

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Pacific-Basin Finance Journal

journal homepage: [www.elsevier.com/locate/pacfin](http://www.elsevier.com/locate/pacfin)

# To fix or not to fix: The representativeness of the WM/R methodology that underpins the FX benchmark rates. A pre-registered report

Matteo Benenchia <sup>a,b,c</sup>, Luca Galati <sup>c,d,e,\*</sup>, Andrew Lepone <sup>b,c</sup>

<sup>a</sup> Department of Economics and Management, University of Trento, Via Calepina 14, Trento, TN, 38122, Italy

<sup>b</sup> Department of Applied Finance, Macquarie University, Balaclava Rd, Macquarie Park, 2109, NSW, Australia

<sup>c</sup> Rozetta Institute (formerly CMCRC-SIRCA), Level 9/1 Macquarie Pl, Sydney, NSW, 2000, Australia

<sup>d</sup> School of Business, University of Wollongong, Northfield Ave, Wollongong, 2500, NSW, Australia

<sup>e</sup> Department of Economics, University of Molise, Via Francesco De Sanctis, Campobasso, 86100, CB, Italy

## ARTICLE INFO

## JEL classification:

G14

G15

## Keywords:

FX benchmark rates

Financial regulation

Market microstructure

WM/Reuters

4 pm window

## ABSTRACT

This paper examines the effectiveness of the WM/Reuters (WM/R) methodology for FX benchmark rates since its revision in 2015. After a series of market manipulation scandals, the WM/R methodology underwent significant changes to enhance its reliability and minimise the risk of market manipulation. However, the dynamic nature of the FX market, characterised by evolving investor behaviour and continuous trading, necessitates a reassessment of this benchmark's quality. Using a comprehensive dataset of proprietary order book data from Refinitiv, our empirical study critically evaluates the WM/R methodology's current standing in the FX market. We focus on representativeness, attainability, and robustness to determine whether lengthening the time window of the benchmark increases its reliability. We then focus on liquidity, trading activity, transaction costs, and price volatility to determine the ability of the WM/R methodology to minimise market impact during the 4 pm calculation window. Our findings aim to inform future modifications to the FX benchmark, ensuring its alignment with the evolving market environment and maintaining its role as a crucial reference point for global financial activities. This study contributes to the ongoing discourse on effective financial benchmarking, particularly in the high-stakes and continuously active realm of foreign exchange markets.

## 1. Introduction

## 1.1. Brief background and context

Benchmarks play a critical role in financial markets, especially Foreign Exchange (FX), which is characterised by high fragmentation and persistence of bilateral trading. The reduction of information asymmetry and their use as a reference in the valuation of portfolios and investments are among the main goals that can be achieved in developing a robust, attainable, and representative benchmark. Unlike many others, the FX market lacks a closing time (and price) and is active 24/7. This peculiarity underpins the need for a benchmark to allow market participants a representative rate each day. As of today, there are three main FX benchmarks available in the market, with each following a different methodological calculation: the WM/Reuters (WM/R) 4 pm

\* Correspondence to: Department of Management, University of Bologna, Via Capo di Lucca 34, Bologna, 40124, BO, Italy.

E-mail address: [luca.galati@unibo.it](mailto:luca.galati@unibo.it) (L. Galati).

<https://doi.org/10.1016/j.pacfin.2024.102311>

Received 31 January 2024; Received in revised form 21 February 2024; Accepted 24 February 2024

Available online 1 March 2024

0927-538X/© 2024 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

fix administered by Refinitiv, the ECB fix, and the BFIX administered by Bloomberg. Nevertheless, as highlighted by the Financial Stability Board in 2014 (FSB, 2014), the daily fix rates published by WM/Reuters at 4 pm London time “... are by far the dominant benchmark being used, not just in FX but also as a key input in multi-currency equity, bond, and credit indices”.

To contextualise the importance of this benchmark in actual figures, in the first two months of 2023, \$85 billion were traded in only one currency pair (GBP/USD) across all currencies for which Refinitiv provides a benchmark, with \$4.6 billion in only the 5 min representing the benchmark window.<sup>1</sup> Moreover, according to the latest report by the Bank of International Settlement on OTC FX turnover (BIS, 2022), trading in over-the-counter FX markets reached \$7.5 trillion per day in April 2022, against the ‘only’ \$551 billion traded in the equity market over the same period of analysis.<sup>2</sup>

Despite its size and the concentration of trades during the benchmark window calculation that many authors have evidenced in their research (Melvin and Prins, 2015; Michelberger and Witte, 2016; Marsh et al., 2017; Evans et al., 2018), proper market surveillance was not implemented after the scandal unfolded in 2013 when investigations on major financial institutions revealed how traders easily manipulated the fix in collusion practices with each other, front-running their clients’ orders. Aquilina et al. (2022) provided an exhaustive overview of the most recent scandals concerning the manipulation of benchmarks in the financial markets, including the WM/R.

In 2015, after a consultation initiated by the Financial Stability Board, the WM/R fix methodology was revised to best minimise the risk of market manipulation, and a dedicated Refinitiv Market and Surveillance team (M&S) was commissioned by the Financial Conduct Authority (FCA) to develop an effective surveillance system to enable the identification of potential market manipulation activities by market participants. The new methodology, which primarily saw a lengthening of the window calculation from the original 1 min around 4 pm London Time to the actual 5 min centred at 4 pm, brought positive effects in reducing market manipulation, malpractices, and those trading strategies negatively affecting investors’ performance (see Evans et al. (2018), Ito and Yamada (2017) for the main findings).

However, given the little existing research available in the literature, and in response to the ongoing industry debate on the imperfection of the FX markets (e.g., Szalay, 2020), further progress on the development of a new benchmark design better representing the underlying market characteristics and conditions is yet to be made.

### 1.2. Current methodology

The WM/R fix benchmark calculation methodology is published and updated regularly by Reuters (RBSL, 2023). All currency pairs in the WM/R world for which a benchmark is provided can be clustered into two segments: the so-called “traded currencies” (TCs) and the “non-traded currencies” (NTCs) or “quoted currencies” (QCs). The distinction is made based on the degree of liquidity for each specific currency on the three trading platforms sourced by WM/R: Refinitiv Matching, Electronic Broking Services, and Currenex. These data sources are reviewed periodically, assessing their representativeness of the underlying market for each currency.

The benchmark for the 39 TCs indicated by WM/R is determined by sampling from each respective reference platform trades information during a 5-minute window centred at 4 pm UK Time, precisely from 3:57:30 pm to 4:02:29.999 pm UK Time. Every second, a random trade is snapped from the window, becoming part, after several validation checks, of the final sample for the calculation of the benchmark. A bid (ask) rate is then computed starting from the ask (bid) rate sampled from the platforms, applying the prevailing order spread at the time of the trade. From the final sample, a median trade bid and trade offer are calculated independently, and the mid-rate deriving from these values will be published as a daily WM/R benchmark for the currency pair of interest, following a further validation check.

The methodology for the calculation of NTCs (or QCs) benchmarks follows similar procedures but is based on limit order data from a multi-contributor instrument code created by Refinitiv. Every 15 s, a best-bid (best-ask) is captured and, along with its respective best-ask (best-bid), is sampled after several validation checks, with the resulting median rate calculated and published as a benchmark following the same procedure as for the TCs.

### 1.3. Research questions

This research aims to provide empirical evidence based on a unique set of proprietary data provided by Refinitiv to ensure that the methodology underpinning the WM/R benchmark calculation remains representative of the market. The study will, therefore, answer the following research questions:

1. *To what extent has the methodology that underpins the WM/R benchmark remained effective since its inception in terms of representativeness of the market, attainability, and robustness?*
2. *Can the WM/R methodology be improved by the lengthening of the time window in which the benchmark is calculated?*
3. *Has the WM/R methodology minimised the FX benchmarks market impact around the 4 pm calculation in terms of liquidity, trading activity, transaction costs, and price volatility?*

<sup>1</sup> Self-produced calculations based on proprietary data provided by Refinitiv.

<sup>2</sup> Source: World Federation of Exchanges (<https://www.world-exchanges.org/>).

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

Building upon the initially approved research pitch in [Appendix A](#), this report further outlines the planned methodology for our “Engagement & Impact” study, which forms Phase 3 of the PBFJ pre-registration publication process, as per [Faff \(2022\)](#). The subsequent section provides an overview of the relevant literature and the study’s motivations ([Section 2](#)), leading to a presentation of the main concept and the value proposition of the key stakeholders ([Section 3](#)). [Section 4](#) describes the empirical framework of the study, including the data and methods (tools). The report concludes in [Section 5](#), highlighting the innovative aspects, impacts, and contributions of the proposed study. Subject to approval, this study will be thoroughly conducted, and its findings will be reported subsequently.

## 2. Literature review and motivation

### 2.1. Key papers

The works of [Evans et al. \(2018\)](#), [Melvin and Prins \(2015\)](#) and [Evans \(2018\)](#) critically underpin this study.

The centrality of the WM/R benchmark in the FX market is emphasised in [Evans \(2018\)](#), who provides evidence of the atypical trading activity and associated patterns surrounding 4 pm UK Time. We use the study of [Evans \(2018\)](#) to motivate our research and address the inefficiency issue of the 4 pm Fix. Further, the study of [Evans et al. \(2018\)](#) focuses on the shift of the Fix benchmark from 1-minute to 5-min. Although our sample period does not include an event of this nature as the FX market still relies on the 2015 methodology reform, we follow the empirical approach of [Evans et al. \(2018\)](#) throughout our analysis through the identification of potential changes in market quality measures’ and benchmark effectiveness’ trends. Compared to both ([Evans, 2018](#)) and [Evans et al. \(2018\)](#), we extend the length and depth of the data used in previous studies by analysing a time period from the implementation of the WM/R fix in 2015 to the present to provide insight into any change in trading behaviour around the 4 pm Fix. This longer sample period allows the examination of recent economic events at the epicentre of the ongoing debates within the FX industry.

The study by [Evans et al. \(2018\)](#) plays a crucial role in understanding the effectiveness of the 4 pm Fix benchmark in the FX market. Their approach identifies measures of representativeness, attainability, and robustness through outlier trades as potential manipulation. However, while their analysis controls for month-end trading, it does not thoroughly investigate trading patterns during this period, which many industry professionals rely on to rebalance their portfolios. [Melvin and Prins \(2015\)](#) emphasise the significance of month-end trading, revealing that trading activity spikes during this period, potentially affecting benchmark rates. This heightened activity at month-end is crucial for understanding market dynamics and the potential for manipulation, as the increased volume and strategic trades can impact the benchmark’s representativeness and reliability. Relating this back to the potential for manipulation, it is evident that the unique trading behaviour at month-end, coupled with the identification of outlier trades, underscore the importance of vigilant benchmark management (and design) and the need for robust methodologies to ensure the integrity of FX benchmarks. We, therefore, use the study of [Melvin and Prins \(2015\)](#), linked to the strong use case within practice, to motivate our additional focus on month-end analysis.

### 2.2. Literature review

[Evans et al. \(2018\)](#) are the first to examine the effectiveness of the WM/R 4 pm Closing Spot Rate by using unique full-order book data and proposing three dimensions for the benchmark evaluations: representativeness, attainability, and robustness. The authors suggested an improvement in the functioning of the reference market during the fixing (representativeness) after the introduction of the new benchmark methodology in 2015, showing that price reversals after the fix mostly disappeared. As predicted by the authors, the long-established trade-off between attainability and robustness holds after the lengthening of the benchmark window in 2015, reducing the likelihood of market manipulation behaviour at the cost of increased tracking error in replicating the benchmark. Earlier research by [Ito and Yamada \(2017\)](#) analyses the change in banks’ trading behaviour after the introduction of the new methodology, anticipating some of the results in [Evans et al. \(2018\)](#). According to [Ito and Yamada \(2017\)](#), banks changed their optimal execution model, placing orders in a more evenly distributed manner during the new 5-minute window. They argue that this happened to minimise any use of private information from their customers, although this created new profit opportunities for non-bank participants.

By analysing 21 currency pairs from 2000 to 2013, [Evans \(2018\)](#) suggests the presence of a negative auto-correlation between pre- and post-fixing rate changes not found during normal FX trading. [Marsh et al. \(2017\)](#) and [Michelberger and Witte \(2016\)](#) support similar evidence in their studies on intraday foreign exchange rates and the inter-dealer order flow relationship around the WM/R 4 pm London Fix. The former confirms that price behaviour and flows observed during the fixing period are not comparable with the same trend at other points in time, both in terms of trading activity and volumes and considering the price reversal pattern in the post-benchmark window. [Michelberger and Witte \(2016\)](#), instead, is the first to question the role of the WM/R benchmark as the best reference point for investors. They document a sharp increase in volatility and in the probability of extreme price movement during the fixing and conclude that the use of the WM/R fixing can reduce investors’ returns and increase their portfolio volatility, contradicting the best execution principle. The lack of literature about recent potential changes in trading behaviour by market participants is a further gap to be filled. This will enable us to understand whether the methodology is still in step with current practices or if traders tend to take positions ahead of the benchmark window, resulting in pre-window price swings not considered in the benchmark fixing.

A further critical peculiarity of FX trading around the benchmark window is rooted in the theory of the “hedging channel of exchange rate adjustment”. The role of the WM/R benchmark as the main reference FX rate for the evaluation of international portfolios results in the adoption of hedging strategies by fund managers that were concentrated at month-end. As highlighted by Melvin and Prins (2015), this behaviour leads to a spike in trading activity during the fixing window on the last trading day of the month, with a consequent exchange rate predictability based on the fund-underlying equity market performance in the month ahead of the hedging. Being fully aware of trade volume dynamics at month-end, this study cannot escape from correctly discerning between end-of-month and tranquil trading days to deliver accurate results about any changes in patterns in the FX trading environment.

### 2.3. Motivation

Considering that previous studies on WM/R FX benchmark rates are primarily concerned with analysing price dynamics surrounding the fixing window, and associated manipulation techniques (Evans, 2018), very little research exists that examines liquidity indicators for such a large inter-dealer venue. Evans et al. (2018) investigate the effectiveness of the Fix by focusing primarily on normal trading days. As mentioned above, Melvin and Prins (2015) demonstrate a spike in trading activity in the spot market around the Fix at month-end. Therefore, prior research could have underestimated the effectiveness of the WM/R benchmark, as trading activity around the Fix at month-end significantly differs from that of non-month-end days. Further, its value substantially exceeds the total of all regular trading days within the referenced month. Benchmark designers need, thus, a robust methodology based on a more accurate representation of the underlying market.

Further, a top central banker from the Reserve Bank of Australia and the Global FX Committee has recently warned that fund managers trading currencies could be in breach of their duty to clients as exchange rate swings have doubled, increasing costs for investors and benchmark users (Szalay, 2020). Given the ongoing debate and recurrent news in major financial media on the imperfection of FX benchmarks, further advancements in the way these are established (and supervised) are required.

## 3. Idea and key stakeholders

### 3.1. Idea

After 27 years of serving investors who value their portfolios and benchmark their foreign exchange trades and 8 years since the last change in the methodology, the Refinitiv Benchmark Oversight Committee is working on a new public consultation. To address the concerns about high trading costs and the negative impact of predatory trading activities during the fixing window, WM/Reuters fixing is carrying out, for the first time in history, an analysis studying the market participants' trading behaviour to support the business in making decisions about whether or not to change the window or develop and enhance the methodology accordingly.<sup>3</sup> The goal is that the methodology that underpins the WM/R numbers stays representative of the market. Stated differently, minimising the market impact around 4 pm implies that portfolio managers who rely on the 4 pm fix are not negatively impacted.

As a direct response to this industry's needs, this research will provide evidence of potential changes in market behaviour around the benchmark window calculation, offering a comprehensive analysis concerning the evolution of liquidity and price impact metrics over time since the inception of the WM/R methodology. This will help support the client consultation with Benchmark user groups and discussions with major central banks such as the Reserve Bank of Australia, the Central Bank FX Committee, and the Bank of Canada.

In April 2020, members of an FX group arranged by the European Central Bank (ECB) expressed concern that ending the month with end-of-month trading at specific FX fixes was anticipated to increase volatility and put pressure on the efficiency of the market. The consultation was therefore prompted by the ECB after the media announced a potential spark alarm for unusual price swings ahead of the FX fixes. We *hypothesise* that the methodology that underpins the WM/R 4 pm benchmark calculation is no longer representative of the market to the extent that trading activity, as measured by trading frequency, spread, transaction costs, volatility, and price impact, around the 5-minute benchmark window, is driven by abnormal moves in currencies well ahead of the Fix calculation. In other words, unusual market behaviour in the run-up to the 4 pm fix is causing the benchmark to no longer be a reliable reference rate, which in turn implies that fund managers are failing to search for better deals by relying on an inefficient WM/R benchmark and thus countering their client's best interests.

### 3.2. Key stakeholders

Keeping as reference the critical work of Duffie et al. (2018) on the design of a benchmark fixing robust against manipulation practices, the primary essence of this research is to inform the creation of an optimal FX benchmark through a description of the WM/R benchmark's efficacy and the market liquidity it refers to. Multiple stakeholders would benefit from the outcome of this research:

<sup>3</sup> See also the news for reference at [www.reuters.com/business/exclusive-refinitiv-examines-changes-key-daily-FX-fixing-window-2021-09-06/](http://www.reuters.com/business/exclusive-refinitiv-examines-changes-key-daily-FX-fixing-window-2021-09-06/) or [www.FX-markets.com/trading/7946921/in-a-bit-of-a-fix-refinitiv-seeks-ideas-to-improve-wmr](http://www.FX-markets.com/trading/7946921/in-a-bit-of-a-fix-refinitiv-seeks-ideas-to-improve-wmr).

1. **Financial institutions.** Fund managers and, more broadly, banks frequently rebalance their portfolios by relying on the WM/R benchmarks or using them as input in the pricing of their investment solutions. Their main efforts are centred on the search for profit maximisation and the development of trading strategies aimed at reducing transaction costs. They do not know whether FX benchmarks lack representativeness of the market and therefore continue to use WM/R rates, although it might be costly and inefficient and go against their clients' best interests. They could have invested additional resources in market monitoring if they knew that the FX benchmark was manipulated. Offering solutions for the design of an efficient benchmark would facilitate investors' strategies in rebalancing their portfolios, minimising their currency risks, monitoring costs, and satisfying their clients' objectives more easily and efficiently.
2. **Regulators.** To guarantee investors' protection and minimise market abuse behaviour, regulators constantly monitor financial markets and therefore invest considerable resources to implement optimal regulations. As of today, the scarcity of literature about the FX benchmarks does not support any regulators' decisions, despite the importance and impact of the use of the WM/R fix. They need solid up-to-date evidence that frequently assesses changes in FX trading behaviour to assist in the development of best solutions for the market.
3. **Central banks.** The FX market is one of the key channels for the transmission of monetary policy, and central banks regularly quantify their main money aggregates. In some cases, central banks also act as providers of FX benchmarks (i.e., the ECB fix rate) or as official setters of their daily official exchange rate. Having a broad, comprehensive view of the health of the FX markets would help ensure that central banks can rely on a clearer picture in the determination of their reference rates.

These stakeholders have different objectives and necessitate the best trade-off solution to deliver the most efficient outcome. Refinitiv is the benchmark provider and contemporaneously the interlink between all the stakeholders as it delivers under the regulators' guidance by asking clients (financial institutions) what their needs are and including central banks and external committees in the conversation (i.e., consultation). Researchers can facilitate discussion to tackle this problem by offering a third-party and independent view on the topic with reliable evidence-based research findings.

For all the key stakeholders mentioned above, this research can provide new insights about changes in market behaviour that would affect the future design of any new FX benchmark methodology. Financial institutions can reduce costs in rebalancing their portfolio based on a benchmark that minimises the market impact of fixed-based transactions. They would also benefit from more representative benchmarks to serve their clients' needs more efficiently. Regulators and central banks would prefer to avoid issues related to FX manipulation. They would benefit from a more comprehensive view of the FX market microstructure based on a deeper analysis of the market liquidity spanning several years, so they could rely on fair and representative rates that would prevent scandals like the LIBOR case.

## 4. Empirical design

### 4.1. Data

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 WM/Reuters 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, the platforms used by the benchmark administrator for the fixings are three, namely 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 CAD/USD), and the most liquid in the Asia-Pacific region (i.e. AUD/USD, NZD/USD, JPY/USD, and CNH/USD). 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 Trading Economics database. Recalling Section 1.2, for all the currencies in the sample, WM/R calculates the benchmark following the TCs procedure, given their high liquidity profile. The sample period available for the analysis covers over 9 years, from February 15, 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 for each respective platform. 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.

### 4.2. Tools

To support FX benchmark designers and benchmark users in assessing the effectiveness of the existing 5-minute benchmark window and informing on new possible changes in the methodology, this analysis uses trading platform data to infer changes in trading behaviour over an almost 9-year period, comparing year-on-year trends across the major and most liquid currency pairs traded in the market. Although the analysis does not run a natural experiment around a change in the benchmark calculation methodology, it will follow the approach in [Evans et al. \(2018\)](#) for assessing the effectiveness of the WM/R benchmark over the sample period, investigating the most relevant FX market microstructure components as deeply described in [Evans and Rime \(2019\)](#).



Differently from Evans et al. (2018), this analysis focuses on discerning trading behaviour during normal trading days against the end-of-month activity in response to hedging strategies developed by institutional investors (Melvin and Prins, 2015).

To address the first and second research questions, we employ a methodology similar to Evans et al. (2018)<sup>4</sup> by examining the effectiveness (i.e. quality) of the 4 pm fix benchmark through their proposed measures of representativeness, attainability, and robustness. Representativeness provides insight as to whether the benchmark is reliable compared to its representation of the underlying market. Attainability describes the degree to which a market participant, following a calculation pattern consistent with the benchmarking procedure, can replicate the benchmark price. Robustness refers to the resiliency of the benchmark to market manipulation (i.e. outlier trades).

Since Evans et al. (2018) find that a change in the benchmarking procedure by the length of the window, when examined by itself, does not have a mechanical effect on the representativeness of the benchmark rates, we first focus on the endogenous effect of increasing the length of the benchmark window from 5- to 10-min, and then to 20-min, for currency  $i$  on day  $d$  driven by changes in the behaviour of market participants, as follows:

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

where  $N$  is the number of trading days,  $b_{id}^{5,10,20}$  are the fix rates calculated following the current WM/R methodology around a 5-, 10- and 20-minute window and  $p_{id}^d$  is the median trade price for currency  $i$  on day  $d$ .<sup>5</sup>

We then focus on the other two measures of the benchmark's effectiveness, namely attainability and robustness, computed as:

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

where  $p_{id}^{5,10,20}$  is the mean price within each of the different time windows selected; and

$$Robustness_{id}^{5,10,20} = \sqrt{N^{-1} \sum_{d=0}^N (b_{id}^{5,10,20} - \tilde{b}_{id}^{5,10,20})^2} \tag{3}$$

where  $\tilde{b}_{id}^{5,10,20}$  are the 'dirty' simulated benchmarks calculated in each window following the current WM/R methodology but without applying, after the sampling, Refinitiv's validation checks to exclude observations classified as outliers which are potentially manipulative in nature (referred to as the 'manipulation term' in Evans et al. (2018)).

For the three measures described above, for each benchmark  $b_5$ ,  $b_{10}$ , and  $b_{20}$ , we test the null hypothesis that the benchmark is representative of the market, attainable, and robust by undertaking  $t$ -tests on whether the measures are equal to zero. When the  $t$ -statistic is significantly greater than the critical value, the null hypothesis is rejected, supporting the ineffectiveness of the benchmark. We then compare each measure of the actual Fix (over the 5-minute window) with the same measures computed on the simulated benchmarks over the extended windows of 10- and 20-min to examine for variability in the distributions taken separately, and test the mean difference between the distributions. In addition, we calculate the difference between the actual benchmark effectiveness measures and the same measures computed for the simulated benchmark on the extended windows (e.g.,  $Representativeness_{id}^{10-5} = Representativeness_{id}^{10} - Representativeness_{id}^5$ ), and plot the variation of the distribution through time to infer whether a benchmark calculated over an extended window would have been more effective.

Consistent with Evans et al. (2018), we also compute a measure of market efficiency by examining price dynamics around the benchmark window through a correlation analysis of short-term reversals, as follows:

$$MarketEfficiency = \text{cor} \left( \frac{p_2 - p_1}{p_1}, \frac{p_3 - p_2}{p_2} \right) \tag{4}$$

where  $p_1$ ,  $p_2$ , and  $p_3$  are the market-wide average prices in the 15 min before the fix, during the fix, and the 15 min after the fix, respectively. We undertake  $t$ -tests on whether the correlation in each quarter of our sample period is equal to zero to test whether the market inefficiently reacts around the benchmark window through time.

To address the third research question, instead, we employ an event study methodology similar to Aspris et al. (2020), thus examining the time directly surrounding each daily 4 pm Fix. We compute different liquidity measures following the mainstream

<sup>4</sup> Although we share a similar aim with the study of Evans et al. (2018), we are unable to fully replicate their analysis given the absence of trader-level information and the lack of an effective window length change that they use to execute an event study.

<sup>5</sup> Our dataset does not provide information on volume amounts for all currencies analysed; we are thus unable to compute volume-weighted average prices (VWAP). However, the actual WM/R methodology clearly states that volumes are not considered in the calculation of the benchmark rates, reducing the relevance of this information. Further, Rinaldo and Somogyi (2021) show that some agents in FX markets are always more informed than others, irrespective of their trade size, which is not consistent with findings in equity markets. This suggests that volume-weighted prices, which are historically used in equity market research, are not particularly representative in FX markets. Finally, a common characteristic of FX market data is that price series have a skewed and leptokurtic distribution and may not be normally distributed (Celick, 2012), especially in Asian FX markets (Chiang et al., 2007), thus making the comparison between a median (i.e., the benchmark) and a mean of two different price series statistically unreliable and inaccurate.

approach in the market microstructure literature. Consistent with [Aspris et al. \(2020\)](#), we measure liquidity in basis points first as the raw bid–ask spread:

$$bas_{it} = a_{it} - b_{it} \tag{5}$$

where  $a_{it}$  is the lowest ask price prevailing on either venue for currency  $i$  at time  $t$ , and  $b_{it}$  is the highest bid price. Consistent with the timestamp used in [Aspris et al. \(2020\)](#), we calculate the measures within 5-second buckets. Although the focus of this paper is on the “raw” bid–ask spread, the empirical literature also refers to the percentage spread in cents calculated as  $rbas_{it} = (a_{it} - b_{it}) / m_{it}$ , where  $m_{it}$  is the prevailing midpoint between the best ask and best bid prices,  $m_{it} = (a_{it} + b_{it}) / 2$ .

Following both [Evans et al. \(2018\)](#) and [Aspris et al. \(2020\)](#), we compute the cost of a round trip transaction in basis points for the liquidity demander as measured by two times the difference between the transaction price and the prevailing quote midpoint:

$$espread_{it} = 2q_{it} (p_{it} - m_{it}) / m_{it} \tag{6}$$

where  $p_{it}$  is the transaction price for currency  $i$  at time  $t$ ,  $m_{it}$  is the midpoint for currency  $i$  at the time of transaction  $t$ , and  $q_{it}$  is the trade direction being 1 for buyer-initiated transactions and  $-1$  for seller-initiated transactions. Consistent with the studies mentioned above, we look at the price impact in basis points, a metric for market resilience, which measures the direction of change in midpoint price after each transaction:

$$pi_{it} = 2q_{it} (m_{it+20} - m_{it}) / m_{it} \tag{7}$$

where  $m_{it}$  is the midpoint for currency  $i$  at the time of transaction  $t$ ,  $m_{it+20}$  is the prevailing midpoint 20 s after the trade, and  $q_{it}$  is the trade direction being 1 for buyer-initiated transactions and  $-1$  for seller-initiated transactions. This metric represents the implicit transaction costs that liquidity demanders have to pay, less the portion of those costs that can be attributed to liquidity supplier revenues.

We also look at the market depth in terms of quoted sizes at the best bid and offer (i.e., BBO), defined as:

$$BBO_{it} = as_{it} + bs_{it} \tag{8}$$

where  $as_{it}$  is the size of the best ask for currency  $i$  at time  $t$  and  $bs_{it}$  is the size of the best bid. Consistent with [Aspris et al. \(2020\)](#), we focus on the value of the market depth in dollar terms, which also includes information about the price of the quotes and is given by  $vBBO_{it} = a_{it} \times as_{it} + b_{it} \times bs_{it}$ . We finally calculate trading activity measured by counting the number of transactions and best quotes within the 5-second buckets, and volatility by computing the standard deviation within each 5-second bucket.

To understand how the underlying market changed and whether the market quality for currency pairs has been affected, we compare all the metrics explored around the fixing window considering a time frame from 30 min before to 30 min after 4 pm UK time, which is consistent with [Aspris et al. \(2020\)](#).<sup>6</sup> We estimate  $t$ -tests on the null of whether each measure is equal to 0. First, we compare each market quality metric on normal trading days with month-end days. We then focus on month-end days and estimate  $t$ -tests on each measure for each year to understand whether the market participants’ trading behaviour changes through time. This allows us to show liquidity, trading activity, and transaction cost changes over the years and provide inputs to inform benchmark designers on whether or not a change in how the 4 pm Fix is calculated should be made.

Although we are unable to replicate the natural experiment analysis in [Evans et al. \(2018\)](#) due to the lack of a formal change in benchmark methodology since 2015, we develop a panel multivariate regression estimation that includes the market quality metrics described above, but, in contrast with prior studies (e.g., [Evans et al. \(2018\)](#) and [Aspris et al. \(2020\)](#)), using them as regressors to explain the benchmark rate, our dependent variable. The control variables’ vector includes the currency pair volatility within the fixing window and a macroeconomic announcement variable taking an intensity value from 1 to 3 as classified in the Trading Economics Database. Currency- and time-fixed effects are also considered in the analysis. The specification of our model in Eq. (9) evaluates the predictive power of changes in market behaviour ahead of the calculation of the benchmark, as follows:

$$Y_{id} = \alpha + \beta_1 X_{id}^{-30} + \beta_2 X_{id}^{-25} + \dots + \beta_7 X_{id}^0 + \beta_8 K_d + \beta_9 K_d X_{id}^{-30} + \dots + \beta_{15} K_d X_{id}^0 + \delta C_{id} + \gamma FE_{id} + \epsilon_{id} \tag{9}$$

where  $Y_{id}$  is the actual 5-minute benchmark rate for the currency pair  $i$  in a given day  $d$ ,  $X_{id}$  is a vector of the seven market quality metrics described above-calculated 30-min before the window through the window itself around 4 pm UK time,<sup>7</sup> and  $K_d$  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 after 4 pm UK time, capturing the marginal predictive power of the benchmark from changes in market behaviour at month-end.  $C_{id}$  is the vector of control variables for currency  $i$  on day  $d$ , and  $FE_{id}$  captures the currency and time fixed effects. The remaining  $\epsilon_{id}$  is the error term while  $\alpha$  is the intercept.

To estimate the model, we first examine whether the market quality metrics (i.e.,  $X_{id}$ ) are correlated with each other at the different time intervals considered (i.e.,  $-30$  minutes,  $-25$  minutes, etc.). This involves testing both the correlation matrix and an indicator of multicollinearity (i.e., VIF). We estimate the model by separating each time interval surrounding the benchmark window.

<sup>6</sup> Event times are examined relative to the start of the Fix. We also considered several alternative windows, both shorter (e.g., 15 min) and longer (e.g., 45 min).

<sup>7</sup> Each time from  $-30$  to 0 refers to a centred range of 5 min (e.g.,  $-30$  indicates the range of minutes from  $-32.5$  to  $-27.5$  relative to 4 pm UK time).

The model specified in Eq. (10) is a simplified version of Eq. (9) that captures the predictive power of market changes in each time interval surrounding the benchmark window taken separately, as follows:

$$Y_{id}^t = \alpha^t + \beta_1^t X_{id}^t + \beta_2^t K_d^t + \beta_3^t K_d^t X_{id}^t + \delta^t C_{id}^t + \gamma^t F E_{id}^t + \epsilon_{id}^t \quad (10)$$

with  $t$  being each 5-minute interval considered for the calculation of the market quality measures (from 30-min before the start of the benchmark window to 4 pm UK time).

## 5. Conclusion

### 5.1. What's new?

The novelty of this research is first in the idea, initiated as a response to an urgent industry need. Additionally, we extend the length and depth of the data used in previous studies by analysing a time period from the implementation of the WM/R fix in 2015 to the present to provide insight on any changes in trading behaviour around the 4 pm fix in this longer sample period, which also accounts for examination of recent economic events at the epicentre of the ongoing debates within the industry. Lastly, we use other tools, compared to the literature, to measure the quality of the FX market during underestimated (by prior research) trading periods. Our analysis will also focus on end-of-month trading behaviour around the Fix, which is the most vulnerable time for investors, corporations, and fund managers alike concerned with balanced portfolios against currency risks, and where most trading activity consequently occurs.

### 5.2. So what?

The global FX market is the largest financial market in the world, with an average daily transaction of almost \$7.5 trillion as of April 2022. Fund managers frequently rebalance their portfolios by relying on FX benchmarks. Additionally, global corporations might also be motivated to value their currency holdings using a standard reference rate. FX benchmarks allow businesses to avoid currency risk and free them from investing resources in market monitoring. Since tying orders to currency fixings is something that both commercial and financial players are interested in, minimising the risk of market manipulation is a priority for WM/R benchmark designers. This research aims to provide empirical evidence that supports FX benchmark administrators in ensuring that the methodology that underpins the WM/R benchmark, considered by the FSB to be the dominant FX benchmark, stays representative of the market. Any negative impacts with the 4 pm Fix affect trillions of dollars, possibly causing a deterioration of related portfolios, which in turn deteriorates retail investors' wealth.

### 5.3. Contribution

The key contribution of our research is to inform the creation of optimal benchmarks through a description of the WM/R benchmark's efficacy and the market liquidity it refers to. We extend the work of Evans et al. (2018) by examining the representativeness of the WM/R methodology that underpins the FX benchmark rates, focusing primarily on month-end trading activity, which accounts for the concentrated whole months' worth of hedging by fund managers. We also significantly add to the body of knowledge on FX microstructure by investigating liquidity measures around the 4 pm Fix during an extended period from the inception of the WM/R methodology to the present and including recent economic events that are the subject of ongoing debate within the industry. The empirical findings of this research will provide practical, business-based insights on the future potential development of the benchmark methodology. This will be based on a comprehensive analysis of the underlying FX trading microstructure rather than merely testing the effectiveness of regulatory changes ex-post as is common in the market microstructure literature. For example, the first specification of our model examines whether a benchmark calculated on a time window extended to 10- or 20-min could be more representative, attainable, and robust than the actual Fix rate (currently based on a 5-minute window). This would provide evidence that the current WM/R methodology is sub-optimal and could be improved by a change in the length of the benchmark window. Alternatively, our results could indicate that longer benchmark windows do not lead to better outcomes and thus do not support changes to the current benchmarking method, which is still an outcome of interest for the industry. Additionally, the multivariate specification, together with the univariate analysis utilising various market quality metrics, can provide evidence of changes in market behaviour around the fixing window that anticipates price movements, suggesting that the current fixing rate is sub-optimal for investors and warrants updating.

### 5.4. Other considerations

The impact perspective (industry goal) is aligned with the academic objective. The collaboration already involves two novel researchers (who are also FX benchmark surveillance specialists) from different Australian universities and an expert senior researcher. Further collaboration with the industry partner, Refinitiv, has been secured with an intellectual property agreement between the two universities, the research centre, and the industry partner. These features will also boost the social impact of the project.

There are no aspects of potential commercialisation, but the industry partner will benefit from the outcome of the research in terms of complying with the regulatory rules and prompts by implementing a new benchmark methodology. The no-results risk



**Table A.1**  
Scholarly pitch.

Research team	Matteo Benenchia, Luca Galati, Andrew Lepone	FoR category	Market microstructure	Date completed	28 April 2023
FOUR	Four aspects of BIG picture framing				
A. Working Title	To fix or not to fix: The representativeness of the WM/R methodology that underpins the FX benchmark rates				
B. Basic Research Question	To what extent has the methodology that underpins the WM/R benchmark remained representative since its inception?				
C. Key Paper(s)	Evans, M., O'Neill, P., Rime, D., & Saakvitne, J. (2018). Fixing the Fix? Assessing the Effectiveness of the 4pm Fix Benchmark. <i>FCA Occasional Paper 46</i> Melvin, M., & Prins, J. (2015). Equity hedging and exchange rates at the London 4 p.m. fix. <i>Journal of Financial Markets</i> , 22, 50-72. Evans, M. (2018). Forex trading and the WMR Fix. <i>Journal of Banking and Finance</i> , 87, 233-247.				
D. Motivation/Puzzle	Considering that previous studies on WM/R FX benchmark rates are primarily concerned with analysing price dynamics surrounding the fixing window, and associated manipulation techniques (e.g., Evans, 2018), very little research exists that examines liquidity indicators for such a large inter-dealer venue. Evans et al. (2018) are the first to examine the effectiveness of the WM/R 4 pm Closing Spot Rate by using unique full order book data. However, they investigate the representativeness of the Fix by only examining normal trading days. From a more practical perspective, Melvin and Prins (2015) demonstrate that there is a spike in trading activity in the spot market around the Fix, particularly around the end-of-month. Therefore, prior research inaccurately assesses the effectiveness of the WM/R benchmark as month-end trading activity around the Fix vastly differs from non-month-end days, and benchmark designers need a robust methodology that incorporates this. Further, recently, a top central banker from the RBA and the Global FX Committee warned that fund managers trading currencies could be in breach of their duty to clients as exchange rate swings have doubled, increasing costs for investors and benchmark users (Financial Times, 2022). Given the ongoing debate and recurrent news in major financial media on the imperfection of FX benchmarks, further advancements in the way these are established (and supervised) are required.				
THREE	Three core aspects of any empirical research project i.e. the "IDioTs" guide				
E. Idea?	After 8 years from its implementation, the Refinitiv Oversight Committee is undertaking a consultation to understand if changes in market behaviour jeopardising the objectives of the current WM/R benchmark methodology. As a direct response to this industry's need, this research will provide evidence of potential changes in market behaviour around the benchmark window calculation, offering a comprehensive analysis concerning the evolution of liquidity and price impact metrics over time since the inception of the WM/R methodology. This will help support the client consultation with Benchmark user groups and discussions with central banks such as the RBA, Central Bank FX Committee, Bank of Canada, etc. The consultation was prompted by the ECB after the media announced a potential spark alarm for unusual price swings ahead of the FX fixes. We hypothesise that the methodology that underpins the WM/R 4pm benchmark calculation is no longer representative of the market to the extent that trading activity, as measured by trading frequency, spread, volatility and price impact, around the 5-minute benchmark window, are driven by abnormal moves in currencies well ahead of the Fix calculation. In other words, unusual market behaviours in the run-up to the 4pm Fix are causing the benchmark to no longer be a reliable reference rate, which in turn implies that fund managers are failing to search for better deals by relying on an inefficient WM/R benchmark and thus counter to their clients' best interests.				
F. Data	<p><b>Unit of Analysis:</b> Individual currency pairs (quotes, trades, and buy/sell direction) and related benchmark (WM/R) <b>Currencies:</b> Major FX pairs, among which three from the Asia-Pacific region, namely JPY/USD, AUD/USD, NZD/USD, and other relevant currencies such as, for example, the Chinese Yuan (where data is available, we will also consider other less liquid currencies); and the two most liquid currency pairs for which WM/R provide a benchmark, namely EUR/USD, GBP/USD.</p> <p><b>Observation Sampling:</b> Intraday trades and quotes (high-frequency tick-by-tick data) <b>Data Type:</b> Currency exchange rates level <b>Sample Period:</b> 2015–2023, Panel data <b>Data source:</b> Proprietary order-book data from Refinitiv Matching &amp; Datascope Select (Refinitiv Tick History Database)</p> <p><b>Limitation:</b> Volumes and counterparties involved in trading are unknown, but the industry partner could provide this information for a subsample. This, however, will not be possible for all the currencies under examination as approximately half of them are traded on other platforms. For example, JPY, CNH, and EUR are mostly sourced from EBS and not from Thomson Reuters Matching. This would allow a perfect replication of the (Evans et al., 2018) metrics of benchmark representativeness, attainability and robustness.</p>				

(continued on next page)

Table A.1 (continued).

G. Tools	To help FX benchmark designers in assessing the relevance of the existing 5-minute benchmark window with a view to reviewing the process (and possibly extending it), we use trading platform data (best bid and ask, trades and trade frequency, available under license from Refinitiv) used to set the benchmarks and look at trading activity (i.e., trading frequency, spread, price impact, price volatility) and changes in those trading activities over a 9-year period and compare year-on-year trends across the major and most liquid currency pairs. Although we do not run a natural experiment around a change in methodology, we follow the approach in <a href="#">Evans et al. (2018)</a> for assessing the effectiveness of the WM/R benchmark over the sample period and investigating liquidity. However, unlike <a href="#">(Evans et al., 2018)</a> , the analysis is undertaken by month-end (where we know the trading behaviour is significantly different) and compared to normal trading days, and we attempt to infer whether a longer window would be beneficial for WM/R benchmark users. Software: R (Open source) for statistical analysis; Excel (licenses available through Macquarie University and Wollongong University) for data cleaning; PowerBI (license available from Refinitiv) for data visualisation.
TWO	Two key questions
H. What's New?	The novelty of this research is first in the idea, initiated as a response to an urgent industry need. Additionally, we extend the length and depth of the data used in previous studies by analysing a time period from the implementation of the WM/R fix in 2015 to the present to provide insight on any change in trading behaviour around the 4 pm fix in this longer sample period, which also accounts for the examination of recent economic events at the epicentre of the ongoing debates within the industry. Lastly, our analysis will focus on end-of-month trading behaviour around the Fix, which is the most vulnerable time for investors, corporations, and fund managers alike concerned with balanced portfolios against currency risks, and where most trading activity consequently occurs.
I. So What?	The global FX market is the largest financial market in the world, with an average daily transaction of almost \$7.5 trillion as of April 2022. Fund managers frequently rebalance their portfolios by relying on FX benchmarks. Additionally, global corporations might also be motivated to value their currency holdings using a standard reference rate. FX benchmarks allow businesses to avoid currency risk and free them from investing resources in market monitoring. Since tying orders to currency fixings is something that both commercial and financial players are interested in, minimising the risk of market manipulation is a priority for WM/R benchmark designers. The aim of this research is to provide empirical evidence that supports FX benchmark administrators in ensuring that the methodology that underpins the WM/R benchmark, considered by the FSB to be the dominant FX benchmark, stays representative of the market. Any negative impacts with the 4 pm Fix affect trillions of dollars, possibly causing a deterioration of related portfolios, which in turn deteriorates retail investors' wealth.
ONE	One bottom line
J. Contribution	The key contribution of our research is to inform the creation of optimal benchmarks through a description of the WM/R benchmark's efficacy and the market liquidity it refers to. We extend the work of <a href="#">Evans et al. (2018)</a> by examining the representativeness of the WM/R methodology that underpins the FX benchmark rates, focusing primarily on month-end trading activity, which accounts for the concentrated whole months' worth of hedging by fund managers. We also significantly add to the body of knowledge on FX microstructure by investigating liquidity measures around the 4 pm Fix during an extended period from the inception of the WM/R methodology to the present and including recent economic events that are the subject of ongoing debate within the industry. The empirical findings of this research will provide insights on the future potential development of the benchmark methodology based on a comprehensive analysis of the underlying FX trading microstructure rather than simply testing the effectiveness of regulatory changes ex-post as is common in the market microstructure literature.
K. Other Considerations	The collaboration already involves two novel researchers (who are also FX benchmark surveillance specialists) from different Australian universities and an expert senior researcher. Further collaboration with the industry partner, Refinitiv, could boost the social impact of the project. The no-results risk is low, as per the aim of the publication initiative. The risk of competition is low given that all the researchers are the only ones involved in this industry-based change. The obsolescence risk is moderate, as previous research provides a good overview of the issue to be analysed. The target journal is the Pacific-Basin Finance journal.

**Table A.2**  
Engagement and impact pitch.

Research team	Matteo Benenchia, Luca Galati, Andrew Lepone	Area of impact	Regulatory, social, and economic – FX market microstructure	Date completed	25 June 2023
FOUR	<b>Four</b> aspects of BIG picture framing				
A. Working Title	To fix or not to fix: The representativeness of the WM/R methodology that underpins the FX benchmark rates				
B. Basic Impact Goal	The aim of this research is to provide empirical evidence that supports FX benchmark administrators to ensure that the methodology that underpins the WM/R benchmark, the dominant benchmark for retail and institutional investors in the largest financial market in the world, remains representative of the market, enabling global corporations and fund managers to rely on a non-manipulated reference rate against currency and macroeconomic risks.				
C. Key Industry/External Triggers	<ul style="list-style-type: none"> <li>• The Refinitiv Oversight Committee is undertaking a consultation to understand if changes in market behaviour are adversely affecting the objectives of the current WM/R benchmark methodology, with the aim of publishing a highly influential industry report.</li> <li>• Over the past 10 years, the Financial Conduct Authority (FCA), the Financial Stability Board (FSB), and the International Oversight Commission (IOSCO) have published several policy reports and occasional papers on the topic.</li> <li>• A leading central banker from the RBA and the Global FX Committee has recently warned that fund managers trading currencies could be in breach of their duty to clients as exchange rate variations have doubled, increasing costs for investors and benchmark users (Financial Times, 2022). This attention to FX markets also has a long-standing history of controversial news in the media, especially on “The Full Fix”, founded and published by Colin Lambert (a well-known journalist and expert in FX markets) who also frequently organises international events.</li> </ul>				
D. Motivation/Problem	The ongoing consultation has been prompted by the ECB after the media announced a potential spark alarm for unusual price swings ahead of the FX fixes. The methodology that underpins the WM/R 4 pm benchmark calculation could no longer be representative of the market to the extent that trading activity is driven by abnormal moves in currencies well ahead of the Fix window. It thus follows that the benchmark may have lost its role of being a reliable reference rate, which in turn implies that fund managers are failing to search for better deals by relying on an inefficient WM/R benchmark and is counter to their clients’ best interests.				
THREE	Three core impact dimensions				
E. Stakeholders	<p>1. <b>Financial institutions.</b> Fund managers and, more broadly, banks frequently rebalance their portfolios by relying on the WM/R benchmarks or using them as input in the pricing of their investment solutions. Their main efforts are centred on the search for profit maximisation and the development of trading strategies aimed at reducing transaction costs. They do not know whether FX benchmarks lack representativeness of the market and therefore continue to use WM/R rates, although it might be costly and inefficient and go against their clients’ best interests. They could have invested additional resources in market monitoring if they knew that the FX benchmark was manipulated. Offering solutions for the design of an efficient benchmark would facilitate investors’ strategies in rebalancing their portfolios, minimising their currency risks, monitoring costs, and satisfying their clients’ objectives more easily and efficiently.</p> <p>2. <b>Regulators.</b> To guarantee investors’ protection and minimise market abuse behaviour, regulators constantly monitor financial markets and therefore invest considerable resources to implement optimal regulations. As of today, the scarcity of literature about the FX benchmarks does not support any regulators’ decisions, despite the importance and impact of the use of the WM/R fix. They need solid up-to-date evidence that frequently assesses changes in FX trading behaviour to assist in the development of best solutions for the market.</p> <p>3. <b>Central banks.</b> The FX market is one of the key channels for the transmission of monetary policy, and central banks regularly quantify their main money aggregates. In some cases, central banks also act as providers of FX benchmarks (i.e., the ECB fix rate) or as official setters of their daily official exchange rate. Having a broad, comprehensive view of the health of the FX markets would help ensure that central banks can rely on a clearer picture in the determination of their reference rates.</p> <p>These stakeholders have different objectives and necessitate the best trade-off solution to deliver the most efficient outcome. Refinitiv is the benchmark provider and contemporaneously the interlink between all the stakeholders as it delivers under the regulators’ guidance by asking clients (financial institutions) what their needs are and including central banks and external committees in the conversation (i.e., consultation). Researchers can facilitate discussion to tackle this problem by offering a third-party and independent view on the topic with reliable evidence-based research findings.</p>				
F. Value Proposition	For <b>all the key stakeholders</b> mentioned above, this research can provide new insights about changes in market behaviour that would affect the future design of any new FX benchmark methodology. <b>Financial institutions</b> can reduce costs by rebalancing their portfolio based on a benchmark that minimises the market impact of fixed-based transactions. They would also benefit from more representative benchmarks to serve their clients’ needs more efficiently. <b>Regulators</b> and <b>central banks</b> would like to avoid issues related to FX manipulation. They would benefit from a more comprehensive view of the FX market microstructure based on a deeper analysis of the market liquidity spanning several years, so they could rely on fair and representative rates that would prevent scandals like the LIBOR case.				

(continued on next page)

is low, as per the aim of the publication initiative. The risk of competition is low given that all the researchers are the only ones involved in this industry-based change. The obsolescence risk is moderate, as previous research provides a good overview of the issue to be analysed. Since the tools are publicly available, the data has been encrypted, and the research team possesses the necessary experience, the study is considered low-risk overall. All associated ethical clearances have been approved.

Table A.2 (continued).

G. Resources	We already have the time (during the PhD programme), funds (PhD scholarships), people (ourselves plus practitioners' expertise from the industry and stakeholder network), intellectual property agreements (between the research centre and industry partner) and skills (after a 2-year training per researcher) to deliver this significant impact. The long-standing collaboration between Refinitiv and the doctoral students' respective Universities, with the intermediation of Rozetta Institute, enabled us to create a network that will facilitate access to resources, data, software, skills, and benchmark end-users.
TWO	Two key impact signals
H. Communication Strategy	The communication with stakeholders will take place through: <ul style="list-style-type: none"> <li>• Partnership: thanks to the collaboration between Refinitiv and Rozetta</li> <li>• Industry publications</li> <li>• Social media: LinkedIn and Twitter</li> <li>• General publishing (The Conversation and The Full FX)</li> <li>• Conference presentations</li> <li>• Face-to-face meetings with the independent FX Oversight Committee and regulators at the FCA</li> </ul>
I. Metrics	Short-term impact metrics: <ul style="list-style-type: none"> <li>• Publication reach</li> <li>• People directly spoken to at the FX Oversight Committee, FCA and Refinitiv itself (at higher employee levels – managerial level)</li> <li>• Sharing on social media and newspapers</li> </ul> Long-term impact metrics: <ul style="list-style-type: none"> <li>• Direct impact on organisations' revenues and risk management</li> <li>• Design of new benchmark methodologies based on the outcomes of this research</li> </ul>
ONE	One bottom line
J. Impact?	The primary outcome is to inform the creation of optimal benchmarks through a description of the WM/R benchmark's efficacy and the market liquidity it refers to. This will enable policy-makers to better defend against market abuse threats, central banks and financial institutions to efficiently rely on representative benchmark rates and therefore efficiently allocate resources and optimally serve their clients' needs, and overall, the industry to comply with stringent rules through better market monitoring and the services they provide worldwide.
K. Other Considerations	The impact perspective (industry goal) is aligned with the academic objective. The collaboration already involves two novel researchers (who are also FX benchmark surveillance specialists) from different Australian universities and an expert senior researcher. Deeper collaboration with the industry partner, Refinitiv, could boost the social impact of the project. We already have an intellectual property agreement between the two universities, the research centre, and the industry partner. There are no aspects of potential commercialisation, but the industry partner will benefit from the outcome of the research in terms of complying with the regulatory rules and prompts by implementing a new benchmark methodology. The risk of competition is low given that all the researchers are the only ones involved in this industry-based change. The obsolescence risk is moderate, as previous research provides a good overview of the issue to be analysed. There might be confidentiality considerations that can, however, be discussed with the industry partner accordingly, given that our aim is of interest to all the stakeholders and Refinitiv. The risk from an impact perspective is low as this is a highly debated topic, and even confirming that the WM/R benchmark is currently representative and is being calculated in a fair and efficient way would provide insights to all relevant stakeholders.

## 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.

## Acknowledgements

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

## Appendix A. Research pitches

The tables below represent the accepted research pitches for Phase 2 of the PBFJ pre-registration publication submission process in the “Engagement & Impact” study. Table A.1 displays the academic overview of the research based on Faff (2015) Pitching Research Framework (PRF). Additionally, Table A.2 illustrates the goals for engagement and impact of this practical research, using the template developed by Faff and Kastle (2016) for pitching research for engagement and impact (PR4EI).

## References

- Aquilina, M., Ibikunle, G., Mollica, V., Steffen, T., 2022. The visible hand: Benchmarks, regulation, and liquidity. *J. Financial Mark.* 61.
- 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.

- BIS, 2022. OTC Foreign Exchange Turnover in April 2022. Technical Report, Bank of International Settlement.
- Celick, S., 2012. The more contagion effect on emerging markets: The evidence of DCC-GARCH model. *Econ. Model.* 29 (5), 1946–1959.
- 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.
- Duffie, D., C., S.J., Piotr, D., 2018. Robust Benchmark Design. National Bureau of Economic Research.
- Evans, M., 2018. Forex trading and the WMR fix. *J. Bank. Financ.* 87, 233–247.
- Evans, M., O'Neill, P., Rime, D., Saakvitne, J., 2018. Fixing the Fix? Assessing the Effectiveness of the 4pm Fix Benchmark, vol. 46, FCA, FCA Occasional Paper.
- Evans, M., Rime, D., 2019. Microstructure of foreign exchange market. SSRN, Working Paper 6.
- Faff, R.W., 2015. A simple template for pitching research. *Account. Finance* 55 (2), 311–336. <http://dx.doi.org/10.1111/acfi.12116>.
- Faff, R.W., Kastle, T., 2016. Pitching Research for Engagement and Impact, 2813096. <http://dx.doi.org/10.2139/ssrn.2813096>.
- FSB, 2014. Foreign Exchange Benchmarks - Final Report. Technical Report, Financial Stability Board.
- Ito, T., Yamada, M., 2017. Did the reform fix the London fix problem? NBER Working Paper No. w23327.
- Marsh, I.W., Panagiotou, P., Richard, P., 2017. The WMR Fix and Its Impact on Currency Markets. Norges Bank Publication.
- Melvin, M., Prins, J., 2015. Equity hedging and exchange rates at the London 4 p.m. Fix. *J. Financial Mark.* 22, 50–72.
- Michelberger, P., Witte, J., 2016. Foreign exchange market microstructure and the WM/Reuters 4pm fix. *J. Finance Data Sci.* (1), 26–41.
- Rinaldo, A., Somogyi, F., 2021. Asymmetric information risk in FX markets. *J. Financ. Econ.* 140 (2), 391–411.
- RBSL, 2023. WM/Refinitiv Fx Benchmarks Spot, Forward and NDF Methodology Guide. Technical Report, Refinitiv, a LSEG company.
- Szalay, E., 2020. Fund Managers Face Scrutiny over Use of FX Benchmarks. *Financial Times*.

## Further reading

- Financial Conduct Authority (FCA), 2014. FCA fines five banks £1.1 billion for FX failings and announces industry-wide remediation programme (Press Release). <https://www.fca.org.uk/news/press-releases/fca-fines-five-banks-%C2%A311-billion-fx-failings-and-announces-industry-wide-remediation-programme>.

**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.