ..., 0.1n\}$ otherwise. We test the main complex-valued wavelet functions: Morlet, $\psi_m(t)$, and Paul, $\psi_p(t)$. The reason for excluding real-valued functions is their lack of phase content, which makes them useful mainly for detecting jumps and discontinuities but not for interpreting lead-lag relationships. Their functional forms and respective e-folding time are:

$$\psi_m(t) = \pi^{-\frac{1}{4}} e^{i\omega t} e^{-\frac{t^2}{2}}, \qquad g_m(a) = \sqrt{2a}, \tag{12}$$

$$\psi_p(t) = \frac{2^{m_0} i^{m_0} m_0!(1-it)^{-(m_0+1)}}{\sqrt{\pi(2m_0)!}}, \qquad g_p(a) = \frac{a}{\sqrt{2}}. \tag{13}$$

where ω and m_0 are nondimensional time parameters set so that the admissibility conditions are sustained.

The intuition behind testing these two specific functions is also to understand what feature of the series carries the most significant information. This is accomplished considering that Morlet, with its sinusoidal formulation, gives us more frequency resolution, while Paul, considering its single peak, tends to highlight time resolution.

Cross Wavelet Transform The Cross Wavelet Transform (XWT) can be employed to assess how two signals interact in the time-frequency domain and to evaluate the lead-lag relationship between them (Torrence and Compo, 1998). Keeping our previous notation, we can denote an XWT between the signals x_t and y_t as $W_{n,xy}$, so $W_{n,xy} = W_{n,x}W^*_{n,y}$. Although this formulation (Torrence and Compo, 1998) has been extensively used, it leads to a bias that underestimates the small-scale (high-frequency) phenomena. For that reason, our XWT, and its consequent power spectrum, is estimated based on (Veleda et al., 2012), adding a normalization by scales that yields $W_{n,xy} = (W_{n,x}W^*_{n,y})2^{j/2}$

To help us weigh each scale's information about the relation between two series, we use the wavelet coherence defined below (Torrence and Compo, 1998). This measure can be understood as a time-frequency correlation, which indicates (normalized to spans over 0 and 1) the degree of similitude of the oscillatory patterns.

$$R_n(a) = \frac{|S(a^{-1}W_{n,xy}(a)|}{S(a^{-1}|W_{n,x}(a)|)^{1/2} S(a^{-1}|W_{n,y}(a)|)^{1/2}}. \tag{14}$$

Naturally, to $R_n(a)$ present some information about the signals, it must be the case that $\exists(x,y) S(x)S(y) \neq S(xy)$, otherwise $R_n(a) = 1$ for all a. To achieve it, S is a smoothing operator defined as convolution with a Gaussian in time and a rectangular window in scale (Aguiar-Conraria et al., 2008).

Finally, following (Torrence and Compo, 1998) and (Aguiar-Conraria and Soares, 2010) formulation of phase ϕ between two signals x_t and y_t to the scale a, we have:

$$\phi_{x,y}(a) = \tan^{-1}\left(\frac{\Im\{W_{n,xy}(a)\}}{\Re\{W_{n,xy}(a)\}}\right). \tag{15}$$

With $\phi_{x,y} \in [-\pi, \pi]$ and 4 possible cases: (i) if $\phi_{x,y} \in (0, \frac{\pi}{2})$, then the series move in phase, but y_t leads; (ii) if $\phi_{x,y} \in (-\frac{\pi}{2}, 0)$, then the series move in phase, but x_t leads; (iii) if $\phi_{x,y} \in (\frac{\pi}{2}, \pi)$, then they move in anti-phase and x_t leads; (iv) if $\phi_{x,y} \in (-\pi, -\frac{\pi}{2})$, then they move in anti-phase and y_t leads

Considering that our wavelet metric to lead-lag is a matrix which presents the variable anticipation for each frequency and each time step, we must define aggregation methods. We aggregate frequency instead of using it in bands as usual to comparability purposes. IS and CS metrics are not decomposed in components, nor point-wise, then, to compare them with a wavelet based metric, we must present it also in an interval and aggregate way.

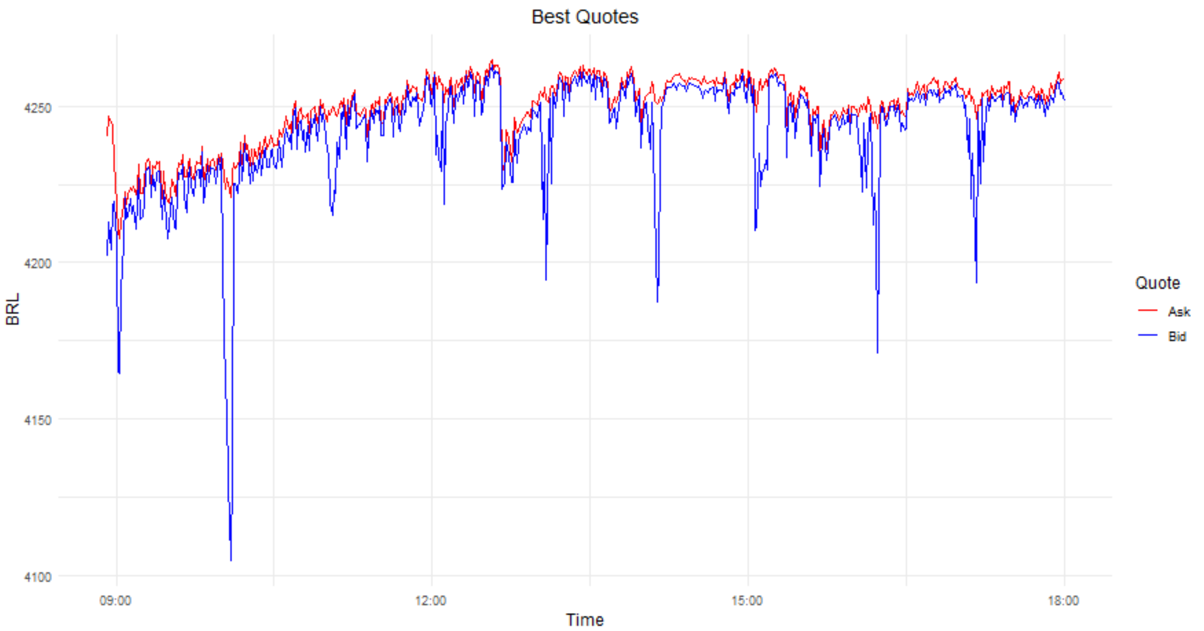
This paper introduces five ways to aggregate frequency content in our phase analysis: coherence weighing, linear discount and three quartile thresholds. The first one uses the wavelet coherence mean for each frequency. The second one attributes 1 to the highest frequency, 0 to the lowest frequency and constant step size to the frequencies in between. The last one uses a simple means to ponder the frequencies in each quartile (ordered from highest to lowest). The emphasis on high-frequency content is due to theoretical purposes. There is no reason to consider that a lead-lag relation would remain too long because of arbitrage opportunities.

3 Data and Calibration

Our high-frequency data comprise all the bid and ask quotes in the FX futures at the Brazilian Stock Exchange Market (B3) on November 27$^{\text{th}}$, 2019. The appendix presents some descriptive statistics of the 5,096,477 sell orders and 9,422,183 buy order prices, and

Figure 1 depicts the best bid and best ask behavior throughout the day. This same day was also analyzed in (Sung, 2021) and (Ferreira et al., 2025), but their focus was on actual trades rather than the order book itself.

Figure 1: Best Quotes



Source: Obtained from the B3 FTP when it was publicly available
Notes: This graph presents the best bid and ask at minute aggregation. The following algorithm calculates the best quotes: (1) order all quotes; (2) remove bids that are greater than the smallest ask; (3) to a fixed second, choose the greatest bid

The period is chosen for three main reasons: data availability, interpolation issues and previous works. The data availability is because B3's clearing house used to provide this database for free. The interpolation issues arise from the WT's necessity to deal with observations that are equally spaced in time. In this sense, some noise would be added due to the interpolation required to ensure temporal uniformity. Regarding previous evidence, (Sung, 2021) analyzed the 26th November trading session and found sell-side prevalence buy-side - although not analyzing the best bid and ask. Finally, (Ferreira et al., 2025) investigated 26th November and 27th November trading sessions and found asymmetric responses to FXI by side (sell or buy) and level of intermediary constraints. Both works used the same tick-by-tick dataset used in our paper.

At the time, the BRL faced significant volatility due to a combination of economic and political factors, as highlighted by (Sung, 2021) and (Ferreira et al., 2025). This included both domestic and international events that influenced exchange rates and futures prices.

Additionally, the trading session featured both pre-announced and unanticipated central bank interventions.

Table 1 displays cointegration statistics for the minute aggregated quotes. As can be seen, results corroborate **H1**, **H2** and **H3**.

Table 1: Engle-Granger Cointegration Test Results

Type	**Lag**	**EG Statistic**	**p-value**
No Trend	5	-5.012	0.01
Linear Trend	5	1.39	0.1
Quadratic Trend	5	0.294	0.1

Source: Own elaboration
Notes: Results of Engle-Granger cointegration test for the bid and ask quotes. A p-value = 0.01 means p-value ≤ 0.01 and a p-value = 0.1 means p-value ≥ 0.1

We use moments of the actual data to calibrate the distribution parameters in our simulated series. To do so, we used the first difference of both quotes as our proxy to the innovation process so that the observed variance and mean approximate the simulated one - considering the total variance of the innovation process, as explained below. As the data can be written as a VECM, we discarded the error-correction portion of the quote variation.

As a way to visualize how the phase analysis captures lead-lag relations, we use our model to simulate three scenarios: (**b** $\Rightarrow$ **a**) bid causes ask, (**a** $\Rightarrow$ **b**) ask causes bid and (**b** $\Leftrightarrow$ **a**) bid and ask have feedback effects. To do so, denote $\hat{\sigma_i}$ as the observed variance of the proxy for the innovation process of the i-th quote. Define the total variance of the innovation process as $\hat{\sigma} = \hat{\sigma_a} + \hat{\sigma_b}$. Therefore, as Table 2 displays, each causation scenario is based on the portion of the total variance that each quote conveys. Additionally, table 3 shows the fixed VAR parameters used to simulate the series.

Table 2: Innovation Process Parameters

Parameter/Scenario	$a \Rightarrow b$	$b \Rightarrow a$	$a \Leftrightarrow b$
σ_a	$0.75\hat{\sigma}$	$0.25\hat{\sigma}$	$0.5\hat{\sigma}$
σ_b	$0.25\hat{\sigma}$	$0.75\hat{\sigma}$	$0.5\hat{\sigma}$

Source: Own elaboration
Notes: Variance parameters of ϵ_a and ϵ_b as a function of the total variance of actual series

Table 3: Simulation Parameters

γ_a	γ_b	ρ_a	ρ_b	δ_a	δ_b
0	0	0.5	0.5	0.5	0.5

Source: Own elaboration
Notes: Simulation parameters based on equations (1) and (2)

As previously mentioned, in order to assess the causality direction by a conventional econometric approach, we use the IS (Hasbrouck, 1995) and CS (Booth et al., 1999). We follow (SCHWERT, 1989) in the lag definition as $\lceil 12 \times (\frac{n}{100})^{\frac{1}{4}} \rceil$.

4 Results

Simulated Series The main way to extract anticipation and, therefore, to unveil price discovery within the wavelet framework is through phase content. In this sense, we built our price discovery metrics by the previous definition of lead-lag intervals in the $\phi_{xy}(a)$. Denoting our frequency aggregations by ϕ_{xy}, we count the time points where $\phi_{xy} \in [0, \pi]$ and attribute them to ask, otherwise ($\phi_{xy} \in [-\pi, 0)$) to bid. The percentage of points to each quote corresponds to its price discovery value. The rationale is to create an interval measure, which can not only be understood in the same way as IS and CS, but also employed continuously, tracking the lead-lag dynamics at each time step.

Tables 4 and 5 present the results by minute and second aggregation, respectively.

As can be seen, IS and CS show considerable sensibility to sample size and internal inconsistency between each other, with IS overcoming CS accuracy. Analyzing the proposed frequency aggregation methods, we see that Coherence Weighing tends to be more reliable, being less sensible to sample size. The probability of falsely assuming an asymmetry in the price formation seems to diminish asymptotically for both VEC and wavelet metrics, especially for the former. In addition, one may note that the CS diverged considerably from the IS in the larger sample. Taking a look at the transition from $a \Rightarrow b$ to $a \Rightarrow b$ in Table 4, we observe, at minute aggregation, more dispersed errors between VEC and wavelet methods. The transition is smoother at the second aggregation, as shown in Table 5, and the difference between VEC and wavelet methods decreases.

Table 4: Price Discovery - simulated Series (by Minute)

Metrics	$a \Rightarrow b$		$b \Rightarrow a$		b $\Leftrightarrow$ a	
	Ask	Bid	Ask	Bid	Ask	Bid
Information Share	84.73%	15.26%	48.07%	51.93%	27.30%	72.70%
Component Share	57.56%	42.44%	62.68%	37.32%	40.49%	59.51%
Coherence Weighing	71.98%	28.02%	28.57%	71.43%	56.78%	43.22%
Linear discount	69.05%	30.95%	35.71%	64.29%	58.80%	41.20%
Quartile Threshold (25)	66.02%	34.98%	47.80%	52.20%	57.15%	42.85%
Quartile Threshold (50)	71.79%	28.21%	31.87%	68.13%	56.41%	43.59%
Quartile Threshold (75)	75.26%	24.73%	29.30%	70.70%	59.89%	40.11%

Source: Own elaboration
Notes: Results of price discovery metrics to simulated series by minute.

Table 5: Price Discovery - simulated Series (by Second)

Metrics	$a \Rightarrow b$		$b \Rightarrow a$		b $\Leftrightarrow$ a	
	Ask	Bid	Ask	Bid	Ask	Bid
Information Share	73.12%	26.87%	26.92%	73.08%	50.25%	49.75%
Component Share	48.71%	51.13%	51.24%	48.76%	50.41%	49.59%
Coherence Weighing	71.05%	28.95%	28.77%	71.23%	50.48%	49.52%
Linear discount	65.38%	34.62%	34.54%	65.56%	50.46%	49.54%
Quartile Threshold (25)	56.82%	43.18%	42.81%	57.19%	50.63%	49.37%
Quartile Threshold (50)	68.85%	31.15%	31.49%	68.51%	50.40%	49.60%
Quartile Threshold (75)	72.57%	27.43%	27.97%	72.03%	50.40%	49.60%

Source: Own elaboration
Notes: Results of price discovery metrics to simulated series by second.

Naturally, we must be conservative in extending this efficiency, given that we use only two observed moments in our simulation, and some more complex dynamics can be hidden in higher-order moments. Moreover, we must highlight that we are simulating a constant direction, i.e. no structural breaks in the innovation process.

Actual Series Table 6 presents the results of the empirical analysis of price discovery to the whole day using IS, CS and wavelet metrics. Coherence weighing was chosen to represent the phase-based price discovery, given its relative performance among the other aggregation alternatives. Apart from the huge difference between minute and second aggregation to the IS and CS results, there seems to be no robust evidence of persistent asymmetry for the given day. As the tables in the appendix show, the minute averaging affects the quotes' variance asymmetrically. Because of this, IS and CS penalize the high variance quote (bid), considering that the adjustment is mainly happening by it and that this quote follows the other.

From the wavelet point of view, considering the small asymmetry at the minute aggregation and posterior shift in the lead when taking into account the second one, it appears that if some asymmetry is assumed, it is probably marginal.

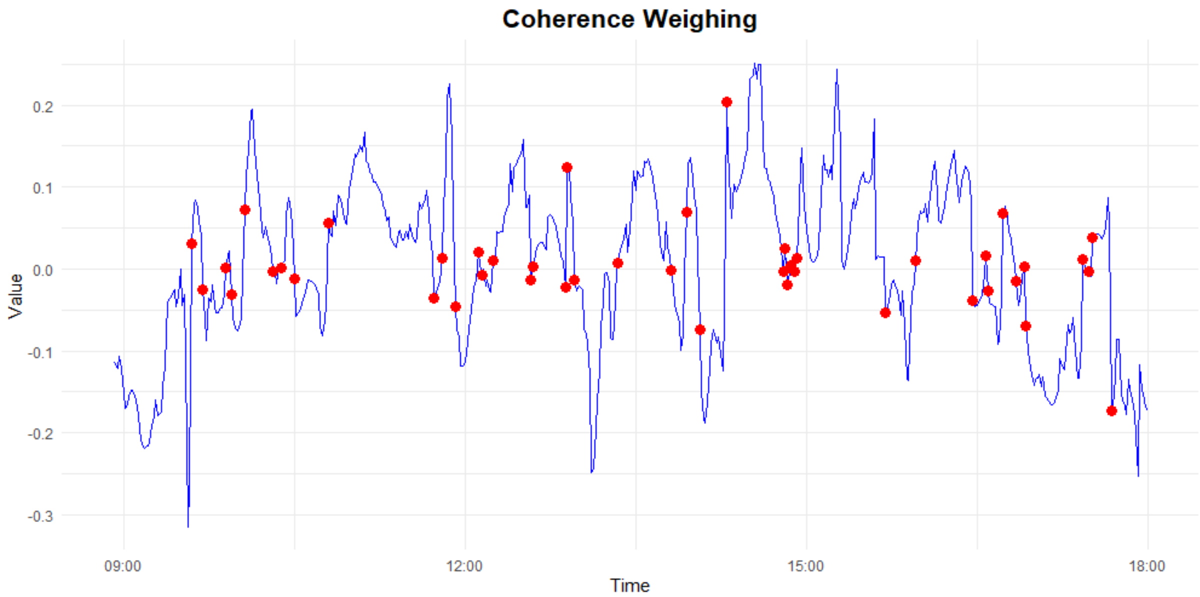
Table 6: Price Discovery - Actual Data

Metrics	Minute		Second	
	Ask	Bid	Ask	Bid
Information Share	50.25%	49.75%	93.72%	6.28%
Component Share	50.41%	49.59%	93.26%	6.74%
Coherence Weighing	55.68%	44.32%	49.4%	50.6%

Source: Own elaboration
Notes: Results of price discovery metrics to actual series to the whole day by minute and by second.

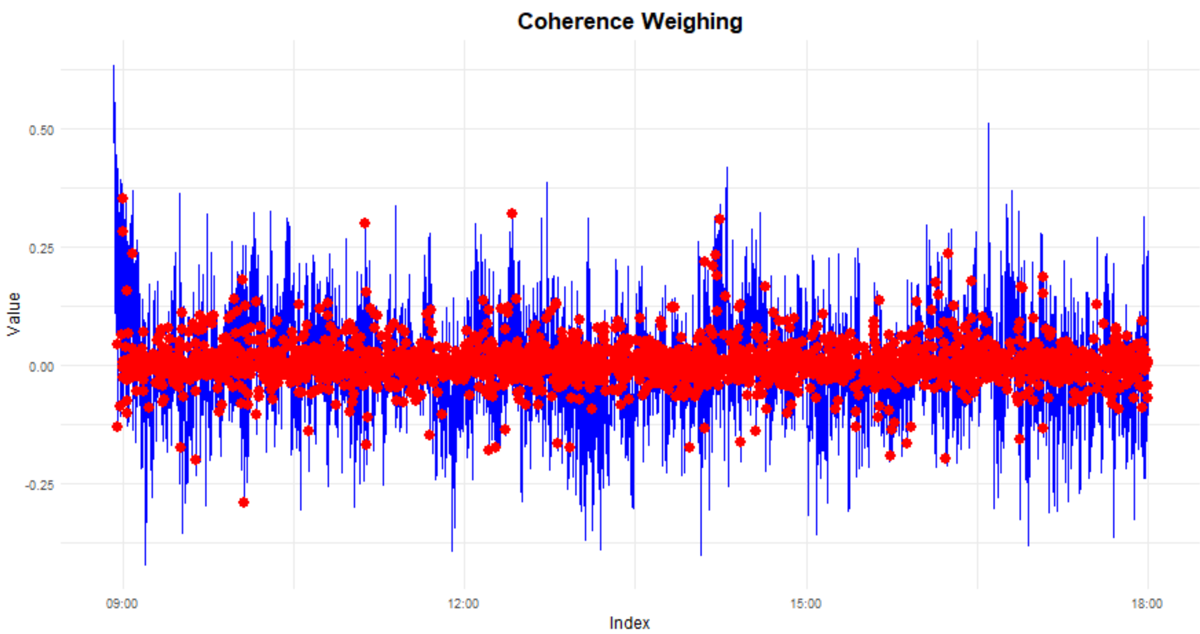
Figures 2 and 3 are beneficial to understand what happens. When we aggregate by minute, some long shifts appear (figure 2), but at a closer look (figure 3), we realize there were so many shifts that the minute aggregation filtered the most relevant. It happens because the shifts at second aggregation occur mostly within a minute, but the highest frequency at minute aggregation is one minute. Therefore, at the minute aggregation, the second dynamics is "diluted" into lower frequencies (minute), and a shift in the minute aggregation comes up as the original shifts in the minute frequencies plus the accumulated and diluted second frequencies[2].

Figure 2: Wavelet Price Discovery Metric (by Minute)



Source: Own elaboration
Notes: Results of Coherence Weighing to price discovery to the whole day by minute. Positive values indicate ask lead and negative values indicate bid lead. Red dots correspond to shift periods, when the lead changes from one quote side to another.

Figure 3: Wavelet Price Discovery Metric (by Second)



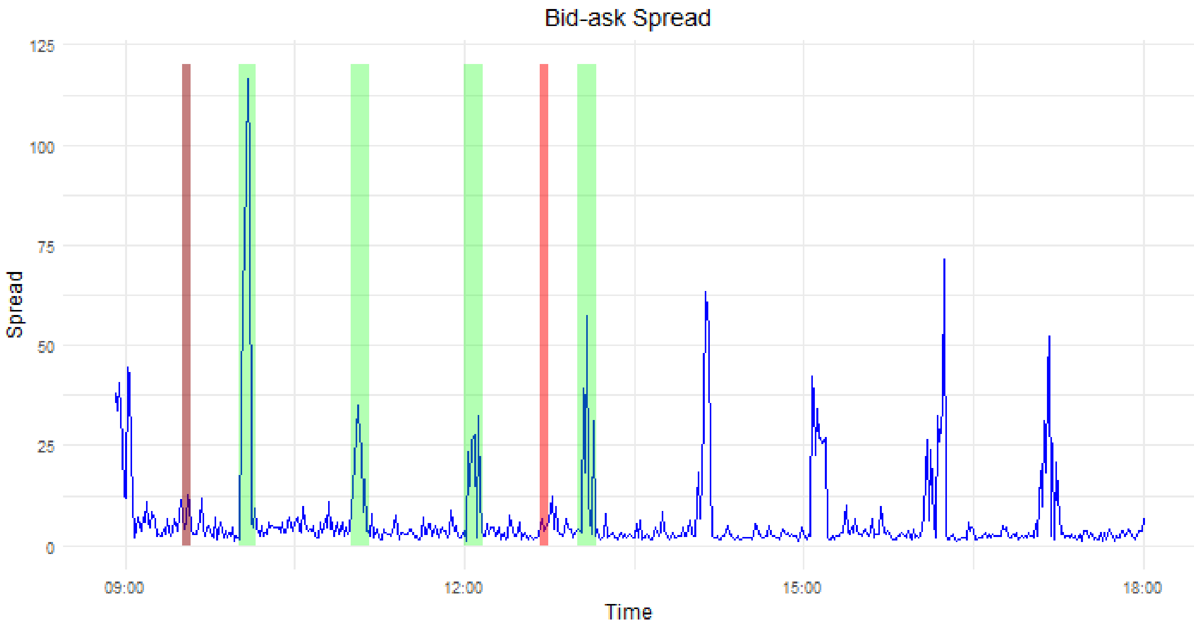
Source: Own elaboration
Notes: Results of Coherence Weighing to price discovery to the whole day by second. Positive values indicate ask lead and negative values indicate bid lead. Red dots correspond to shift periods when the lead changes from one quote side to another.

[2]For further details, see (Torrence and Compo, 1998)

As a matter of course, taking the whole day at once is not the only way to analyze it. It is not impossible that some noisy trading in the quotes may account for the disturbances in the price discovery metrics. For that reason, we proceed to partition our sample. Once the agents witness a trading session with high volatility and FXIs, it is natural to look at some events.

As Figure 4 illustrates, the FXIs (light red and red) and the PTAX survey windows (green) match moments with low liquidity (high bid-ask spread). In that sense, we divided our minute aggregated sample in the following way: (t_1) before the first FXI; (t_2) after the first and before the second FXI; (t_3) after the second FXI. Our second aggregated sample was divided into the 6-time windows of each event - 5 minutes to FXIs and 10 minutes to PTAX surveys : (p_1) first FXI; (p_2) first PTAX survey; (p_3) second PTAX survey; (p_4) third PTAX survey; (p_5) second FXI; and (p_6) fourth PTAX survey.

Figure 4: Bid-ask Spread (by Minute)



Source: Obtained from the B3 FTP when it was publicly available
Notes: The best quotes are calculated by the following algorithm: (1) order all quotes; (2) remove bids that are greater than the smallest ask; (3) to a fixed second, choose the greatest bid. The bid-ask spread is calculated as the best ask minus the best bid.
The shaded regions correspond to the anticipated FXI (dark red), PTAX survey windows (green) and unanticipated FXI (red).

Table 7 presents results to the minute aggregated partitioned samples. For convenience, we show the price discovery only to the ask, given that to IS, CS and wavelet metric sum to one. As we see, the t_3, which accounts for more than half of the trading session, was leaded by the sell-side, as unanimously pointed by the metrics. To the two other periods, the mixed results do not allow any clear dominance.

Table 7: Price Discovery - Partitioned Samples (by Minute)

Metric	Before 1º FXI (t_1)	Between FXIs (t_2)	After 2º FXI (t_3)
Coherence Weighing	22.86%	33.15 %	50.80 %
Information Share	98.73 %	91.51%	97.36 %
Component Share	38.20 %	85.68%	94.69 %

Source: Own elaboration
Notes: Results of price discovery metrics (only ask) to actual series to the whole day by minute and by second. Where t_1 is before the first FXI; t_2 after the first and before the second FXI; and t_3 after the second FXI

Table 8 disposes results to the second aggregated partitioned samples. At first glance, we notice some predominance of the sell-side throughout these illiquid moments. Considering only unanimous results, the first and third PTAX survey windows and the unanticipated FXI was characterized by sellers leadership. Figure 5 discloses the partitioned samples' informational dynamics based on the wavelet metric. For better imaging, the first column displays only the FXIs (p_1 and p_5) and columns 2 and 3 show the PTAX survey windows.

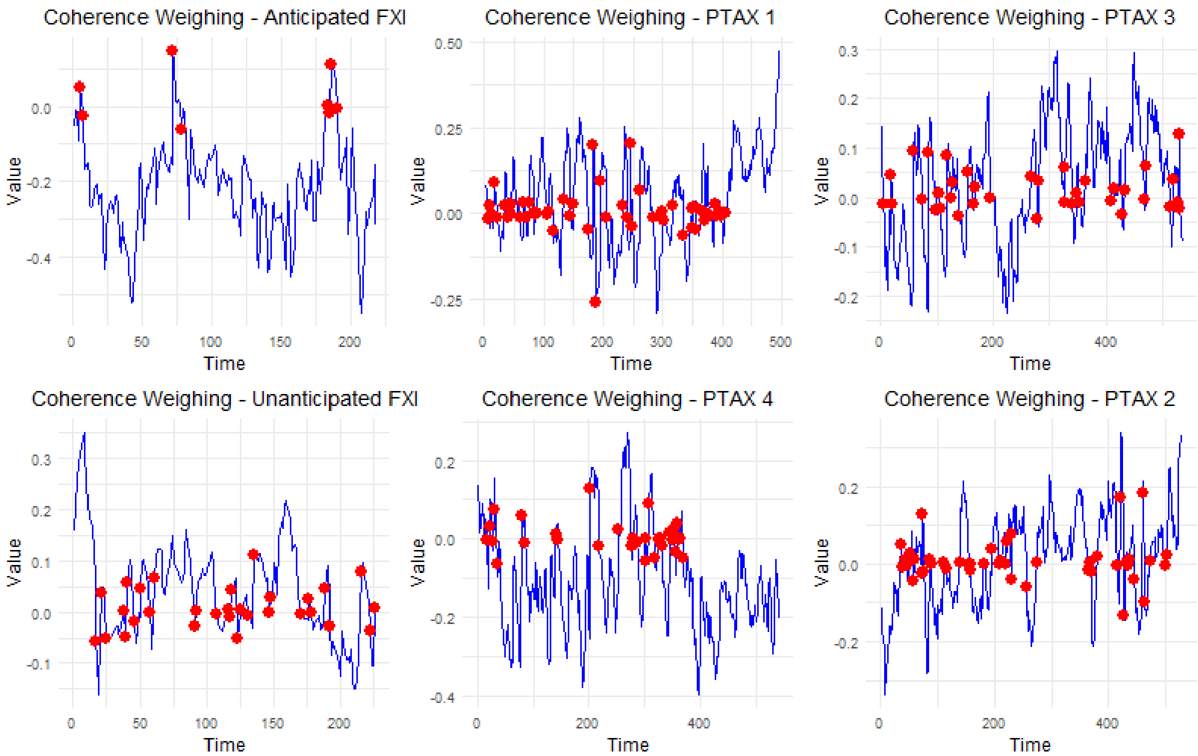
Table 8: Price Discovery - Partitioned Samples (by Second)

Metric	1º FXI (p_1)	1º PTAX (p_2)	2º PTAX (p_3)	3º PTAX (p_4)	2º FXI (p_5)	4º PTAX (p_4)
Coherence Weighing	5.96%	60.89 %	59.47 %	59.55%	60.35 %	21.44 %
Information Share	60.43%	51.17 %	00.25 %	99.25%	76.42 %	2.29 %
Component Share	63.89%	65.47 %	79.58 %	98.48%	77.06 %	82.99 %

Source: Own elaboration
Notes: Results of price discovery metrics (only ask)to actual series to the whole day by minute and by second. Where p_1 is the first FXI; p_2 first PTAX survey; p_3 second PTAX survey; p_4 third PTAX survey; p_5 second FXI; and p_6 fourth PTAX survey

A comparison between table 8 and table 11 (appendix) can clarify the variance of quotes' impact on IS and CS. As we can see, they penalize the high variance quote.

Figure 5: Wavelet Price Discovery (by Second)



Source: Own elaboration
Notes: Results of Coherence Weighing to price discovery to specific moments by second: (p_1) first FXI; (p_2) first PTAX survey; (p_3) second PTAX survey; (p_4) third PTAX survey; (p_5) second FXI; and (p_6) fourth PTAX survey. Positive values indicate ask lead and negative values indicate bid lead. Red dots correspond to shift periods, when the lead changes from one quote side to another.

Looking particularly at p_1 and p_3, we can see that there were way more bid quotes (considering the entire book) than ask quotes. For that reason, we focus our discussions on the coherence weighing measure.

The wavelet metric suggests that the anticipated FXI (p_1) was leaded by the buy-side with few shifts where the ask dominated. The second (unanticipated) FXI (p_5) was mostly dominated by the sell-side with more shifts and less persistent leads by both sides. The first (p_2) and last (p_6) PTAX survey windows showed a similar dynamics with opposite directions: a persistent lead by the buy-side at p_2 at the end of the period and the same to the sell-side at the end of p_6. Finally, p_3 and p_4, both leaded by the ask, showed a similar feature of a nearly persistent lead in the middle of the survey window.

5 Final Remarks

Our study investigates price discovery in the Brazilian FX futures market, focusing on the role of asymmetric information in shaping leadership dynamics between bid and ask quotes. The dataset includes over 14 million observations from a volatile trading session marked by central bank interventions on November 27th, 2019. Previous studies using the same dataset, such as (Sung, 2021) and (Ferreira et al., 2025), examined trading behavior and the effects of FX interventions but did not directly address the dynamics between the bid and ask quotes. These studies found evidence of sell-side prevalence and asymmetric responses to FX interventions, motivating our deeper investigation of quote-level interactions.

Our main finding is that price discovery does not consistently favor bid or ask. Instead, leadership alternates frequently, reflecting a diffuse process where both informed and uninformed traders influence price formation intermittently. This dynamic interaction challenges the notion of a persistent informational advantage on either side of the order book.

Leadership alternates rapidly during the entire trading session, particularly when analyzed at the second level, highlighting that minute aggregation may obscure important short-term shifts. Low-liquidity events, such as FX interventions (FXIs) and PTAX survey windows, reveal distinct leadership patterns: the anticipated FXI shows buy-side dominance, while the unanticipated FXI exhibits balanced shifts. PTAX windows show stable leadership toward the end, while intermediate windows display more consistent mid-session dominance.

These results suggest that asymmetric information is not consistently concentrated on one quoting side; instead, it fluctuates as market conditions evolve, with informed traders appearing on both sides of the order book. Both bid and ask sides may momentarily lead in response to evolving information, particularly during liquidity shocks or intervention periods. Future research should explore more trading sessions to assess the robustness of these patterns.

References

Agarwal, S., Singh, O., and Nagaria, D. (2017). Analysis and comparison of wavelet transforms for denoising mri images. *Biomed Pharmacol J*, 10(2):831–836.

Aguiar-Conraria, L., Azevedo, N., and Soares, M. J. (2008). Using wavelets to decompose the time–frequency effects of monetary policy. *Physica A: Statistical Mechanics and its Applications*, 387(12):2863–2878.

Aguiar-Conraria, L. and Soares, M. J. (2010). Oil and the macroeconomy: Using wavelets to analyze old issues. *Empirical Economics*, 40:645–655.

Baillie, R. T., Geoffrey Booth, G., Tse, Y., and Zabotina, T. (2002). Price discovery and common factor models. *Journal of Financial Markets*, 5(3):309–321.

Booth, G. G., So, R. W., and Tse, Y. (1999). Price discovery in the german equity index derivatives markets. *Journal of Futures Markets*, 19(6):619–643.

Cabrera, J., Wang, T., and Yang, J. (2009). Do futures lead price discovery in electronic foreign exchange markets? *Journal of Futures Markets*, 29(2):137–156.

Chakravarty, S., Gulen, H., and Mayhew, S. (2004). Informed trading in stock and option markets. *The Journal of Finance*, 59(3):1235–1257.

Daubechies, I. (1994). *Ten Lectures on Wavelets*. Springer-Verlag.

De Boyrie, M., Pavlova, I., and A.M., P. (2012). Price discovery in currency markets: Evidence from three emerging markets. *International Journal of Economics and Finance*, 4(12):61–75.

de Jong, F. (2002). Measures of contributions to price discovery: a comparison. *Journal of Financial Markets*, 5(3):323–327.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Farge, M. (1992). Wavelet transforms and their applications to turbulence. *Annual Review of Fluid Mechanics*, 24:395–458.

Ferreira, A., Mullen, R., Ricco, G., Viswanath-Natraj, G., and Wang, Z. (2025). Foreign exchange interventions and intermediary constraints. *Working Paper*.

Figuerola-Ferretti, I. and Gonzalo, J. (2010). Modelling and measuring price discovery in commodity markets. *Journal of Econometrics*, 158(1):95–107. Twenty Years of Cointegration.

Garbade, K. D. and Silber, W. L. (1983). Price movements and price discovery in futures and cash markets. *The Review of Economics and Statistics*, 65(2):289–297.

Garcia, M., Medeiros, M., and Santos, F. (2015). Price discovery in brazilian fx markets. *Brazilian Review of Econometrics*, 35(1):65–94.

Ghosh, I. and Chaudhuri, T. D. (2017). Fractal investigation and maximal overlap discrete wavelet transformation (modwt)-based machine learning framework for forecasting exchange rates. *Studies in Microeconomics*, 5(2):105–131.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, 50(4):1175–1199.

Joseph, A., Sisodia, G., and Tiwari, A. K. (2015). The inter-temporal causal nexus between indian commodity futures and spot prices: A wavelet analysis. *Theoretical Economics Letters*, 5(2):312–324.

Kumar, S. (2018). Price discovery in emerging currency markets. *Research in International Business and Finance*, 46:528–536.

Kumar, S., Pathak, R., Tiwari, A. K., and Yoon, S.-M. (2016). Are exchange rates interdependent? evidence using wavelet analysis. *Applied Economics*, 49(33):3231–3245.

Laurini, M. P., Furlani, L. G. C., and Portugal, M. S. (2008). Empirical market microstructure: An analysis of the brl/us exchange rate market. *Emerging Markets Review*, 9(4):247–265.

Li, F., Meng, G., Kageyama, K., Su, Z., and Ye, L. (2009). Optimal mother wavelet selection for lamb wave analyses. *Journal of Intelligent Material Systems and Structures*, 20(10):1147–1161.

Matz, V., Smid, R., Starman, S., and Kreidl, M. (2009). Signal-to-noise ratio enhancement based on wavelet filtering in ultrasonic testing. *Ultrasonics*, 49(8):752–759.

Nigatu, G. and Adjemian, M. (2020). A wavelet analysis of price integration in major agricultural markets. *Journal of Agricultural and Applied Economics*, 52(1):117–134.

Osler, C. L., Mende, A., and Menkhoff, L. (2011). Price discovery in currency markets. *Journal of International Money and Finance*, 30(8):1696–1718.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4):1127–1139.

SCHWERT, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, 44(5):1115–1153.

Sung, G. (2021). The effects of the central bank intervention auctions on the foreign exchange futures market: evidence from brazil. Dissertação de mestrado, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil.

Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

Veleda, D., Montagne, R., and Araújo, M. (2012). Cross-wavelet bias corrected by normalizing scales. *Journal of Atmospheric and Oceanic Technology*, 22(9):1401–1408.

A Additional tables and figures

Table 9: Descriptive statistics of quotes

	Min	Q1	Median	Mean	Q3	Max	Var	Skew	Kurtosis
All Quotes - Tick-by-tick									
Buy	3994	4229	4238	4220	4250	4335	5013	-2.77	6.03
Sell	4078	4261	4277	4271	4282	4500	197	-0.917	1.05
Best Quotes - Minute									
Buy	4105	4235	4248	4242	4254	4263	333	-2.76	12.5
Sell	4207	4245	4252	4249	4257	4265	117	-1.22	0.879
Best Quotes - Second									
Buy	3994	4238	4252	4242	4260	4278	1375	-5.32	32.3
Sell	4178	4241	4254	4250	4260	4301	186	-0.908	0.804

Source: Obtained from the B3 FTP when it was publicly available
Notes: The best quotes are calculated by the following algorithm: (1) order all quotes; (2) remove bids that are greater than the smallest ask; (3) to a fixed second, choose the greatest bid

Table 10: Descriptive statistics of quotes by sub-samples

		Obs	Min	Q1	Median	Mean	Q3	Max	Var	Skew	Kurtosis
All Quotes - Tick-by-tick											
t_1	Buy	1,719,759	3994	4221	4228	4209	4230	4300	3998	-3.0553	7.4909
	Sell	229,643	4100	4238	4242	4244	4258	4444	147	0.2644	3.1045
t_2	Buy	2,597,892	3994	4233	4246	4224	4257	4328	5406	-2.6879	5.6041
	Sell	2,519,390	4172	4262	4271	4271	4282	4500	177	-0.6575	0.6193
t_3	Buy	4,847,760	3994	4236	4241	4221	4251	4335	5360	-2.7271	5.6329
	Sell	2,264,216	4079	4265	4279	4275	4283	4500	121	-1.0525	2.5370
Best Quotes - Minute											
t_1	Buy	35	4164	4212	4218	4215	4223	4230	233	-1.8886	3.7130
	Sell	35	4207	4222	4225	4227	4232	4247	81	0.2898	-0.0701
t_2	Buy	184	4105	4231	4243	4239	4250	4263	442	-3.4260	16.0799
	Sell	184	4221	4237	4248	4246	4254	4265	111	-0.3569	-0.9462
t_3	Buy	315	4171	4245	4252	4248	4255	4261	151	-2.6703	9.3463
	Sell	315	4234	4251	4255	4254	4258	4263	26	-0.7731	0.5326

Source: Obtained from the B3 FTP when it was publicly available
Notes: The best quotes are calculated by the following algorithm: (1) order all quotes; (2) remove bids that are greater than the smallest ask; (3) to a fixed second, choose the greatest bid

Table 11: Descriptive statistics of quotes by events

		Obs	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**	**Var**	**Skew**	**Kurtosis**
All Quotes - Tick-by-tick											
1º FXI	Buy	231,491	4011	4224	4227	4226	4228	4292	48	0.0321	24.4122
	Sell	38,929	4185	4237	4239	4241	4243	4400	108	0.4834	5.7574
1º PTAX	Buy	248,901	3994	3994	4222	4131	4239	4299	14,241	-0.2656	-1.9075
	Sell	124,516	4192	4248	4265	4258	4267	4500	124	-0.6815	3.9289
2º PTAX	Buy	151,671	3994	3994	4240	4126	4259	4313	17,053	-0.0225	-1.9919
	Sell	186,287	4210	4262	4281	4274	4282	4500	96	-0.7014	2.8250
3º PTAX	Buy	158,301	3994	3994	4255	4172	4265	4318	15,636	-0.7172	-1.4728
	Sell	167,387	4212	4269	4285	4280	4286	4443	82	-1.0750	1.7041
2º FXI	Buy	22,931	4050	4249	4252	4253	4257	4308	129	-0.9076	18.0567
	Sell	41,301	4079	4253	4259	4256	4260	4375	81	-0.3619	12.3057
4º PTAX	Buy	215,942	3994	3994	4242	4179	4252	4335	13,059	-0.9844	-1.0053
	Sell	98,113	4208	4264	4283	4276	4284	4365	105	-0.7431	0.0237
Best Quotes - Second											
1º FXI	Buy	218	4011	4213	4217	4215	4221	4238	637	-5.2150	37.5560
	Sell	218	4194	4217	4222	4223	4223	4239	68	0.5817	-0.0668
1º PTAX	Buy	496	3994	4186	4223	4171	4237	4248	10,146	-1.1491	-0.6131
	Sell	496	4192	4223	4227	4228	4239	4249	140	-0.8001	1.1626
2º PTAX	Buy	528	3994	4240	4242	4231	4260	4262	3,760	-3.4184	10.2843
	Sell	528	4215	4243	4246	4249	4260	4262	102	-0.6251	1.0039
3º PTAX	Buy	534	3994	4245	4260	4241	4264	4267	3,515	-3.7727	12.8584
	Sell	534	4219	4249	4261	4256	4265	4268	89	-0.7075	0.0528
2º FXI	Buy	227	4152	4228	4232	4232	4246	4252	253	-1.6142	4.9712
	Sell	227	4200	4231	4236	4237	4247	4253	100	-0.4144	0.9865
4º PTAX	Buy	541	3994	4243	4254	4235	4262	4264	4,278	-3.2967	9.2895
	Sell	541	4217	4247	4258	4254	4263	4266	94	-0.6968	0.2683

Source: Obtained from the B3 FTP when it was publicly available
Notes: The best quotes are calculated by the following algorithm: (1) order all quotes; (2) remove bids that are greater than the smallest ask; (3) to a fixed second, choose the greatest bid