Pacific-Basin Finance Journal 93 (2025) 102652

Contents lists available at ScienceDirect

Pacific-Basin Finance Journal

journal homepage: www.elsevier.com/locate/pacfin

To fix or not to fix, the Fix: Reassessing the effectiveness of the 4 pm Fix. A pre-registered study

Matteo Benenchia [a,b,c], Luca Galati [d,e,*], Andrew Lepone [b,c]

[a] Department of Economics and Management, University of Trento, Via Calepina 14, Trento, TN, 38122, Italy
[b] Department of Applied Finance, Macquarie University, Balaclava Rd, Macquarie Park, 2109, NSW, Australia
[c] Rozetta Institute (formerly CMCRC-SIRCA), Level 9/1 Macquarie Pl, Sydney, NSW, 2000, Australia
[d] Department of Management, University of Bologna, Via Capo di Lucca 34, Bologna, 40124, BO, Italy
[e] School of Business, University of Wollongong, Northfield Ave, Wollongong, 2500, NSW, Australia

ARTICLE INFO

Dataset link: 10.17632/zbdgtt8mmt.1

JEL classification:
G14
G15

Keywords:
FX benchmark rates
Financial regulation
Market microstructure
WM/Refinitiv
4pm window

ABSTRACT

This study critically re-evaluates the effectiveness of the WM/Refinitiv (WMR) methodology for Foreign Exchange (FX) benchmark rates following its revision in 2015. Through the use of proprietary high-frequency trading data provided by Refinitiv, the research assesses the representativeness, attainability, and robustness of the WMR 4 pm fix. The study examines the benchmark's ability to accurately reflect market conditions and the implications of potential methodological enhancements, such as extending the calculation window. Findings indicate that while limited improvements in robustness can be achieved with longer windows, the current 5-min window remains broadly effective. However, the increased complexity and cost of changing the length of the window could outweigh the benefits. Additional results underscore the need for ongoing adaptation to evolving market dynamics and provide critical insights for financial institutions, regulators, and central banks in maintaining reliable and manipulation-resistant benchmarks.

1. Introduction

In the dynamically complex Foreign Exchange (FX) market, which operates continuously around the clock and is characterized by its high fragmentation and bilateral trading, benchmarks play a pivotal role. They reduce information asymmetry and are essential for the accurate valuation of portfolios and investments. The FX market's distinct lack of a closing time intensifies the need for a dependable daily benchmark to ensure market participants can access a representative rate each day. The WM/Refinitiv (WMR) 4 pm fix, along with other benchmarks like the ECB fix and the BFIX administered by Bloomberg, are foundational to financial activities spanning continents and currencies. However, the WMR 4 pm fix (or simply the Fix), as highlighted by the Financial Stability Board, remains the dominant benchmark, critical not just in FX but as a key component in multi-currency equity, bond, and credit indices. This underscores its expansive influence and the substantial financial implications tied to its accuracy and integrity.

In early 2023, $85 billion in trading volume[1] in just one currency pair (GBP/USD) underscored the substantial scale of these benchmarks, with billions traded within mere minutes. Such volumes highlight the critical nature of the benchmark window and the vast sums that depend on this brief yet intensely scrutinized period.

* Corresponding author at: Department of Management, University of Bologna, Via Capo di Lucca 34, Bologna, 40124, BO, Italy.
E-mail address: luca.galati@unibo.it (L. Galati).

[1] Self-produced calculations based on proprietary data provided by Refinitiv.

https://doi.org/10.1016/j.pacfin.2024.102652
Received 13 September 2024; Received in revised form 17 December 2024; Accepted 18 December 2024
Available online 25 December 2024

The importance of FX benchmarks is paralleled by their vulnerability. The 2013 scandal exposed severe manipulation by traders, who exploited the benchmark's calculation mechanisms, often at the expense of their clients.[2] This revelation led to a comprehensive review by the Financial Stability Board, culminating in a reformed methodology introduced in 2015. The adjustments included extending the benchmark calculation window[3] and establishing a dedicated surveillance team to monitor and prevent potential manipulation. These changes aimed to fortify the benchmark against malpractices and restore trust among global market participants.

Despite these improvements, the continuous evolution of the FX market presents ongoing challenges. The methods of trading, the instruments used, and the very structure of the market continue to transform, prompting a perpetual reassessment of the benchmark's methodology to ensure it remains representative of the market it serves.

The methodology for calculating the WMR benchmark, which is detailed and updated by Refinitiv, serves as a critical mechanism for ensuring market representativeness. It involves sampling transaction data from prominent trading platforms within a defined 5-min window, applying checks to each data point before determining the daily benchmark rate. This method is designed to capture a snapshot of market activity that reflects market conditions as accurately as possible. However, as the market evolves, so too must the methodologies that underpin these critical benchmarks. This study, utilizing a unique dataset provided by Refinitiv, aims to critically assess the current WMR methodology's effectiveness and explore potential enhancements to ensure its ongoing relevance and reliability.

The "Engagement & Impact" pre-registered report previously published (Benenchia et al., 2024) outlines the foundational research questions underpinning the current study. This study addresses the following research questions:

1. *How has the market impacted the 4 pm Fix in terms of liquidity, trading activity, transaction and adverse selection costs, and price volatility?*[4]
2. *To what extent has the methodology that underpins the WMR benchmark remained effective since its inception in terms of representativeness of the market, attainability, and robustness?*
3. *Can the WMR methodology be improved by the lengthening of the time window in which the benchmark is calculated?*

Answering these questions will enable global corporations and fund managers to rely on a more resilient and non-manipulated reference rate, thereby providing a benchmark of better quality against currency and macroeconomic risks.

As part of the pre-registration publication process (Faff, 2023), this "Engagement & Impact" study is structured as follows. Section 2 outlines the connection with the initial research proposal (Benenchia et al., 2024). Section 3 details the results of the applied analysis. The remaining Sections 4 and 5 discuss and conclude the study.

2. Pre-registered report articulation

2.1. Background and motivation

Building upon the initially approved research pitch and report pertaining to Phase 3 of the PBFJ pre-registration publication process (Benenchia et al., 2024), this section outlines the relevant literature and the study's motivations.

As outlined in Benenchia et al. (2024), the works of Evans et al. (2018), Melvin and Prins (2015) and Evans (2018) critically underpin this study. Evans (2018) emphasize the centrality of the WMR benchmark in the FX market, highlighting atypical trading activity surrounding 4 pm UK Time. Melvin and Prins (2015) motivates our research to address inefficiencies in the 4 pm Fix. Evans et al. (2018) focus on the shift of the Fix benchmark from 1-min to 5-min, and while our sample does not include such an event, we follow their empirical approach. By extending the data period from the WMR fix implementation in 2015 to the present, we examine changes in trading behavior around the 4 pm Fix, considering recent economic events at the epicenter of ongoing debates within the FX industry.

Evans et al. (2018) provide a framework for understanding the effectiveness of the 4 pm Fix benchmark, focusing on representativeness, attainability, and robustness. While their analysis controls for month-end trading, it does not thoroughly investigate trading patterns during this period, which is crucial for rebalancing portfolios. Melvin and Prins (2015) reveals significant trading activity spikes during month-end, potentially affecting benchmark rates. This behavior underscores the importance of vigilant benchmark management and robust methodologies to ensure the integrity of FX benchmarks. This study builds on Melvin and Prins (2015) to focus on month-end analysis.

This study is crucial for financial institutions, regulators, and central banks. Financial institutions rely on WMR benchmarks for portfolio rebalancing, and an efficient benchmark minimizes currency risks and transaction costs. Regulators need up-to-date evidence to develop optimal regulations, and central banks require a clear view of FX markets for monetary policy transmission and setting reference rates. Therefore, this research aims to provide new insights into market behavior changes, informing future FX benchmark methodology design. The findings will help financial institutions reduce costs, regulators prevent manipulation, and central banks maintain fair and representative rates.

2 See, e.g., Financial Conduct Authority (FCA), 2014. FCA fines five banks £1.1 billion for FX failings and announces industry-wide remediation programme (Press Release).

3 Media news still argues against the effectiveness of a 5-min benchmark window calculation compared to a longer 20-min window. See, e.g., The Full FX at https://thefullfx.com/.

4 This research question slightly differs from that of the pre-registered report as we realized that the data/tools available only appropriately address the question stated above.

2.2. Methodology

To address the *first* research question, we employ a multivariate panel regression model examining liquidity measures to compare these metrics on normal trading days versus month-end days.[5] These include: (i) bid–ask spread — calculated as the difference between the ask and the bid prices; (ii) effective spread — computed as two times the difference between the transaction price and the prevailing quote midpoint; (iii) price impact — measured as two times the difference between the prevailing midpoint 20 s after the trade and the quote midpoint, all being consistent with mainstream microstructure literature (e.g., Galati et al., 2024). This allows us to understand changes in trading behavior and inform benchmark design improvements.

Initially, we intended to use the model specified in Equation 9 of the pre-registered report (Benenchia et al., 2024) to evaluate the predictive power of changes in market behavior ahead of the benchmark calculation. However, as predicted, analysis reveals multicollinearity issues among the same market quality metrics computed in different time horizons. Therefore, we use the simplified model in Equation 10 of the pre-registered report (Benenchia et al., 2024), which captures the predictive power of market changes in each time interval surrounding the benchmark window separately,[6] as follows:

$$Y_{id}^{t} = \alpha^{t} + \beta_{1}^{t} X_{id}^{t} + \beta_{2}^{t} K_{d}^{t} + \beta_{3}^{t} K_{d}^{t} X_{id}^{t} + \delta^{t} C_{id}^{t} + \gamma^{t} FE_{id}^{t} + \epsilon_{id}^{t} \tag{1}$$

where t represents each 5-min interval considered for the calculation of market quality measures, from 30 min before the start of the benchmark window to 4 pm UK time.[7] Y_{id}^{t} represents the natural logarithm of the actual published 4 pm Fix benchmark rate for the currency pair i in a given day d. X_{id} is a vector of the three market quality metrics (i.e., bid–ask spread, effective spread, and price impact).[8] K_d is a dummy variable taking the value of 1 when the observation refers to the last trading weekday of the month and 0 otherwise. Our variables of interest are the interaction terms between K_d and X_{id} at each time 30-min before the window through to 5 min around 4 pm UK time, capturing the marginal predictive power of the benchmark from changes in market behavior at month-end. C_{id} is the vector of control variables for currency i on day d (i.e., currency pairs volatility measured as the standard deviation of prices within the fixing window, an economic event dummy variable taking the value of 1 when there is a macroeconomic announcement for the underlying US Dollar or for the currency in question during day d and 0 otherwise,[9] and the order-to-trade ratio calculated as the total number of limit orders divided by that of the market orders within the benchmark window), and FE_{id} captures the currency and time fixed effects. The remaining ϵ_{id} is the error term while α is the intercept.

To address the *second* and *third* research questions, we use a methodology similar to Evans et al. (2018), examining the 4 pm Fix benchmark's quality through measures of representativeness, attainability, and robustness. Representativeness evaluates the benchmark's reliability in reflecting the underlying market, attainability assesses the degree to which market participants can replicate the benchmark price, and robustness measures the benchmark's resilience to market manipulation.

As outlined in Benenchia et al. (2024), for representativeness, we calculate the root mean square error between the fix rates and the mean trade prices for different time windows (5, 10, and 20 min):

$$Representativeness_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - p_{id}^{d} \right)^{2}} \tag{2}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows,[10] and p_{id}^{d} is the mean trade price for currency i on day d.

Attainability and robustness are computed similarly, with attainability comparing the fix rates to the mean prices within each benchmark window:

$$Attainability_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - p_{id}^{5,10,20} \right)^{2}} \tag{3}$$

where $p_{id}^{5,10,20}$ is the mean price within each time window.

Robustness assesses the difference between the fix rates and dirty simulated benchmarks that exclude Refinitiv's validation checks for outliers:

$$Robustness_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - \tilde{b}_{id}^{5,10,20} \right)^{2}} \tag{4}$$

---

[5] We did not perform the event study analysis similar to Aspris et al. (2020) as planned in the pre-registered report because the multivariate analysis performed permits a better understanding of how changes in market behavior cause changes in the 4 pm fixes through a more sophisticated approach. This analysis is also novel compared to the literature. Additionally, our setting is less appropriate for an event study analysis given the absence of major events or a change that can be used as exogenous shock.

[6] The only exception are the measures at time 0, which are contemporaneous to the dependent variables.

[7] Each time from −30 to 0 refers to a centered range of 5 min (e.g., −30 indicates the range of minutes from −32.5 to −27.5 relative to 4 pm UK time).

[8] To avoid multicollinearity issues, we excluded the realized spread measure. For the same reason, we ran the same set of regressions with the relative spread measure instead of the raw bid–ask spread and obtained qualitatively similar results in untabulated tests.

[9] We calculate the *economic_event* dummy variable that incorporates important news but excludes announcements released before midday (i.e., 12 pm UK time), being news unlikely to significantly influence the benchmark rates. The results for the untabulated tests using this revised dummy variable are quantitatively similar to those presented in the paper, with identical conclusions reached.

[10] As we replicated a potential benchmark based on a 10- and 20-min window, we decide not to use the actual published benchmark when considering the 5-min window for consistency and computed b_{id}^{5} in an identical way as per b_{id}^{10} and b_{id}^{20}. The benchmark replication cannot be disclosed in the code as per the legal agreement with Refinitiv.

where $\tilde{b}_{id}^{5,10,20}$ are the simulated benchmarks without outlier exclusion.

To test these measures, we conduct t-tests to determine if they are significantly different from zero. We also compare the actual Fix measures over a 5-min window with the extended windows (10 and 20 min) to examine variability and infer the effectiveness of a longer benchmark window. We finally follow Benenchia et al. (2024) and also test the difference between the actual benchmark effectiveness measures and the same measures computed for the simulated benchmark on the extended windows (e.g., Re_{iy}^{10-5}), and plot the variation of the distribution through time to infer whether a benchmark calculated over an extended window would have been more effective.

For market efficiency, we analyze price dynamics around the benchmark window using correlation analysis of short-term reversals:

$$MarketEfficiency = \operatorname{cor}\left(\frac{p_2 - p_1}{p_1}, \frac{p_3 - p_2}{p_2}\right) \tag{5}$$

where p_1, p_2, and p_3 are the average prices 15 min before, during, and 15 min after the fix, respectively.

3. Empirical analysis

3.1. Sample descriptive statistics

Our analysis relies on proprietary high-frequency tick-by-tick data provided by Refinitiv, an LSEG (London Stock Exchange Group) business, the benchmark administrator of the WMR fix. The data consists of all order book events from all the platforms used as data sources for the benchmarks calculated by Refinitiv. As of today, there are three platforms used by the benchmark administrator for the fixings; Refinitiv Matching (TR), Electronic Broking Services (EBS) and Currenex (CRNX). The unit of analysis consists of the major currency pairs traded on the FX market, including the most liquid across the entire market (i.e. GBP/USD, EUR/USD, and USD/CAD) and the most liquid in the Asia-Pacific region (i.e. AUD/USD, NZD/USD, and JPY/USD).[11] Compared to the studies of Evans et al. (2018) and Marsh et al. (2017), which respectively focus on 5- and 3-currency pairs only, we undertake analysis on several additional currencies that are highly liquid and important for the Asia-Pacific region. We also access macroeconomic news information from the API of the Refinitiv Eikon Economic Monitor. For all the currencies in the sample, WMR calculates the benchmark following the TCs procedure, given their high liquidity profile. The sample period available for the analysis extends over 9 years, from 15 February 2015, the day in which the current methodology became effective, to the end of 2023.

The order book events available for the analysis include all trading information (new orders, cancellations, amended orders) and characteristics (execution price and trade direction), along with the best bid and the best ask prices for each respective platform.[12] All events are ordered sequentially at millisecond-time precision. For each trading day of analysis, key economic events highly affecting the volatility of the involved currencies are catalogued from the Eikon Economic Monitor, which indicates a level of market impact on a scale from 1 to 3.

3.1.1. Multivariate analysis

Regarding the multivariate analysis on the actual published 4 pm Fix benchmark, Table 1 presents the descriptive statistics for the liquidity variables of interest (i.e., bas, $espread$, and pi) at each time horizon, as well as the control variables (i.e., $econ\_event$, otr, $month\_end$, and $volatility$).

The statistics for the bid–ask spread (bas) show a mean value ranging from 1.0987 at the 0-min horizon (bas_0 within the 5-min benchmark window) to 1.3541 at the 30-min horizon (bas_{30} which is calculated 30 min before the benchmark window begins). The standard deviation increases over time as expected, indicating greater variability in the bid–ask spread as the time horizon extends. The median values are relatively stable, with a slight increase over time, while the minimum and maximum values show substantial variation, particularly at bas_{25} with a maximum of 15.9643. Skewness and kurtosis values are notably high at certain horizons, especially for bas_{25}, suggesting a distribution with extreme outliers.

For the effective spread ($espread$), the mean values are close to zero across all time horizons, with $espread_0$ at 0.0069 and $espread_{30}$ at −0.0003. The standard deviations are higher than the mean values, indicating significant variability in the effective spread. The skewness and kurtosis values for $espread$ are extreme, particularly for $espread_{25}$, which shows a skewness of −5.8485 and kurtosis of 262.2838, indicating a highly skewed distribution with heavy tails.

The price impact (pi) exhibits mean values of approximately 0.006, with pi_0 at 0.0060 and pi_{30} at 0.0036. The standard deviations increase significantly at longer horizons, reflecting greater variability. The skewness and kurtosis values for pi are exceptionally high, particularly at pi_{15} with skewness of 33.8447 and kurtosis of 2775.1285, indicating a distribution with extreme outliers.

The control variables also display interesting statistics. The economic event indicator ($econ\_event$) has a mean of 0.7918 and a standard deviation of 0.4061, suggesting frequent economic events during the sample period (i.e., on 79% of trading days). The order-to-trade ratio (otr) shows a mean of 1.0047 with a standard deviation of 0.7468, indicating the distribution of trading activity relative to orders placed. The month-end indicator ($month\_end$) has a mean of 0.0467, reflecting the lower number of month-end days in the dataset, which are of particular interest for examining benchmark design effectiveness. Lastly, the volatility measure ($volatility$) has a mean of 0.0007 and a very low standard deviation of 0.0005, with skewness and kurtosis indicating the presence of rare but extreme volatility spikes.

[11] We exclude the CNH/USD given its very low liquidity compared to the other currencies.

[12] We are unable to compute the market depth variable of interest given the absence of relevant data.

Table 1
Descriptive statistics for multivariate analysis. This table summarizes the main descriptive statistics for the liquidity variables and for the control variables used in the multivariate analysis. The bid–ask spread (*bas*), the effective spread (*espread*) and the price impact (*pi*) have been calculated at different time horizons, from 0 referring to the centered range of 5 min around 4 pm UK time, to 30 referring to the range −32.5 min to −27.5 min relative to 4 pm UK time. *bas* is calculated as the difference between the ask and the bid prices; *espread* is computed as two times the difference between the transaction price and the prevailing quote midpoint; *pi* is measured as two times the difference between the prevailing midpoint 20 s after the trade and the quote midpoint; $month\_end$ is a dummy variable taking the value of 1 when the observation refers to the last trading weekday of the month and 0 otherwise; $volatility$ is measured as the standard deviation of prices within the fixing window; $econ\_event$ is a dummy variable taking the value of 1 when there is a macroeconomic announcement for the underlying US Dollar or for the currency in question during day d and 0 otherwise; and otr is the order-to-trade ratio calculated as the total number of limit orders divided by that of the market orders within the benchmark window. The sample period spans from 15 February 2015 to 31 December 2023.

Statistic	bas_0	bas_5	bas_{10}	bas_{15}	bas_{20}	bas_{25}	bas_{30}
Mean	1.0987	1.2598	1.2662	1.2657	1.2695	1.2708	1.3541
St. Dev.	0.3695	0.4751	0.4839	0.4883	0.4864	0.5228	0.5584
Median	1.0755	1.1658	1.1667	1.1563	1.1667	1.1667	1.2368
Min	0.3422	0.3495	0.3000	0.3000	0.3000	0.3103	0.3400
Max	4.5836	6.1818	5.8889	7.6000	6.0000	15.9643	7.7500
Q25	0.7945	0.9103	0.9167	0.9127	0.9167	0.9167	0.9643
Q75	1.3198	1.5439	1.5417	1.5455	1.5556	1.5392	1.6818
Skewness	1.0235	1.1759	1.2565	1.3398	1.2567	3.9771	1.2566
Kurtosis	5.7776	6.1830	6.6537	8.2008	6.6639	79.8405	6.4240
	$espread_0$	$espread_5$	$espread_{10}$	$espread_{15}$	$espread_{20}$	$espread_{25}$	$espread_{30}$
Mean	0.0069	0.0065	0.0067	0.0068	0.0037	0.0002	−0.0003
St. Dev.	0.2770	0.3204	0.3732	0.4172	0.3603	0.3714	0.4373
Median	0.0074	0.0080	0.0077	0.0076	0.0076	0.0077	0.0074
Min	−8.0165	−6.9885	−8.4491	−11.3881	−7.4090	−11.9938	−9.8293
Max	4.6359	7.4417	7.6472	10.5345	7.6704	7.6372	8.7341
Q25	0.0031	0.0024	0.0020	0.0020	0.0019	0.0020	−0.0055
Q75	0.0105	0.0120	0.0120	0.0120	0.0120	0.0122	0.0200
Skewness	−3.8352	−1.8904	−1.1803	−0.9956	−0.8533	−5.8485	−2.1324
Kurtosis	207.3052	148.2802	191.7109	213.4640	186.1406	262.2838	112.9515
	pi_0	pi_5	pi_{10}	pi_{15}	pi_{20}	pi_{25}	pi_{30}
Mean	0.0060	0.0065	0.0061	0.0058	0.0055	0.0062	0.0036
St. Dev.	0.0073	0.0155	0.0150	0.0259	0.0325	0.0535	0.1049
Median	0.0048	0.0043	0.0038	0.0035	0.0035	0.0035	0.0024
Min	−0.0666	−0.7409	−0.1595	−0.7007	−2.0037	−1.6436	−2.1374
Max	0.0999	0.1480	0.5168	2.0051	0.7574	1.6711	1.4378
Q25	0.0011	0.0000	−0.0006	−0.0008	−0.0008	−0.0009	−0.0026
Q75	0.0096	0.0116	0.0115	0.0110	0.0110	0.0112	0.0130
Skewness	1.1816	−14.4694	4.2806	33.8447	−24.7034	2.3604	−2.4936
Kurtosis	12.8135	712.6879	126.0071	2775.1285	1487.9242	403.7888	79.7447
	otr	$econ\_event$	$month\_end$	$volatility$			
Mean	1.0047	0.7918	0.0467	0.0007			
St. Dev.	0.7468	0.4061	0.2111	0.0005			
Median	0.9383	1.0000	0.0000	0.0006			
Min	−1.5861	0.0000	0.0000	0.0000			
Max	5.4765	1.0000	1.0000	0.0123			
Q25	0.5197	1.0000	0.0000	0.0004			
Q75	1.4043	1.0000	0.0000	0.0009			
Skewness	0.7699	−1.4370	4.2944	4.7019			
Kurtosis	4.9734	3.0651	19.4416	60.8783			

3.1.2. Univariate analysis

The descriptive statistics for the univariate analysis, which includes the measures of representativeness, attainability, and robustness across different currency pairs expressed in pips, are presented in Table 2. Representativeness, as indicated by the mean values, suggests varying degrees of alignment with the underlying market across different currencies. For instance, the AUDUSD pair shows a mean of −0.4609 for Re_{id}^{10}, −0.4639 for Re_{id}^{20}, and −0.4674 for Re_{id}^{5}, indicating a consistent but slightly negative deviation from the benchmark. The variability of this measure is high, as shown by the standard deviations exceeding 19, implying significant fluctuations in how well the benchmark represents the market.

Attainability reflects the ease with which market participants can replicate benchmark prices. For AUDUSD, the standard deviation values for A_{id}^{10}, A_{id}^{20}, and A_{id}^{5} are 0.5804, 0.8339, and 0.4587, respectively, suggesting moderate to high variability in replication accuracy. The skewness and kurtosis values indicate the distribution characteristics of these measures, with AUDUSD showing positive skewness and high kurtosis, particularly for A_{id}^{5}, where skewness is 1.7844, and kurtosis is 37.2867, indicating a heavy-tailed distribution.

Robustness, as assessed by the robustness measures (Ro_{id}), reveals the benchmark's resilience to market manipulation. The mean values for AUDUSD are positive or near zero, with a mean of 0.0196 for Ro_{id}^{10}, −0.0062 for Ro_{id}^{20}, and 0.0102 for Ro_{id}^{5}. The high

Table 2

Descriptive statistics for benchmark effectiveness measures. This table summarizes the main descriptive statistics for the benchmark dimensions, with $Re_i^{5,10,20}$, $A_i^{5,10,20}$, $Ro_i^{5,10,20}$ being respectively the benchmark representativeness, attainability and robustness at 5, 10 and 20 min by currency i, and calculated as the root mean square error between the benchmark rates and the mean trade price of day d, the mean trade price within each time window, and the simulated benchmarks without outlier exclusion, respectively. The sample period spans from 15 February 2015 to 31 December 2023.

Statistic	Re_i^5	Re_i^{10}	Re_i^{20}	A_i^5	A_i^{10}	A_i^{20}	Ro_i^5	Ro_i^{10}	Ro_i^{20}
Panel A: AUDUSD									
Mean	−0.4674	−0.4609	−0.4639	0.0068	0.0240	0.0038	0.0102	0.0196	−0.0062
St. Dev.	19.1279	19.0741	19.0207	0.4587	0.5804	0.8339	0.3886	0.5762	0.8450
Median	−0.3080	−0.2456	−0.3080	0.0040	0.0137	0.0128	0.0081	0.0116	0.0123
Min	−91.5487	−90.2987	−89.2857	−2.5081	−3.7942	−12.0616	−2.2795	−3.9651	−10.6722
Max	119.5350	127.2850	130.7850	7.5719	7.0031	5.9403	6.8885	6.7920	6.3707
Q25	−11.5022	−11.5251	−11.5016	−0.2179	−0.2576	−0.3860	−0.1698	−0.2549	−0.3911
Q75	10.8435	10.6815	10.6637	0.2322	0.3035	0.4009	0.1977	0.3023	0.4000
Skewness	−0.0387	−0.0066	0.0168	1.7844	0.9991	−1.9509	2.3261	0.7337	−1.1285
Kurtosis	5.0872	5.2323	5.2880	37.2867	19.8522	32.6697	48.8856	16.2513	21.9109
Panel B: USDCAD									
Mean	0.7088	0.7329	0.7238	0.0037	0.0064	0.0411	−0.0027	0.0089	0.0327
St. Dev.	19.1226	19.0060	18.8298	0.7039	0.8698	1.2854	0.6119	0.8958	1.3029
Median	0.3518	0.4160	0.2334	0.0136	0.0000	−0.0025	0.0055	0.0000	−0.0074
Min	−93.0634	−90.5634	−89.0634	−4.0253	−6.2538	−9.9806	−3.8030	−7.3598	−7.5909
Max	88.9699	86.8317	85.3317	8.2495	8.4541	11.0288	6.8974	9.9561	9.4662
Q25	−10.6429	−10.4919	−10.3363	−0.3571	−0.4411	−0.6062	−0.2931	−0.4341	−0.6257
Q75	11.4641	11.3462	11.2711	0.3641	0.4157	0.6511	0.2857	0.4320	0.6639
Skewness	0.0875	0.0786	0.0655	0.7624	0.4214	0.3250	0.8431	0.5859	0.5587
Kurtosis	4.5660	4.5302	4.4818	15.6425	11.0536	9.7305	16.1633	13.8574	8.1734
Panel C: EURUSD									
Mean	−1.2916	−1.2668	−1.1939	0.0444	0.0169	−0.0457	0.0231	−0.0275	−0.0834
St. Dev.	20.6642	20.5393	20.3364	0.5497	0.7346	1.0122	0.4534	0.7359	1.0681
Median	−1.0335	−1.0429	−0.7686	0.0567	0.0479	0.0103	0.0312	0.0060	−0.0207
Min	−142.5822	−142.8072	−143.4572	−10.6155	−10.8807	−8.8285	−3.8850	−7.4530	−10.1210
Max	123.9984	123.9984	121.2484	2.8181	4.2526	5.2092	2.8268	5.2180	5.2627
Q25	−12.8337	−12.7747	−12.5390	−0.1820	−0.2762	−0.4494	−0.1772	−0.3581	−0.5552
Q75	10.5566	10.5566	10.6176	0.2877	0.3578	0.4368	0.2381	0.3449	0.4333
Skewness	−0.1336	−0.1255	−0.1307	−3.5725	−2.7858	−1.2800	−0.3445	−0.9794	−0.7910
Kurtosis	6.2786	6.3460	6.3519	69.0132	38.7408	13.9390	9.1311	13.7363	11.0096
Panel D: GBPUSD									
Mean	−1.1678	−1.1565	−1.1145	0.0337	0.0153	−0.0726	0.0064	−0.0196	−0.0591
St. Dev.	27.1643	27.0355	26.8180	0.8349	1.0907	1.6280	0.7164	1.1316	1.6388
Median	−0.5054	−0.4920	−0.3961	0.0117	0.0423	0.0044	0.0086	0.0122	0.0000
Min	−204.0105	−200.0105	−197.2605	−6.2387	−15.1253	−21.6268	−8.5602	−15.5187	−23.3351
Max	111.7004	111.2004	109.7004	6.5257	5.1784	7.7393	4.9605	5.5070	8.7963
Q25	−16.9465	−16.7248	−16.4022	−0.3793	−0.4593	−0.6678	−0.3023	−0.4632	−0.6883
Q75	14.3618	14.3279	14.4663	0.4385	0.5272	0.6475	0.3050	0.4597	0.6387
Skewness	−0.2968	−0.2801	−0.2767	−0.0157	−1.8170	−3.4847	−0.8173	−2.4182	−2.9334
Kurtosis	6.0828	5.9976	5.9456	10.2927	27.1791	42.6794	18.2511	35.5837	37.1498
Panel E: JPYUSD									
Mean	1.3156	1.3027	1.2983	0.0401	0.0405	0.0666	0.0195	0.0298	0.0388
St. Dev.	27.4604	27.3401	27.1744	0.5605	0.6864	1.0900	0.4698	0.7079	1.0741
Median	1.5244	1.5139	1.5815	0.0444	0.0214	0.0357	0.0200	0.0147	0.0070
Min	−191.4123	−187.6623	−185.6623	−10.4003	−4.5656	−8.6954	−8.0702	−4.0231	−5.5517
Max	128.4154	128.4154	128.2904	3.7917	5.5862	22.6815	2.8249	5.5815	20.0794
Q25	−12.5237	−12.5444	−12.4770	−0.2052	−0.2949	−0.4162	−0.1877	−0.3212	−0.4435
Q75	15.1530	15.2334	15.2462	0.2798	0.3532	0.4898	0.2286	0.3615	0.4675
Skewness	−0.3362	−0.3311	−0.3415	−2.3443	0.4016	4.1164	−2.0794	0.3835	3.2591
Kurtosis	7.6392	7.6013	7.5492	59.2836	10.1989	88.7180	44.8463	8.4220	60.8549
Panel F: NZDUSD									
Mean	−0.0023	−0.0113	−0.0238	0.0139	0.0182	0.0153	0.0057	0.0137	−0.0018
St. Dev.	17.7290	17.6778	17.6643	0.5957	0.6885	0.9543	0.4313	0.5910	0.8171
Median	−0.2901	−0.2744	−0.3778	0.0196	0.0166	0.0390	0.0000	0.0000	0.0189
Min	−95.0401	−94.5401	−94.2901	−5.4227	−7.3711	−11.6199	−4.8000	−6.7336	−10.3785
Max	142.9298	145.1798	147.9298	4.1667	5.6067	6.4488	2.8119	3.8860	5.2626
Q25	−10.3849	−10.3714	−10.4263	−0.3333	−0.3542	−0.4204	−0.2162	−0.2692	−0.3524
Q75	10.2019	10.2093	10.2415	0.3556	0.3961	0.4545	0.2469	0.3000	0.3750
Skewness	0.3778	0.3917	0.4023	−0.4090	−0.3901	−1.2864	−0.9622	−0.7418	−1.6716
Kurtosis	6.3979	6.4924	6.5931	10.3513	13.4392	23.7527	16.0851	15.5597	26.1588

Table 3
OLS panel regression results. This table reports the results of the following OLS panel regression analysis:

$$Y_{id}^{t} = \alpha^{t} + \beta_1^{t} X_{id}^{t} + \beta_2^{t} K_d^{t} + \beta_3^{t} K_d^{t} X_{id}^{t} + \delta^{t} C_{id}^{t} + \gamma^{t} FE_{id}^{t} + \epsilon_{id}^{t}$$

where t represents each 5-min interval considered for the calculation of market quality measures, from 30 min before the start of the benchmark window to 4 pm UK time. Y_{id}^{t} represents the natural logarithm of the actual published 4 pm Fix benchmark rate for the currency pair i in a given day d. X_{id} is a vector of the three market quality metrics (i.e., bid–ask spread, effective spread, and price impact). K_d is a dummy variable taking the value of 1 when the observation refers to the last trading weekday of the month and 0 otherwise. Our variables of interest are the interaction terms between K_d and X_{id} at each time 30-min before the window through to 5 min around 4 pm UK time, capturing the marginal predictive power of the benchmark from changes in market behavior at month-end. C_{id} is the vector of control variables for currency i on day d (i.e., volatility, order-to-trade ratio and a dummy variable capturing major macroeconomic news influencing the benchmark rates). The sample period spans from 15 February 2015 to 31 December 2023. White (1980) standard errors are reported in parentheses, while ***, ** and * denote significance at 1%, 5% and 10%, respectively. All currencies are pooled together.

Variables	4 pm Fix benchmark (ln)						
	$t0$	$t5$	$t10$	$t15$	$t20$	$t25$	$t30$
bas (pips)	0.1056 (0.058)*	0.0600 (0.042)	0.0549 (0.039)	0.0508 (0.039)	0.0519 (0.040)	0.0393 (0.031)	0.0414 (0.031)
$espread$ (bps)	−0.0022 (0.000)***	−0.0017 (0.001)**	−0.0030 (0.001)***	−0.0006 (0.000)*	−0.0010 (0.001)	−0.0017 (0.001)**	−0.0004 (0.001)
pi (bps)	−0.5092 (0.168)***	−0.1752 (0.075)**	−0.2682 (0.097)***	−0.0572 (0.037)	−0.0265 (0.021)	−0.0115 (0.010)	−0.0021 (0.004)
$bas \times month\_end$	0.0068 (0.005)	0.0201 (0.007)***	0.0105 (0.008)	0.0142 (0.008)*	0.0147 (0.007)**	0.0208 (0.012)*	0.0088 (0.003)***
$espread \times month\_end$	−0.0072 (0.002)***	0.0023 (0.006)	0.0121 (0.007)*	−0.0044 (0.002)***	−0.0076 (0.001)***	−0.0103 (0.004)***	−0.0023 (0.007)
$pi \times month\_end$	0.7033 (0.265)***	−0.2627 (0.143)*	0.2378 (0.179)	−0.2608 (0.197)	−0.2124 (0.107)**	−0.6533 (0.218)***	0.0017 (0.025)
otr (ln)	−0.0031 (0.004)	0.0067 (0.005)	0.0066 (0.006)	0.0072 (0.006)	0.0073 (0.006)	0.0081 (0.007)	0.0077 (0.006)
$volatility$	−23.3181 (2.341)***	−16.3762 (3.947)***	−14.5386 (5.223)***	−15.0000 (5.302)***	−15.2573 (4.784)***	−14.0469 (5.800)**	−14.0334 (5.521)**
$economic\_event$	−0.0021 (0.001)	−0.0018 (0.002)	−0.0020 (0.002)	−0.0021 (0.002)	−0.0014 (0.001)	−0.0017 (0.002)	−0.0021 (0.002)
Currency-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	13.48%	8.82%	8.35%	7.37%	7.59%	6.04%	6.66%
# Obs.	13,722	13,482	13,377	13,324	13,284	13,340	13,335

standard deviation values (e.g., 0.5762 for Ro_{id}^{10}) and extreme kurtosis values (e.g., 48.8856 for Ro_{id}^{5}) indicate significant outlier influence, which might affect the benchmark's robustness.

For the USDCAD pair, the mean values for attainability measures are generally positive, with A_{id}^{10} at 0.0064, A_{id}^{20} at 0.0411, and A_{id}^{5} at 0.0037, suggesting that the benchmarks are attainable to a certain degree. However, the standard deviation values indicate substantial variability in this attainability. Representativeness measures for USDCAD show a mean of approximately 0.7 for different horizons, with a standard deviation close to 19, similar to AUDUSD.

EURUSD and GBPUSD pairs exhibit different statistical profiles. EURUSD's mean values for representativeness are notably negative, indicating a significant deviation from the underlying market, with standard deviations of approximately 20, reflecting high variability. GBPUSD shows a mean of −1.1565 for Re_{id}^{10} with a standard deviation of 27.0355, indicating both a substantial deviation and high variability. For the JPYUSD pair, the attainability and robustness measures indicate positive mean values and relatively high standard deviations, suggesting a complex interaction between market conditions and benchmark replication. The skewness and kurtosis for A_{id}^{20} are particularly high (4.1164 and 88.7180, respectively), indicating a significant presence of outliers and a skewed distribution. Finally, the NZDUSD pair presents a mix of positive and negative mean values across the different measures. The standard deviations are lower compared to other pairs, suggesting less variability linked to its lower levels of liquidity.

It is important to note that such findings are consistent with the established characteristics of FX markets. The literature demonstrates that FX markets are often characterized by excess kurtosis and skewness, reflecting the presence of heavy tails and asymmetry in return distributions (Chiang et al., 2007; Celick, 2012). This is particularly true during high-volatility periods and around 4 pm, where sharp price movements and liquidity imbalances can occur. Further, large standard deviations and extremes in the minimum and maximum values are expected due to the highly volatile nature of the FX market. In particular, the representativeness measure employed compares the benchmark rate (calculated at 4 pm) to the mean price of the day, which can fluctuate significantly throughout the trading day. Therefore, it is not uncommon for the mean price of the day to differ vastly from the 4 pm benchmark, resulting in higher variability in our data. The other two measures present consistent variation with prior studies (e.g., Evans, 2018). Finally, the metrics show consistent variations across all currencies, providing further evidence that the data do not contain errors but instead confirm the inherent characteristics of the FX market, supported by the descriptive statistics.

3.2. Main confirmatory analysis (pre-registered)

3.2.1. Multivariate analysis

The multivariate analysis on the actual published 4 pm Fix benchmark presented in Table 3 examines the relationship between various market quality metrics and the benchmark rate at different time intervals surrounding the benchmark window. The model developed for this analysis uses a panel multivariate regression estimation, incorporating market quality metrics including the bid–ask spread (bas), effective spread ($espread$), and price impact (pi) as regressors. Additionally, the model includes control variables including the order-to-trade ratio (otr), volatility ($volatility$), and a macroeconomic announcement variable ($economic\_event$), along with fixed effects for currency and time. Following Galati (2024), robust standard errors are computed through the heteroskedasticity-consistent covariance matrix approach of White (1980).

The bid–ask spread (bas) shows a positive relationship with the benchmark rate across all time intervals, with coefficients ranging from 0.1056 at $t0$ to 0.0414 at $t30$. The effective spread ($espread$), a proxy for implicit transaction costs for liquidity takers, on the other hand, exhibits a negative relationship with the benchmark rate, with coefficients ranging from −0.0022 at $t0$ to −0.0004 at $t30$. The coefficients are highly significant at various levels. The price impact (pi), a proxy for adverse selection costs between market participants, also exhibits coefficients decreasing from −0.5092 at $t0$ to −0.0021 at $t30$, with significant coefficients at the initial time intervals. These results, altogether, suggest that transaction and adverse selection costs can provide significant information regarding the level of the published benchmark rates only in periods close to the benchmark window, and is statistically significant within that proximity.

The interaction terms, our variables of interest, between the bid–ask spread and the month-end ($bas \times month\_end$) reveal significant coefficients, indicating that the month-end effect amplifies the predictive power of liquidity on the benchmark rate. Similarly, the interaction terms for the effective spread and month-end ($espread \times month\_end$) show mixed results, with significant coefficients at some intervals. The interaction terms for the price impact and month-end ($pi \times month\_end$) also present significant coefficients at various intervals. These findings suggest a complex relationship between transaction and adverse selection costs, month-end effects, and the benchmark rate.

The order-to-trade ratio (otr) within the benchmark 5-min window shows mostly positive but insignificant coefficients, indicating that it does not have substantial predictive power on the benchmark rate. Volatility ($volatility$) within the benchmark 5-min window, however, exhibits highly significant coefficients across all intervals, implying that volatility during the window is a significant component, as expected, of the benchmark rate. The macroeconomic announcement variable ($economic\_event$) shows mostly insignificant coefficients, indicating that macroeconomic announcements do not have a consistent or significant impact on the benchmark rate.[13]

The model also includes fixed effects to control for currency and time, ensuring that the observed relationships are not driven by unobserved heterogeneity. The R^2 values range from 6.04% to 13.48%, indicating the proportion of variance explained by the model at different time intervals. The number of observations varies slightly across the time intervals, ranging from 13,284 to 13,722, reflecting the diverse liquidity of currencies and trading activity in different time horizons. Overall, this multivariate analysis provides insights into the predictive power of market quality metrics on the benchmark rate, highlighting the significance of the bid–ask spread, effective spread, price impact, and volatility, while accounting for month-end effects, the order-to-trade ratio, and macroeconomic announcements.

3.2.2. Univariate analysis

In the univariate analysis, we perform a series of Student t-tests on the differences among each of the three benchmark dimensions (representativeness, attainability, and robustness) calculated around a 10- and 20-min benchmark window with respect to the same measure within the actual 5-min benchmark window. Each analysis is executed with and without distinction between normal and end-of-month trading days to assess the critical end-of-month trading behavior.

Tables 4, 5, and 6 present the results of the t-tests on the representativeness, attainability and robustness differences, respectively, disentangling between all trading days, normal trading days, and end-of-month trading days. For all currencies, the results focusing on all trading days are consistent with the results on normal trading days.

All currencies tend to show non-significant results regarding the differences in the representativeness dimension, leading us to accept the null hypothesis that the differences are equal to 0, supporting the effectiveness of the benchmark calculated with the current 5-min window methodology. We only observe a few exceptions for EURUSD and GBPUSD in the differences between the measures calculated through the 20- and 5-min periods. End-of-month results provide evidence that the simulated benchmarks based on 10- and 20-min windows do not deviate from the 4 pm Fix rate calculated following the current 5-min method, suggesting that the benchmark representativeness does not improve as the window widens, even when considering the crucial and substantially different end-of-month trading activity.

When examining differences in the attainability dimension, we obtain significant results. This proxy measures how easily traders can replicate the benchmark by observing market prices. During normal trading days, and with only very few exceptions, all currencies show statistically significant results on the tested differences, with some variability in the sign of the estimates across currencies. The positive sign of the estimates for AUDUSD, USDCAD, JPYUSD and NZDUSD indicates that attainability worsens as

[13] We conduct a correlation analysis to assess the relationship between the macroeconomic news variable and the other variables in the model. The analysis indicates that the macroeconomic news variable is not strongly correlated with the other variables included in the model.

Table 4
This table reports the t-test results by currency on the difference between the benchmark representativeness calculated on a 10-min window and on a 20-min window relative to the representativeness calculated on a 5-min window. Representativeness is measured as

$$Representativeness_{iy}^{5,10,20} = \sqrt{N^{-1}\sum_{d=0}^{N}\left(b_{id}^{5,10,20} - p_{id}^{d}\right)^2}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows, and p_{id}^{d} is the mean trade price for currency i on day d. The sample period spans from 15 February 2015 to 31 December 2023. ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Representativeness (Re_i)					
	All trading days		Normal trading days		End-of-Month trading days	
	10 - 5	20 - 5	10 - 5	20 - 5	10 - 5	20 - 5
Panel A: AUDUSD						
Mean	0.0065	0.0035	0.0109	0.0119	−0.0829	−0.1682
t-test	0.2753	0.1483	0.4379	0.4799	−0.1790	−0.3660
# Obs.	4578	4578	4364	4364	214	214
Panel B: USDCAD						
Mean	0.0241	0.0150	0.0188	0.0063	0.1320	0.1928
t-test	1.0223	0.6404	0.7644	0.2575	0.2373	0.3490
# Obs.	4578	4578	4364	4364	214	214
Panel C: EURUSD						
Mean	0.0248	0.0977	0.0358	0.1205	−0.2000	−0.3675
t-test	0.9730	3.8574***	1.3432	4.5419***	−0.3520	−0.6600
# Obs.	4578	4578	4364	4364	214	214
Panel D: GBPUSD						
Mean	0.0114	0.0533	0.0368	0.1113	−0.5082	−1.1297
t-test	0.3392	1.5980	1.0448	3.1692**	−0.8032	−1.7938
# Obs.	4578	4578	4364	4364	214	214
Panel E: JPYUSD						
Mean	−0.0129	−0.0173	−0.0211	−0.0347	0.1540	0.3383
t-test	−0.3806	−0.5122	−0.5945	−0.9828	0.2030	0.4473
# Obs.	4578	4578	4364	4364	214	214
Panel F: NZDUSD						
Mean	−0.0090	−0.0215	−0.0047	−0.0077	−0.0900	−0.2944
t-test	−0.4091	−0.9801	−0.2030	−0.3315	−0.2073	−0.6848
# Obs.	4561	4561	4347	4347	214	214

the window considered for the benchmark calculation widens. The opposite can be inferred for GPBUSD and EURUSD. Focusing on the end-of-month days, instead, only EURUSD and GBPUSD continue to report a strong significance, again confirming a negative sign of the estimates and the consequent improvement in attainability as the window is widened.

The significant results on the robustness dimension on normal trading days almost completely lose significance when focusing only on the end-of-month days. On normal trading days, t-tests indicate statistically significant results with positive estimates for AUDUSD, USDCAD, JPYUSD and NZDUSD, suggesting a greater robustness of the benchmark if calculated over a wider benchmark window. Both EURUSD and GBPUSD, instead, indicate significant but negative estimates, suggesting that the benchmark calculated on a 5-min window is more robust than those computed on longer time windows. End-of-month results, with the exception of GBPUSD and NZDUSD, confirm the interpretations from normal trading days. These end-of-month results are statistically significant, and so do not exclude, on average, the possibility of the same degree of robustness regardless of the window length.

Existing literature highlights distinct price dynamics around the 4 pm fix, characterized by short-term spikes and subsequent reversals (Evans, 2018). Evans et al. (2018) investigate these reversals through correlation analysis and find that market efficiency around the 4 pm fix has significantly improved during their sample period. According to Evans et al. (2018), one possible explanation is the elimination of collusive behavior, but another factor could be the extended duration of the fix, which provides additional time for liquidity shocks to subside. While this provides compelling evidence supporting the extension of the window within which benchmark rates are measured, the authors do not seek to establish any causal inferences for this improvement, as it falls beyond the scope of their analysis. Accordingly, it is not the objective of this study to do so either.

Fig. 7 presents the results of replicating the autocorrelation analysis performed by Evans et al. (2018), extending the calculation to date and to other relevant currencies such as those of the Asia-Pacific region. The market efficiency measure computed, as stated in Section 2.2, captures the presence of short-term price reversals surrounding the fixing window. The absence of serial dependence on price changes is driven by efficient market functioning, as evidenced by Evans et al. (2018) after the implementation of the new methodology in 2015. Our analysis confirms the findings of Evans et al. (2018) in every quarter of the analysis, excluding a few exceptions mainly in 2020 and 2021. These quarters report an autocorrelation coefficient that is not statistically significant, supporting the efficiency of the benchmark calculated following the current WMR methodology.

Table 5
This table reports the t-test results by currency on the difference between the benchmark attainability calculated on a 10-min window and on a 20-min window relative to the attainability calculated on a 5-min window. Attainability is measured as

$$Attainability_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - p_{id}^{5,10,20} \right)^2}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows, and $p_{id}^{5,10,20}$ is the mean price within each time window. The sample period spans from 15 February 2015 to 31 December 2023. ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Attainability (A_i)					
	All trading days		Normal trading days		End-of-Month trading days	
	10 - 5	20 - 5	10 - 5	20 - 5	10 - 5	20 - 5
Panel A: AUDUSD						
Mean	0.0172	−0.0031	0.0199	−0.0021	−0.0385	−0.0234
t-test	26.5993***	−3.7122***	30.6585***	−2.4341*	−1.7069*	−1.0331
# Obs.	4578	4578	4364	4364	214	214
Panel B: USDCAD						
Mean	0.0027	0.0374	−0.0010	0.0362	0.0777	0.0610
t-test	2.7416**	29.1677***	−1.0014	27.5227***	2.6407**	1.6433
# Obs.	4578	4578	4364	4364	214	214
Panel C: EURUSD						
Mean	−0.0275	−0.0901	−0.0230	−0.0888	−0.1192	−0.1164
t-test	−34.3179***	−89.5258***	−28.1793***	−86.8953***	−4.7155***	−3.5631***
# Obs.	4578	4578	4364	4364	214	214
Panel D: GBPUSD						
Mean	−0.0184	−0.1063	−0.0124	−0.1000	−0.1399	−0.2365
t-test	−15.3085***	−66.5143***	−10.2087***	−61.9874***	−3.5507***	−4.3370***
# Obs.	4578	4578	4364	4364	214	214
Panel E: JPYUSD						
Mean	0.0004	0.0265	0.0003	0.0298	0.0023	−0.0421
t-test	0.5036	24.7177***	0.3800	27.1373***	0.0902	−1.3463
# Obs.	4578	4578	4364	4364	214	214
Panel F: NZDUSD						
Mean	0.0043	0.0014	0.0032	0.0006	0.0280	0.0197
t-test	5.4207***	1.4513	3.8200***	0.5446	1.5688	1.0452
# Obs.	4561	4561	4347	4347	214	214

3.3. *Additional exploratory analysis (unregistered)*

3.3.1. *Change in the research questions*

During the course of this study, we encountered a pivotal issue related to the formulation of the third research question as outlined in the pre-registered report (Benenchia et al., 2024). Initially, our research aimed to address the impact of the WMR methodology on market behavior, specifically asking whether it minimized the market impact surrounding the 4 pm Fix. However, as our analysis progressed, it became clear that our dataset and the tools we utilize provide additional insight into how market behavior influences the benchmarks, rather than isolating whether the methodology itself controls for abnormal movements.

In this pivot, we re-positioned the third research question as the first, so that the natural consequence of that analysis was to then look at whether the WMR methodology was effective. Specifically, we observe that liquidity, trading activity, transaction costs, and adverse selection costs all exhibit significant variability around the 4 pm Fix window, with notable differences between month-end and normal trading days. These findings underscore that the 4 pm Fix is not only a product of the methodology but a reflection of market behavior, especially under the influence of trading volume surges and liquidity changes on critical days like month-end.

Given this, we re-defined our research question to the following: “How has the market impacted the 4 pm Fix in terms of liquidity, trading activity, transaction and adverse selection costs, and price volatility?” This new formulation more accurately captures the nuances of our analysis, as it reflects the influence of market factors on the benchmark calculation rather than attributing all observed changes to the methodology alone. The interaction terms, particularly those distinguishing between normal and month-end trading days, provide valuable insights into how market conditions shape the 4 pm Fix.

This adjustment is driven by the way the industry often phrases discussions around the WMR methodology, particularly the common question, “Has the WMR minimized market impact?” While this is a frequent way of discussion within FX markets, it does not fully capture the complexities of how the benchmark interacts with market behavior. This industry language led us to initially formulate our research question in a way that does not align with the deeper dynamics at play. Our analysis contributes to clarifying this by highlighting the significance of market behavior in determining the accuracy of the 4 pm Fix, suggesting that the methodology’s effectiveness cannot be fully assessed without considering the broader market context in which it operates.

Table 6
This table reports the t-test results by currency on the difference between the benchmark robustness calculated on a 10-min window and on a 20-min window relative to the robustness calculated on a 5-min window. Robustness is measured as

$$Robustness_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - \tilde{b}_{id}^{5,10,20} \right)^2}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows, and $\tilde{b}_{id}^{5,10,20}$ are the simulated benchmarks without outlier exclusion. The sample period spans from 15 February 2015 to 31 December 2023. ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Robustness (Ro_i)					
	All trading days		Normal trading days		End-of-Month trading days	
	10 - 5	20 - 5	10 - 5	20 - 5	10 - 5	20 - 5
Panel A: AUDUSD						
Mean	0.0094	−0.0164	0.0093	−0.0183	0.0118	0.0219
t-test	15.5113***	−20.1453***	15.3270***	−22.1211***	0.5366	0.8351
# Obs.	4578	4578	4364	4364	214	214
Panel B: USDCAD						
Mean	0.0116	0.0354	0.0102	0.0357	0.0398	0.0295
t-test	12.2468***	28.1599***	10.7341***	28.2520***	1.2059	0.6695
# Obs.	4578	4578	4364	4364	214	214
Panel C: EURUSD						
Mean	−0.0506	−0.1066	−0.0506	−0.1066	−0.0307	0.0359
t-test	−66.9986***	−105.1156***	−67.9047***	−112.6255***	−1.1625	0.9454
# Obs.	4578	4578	4364	4364	214	214
Panel D: GBPUSD						
Mean	−0.0261	−0.0656	−0.0346	−0.0764	0.1473	0.1553
t-test	−22.2625***	−41.9619***	−29.5423***	−49.8590***	3.5069***	2.4669*
# Obs.	4578	4578	4364	4364	214	214
Panel E: JPYUSD						
Mean	0.0103	0.0192	0.0114	0.0235	−0.0118	−0.0682
t-test	13.8834***	18.7636***	15.1882***	22.6756***	−0.4687	−1.9956*
# Obs.	4578	4578	4364	4364	214	214
Panel F: NZDUSD						
Mean	0.0080	−0.0075	0.0026	−0.0146	0.1190	0.1377
t-test	12.5146***	−9.2137***	3.8605***	−17.1869***	7.8867***	7.8079***
# Obs.	4561	4561	4347	4347	214	214

As a result, the decision to modify our research question reflects the natural evolution of our analysis and the importance of providing a clearer, more accurate representation of our findings. The original question posed in the pre-registered report assumes that the WMR methodology can be isolated from broader market influences, but our results demonstrate that these influences are too significant to ignore. Hence, this change was both necessary and reflective of the insights gained throughout the research process. In Table 3, we provide the detailed results of the interaction analysis, illustrating how market-driven factors, particularly those related to liquidity and trading costs, significantly affect the benchmark rates during month-end versus normal trading days. These results confirm the need to shift the research question and offer a more comprehensive understanding of the forces shaping the 4 pm Fix.

3.3.2. Year-on-year univariate analysis on the overall WMR effectiveness

We further undertake the same analysis as in Section 3.2.2, considering all currencies together in the sample but focusing on the trend year by year. Tables 7, 8, and 9 report the results of the t-tests on the benchmark representativeness, attainability and robustness, respectively, separating all trading days, normal trading days and end-of-month trading days. Again, the generalization of the results between all days and normal days holds. Overall, the statistical significance of the results year-over-year is low compared to the results obtained by looking at each currency separately, highlighting potential currency-specific trading behavior that impacts the benchmark dimensions more significantly.

The differences in representativeness lose significance starting from 2017. The estimated means vary around zero, but all remain in an increasingly narrow range as we approach 2023, the end of the sample period. Attainability and robustness exhibit some level of significance until 2018, and, similar to representativeness, they tend to approach zero towards the end of the sample period. The results of the end-of-month days provide limited contribution in terms of statistical significance. Except in 2017, when all dimensions are significant, all other years have t-statistics close to 0.

We further investigate our research questions with the support of illustrations similar to Evans et al. (2018). The representativeness trend over time, calculated through a window of 5-, 10-, and 20-minutes, follows similar patterns for all currencies. Except for GBPUSD, Fig. 1 evidences a decreasing trend in the standard deviation of the tracking error underlying an improvement in the benchmark representativeness.

Table 7
This table reports the t-test results by year on the difference between the benchmark representativeness calculated on a 10-min window and on a 20-min window relative to the representativeness calculated on a 5-min window. Representativeness is measured as

$$Representativeness_{iy}^{5,10,20} = \sqrt{N^{-1}\sum_{d=0}^{N}\left(b_{id}^{5,10,20} - p_{id}^{d}\right)^{2}}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows, and p_{id}^{d} is the mean trade price for currency i on day d. The sample period spans from 15 February 2015 to 31 December 2023. ***, ** and * denote significance at 1%, 5% and 10%, respectively. All currencies are pooled together.

	Representativeness (Re_y)					
	All trading days		Normal trading days		End-of-Month trading days	
	10 - 5	20 - 5	10 - 5	20 - 5	10 - 5	20 - 5
Panel A: 2015						
Mean	0.05973	0.09117	0.06357	0.09700	−0.01515	−0.02273
t-test	2.5549**	1.9904*	2.7101***	2.1321**	−0.1051	−0.0733
# Obs.	1356	1356	1290	1290	66	66
Panel B: 2016						
Mean	0.04567	0.13749	0.06747	0.18663	−0.40104	−0.86910
t-test	1.6330	2.3745*	2.6983***	3.4867***	−1.2874	−1.5010
# Obs.	1547	1547	1475	1475	72	72
Panel C: 2017						
Mean	−0.00029	0.00440	0.01853	0.01480	−0.38437	−0.84965
t-test	−0.0159	0.1192	1.0180	1.1068	−3.1445***	−3.8412***
# Obs.	1541	1541	1469	1469	72	72
Panel D: 2018						
Mean	0.00275	0.005677	0.04626	0.00295	−0.10590	−0.26711
t-test	0.1541	1.5640	1.2568	0.0946	−0.9676	−1.3111
# Obs.	1548	1548	1469	1469	72	72
Panel E: 2019						
Mean	0.01300	0.01909	−0.01393	−0.01030	0.01389	−0.10243
t-test	0.7787	0.3389	−0.6475	−0.2338	0.1623	−0.6160
# Obs.	1548	1548	1481	1481	72	72
Panel F: 2020						
Mean	−0.01288	−0.01151	−0.00686	−0.00797	0.00868	−0.03646
t-test	−0.5849	−0.2592	−0.4961	−0.3287	0.0499	−0.1161
# Obs.	1553	1553	1475	1475	72	72
Panel G: 2021						
Mean	−0.00436	−0.00695	−0.03189	−0.06182	0.04687	0.01389
t-test	−0.3013	−0.2716	−2.0081*	−1.8358	0.3635	0.0584
# Obs.	1547	1547	1474	1474	72	72
Panel H: 2022						
Mean	−0.03275	−0.05846	−0.00726	−0.01494	−0.05035	0.01042
t-test	−1.9337*	−1.6958	−0.5023	−0.6054	−0.3070	0.0384
# Obs.	1546	1546	1464	1464	72	72
Panel I: 2023						
Mean	−0.00684	−0.01668	−0.00726	−0.01494	0.00174	−0.05208
t-test	−0.4813	−0.6858	−0.5023	−0.6054	0.0233	−0.3915
# Obs.	1536	1536	1464	1464	72	72

In 2020, possibly due to the unprecedented market conditions following the COVID-19 pandemic, the measure increased significantly, reversing soon after but not sufficiently to reach the levels in 2019. JPYUSD represents the only exception, with a general deterioration in market representativeness since 2019. The window length over which the measure is computed contributes little to improving this benchmark effectiveness dimension, with all currencies showing the three representativeness trends in a very tight range.

Focusing only on the end-of-month days, the trend does not differ significantly from the normal days, with just a minor degree of variation between the measures calculated over different time windows. In these cases, the 5-min representativeness appears slightly less effective than the measure calculated over the 10- and 20-minutes (mainly for EURUSD, GBPUSD, and AUDUSD).

The results on attainability deliver critical insights (see Fig. 2). Focusing on normal trading days, all currencies exhibit a generalized higher effectiveness of the 5-min window compared to the alternative 10- and 20-min windows. Over the 9 years analyzed, excluding minor and rare exceptions, the 5-min window allows traders to replicate the benchmark more accurately than what would have been possible if the benchmark was calculated over a window of 10- or 20-min. Only JPYUSD experiences an increasing (deteriorating) trend in the last three years, but the 5-min window still remains at least as effective as the 10-min window.

Table 8
This table reports the t-test results by year on the difference between the benchmark attainability calculated on a 10-min window and on a 20-min window relative to the attainability calculated on a 5-min window. Attainability is measured as

$$Attainability_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - p_{id}^{5,10,20}\right)^2}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows, and $p_{id}^{5,10,20}$ is the mean price within each time window. The sample period spans from 15 February 2015 to 31 December 2023. ***, ** and * denote significance at 1%, 5% and 10%, respectively. All currencies are pooled together.

	Attainability (A_y)					
	All trading days		Normal trading days		End-of-Month trading days	
	10 - 5	20 - 5	10 - 5	20 - 5	10 - 5	20 - 5
Panel A: 2015						
Mean	−0.00177	0.00703	−0.00450	−0.00127	0.01550	0.16919
t-test	−0.09186	0.2214	−0.2335	−0.0406	0.3480	0.7521
# Obs.	1356	1356	1290	1290	66	66
Panel B: 2016						
Mean	−0.04257	−0.08025	−0.03483	−0.06058	−0.20120	−0.48322
t-test	−2.2192*	−2.3943**	−1.8939*	−1.9409*	−1.2031	−1.4646
# Obs.	1547	1547	1475	1475	72	72
Panel C: 2017						
Mean	0.02188	0.01664	0.01540	−0.00610	0.16633	0.29592
t-test	1.6413	0.5444	1.0248	−0.2038	2.0236**	1.9548*
# Obs.	1541	1541	1469	1469	72	72
Panel D: 2018						
Mean	−0.03139	−0.07657	−0.00610	0.07012	−0.15964	−0.29999
t-test	−2.0172*	−2.7182**	−0.2038	1.9014*	−2.0924**	−1.8559*
# Obs.	1548	1548	1476	1476	72	72
Panel E: 2019						
Mean	0.01487	0.00462	0.00838	−0.04773	0.02564	0.03479
t-test	1.0909	0.2107	0.4570	−1.3910	0.2586	0.3149
# Obs.	1548	1548	1481	1481	72	72
Panel F: 2020						
Mean	0.01197	−0.03602	0.02270	0.02375	0.08592	0.20486
t-test	0.6557	−1.0532	2.0095**	1.2157	0.7542	0.9568
# Obs.	1553	1553	1475	1475	72	72
Panel G: 2021						
Mean	0.01432	0.00505	−0.00909	−0.04856	−0.15735	−0.36745
t-test	1.2293	0.2580	−0.5840	−1.5687	−1.6792*	−1.6147
# Obs.	1547	1547	1474	1474	72	72
Panel H: 2022						
Mean	−0.01328	−0.04825	−0.00598	0.01004	−0.09917	−0.04109
t-test	−0.8563	−1.5984	−0.4549	0.4614	−1.0109	−0.3605
# Obs.	1546	1546	1464	1464	72	72
Panel I: 2023						
Mean	−0.00521	0.00916	−0.00598	0.01004	0.01036	−0.00885
t-test	−0.4094	0.4312	−0.4549	0.4614	0.2118	−0.0908
# Obs.	1536	1536	1464	1464	72	72

The same conclusions cannot be drawn, instead, if we focus on end-of-month trading days. The ease of replicating the benchmark during the end-of-month trading activity does not seem to be strictly related to window length. Across all currencies, the chart shows the 5-min attainability line crossing multiple times the measures of the 10- and 20-min windows and does not follow a defined trend. This result stresses again the peculiarity of the end-of-month trading activity when the greatest share of volume is exchanged on the main FX platforms.

The robustness results in Fig. 3 complement the attainability results just discussed, confirming the trade-off between the two dimensions (Evans et al., 2018). Looking at normal trading days, the 5-min benchmark tends to be less robust, losing any preference ranking when we analyze the end-of-month pattern, which again shows the high variability among the three measures for all currencies.

The univariate analysis discussed in Section 3.2.2 finds supporting evidence when examining Figs. 4, 5, and 6. They show the boxplot trend of the differences in the three dimensions at 10- and 20 min compared to the 5-min window, grouping all currencies together and still disentangling for all days, normal days and end-of-month days. The boxplots for representativeness and attainability on normal trading days demonstrate a trend of increasingly concentrated distributions around zero as we approach 2023. This

Table 9
This table reports the t-test results by year on the difference between the benchmark robustness calculated on a 10-min window and on a 20-min window relative to the robustness calculated on a 5-min window. Robustness is measured as

$$Robustness_{iy}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^{N} \left(b_{id}^{5,10,20} - \tilde{b}_{id}^{5,10,20} \right)^2}$$

where N is the number of trading days in year y, $b_{id}^{5,10,20}$ are the fix rates for 5, 10, and 20-min windows, and $\tilde{b}_{id}^{5,10,20}$ are the simulated benchmarks without outlier exclusion. The sample period spans from 15 February 2015 to 31 December 2023. ***, ** and * denote significance at 1%, 5% and 10%, respectively. All currencies are pooled together.

	Robustness (Ro_y)					
	All trading days		Normal trading days		End-of-Month trading days	
	10 - 5	20 - 5	10 - 5	20 - 5	10 - 5	20 - 5
Panel A: 2015						
Mean	−0.00830	0.01870	−0.01454	0.00833	0.11366	0.22151
t-test	−0.3934	0.5640	−0.6949	0.2584	0.7897	0.8496
# Obs.	1356	1356	1290	1290	66	66
Panel B: 2016						
Mean	−0.06216	−0.10187	−0.06413	−0.09568	−0.02194	−0.22882
t-test	−2.9588**	−3.0124**	−3.0946***	−2.9793***	−0.1402	−0.7377
# Obs.	1547	1547	1475	1475	72	72
Panel C: 2017						
Mean	0.03396	0.02432	−0.02513	−0.03310	0.41266	0.64490
t-test	2.2118*	0.8060	−1.5829	−1.7866	3.7585***	3.2384***
# Obs.	1541	1541	1476	1476	72	72
Panel D: 2018						
Mean	−0.03665	−0.06761	−0.06611	−0.06436	−0.10939	−0.13430
t-test	−1.9892*	−2.6755**	−2.3179**	−2.5855***	−0.9685	−0.7165
# Obs.	1548	1548	1476	1476	72	72
Panel E: 2019						
Mean	0.02223	0.01434	−0.00156	−0.05240	0.10769	0.10969
t-test	1.6048	0.6533	−0.0797	−1.5368	0.8391	0.8391
# Obs.	1548	1548	1481	1481	72	72
Panel F: 2020						
Mean	0.00837	−0.03360	0.01561	0.01604	0.21274	0.35314
t-test	0.4207	−0.9659	1.2449	0.5159	1.4698	1.3359
# Obs.	1553	1553	1475	1475	72	72
Panel G: 2021						
Mean	0.00954	−0.00937	−0.00024	−0.04460	−0.19219	−0.41939
t-test	0.4466	−0.4050	−0.0150	−1.5123	−1.5561	−1.6211
# Obs.	1547	1547	1474	1474	72	72
Panel H: 2022						
Mean	−0.00626	−0.04603	−0.01579	−0.00549	−0.12939	−0.07534
t-test	−0.3880	−1.5907	−1.1164	−0.2614	−0.9685	−0.5098
# Obs.	1546	1546	1464	1464	72	72
Panel I: 2023						
Mean	−0.01388	−0.00473	−0.01579	−0.00549	0.02498	0.01077
t-test	−1.0076	−0.2284	−1.1164	−0.2614	0.4123	0.0953
# Obs.	1536	1536	1464	1464	72	72

indicates a consistent improvement in these two dimensions for the benchmark calculated over the current 5-min window. Although there is greater variability in the distributions at the end of each month, a more tightly clustered distribution around zero is still evident for the two dimensions mentioned previously.

3.4. Future research directions

In this section, we focus on further analyses that can be carried out in future research to extend our study. One promising avenue involves investigating the influence of alternative benchmarking procedures, such as different statistical estimators, on the 4 pm Fix. The findings from Evans et al. (2018), as represented in Figure 1.9, demonstrate varying efficiencies across different benchmarking procedures. Notably, the median, often a favored estimator for its robustness against outliers, appears less efficient compared to methods such as the trimmed mean or the winsorized mean in capturing price stability. The results suggest that while the length of the fix window (e.g., 60 or 300 s) plays a significant role in benchmark effectiveness, the choice of benchmarking procedure itself could be just as critical.

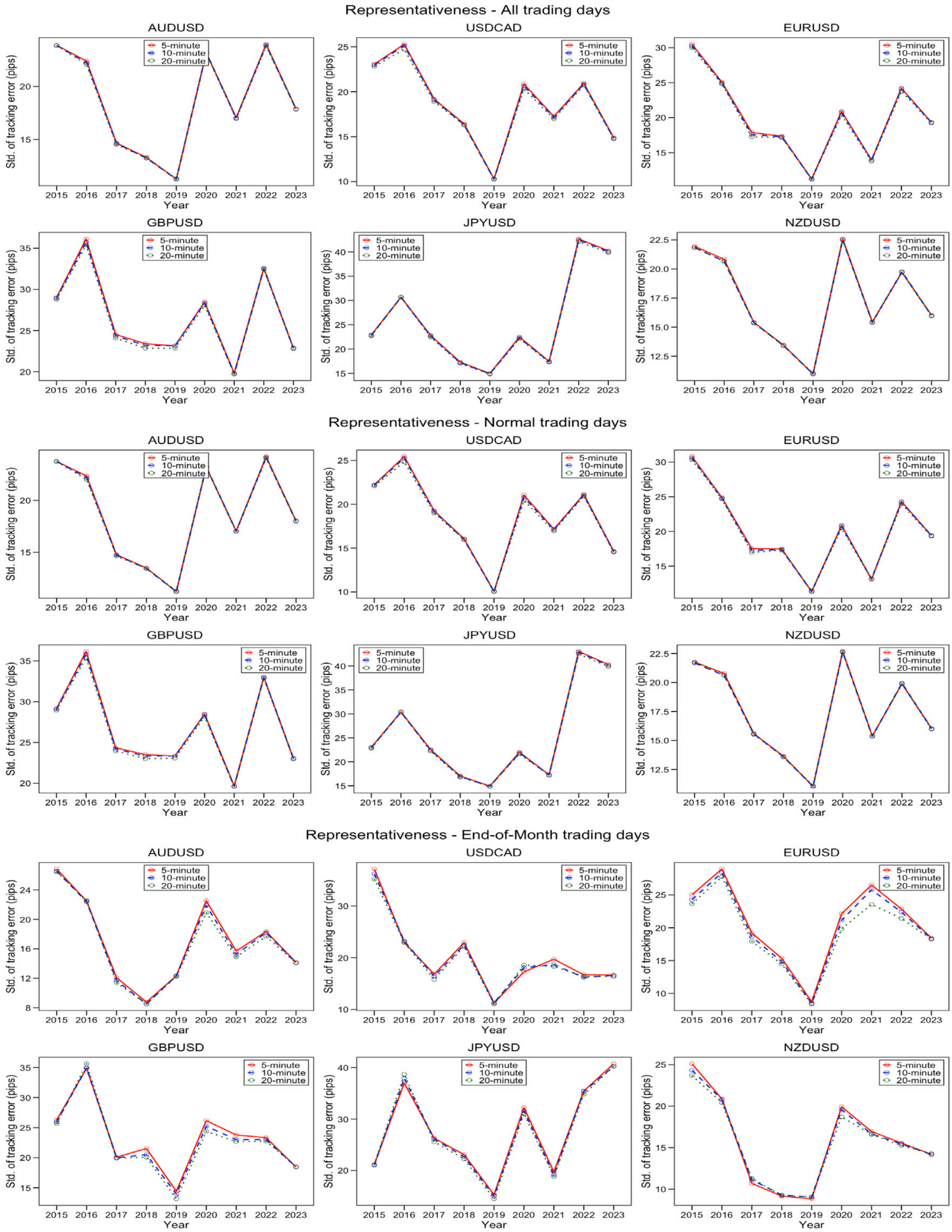

Fig. 1. Benchmark Representativeness Measures by Currency and Year. This figure presents the trend over the years of the standard deviation of the representativeness difference between a 10-min and 20-min benchmark relative to the 5-min benchmark by currency and trading days.

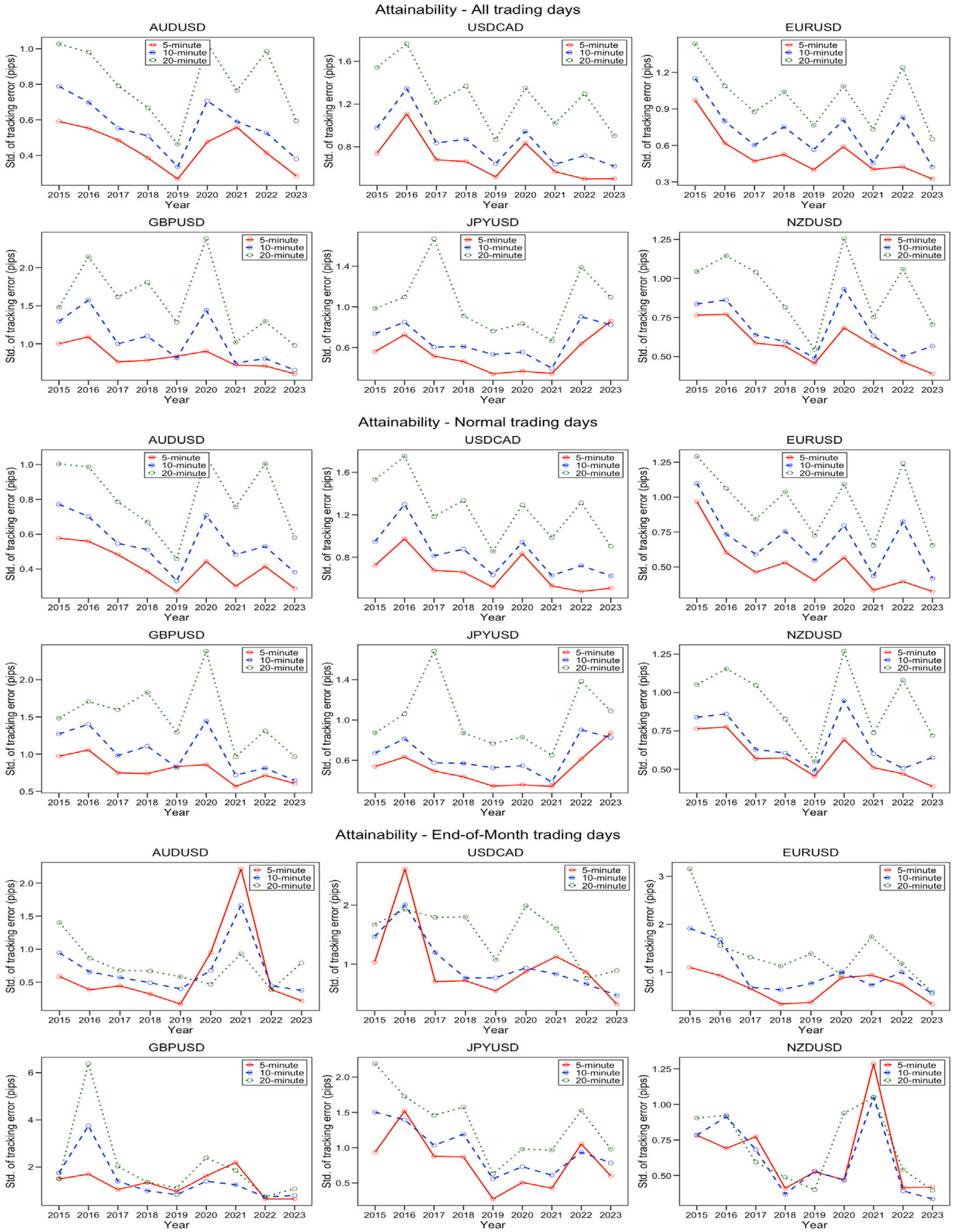

Fig. 2. Benchmark Attainability Measures by Currency and Year. This figure presents the trend over the years of the standard deviation of the attainability difference between a 10-min and 20-min benchmark relative to the 5-min benchmark by currency and trading days.

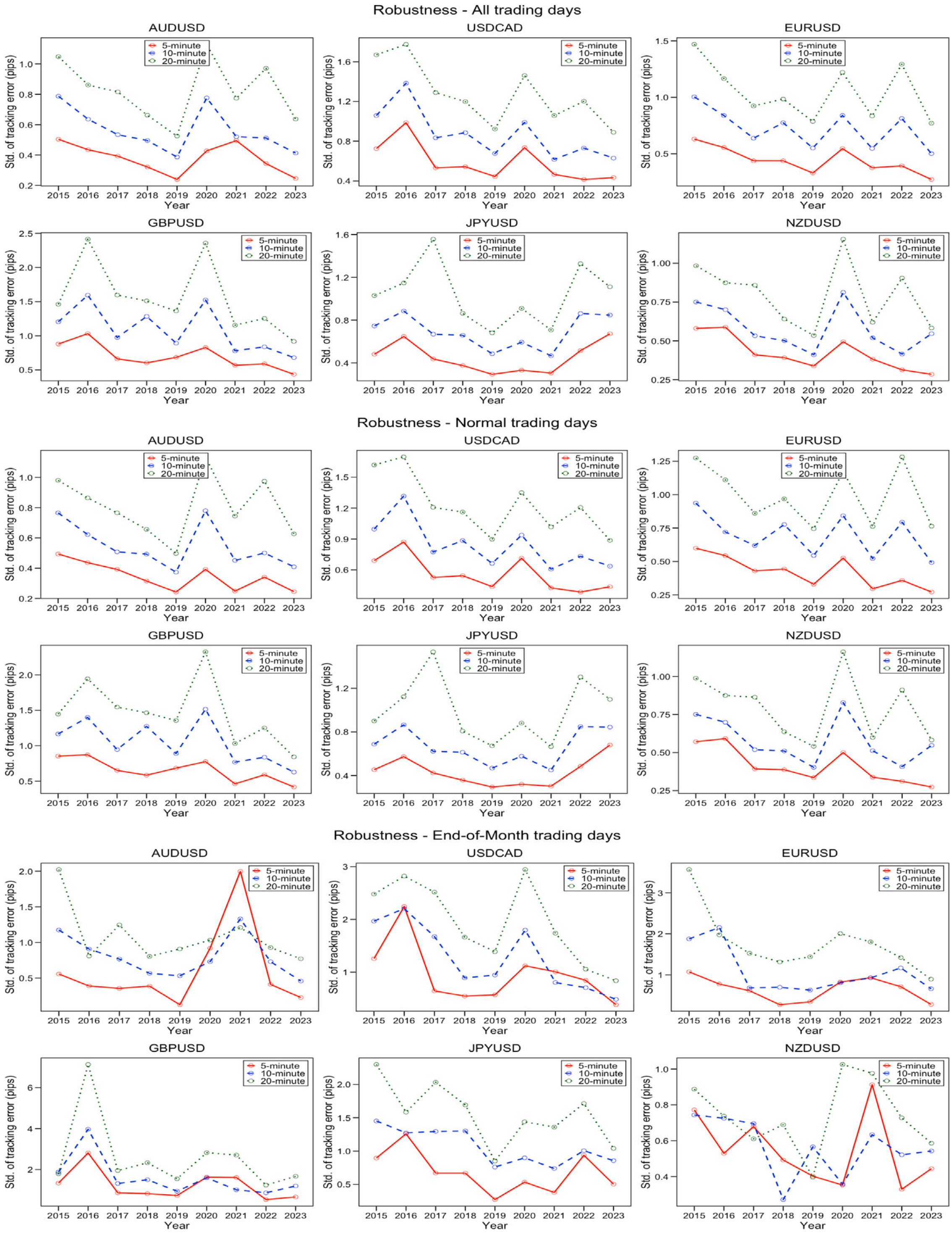

Fig. 3. Benchmark Robustness Measures by Currency and Year. This figure presents the trend over the years of the standard deviation of the robustness difference between a 10-min and 20-min benchmark relative to the 5-min benchmark by currency and trading days.

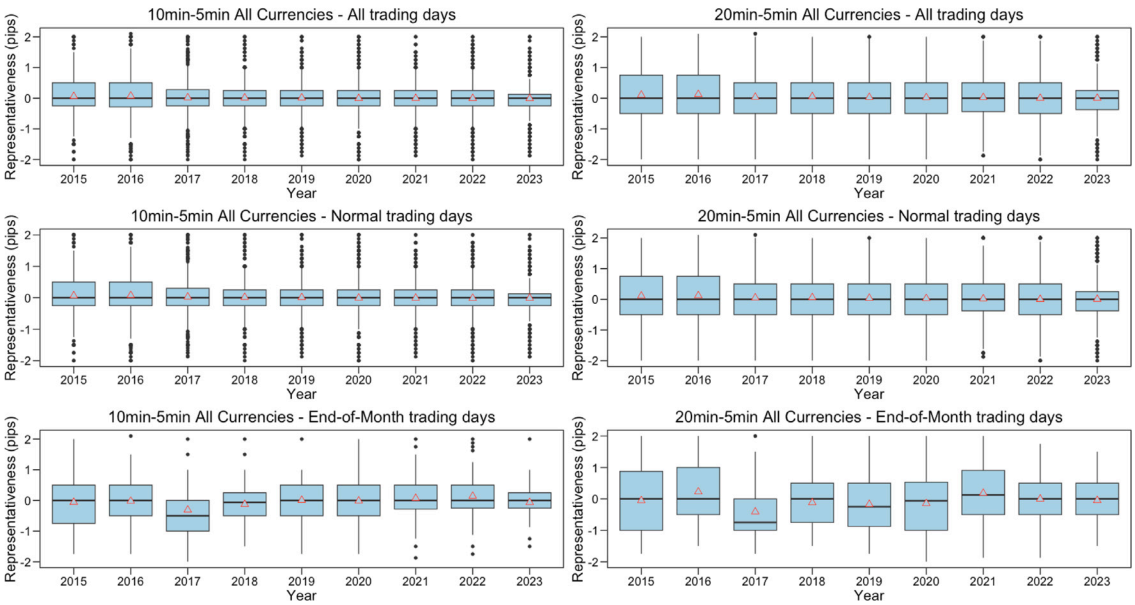

Fig. 4. Benchmark Representativeness Difference in Window by Year. This figure presents the trend over the years of the distribution (box plot) of the representativeness difference between a 10-min and 20-min benchmark relative to the 5-min benchmark by trading days. All currencies are poled together.

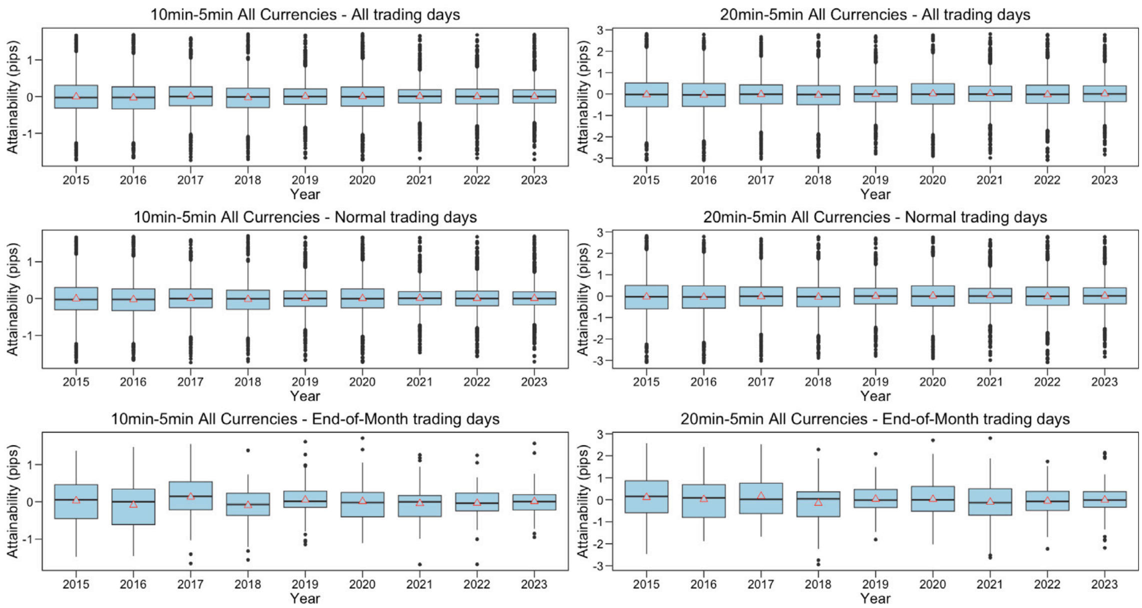

Fig. 5. Benchmark Attainability Difference in Window by Year. This figure presents the trend over the years of the distribution (box plot) of the attainability difference between a 10-min and 20-min benchmark relative to the 5-min benchmark by trading days. All currencies are poled together.

Future research could explore this potential oversight in the industry, where the focus has largely been on adjusting the calculation window, without sufficiently addressing the statistical techniques underlying the benchmark calculations. For example, it may be beneficial to assess how different data filtering techniques (e.g., further variations in trimmed or winsorized means) or even more sophisticated, robust estimators like Hodges–Lehmann could improve the benchmark's ability to reflect true market conditions. These alternative approaches may be especially relevant in periods of high market volatility, such as month-end trading days, where extreme values can distort price estimates. By shifting focus from the window length to a broader exploration of benchmarking methodologies, future studies could provide new insights into improving the robustness and accuracy of the WMR methodology.

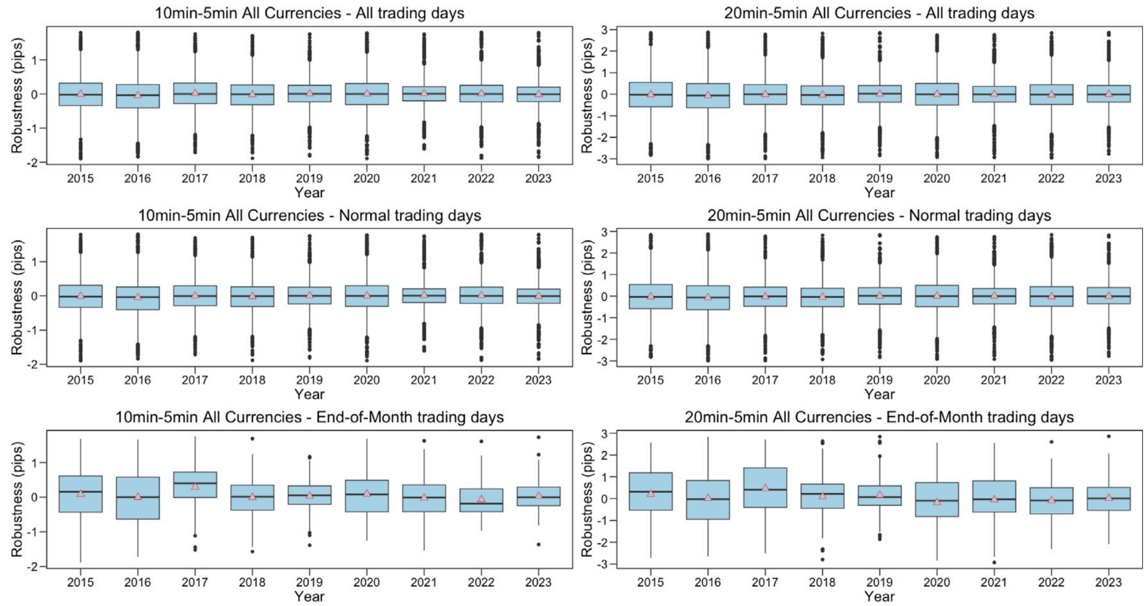

Fig. 6. Benchmark Robustness Difference in Window Lengths by Year. This figure presents the trend over the years of the distribution (box plot) of the robustness difference between a 10-min and 20-min benchmark relative to the 5-min benchmark by trading days. All currencies are poled together.

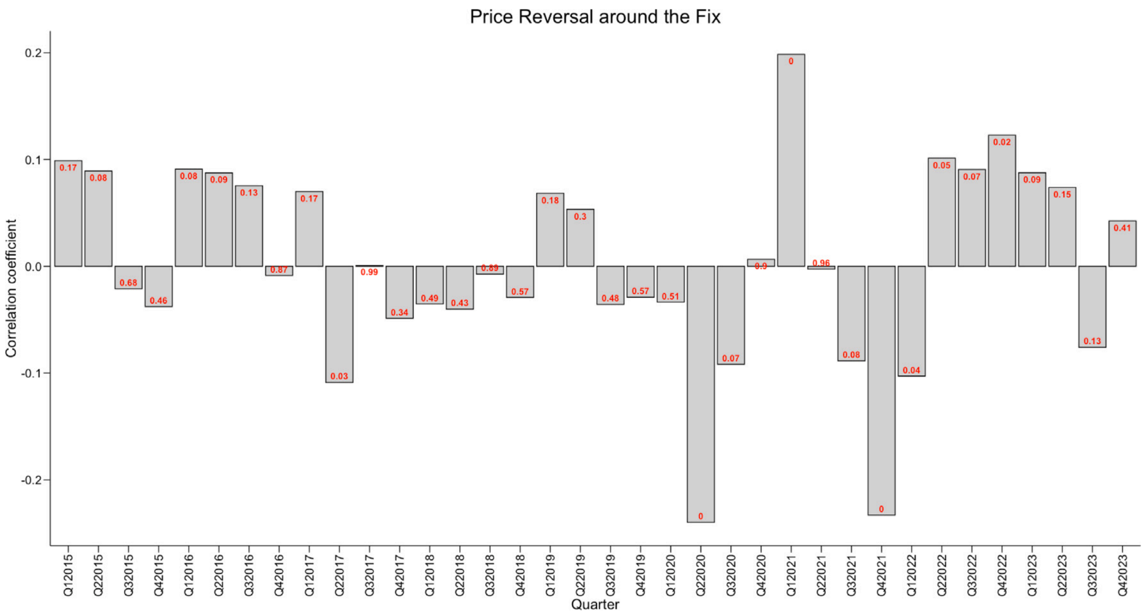

Fig. 7. Price Reversal around the Fix. The figure shows the correlation coefficient of currency returns around the fix and the associated p-values (in red) by quarter, computed from a Z-Fisher transformation. The market efficiency metric is computed as follows:

$$MarketEfficiency = \text{cor}\left(\frac{p_2 - p_1}{p_1}, \frac{p_3 - p_2}{p_2}\right)$$

where p_1, p_2, and p_3 are the average prices 15 min before, during, and 15 min after the fix, respectively.
All currencies are pooled together.

4. Discussion

In this section, we discuss the results presented throughout the paper to provide an interpretation of their economic significance and possible implications for industry practice.

4.1. How has the market impacted the 4 pm fix in terms of liquidity, trading activity, transaction and adverse selection costs, and price volatility?

In terms of the multivariate analysis, our interpretation focuses on the predictive power of the independent variables rather than the specific coefficients' signs. The sign of the dependent variable (the benchmark rate), in response to changes in the independent variables, does not carry direct economic meaning. This is because the benchmark rate itself is a market-derived measure and cannot inherently 'increase' or 'decrease' in a causal sense relative to these variables. Instead, the key insight lies in understanding how variations in market conditions — such as the bid–ask spread — are related to the behavior of the benchmark rate, particularly during high-intensity trading periods like month-end. Thus, the focus remains on these relationships and the implications for the robustness and reliability of the benchmark rate formation process.

The results suggest that the bid–ask spread, while not influencing the published benchmark rate during normal trading periods, does have a noticeable impact during end-of-month trading days. This insight sheds light on the underlying market dynamics, emphasizing that the spread adjusts frequently during high-volume trading periods, which in turn affects the overall published benchmark rate. Variations in market conditions, such as liquidity and volatility, influence the benchmark rate differently during normal and high-intensity trading periods, highlighting the interplay between market dynamics and benchmark rate formation. However, these results do not necessarily point towards the ineffectiveness of the WMR methodology underlying the 4 pm Fix. Whether the 5-min benchmark window remains a robust and reliable benchmarking calculation procedure for assessing market conditions across different currencies and trading periods remains a question that our univariate analysis seeks to address.

4.2. Can the WMR methodology be improved by the lengthening of the time window in which the benchmark is calculated?

The results of the univariate analysis provide compelling evidence regarding the effectiveness of the 5-min WMR benchmark window in capturing market representativeness. Notably, the statistical analysis across multiple currencies shows that representativeness is almost never significant when tested against that of benchmarks computed over longer time windows, suggesting that the 5-min window provides an accurate reflection of the market, comparable to the broader 10- or 20-min windows. This finding holds across all currencies except for GBPUSD or EURUSD, where a slight deviation is observed with the 20-min benchmark window. However, this exception does not warrant a comprehensive change in methodology, as reliance on one currency's behavior is insufficient to drive global methodological changes. Importantly, the lack of significant differences supports the current methodology's effectiveness without the need for complex and costlier adjustments.

Further, the analysis highlights a consistent trade-off between the attainability and robustness of the benchmark rates, consistent with the literature (Evans et al., 2018). Although longer windows could slightly enhance robustness, they simultaneously reduce attainability, as traders find it more difficult to replicate the rates based on extended windows. This trade-off is particularly evident in the case of currencies like AUDUSD and USDCAD, where longer windows offer diminishing returns in terms of market reflection. These findings reinforce the notion that the 5-min window strikes a balanced compromise between being attainable and robust, making it an optimal choice for the benchmark.

4.3. To what extent has the methodology that underpins the WMR benchmark remained effective since its inception in terms of representativeness of the market, attainability, and robustness?

Our results also demonstrate that when analyzing all currencies together, the overall trends confirm the efficacy of the WMR methodology. Except for the year 2016 during normal trading days and 2017 during end-of-month trading days, the effectiveness of the 5-min window is consistently reaffirmed across multiple years and currencies. Even during end-of-month trading days, a period that typically exhibits higher volatility and market activity, the statistical differences between the 5-min window and longer windows remain minimal. These findings are crucial, as end-of-month transactions are particularly important for portfolio rebalancing, and the consistency of the results suggests that the WMR methodology remains reliable even during these critical periods.

Despite some outlier results for EURUSD and GBPUSD during end-of-month days, these deviations do not provide sufficient evidence to advocate for a significant change in the benchmark methodology, especially when counterbalanced by the other effectiveness measures. A singular focus on these currencies does not justify altering a system that works effectively across a broader range of currencies. This is further corroborated by the figures that illustrate the relative consistency of the 5-min window across the majority of currency pairs and years. It is evident that the WMR methodology is well-suited to its purpose, providing a stable and manipulation-resistant benchmark that reflects market realities without the need for extending the benchmark window or altering the methodology substantially unless revised in other aspects that are currently not at the center of industry debates.

5. Conclusion

This study emerges in response to pressing needs within the global financial industry, particularly addressing the complexities surrounding the FX benchmark process. Our research primarily focuses on an extensive analysis of the WMR fix, from its inception in 2015. This extended scrutiny reveals insights into trading behavior, particularly during end-of-month transactions, a period crucial for investors, corporations, and fund managers who seek to mitigate currency risks.

In the vast realm of the global FX market — a market processing transactions nearing $7.5 trillion daily — benchmarks like the WMR fix play a critical role. They not only facilitate portfolio rebalancing for fund managers, but also allow corporations to value currency holdings consistently, thereby circumventing the need for continuous market monitoring. This capability is crucial for minimizing the risk of market manipulation, which is a top priority for both commercial and financial stakeholders relying on these benchmarks.

Our study extends and complements existing literature by not only evaluating the WMR benchmark's efficacy and its reflection of market liquidity, but also by proposing and testing for potential improvements. By examining month-end trading activity, we offer a distinct understanding of how FX benchmarks can better serve to hedge entire months' worth of transactions. Further, our empirical findings suggest that the current WMR methodology might not benefit from adjustments in the benchmark window duration. Results indicate that extending the time window for the WMR fix from the standard 5-min duration to potentially 10- or 20-min could enhance its representativeness and robustness minimally, thereby supporting the proposition that it is not necessary or cost-effective to change how the entire business currently functions. Extending the length of the window would necessitate the re-assessment of algorithms aimed at identifying manipulative behavior within surveillance systems and the reallocation of multiple resources. Therefore, a change in the length of the window does not imply more stable portfolio valuations and a reduction of the susceptibility to sudden market shifts.

The multivariate analyses provide a somewhat different perspective, combining various market quality metrics to detect changes in market behavior around the fixing window that could anticipate price movements. This evidence supports the notion that the current fixing rate may be sub-optimal on month-end days and highlights the need for ongoing revisions to ensure it remains representative of market dynamics. However, the bid–ask spread is found to not affect the 4 pm Fix except during the window itself, thereby supporting the view that abnormal price movements ahead of the window do not significantly influence the current published rates.

In conclusion, our research contributes significant, actionable insights that can assist FX benchmark administrators and market participants in refining their strategies and methodologies. By ensuring that the benchmarking process accurately reflects the evolving market landscape, the financial industry can safeguard against potential vulnerabilities, thereby protecting investor assets and maintaining market integrity. Our suggestion is that changing the length of the benchmark window may not provide significant improvements overall, and therefore it is too costly for the industry to re-adapt to such a change. Further research could focus on other aspects of the methodology underpinning the 4 pm fix to provide new insights as to whether those different factors could perhaps improve the latter.

CRediT authorship contribution statement

Matteo Benenchia: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Investigation, Formal analysis, Data curation. **Luca Galati:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrew Lepone:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the helpful comments by Robert Faff (the editor), an anonymous reviewer, Claudio Lupi, and Keith Elliott. This work was supported by the Rozetta Institute (formerly CMCRC-SIRCA).

Data availability

Data not available / The data that has been used is confidential. To access the code used to produce the results of this study, please refer below: Benenchia, Matteo; Galati, Luca; Lepone, Andrew (2024), "To fix or not to fix, the Fix: Reassessing the effectiveness of the 4pm Fix. A pre-registered study", Mendeley Data, V1, doi: 10.17632/zbdgtt8mmt.1.

References

Aspris, A., Foley, S., O'Neill, P., 2020. Benchmarks in the spotlight: The impact on exchange traded markets. J. Futures Mark. 40 (11), 1691–1710. http://dx.doi.org/10.1002/fut.22120.

Benenchia, M., Galati, L., Lepone, A., 2024. To fix or not to fix: The representativeness of the WM/R methodology that underpins the FX benchmark rates. A pre-registered report. Pac.-Basin Finance J. 84, 102311. http://dx.doi.org/10.1016/j.pacfin.2024.102311.

Celick, S., 2012. The more contagion effect on emerging markets: The evidence of DCC-GARCH model. Econ. Model. 29 (5), 1946–1959. http://dx.doi.org/10.1016/j.econmod.2012.06.011.

Chiang, T., Bang, N., L., H., 2007. Dynamic correlation analysis of financial contagion: Evidence from Asian markets. J. Int. Money Finance 26 (7), 1206–1228. http://dx.doi.org/10.1016/j.jimonfin.2007.06.005.

Evans, M., 2018. Forex trading and the WMR fix. J. Bank. Financ. 87, 233–247. http://dx.doi.org/10.1016/j.jbankfin.2017.09.017.

Evans, M., O'Neill, P., Rime, D., Saakvitne, J., 2018. Fixing the fix? Assessing the effectiveness of the 4pm fix benchmark. FCA Occas. Pap. 46.

Faff, R., 2023. *PBFJ* editorial . . . engaging with responsible science. "OPEN FOR BUSINESS" – Launching the *PBFJ* pre-registration publication initiative. Pac.-Basin Finance J. 79, 101837. http://dx.doi.org/10.1016/j.pacfin.2022.101837.

Galati, L., 2024. Exchange market share, market makers, and murky behavior: The impact of no-fee trading on cryptocurrency market quality. J. Bank. Financ. 165, 107222. http://dx.doi.org/10.1016/j.jbankfin.2024.107222.

Galati, L., Webb, A., Webb, R.I., 2024. Market behaviors around bankruptcy and frozen funds withdrawal: Trading stranded assets on FTX. J. Econ. Bus. 106196. http://dx.doi.org/10.1016/j.jeconbus.2024.106196.

Marsh, I.W., Panagiotou, P., Richard, P., 2017. The WMR fix and its impact on currency markets. Norges Bank Publ..

Melvin, M., Prins, J., 2015. Equity hedging and exchange rates at the London 4 p.m. Fix. J. Financial Mark. 22, 50–72. http://dx.doi.org/10.1016/j.finmar.2014.11.001.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838. http://dx.doi.org/10.2307/1912934, arXiv:1912934.

Matteo Benenchia is a Ph.D. Candidate in Finance at the University of Trento (Italy) and Macquarie University (Australia). He is also an Industrial Ph.D. Researcher in Market Microstructure at the Rozetta Institute (Australia), and a Foreign Exchange Benchmark Surveillance Specialist at Refinitiv, an LSEG (London Stock Exchange Group) business (UK). His research interests include market microstructure, FX trading, and time series analysis. He holds a Bachelor of Science in Economics from the University of Marche and a Master of Science in Finance from the University of Trento (Italy).

Luca Galati is a Research Fellow at the University of Bologna. He was previously a Ph.D. Candidate in Finance and Accounting at the University of Wollongong and the University of Molise, an Industrial Ph.D. Researcher at Rozetta Institute, an FX Benchmark Surveillance Specialist at Refinitiv, and the co-founder of 110 Laude, a University of Molise Fintech spin-off. His research interests include empirical financial market microstructure, fintech economics, and forensic accounting. He was also a visiting research fellow at Florida Polytechnic University in 2023.

Andrew Lepone is a Professor of Finance at Macquarie University (Australia) and the Chief Research Officer at Rozetta Institute (formerly Capital Markets CRC-SIRCA). He is actively involved in industry-oriented research in equity and futures market microstructure, particularly related to the areas of market design and market quality. As Rozetta Institute's Chief Research Officer, Prof Lepone manages all matters associated with the interaction between Rozetta Institute and its university and industry partners.