Benchmarking benchmarks*

James Brugler[1], Marta Khomyn[2], Tālis Putniņš[3,4,5],

[1] The University of Melbourne

[2] The University of Adelaide

[3] University of Technology Sydney

[4] Stockholm School of Economics in Riga

[5] Digital Finance Co-operative Research Centre

July 10, 2024

Abstract

Financial benchmarks such as LIBOR underpin the pricing of trillions of dollars of contracts around the world. We evaluate the quality of benchmark prices using a state-space model to separate information from noise. Applying the method to LIBOR benchmarks and their replacements, we find that alternative reference rates (ARRs) are less noisy in four of the five currencies. However, the USD ARR is considerably more noisy, resulting in billions of dollars of noise-related wealth transfers between contract counterparties. We show that benchmark reforms such as expanding the reference market and using a trimmed mean can reduce noise in ARRs.

Keywords: benchmark prices, information shares, price discovery

JEL Codes: G11, G12, G14

*We thank Philipp Schnabl (editor), an anonymous referee, Charlie Cai, Philip Drummond, Huu Duong, Paul Kofman, Paul Lajbcygier, Shane Miller, Murrough O'Brien, Jiri Svec, David Rabinowitz, Kumar Venkataraman, seminar participants at Monash University, Kyiv School of Economics, Baltic International Centre for Economic Policy Studies, 2019 International Accounting and Finance Consortium, 32nd ESA Conference in Economics and Business, 2019 FIRN Annual Conference, 2019 Behavioral Finance and Capital Markets Conference, 2019 Australasian Finance and Banking Conference for helpful comments. We also thank Jiaqi Wang for insightful discussions that helped this paper. Marta Khomyn was a visiting researcher at Chi-X Australia during the time spent on this study. Tālis J. Putniņš gratefully acknowledges funding from the Australian Research Council (ARC DP200101445).

Corresponding author: Marta Khomyn. E-mail addresses: james.brugler@unimelb.edu.au (J. Brugler), marta.khomyn@adelaide.edu.au (M. Khomyn) and talis.putnins@uts.edu.au (T. Putniņš)

Benchmarking benchmarks

Abstract

Financial benchmarks such as LIBOR underpin the pricing of trillions of dollars of contracts around the world. We evaluate the quality of benchmark prices using a state-space model to separate information from noise. Applying the method to LIBOR benchmarks and their replacements, we find that alternative reference rates (ARRs) are less noisy in four of the five currencies. However, the USD ARR is considerably more noisy, resulting in billions of dollars of noise-related wealth transfers between contract counterparties. We show that benchmark reforms such as expanding the reference market and using a trimmed mean can reduce noise in ARRs.

1 Introduction

Noise makes financial markets possible, but also makes them imperfect.

Fischer Black (1986) "Noise" [p. 530]

Benchmark prices are important. In financial markets, benchmarks such as the London Inter-Bank Offered Rate (LIBOR) underpin hundreds of trillions of dollars worth of contracts. Good benchmarks substantially increase welfare (Duffie, Dworczak and Zhu, 2017), promote efficient resource allocation, and reduce costs for market participants. But how does one tell whether a benchmark is any good? How much information is in a particular benchmark price? How noisy are benchmarks? We address these issues by decomposing a benchmark price into information and noise components. Using a state-space model, we estimate the noise in LIBOR and compare that to alternative reference rates (ARRs) that are replacing LIBOR. We show that on the whole, ARRs tend to be less noisy than LIBOR, but with economically important exceptions such as the USD ARR.

Following multiple LIBOR manipulation scandals and decreasing liquidity in the inter-bank market after the 2008-2009 crisis, these systemically important interest rate benchmarks have undergone a major transformation, phasing out LIBOR by the end of 2021, replaced by ARRs. Given that trillions of dollars of transactions and contracts are linked to these benchmarks, including swaps, futures, forwards, options, and consumer and business loans, it is critical to understand their information/noise content and the resulting wealth transfers.[1] For example, a mere 1 basis point benchmark error due to a temporary distortion or noise when $10 trillion of contracts fix on that benchmark, could lead to $1 billion incorrectly transferred between contract counterparties. Noise also interferes with the market's ability to price risks accurately and allocate them efficiently.

To estimate the noisiness of benchmarks, we build on the approach of Menkveld, Koopman and Lucas (2007) by using the state-space representation of sequentially set benchmark rates. The efficient benchmark rate is an unobserved state variable that follows a random walk. The observed rate also includes a stationary component that is a noise term uncorrelated with the state

[1] In the US the LIBOR market footprint (as of Q4 2020) is over ten times the value of the US GDP for the year 2020: $223 trillion vs. $21 trillion (New York Fed, 2021)

innovations. This modeling approach allows for sequential rate-setting regimes during different intraday time periods. We use Monte Carlo simulations to validate our empirical model before applying it to evaluate benchmark quality.

Our analysis of LIBOR and ARR benchmarks of five major currencies yields several key results. First, we find that in four out of five currencies, moves from LIBOR to transaction-based benchmarks result in less noisy benchmarks. This finding is in line with the predictions of Duffie and Dworczak (2021), who advocate for transaction-based benchmarks to replace submission-based LIBOR. Specifically, the noise as a portion of the total benchmark variance, or the “noise share”, in CHF SARON, EUR ESTR, GBP SONIA, and JPY TONAR is lower than in the corresponding LIBOR counterpart.

In contrast, we find that the newly established USD ARR (SOFR) is much noisier than USD LIBOR. Unlike the other benchmarks, SOFR is based on collateralized repo transactions, which makes it vulnerable to extreme spikes based on the supply and demand of the underlying collateral. For example, there are instances in the data where SOFR spikes more than a full percentage point in a single day before reverting to its prior level.

To assess the economic significance of noise in interest rate benchmarks, we use our model to calculate the wealth transfers between contract counterparties that occur because of temporary pricing errors (noise) in the benchmark rates. Each day, we extract the estimated noise component of the reference rates and match it to the notional value of overnight interest rate swaps (OIS) fixing on that rate. We find that the value of noise-related transfers is substantial. For example, if during 2020, all OIS were based on LIBOR, a total of \$345 billion would change hands between contract counterparties as a direct result of noise in the benchmark. In contrast, if the ARRs were used instead of LIBOR, the equivalent number would be smaller in most currencies, but much larger overall (total of \$1.1 trillion) because of the high level of noise in SOFR. These estimates are adjusted for the netting of long and short positions using entity-netted notional values. The annual value of these noise-related wealth transfers using ARRs equates to approximately 2.5% of the total outstanding notional value of OIS contracts and an even bigger proportion of entity-netted notional value.

We also find that well-designed reforms can significantly reduce noise in benchmarks. For example, GBP SONIA experienced a decline in noise share from 54.7% to 11.7% following the SONIA reform of April 23, 2018. In this reform, the reference market was expanded so that the benchmark would be based on a higher volume of transactions and mean trimming was introduced. However, many benchmark design choices involve trade-offs. For example, a broad reference market can benefit a benchmark by being more liquid, but can be detrimental if it increases the heterogeneity of transactions.

While our results quantify one of the advantages of the ARRs, namely the tendency for most of the ARRs to be less noisy than the LIBORs that they replaced, there are downsides of the ARRs. One that has been highlighted in recent papers is that the ARRs eliminate or at least substantially reduce the bank credit risk component that was present in LIBOR. For example, SOFR is based on secured repo funding transactions collateralized by US treasuries and all ARRs are overnight rates. Having credit risk in LIBOR helped banks transfer funding risk to borrowers, thereby lowering loan rates — as bank credit risk rose, LIBOR increased, raising the cost of funds for borrowers with floating rate loans or lines of credit (Kirti, 2022). In contrast, "risk-free" ARRs such as SOFR tend to fall when markets are stressed, encouraging borrowers to draw on their credit lines precisely when bank funding costs spike. That effect may reduce ex-ante incentives to provide bank credit (Duffie, Harry, Luck, Wang, Yang et al., 2022). From a policy perspective, such downsides of ARRs must be weighed up against the pricing efficiency benefits that we document and the robustness against manipulation that motivated the transition to ARRs.

As a practical application, measuring the noise content of different benchmark prices or rates can help inform better benchmark design. While Duffie and Dworczak (2021) propose a theory of optimal benchmark design, our study tackles the empirical evaluation of benchmarks. Because each of the major currencies (CHF, EUR, GBP, JPY, USD) has a slightly different ARR design, comparing how these currencies' LIBOR-to-ARR transitions fared under each currency's unique regime provides insights for benchmark design. Moreover, our approach could be applied to evaluating benchmarks in other settings, like FX fixings, equity price benchmarks, futures settlement benchmarks, and so on.

This paper contributes to several strands of literature. We build on the theoretical papers on optimal benchmark design. Duffie and Stein (2015) argue that LIBOR manipulation scandals highlight the need for LIBOR reforms. The specifics of those reforms are examined in the theoretical papers on robust benchmark design (Duffie et al., 2017; Duffie and Dworczak, 2021), as well as in the empirical papers analyzing whether LIBOR alternatives are better than the original LIBOR rates (Schrimpf and Sushko, 2019; Klingler and Syrstad, 2021). Several papers focus on US LIBOR markets alone. For example, Fassas (2021) studies price discovery in US money markets using Hasbrouck (1995) information shares. Indriawan, Jiao and Tse (2021) show that the SOFR aligns with the Federal Reserve's policy target more closely than LIBOR. To the best of our knowledge, our paper is the first to analyze the information and noise content of all the major LIBOR and ARR rates, and estimate the noise-related wealth transfers in those markets.

Several practitioner- and policymaker-focused studies consider the problem of constructing the term structure of ARRs from the overnight rates. Because ARRs, unlike LIBORs, are not forward-looking, Bai et al. (2022) highlight the challenges involved in publishing and referencing term ARRs. Heitfield and Park (2019) and Skov and Skovmand (2021) propose alternative models that rely on SOFR futures prices to construct forward-looking term reference rates that are conceptually similar to the term LIBOR rates. While we focus on overnight, rather than term rates, we highlight that the noise component of overnight rate enters the calculation of any constructed term rates based on ARRs. Therefore, our inference about noise in overnight ARRs applies to term ARRs as well.

The method in this paper uses the state-space model framework, adapted to the specific issue of separating noise from information in sequentially set prices and determining each price's contribution to the efficient price. The empirical market microstructure literature has developed empirical methods to separate information from noise and attribute information shares in *parallel* markets (rather than *sequential*), most often using vector autoregressive models (see Baillie et al., 2002; Yan and Zivot et al., 2010; Putnins, 2013; and Hasbrouck, 2021). The state-space specification for sequential markets has the advantage of not requiring that each day evolves as a sequence of sub-periods, with each one corresponding to a benchmark. Rather, the model in state-space form keeps the different rate regimes distinct, similar to the 24-hour price discovery setting in Menkveld et al. (2007). Other market microstructure applications using state-space models include Hasbrouck (1999); Figuerola-Ferretti and Gonzalo (2010); Durbin and Koopman (2012); Hendershott and Menkveld (2014). The present study is also related to literature on price informativeness, which uses a variety of different models (e.g., Campbell & Shiller, 1988; Morck, Yeung, & Yu, 2000; Hasbrouck, 1993; and Brogaard, Nguyen, Putnins, & Wu, 2021).

2 State-space estimation of sequential benchmarks

Consider N benchmark interest rates ("reference rates") that are published sequentially, splitting the 24-hour period t into N phases, indexed by τ. For example, the EUR ARR and EUR LIBOR are published on the same day t, but at different times τ (7 am and 11.55 am respectively). The set of benchmark interest rates share a common underlying *efficient* interest rate process, $m_{t,\tau}$, that evolves according to a random walk with volatility that depends on the phase of the day, τ:

$$m_{t,\tau+1} = m_{t,\tau} + w_{t,\tau} \tag{1}$$

$$w_{t,\tau} \sim \mathcal{N}\left(0, \sigma^2_{w_\tau}\right). \tag{2}$$

Let $y_{t,\tau}$ denote the interest rate for the benchmark published at period t, τ. This benchmark interest rate is the sum of the efficient interest rate at that period, defined in Equations (1) and (2), and a pricing error (the noise component):

$$\begin{aligned} y_{t,\tau} &= m_{t,\tau} + s_{t,\tau} \\ &= m_{t,\tau-1} + w_{t,\tau} + s_{t,\tau} \end{aligned} \tag{3}$$

where $s_{t,\tau} \sim N(0, \sigma^2_{s_\tau})$ is a pricing error for the τ^{th} benchmark and $m_{t,\tau}$ and $w_{t,\tau}$ are defined as above.[2] In this structural model, each benchmark shares a common efficient interest rate process ($m_{t,\tau}$) that has a different efficient innovation variance ($\sigma^2_{w_\tau}$) and idiosyncratic pricing error variance ($\sigma^2_{s_\tau}$) depending on the time of day, τ. These variance terms are the key quantities of interest that characterize the quality of different benchmark interest rates.

Following Menkveld et al. (2007), the structural model in Equation (3) has a state-space representation. The state vector is the efficient interest rate process $m_{t,\tau}$ with time-varying volatility. The observation vector $\mathbf{y}_{t,\tau}$ has N elements, one for each of the benchmarks published throughout a single trading day t. In a given time period, t, τ, only the τ^{th} element of $\mathbf{y}_{t,\tau}$ is observed. All other elements of $\mathbf{y}_{t,\tau}$ contain missing values. Noting that $N = 2$ in our application of this model to LIBOR vs. ARRs, the remaining notation focuses on this case, but as per Menkveld et al. (2007),

[2] Note that $1 \leqslant \tau \leqslant N$ and when $\tau = 1$, $\tau - 1$ refers to benchmark N of the previous day $t - 1$.

this can be easily extended to cases with $N > 2$. The observation vector is:

$$\mathbf{y}_{t,\tau} = \begin{cases} \begin{pmatrix} y_{t,1} \\ \cdot \end{pmatrix} & \text{if } \tau = 1 \\ \begin{pmatrix} \cdot \\ y_{t,2} \end{pmatrix} & \text{if } \tau = 2. \end{cases} \tag{4}$$

Using $s = t, \tau$ to index time, Equations (3) to (4) can be written in state-space form as:

$$m_{s+1} = m_s + w_s \tag{5}$$

$$\mathbf{y}_s = I_2 \times m_s + \varepsilon_s \tag{6}$$

where m_s is the unobserved state in period s, w_s is the state innovation (updating of the efficient interest rate) with distribution $w_s \sim \mathcal{N}(0, \sigma^2_{w_s})$ and $\sigma^2_{w_s} = (\sigma^2_{w1}, \sigma^2_{w2}, \sigma^2_{w1}, \sigma^2_{w2}, ...)$, $\mathbf{y}_s$ is the 2×1 observation vector containing missing values in alternating positions each period, and ε_s is the 2×1 noise disturbance vector with distribution $\varepsilon_s \sim \mathcal{N}(0, H)$ where H is a 2×2 diagonal matrix with $(\sigma^2_{s_1}, \sigma^2_{s_2})$ on the diagonal elements. The state innovation and noise disturbance vectors are assumed to be mutually independent at all leads and lags. The design matrix is the two-dimensional identity matrix. The transition matrix and selection matrix are both equal to 1.

The system described by Equations (5) and (6) can be extended to include terms that allow benchmarks to differ by a fixed amount in each period, reflecting possible time-invariant differences in credit risk or liquidity. It can also be extended to capture the effects of control variables, such as each country's central bank policy rate, or other time-varying controls, on each benchmark. These are captured respectively by incorporating a constant term $\mu = (\mu_1, \mu_2)'$ and $\beta'\mathbf{x_t}$ in the observation equation:

$$m_{s+1} = m_s + w_s \tag{7}$$

$$\mathbf{y}_s = \mu + I_2 \times m_s + \beta'\mathbf{x_t} + \varepsilon_s \tag{8}$$

where one element of μ is normalized to zero without loss of generality, such that the non-zero term captures the average spread between the two benchmarks.[3] The parameter vector that loads onto the conditioning variables $\mathbf{x}_t$ contains different estimates of the effect of these control variables on each of the N benchmarks. Equations (7) and (8) represent the econometric model we use to estimate the variances of the innovations to the efficient interest rate process and noise for each benchmark.

The key parameters of the system represented in Equations (7) and (8) are the variances of the efficient interest rate innovations and noise in each publication period: $\sigma^2_{w_\tau}$ and $\sigma^2_{s_\tau}$, $\tau \in \{1, 2, .., N\}$. Using these variance terms, we define the information share of benchmark reference rate τ as the proportion of total variation in the efficient interest rate impounded by benchmark τ:

$$IS_\tau = \frac{\sigma^2_{w_\tau}}{\sum_{i=1}^{N} \sigma^2_{w_i}}. \tag{9}$$

Analogously, we define the noise share of benchmark reference rate i as the noise in rate τ normalized by sum of noise in all reference rates:

$$NS_\tau = \frac{\sigma^2_{s_\tau}}{\sum_{i=1}^{N} \sigma^2_{s_i}}. \tag{10}$$

For each benchmark τ, we also define an information-to-noise ratio (IN):

$$IN_\tau = \frac{\sigma^2_{w_\tau}}{\sigma^2_{w_\tau} + \sigma^2_{s_\tau}} \tag{11}$$

2.1 Model discussion

The state-space model defined in Section 2 is a version of the common levels model in Durbin and Koopman (2012), extended to allow time-varying volatility of the state vector disturbances. Menkveld et al. (2007) use a similar state-space approach to analyze the price discovery contributions of different markets trading cross-listed equities. In their case, the efficient price of each stock is an unobserved state variable.

In our context of benchmark interest rates for overnight borrowing, the efficient interest rate

[3] This normalization is equivalent to a level shift in the state vector equal to the omitted term of μ. Incorporating stochastic, time-varying credit risk and liquidity effects would lead to a more complicated model. Instead, we include control variables for credit risk and liquidity effects in our empirical implementation.

process captures the evolution of the (unobserved) overnight risk-free rate. Klingler and Syrstad (2021) note that theoretically, overnight interest rates are virtually risk-free but in practice are subject to frictions that create deviations. We model this process by assuming each observed benchmark interest rate is the sum of the unobserved risk-free rate and additive noise components reflecting liquidity, regulatory, and (potentially) credit factors that can differ across benchmarks.[4] This underlying data generating process implies that both elements of the design matrix are equal to one, rather than just the first element, as would be the case in the standard common levels model.[5]

Expressing the data generating process in state-space form with Gaussian disturbances demonstrates how estimation can be carried out using Maximum Likelihood techniques and the Kalman filter. A particular advantage of this approach is that the Kalman filter directly handles the missing observations in each reference rate period. A necessary condition for the validity of the Kalman Filter is that the state disturbance and noise disturbance vector are independent of each other and serially independent at all leads and lags. In our model, this implies that the noise process in each reference rate contains no information about the underlying efficient process and vice versa.[6] Additionally, unlike a VAR representation, the Kalman filter generally requires Gaussian disturbances.

We use a diffuse prior for the initial conditions and the L-BFGS algorithm for maximizing the log-likelihood function. We constrain all variance terms to be no smaller than 1e-7 to avoid regions where the likelihood function is undefined. We estimate all models twice, once using a maximum number of iterations of 1,000 to obtain start parameters for a second estimation using a maximum number of iterations of 10,000. In all cases, the estimation converges.

To validate that the proposed empirical methodology can reliably recover the structural model parameters, we simulate benchmark rates using a Monte Carlo procedure and then compare

[4]Although credit risk on overnight borrowing and lending is close to zero (see, e.g., Klingler and Syrstad, 2021), we allow for credit risk disturbances caused by fluctuations in overall economic conditions, which we capture through government bond returns, changes in CDS spreads, and the daily change in the ICE BofA High Yield Option-Adjusted Spreads (OASs).

[5]The unrestricted common levels model also nests a multivariate Local Level model with separate, cointegrated state variables for each benchmark interest rate. In such a case, one of the state variables can be re-expressed as a linear function of the other state variable plus a noise term that cannot be separately identified from the disturbance term in the noise process.

[6]Identification of permanent and noise components of time series under varying assumptions are discussed in Watson (1986) and Hasbrouck (2007).

the estimated IS, NS, and IN ratios to their theoretical true values. The baseline Monte Carlo simulations follow a structural model described in Section 2. Details of the simulation are presented in the Internet Appendix.

3 Assessing the quality of LIBOR vs. ARRs

3.1 Institutional details of the transition from LIBOR to ARRs

Bloomberg calls LIBOR "The world's most important number", given its prominence as (i) the reference rate for financial contracts like swaps, loans, or mortgages and (ii) the benchmark rate used to gauge bank borrowing costs (Bloomberg, 2021). However, LIBOR fell out of favor following a series of rate manipulations by the rate-submitting banks. The 2012–2015 litigation surrounding the LIBOR scandals sped up the coordinated effort by central banks, the International Organization of Securities Commissions (IOSCO), and the Bank for International Settlements (BIS) to replace LIBOR with alternative reference rates (ARRs). Policymakers set December 31, 2021 as the deadline for transition away from LIBOR.

Besides manipulation, regulators also raised concerns about the decreasing liquidity in LIBOR rate-setting transactions. For example, Duffie and Stein (2015) highlight that the market for interbank unsecured borrowing has been in decline in the aftermath of the Global Financial Crisis of 2008/09, mainly because regulators decided to limit unsecured inter-bank borrowing/lending activity. This in turn limited liquidity in the LIBOR rate-setting transactions. These dynamics created a gap between the relatively low value of transactions that *determined* LIBOR and the extremely high value of outstanding contracts that *referenced* LIBOR.[7]

The shortcomings of LIBOR — related to both the methodology and lack of underlying liquidity/reference transactions — are the key reasons for the transition from LIBOR to ARRs. ARR design relies on the principles for robust benchmarks published by IOSCO in 2013. These principles recommend calculating benchmark rates using actual transactions data from deep, liquid

[7] For example, the 2018 presentation on GBP LIBOR transition mentions that in the 3-month GBP LIBOR (the most widely used of the GBP tenors) there was only GBP 187 million in daily deposits value, while the value of financial contracts referencing this 3-month borrowing benchmark was around GBP 30 trillion (Bank of England, 2018). In USD LIBOR, the Duffie and Stein (2015) report suggests that the 3-month USD LIBOR had an average of 25-30 rate-setting transactions, but on the lowest-volume days, there could be as few as three to eight transactions. At the same time, the value of transactions referencing the 3-month USD LIBOR was in the order of trillions of dollars (e.g., USD 107 trillion in gross national value of OTC interest-rate swaps, 65% of which are linked to LIBOR).

markets as opposed to survey data from market participants. The benchmark administrators therefore designed ARRs to (i) rely on overnight (O/N) money markets, which have greater volumes and more depth than longer-dated tenors such as three months, (ii) include transactions by non-bank wholesale counterparties including investment funds and insurance companies, and (iii) in some cases (e.g., SARON and SOFR) draw on secured rather than unsecured transactions. For a detailed overview of the market microstructure of LIBOR and ARR benchmarks, see the Internet Appendix Section 1.

The design of ARRs involves striking a balance between various trade-offs. For example, ARRs based on secured overnight lending markets, as is the case for SOFR and SARON, may be subject to additional rate volatility due to market conditions in investment-grade debt securities that are used as collateral in overnight borrowing. In other words, ARRs may be noisy, as they change for reasons not related to the actual cost of overnight borrowing. Policymakers in the various countries in our sample made different decisions regarding the design of their respective ARRs. Therefore, assessing the information and noise in the different ARRs compared to LIBOR presents a natural experiment that can inform benchmark design more broadly.

Table 1 summarizes the key details of LIBOR and the ARRs, and the relevant transition dates.[8] We identify the key transition milestone for each ARR based on regulatory documents referenced in the notes to Table 1.

[Table 1 about here.]

3.2 Assessing the quality of LIBOR and its replacements

ICE LIBOR, the LIBOR administrator,[9] calculates LIBOR rates in five currencies: CHF, EUR, GBP, JPY, and USD. Hence, we focus on these five currencies. The relevant tenor is the O/N rate, as it matches the tenor of the ARRs (SARON, ESTR, SONIA, TONAR, and SOFR). We obtain daily data on these reference rates from Factset.

For each of the five currencies, we estimate the state-space model described in Section 2 with two interest rate series (LIBOR and the corresponding ARR), using daily data. Our main sample period starts two years prior to transition and ends on December 31, 2021, when LIBOR

[8] We rely on the presentation from the Bank of England (2018).
[9] See ICE (2018)

rates stop being published, except for USD. Overnight reference rates in all currencies closely track the prevailing policy rate set by the central bank in that currency. We therefore include the current level of the policy rate in the measurement equation, as discussed in Section 2, and provide robustness checks where we expand the controls to include government bond returns, credit risk and trading activity in OIS markets (discussed at the end of this section). We produce two estimates of information shares (IS), noise shares (NS), and information-to-noise ratios (IN): pre- and post- the major transition milestone. For the USD and EUR rates, only the post-transition estimate exists, given that SOFR and ESTR are newly created ARRs and therefore do not have pre-transition data.

Table 2 Panel A presents these results using our main sampling periods. Our main sampling approach generates pre- and post-transition windows that differ in duration across currencies. To ensure our results do not reflect differences in sample periods, we conduct a robustness test in Panel B of Table 2 by using pre- and post-transition windows that are identical for all currencies. An additional reason for using an alternative pre-post window is that the calculation of LIBOR itself changed between April 25, 2018 and April 1, 2019. From April 25, 2018 to April 1, 2019 IBA (ICE Benchmark Administration Limited) transitioned from the panel banks to making LIBOR submissions according to the updated VWAP (volume-weighted average price) methodology. Therefore, in the alternative estimation, we use the pre-transition period from July 1, 2017 (when the FCA announced the commitment to phasing out LIBOR) to April 25, 2018 (the date of the LIBOR methodology change). The post-transition period starts on October 1, 2019 (the date of ESTR launch) and finishes on December 31, 2021 (the end of our sample, at which point LIBOR ceases to be published, except for in USD). The results using this alternative pre-post window are presented in Panel B of Table 2. Estimates using these alternative sample dates are very similar to those in Panel A of Table 2.

[Table 2 about here.]

As a further robustness test, we expand the list of control variables to include proxies for credit risk and liquidity risk. Specifically, the set of conditioning variables includes government bond returns, changes in credit default swap (CDS) prices, changes in ICE BofA High Yield Option-Adjusted Spreads (OASs), and the weekly traded notional amount in OIS contracts. The estimated information shares and noise shares are similar using either set of variables (policy rates, bond returns, changes in CDS prices, OASs, and OIS trading activity, or just policy rates). These additional results are available in the Internet Appendix Section 3. The parameter estimates from the state-space model (rather than the quantities in Equations (9) - (11)) are also reported in the Internet Appendix Section 4.

In addition to period-specific estimates, we estimate and plot the time series evolution of noise shares (NS) in ARRs. The estimation procedure is as described above, except that we use a one-year rolling window to generate the estimates of NS. For example, to obtain the NS estimates for January 15, 2018, we use data from January 16, 2017 to January 15, 2018; to obtain the NS estimates for January 16, 2018, we use the data from January 17, 2017 to January 16, 2018, and so on. Figures 1-3 contain these time series. Figure 1 presents results for CHF (SARON) in Panel A and GBP (SONIA) in Panel B. Figure 2 presents results for EUR (ESTR) in Panel A and JPY (TONAR) in Panel B. Figure 3 presents results for USD (SOFR). The lines in the plot are smoothed using Locally Weighted Scatter-plot Smoothing (LOWESS) with tuning parameter of one-fourth.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

3.3 Designing robust benchmarks: From theory to practice

How noisy are the ARRs that replace LIBOR? Do ARRs become less noisy once they are set as official replacements of LIBOR (i.e., post-transition)? While prior literature proposes optimal benchmark design principles, we empirically evaluate the noise in the different benchmarks and discuss the trade-offs between various design features in practice.

Table 3 summarizes the design features of ARRs and provides a reference for our further discussion of how those trade-offs affect ARR noisiness.

[Table 3 about here.]

The optimal benchmark design principles outlined in Duffie and Dworczak (2021) suggest that all else equal, more robust benchmarks are those that (i) are more *costly* to manipulate, and

(ii) offer *fewer incentives* for manipulation. In practice, the cost of manipulation is greater when a benchmark is based on a more liquid underlying market.

Greater volumes in the underlying benchmark-setting market are also likely to reduce the noise in the benchmark because (i) in a deep market, individual trades that could cause temporary price distortions tend to have a smaller impact and (ii) averaging across many trades results in a smaller average noise term.

Consequently, given that ARRs have been designed to draw on deeper markets than the original LIBOR benchmarks, our **First Hypothesis** is that ARRs should be less noisy than the corresponding LIBOR benchmark. For similar reasons, we expect ARRs that are based on more liquid underlying markets to be less noisy than ARRs based on less liquid underlying markets.

Our main results in Table 2 support the First Hypothesis. Four out of five ARRs are less noisy than LIBOR, suggesting that trade-based benchmarks based on deep underlying markets tend to be more robust. This finding supports the theory and main design principles in (Duffie and Stein, 2015; Duffie and Dworczak, 2021).

The main exception is USD SOFR, which according to our estimates is a much noisier benchmark rate than USD LIBOR. It appears that the high level of noise in SOFR is linked to specific design choices. In particular, the decision to base the benchmark on collateralized repo transactions makes it vulnerable to extreme spikes based on the supply and demand of the underlying collateral (US treasuries and other investment-grade debt). One prominent example is the "SOFR surge event" on September 17, 2019 when SOFR jumped 282 bps in a single day. Because of high volumes of issuance and trading in investment-grade debt (USD 115 billion in the first half of September 2019), dealers accumulated excess long positions in investment-grade bonds (usually used as collateral in repo transactions). This increased dealers' demand for short positions (e.g., borrowing overnight), and pushed the repo rate up. This imbalance was reflected in a sharp jump in SOFR the next morning, as SOFR takes the median value of repo rates from the previous day. As a result of this design choice, SOFR has the highest noise share among the ARRs that we examine (see Table 2).

The high level of noise in SOFR is even apparent in a simple visual inspection of SOFR vs. USD LIBOR. For example, in Figure 3, which presents the time series of SOFR, one can see that SOFR often and substantially departs from USD LIBOR in temporary spikes before returning back to the USD LIBOR level.

Our **Second Hypothesis**, based on Duffie and Stein (2015) is that all else equal, more homogeneous underlying markets lead to less noise-prone benchmarks. The simple intuition is that a benchmark of apples and oranges could fluctuate between the price of apples and the price of oranges on a given day depending on how many transactions happen in apples vs. oranges, in contrast to a benchmark of apples only or oranges only. This Second Hypothesis highlights one of the benchmark design trade-offs — defining a very broad underlying market so that it includes many transaction types may increase the overall volume of the underlying reference market (beneficial according to Hypothesis 1), but in doing so increase the heterogeneity of the transactions (detrimental according to Hypothesis 2).

Consistent with our Second Hypothesis, SARON, which has one of the most homogenous underlying markets because the underlying transactions are completed by banks only, unlike in other rates, has one of the lowest noise shares across benchmarks (Table 2). This is particularly the case in the post-transition period when underlying volumes grew.

Our **Third Hypothesis** is that the statistical methods used in benchmark calculation matter and trimming outliers when computing the benchmark rate leads to less noisy benchmarks (e.g., Youle (2014) and Eisl, Jankowitsch and Subrahmanyam (2017)). A test of this hypothesis, although not a perfect test, is the reform of GBP SONIA. In the reform, GBP SONIA added outlier trimming in its benchmark methodology post-transition. Table 2 shows that from pre-reform to post-reform, the noise share in SONIA went down from 54.7% to 11.7%, consistent with our Third Hypothesis. One reason this is not a perfect test of the hypothesis is that the reform also included a wider range of transactions in the benchmark to make a deeper underlying market. According to our First Hypothesis, the deeper underlying market could also have contributed to the decrease in noise and we are unable to separate these two potential drivers of the decrease.

The last hypothesis that follows from theory is about manipulation incentives. The incentive to manipulate a benchmark is directly proportional to potential profits to be made from ARR-linked derivatives (e.g., (Duffie and Stein, 2015)). Therefore, our **Fourth Hypothesis** is the greater the ratio of derivatives volume to underlying benchmark market volume, the greater the incentives to manipulate and thus the higher the level of noise in the benchmark rate. Some ARRs have developed large and liquid derivative markets and therefore would offer plentiful incentives for manipulation.

Testing this hypothesis, however, is challenging. The main impediment is that the LIBOR manipulation litigation was an extremely high-profile event, with extensive media coverage and billions of dollars of fines (even some jail sentences). Given the deterrence effect, it is unlikely that traders would attempt to manipulate the LIBOR replacements in the near term following the LIBOR manipulations.

3.4 Benchmark implementation trade-offs

As noted in the previous subsection, the practical implementation of benchmarks involves trade-offs between multiple factors. In this subsection, we discuss the main design choices for interest rate benchmarks and, where possible, what our results imply about these choices. Many of these design choices apply to financial benchmarks more broadly.

The design choices include the benchmark type (trade-based or not), the choice of underlying market (bank-only or bank-and-wholesale), the collateralization (relying on collateralized transactions or not), the aggregation window (overnight or other tenors), statistical stabilization (whether outliers are excluded or not), whether to discard small trades or not, the weighting scheme (equal-weighted, volume-weighted or other), the multiplicity of benchmarks (whether there are multiple alternative benchmarks or not), and the role of regulatory clarity.

Benchmark type. The move to trade-based benchmarks and away from survey-based ones is at the core of the transition from LIBOR to ARRs. In line with (Duffie and Dworczak, 2021), we find that trade-based benchmarks are less noisy, especially as underlying transaction markets become more liquid over time. Indeed, the time series of ARR noise shares (Figures 1-3) show a downward trend in ARR noise post-transition, as ARRs accumulate more depth in their underlying markets. All ARRs except SOFR experience a significant reduction in noise post-transition, while SOFR noise goes down only slightly.

Choice of underlying market. The choice of which transactions to include in calculating the benchmark affects the volume of rate-setting transactions (the broader the range of transactions, the greater the volume), but also affects the heterogeneity of counterparties and reasons to transact. Therefore, broadening the transaction base by including wholesale and interbank transactions in the transaction pool may introduce additional noise to ARRs. Among the rates we study, the CHF

SARON is the only rate that only includes bank transactions. Other rates include both bank and wholesale transactions, and have on average higher noise shares than SARON (see Table 2). Our empirical findings suggest that the narrower pool of counterparties in the SARON rate does not substantially increase the noisiness of SARON.

Collateralization. Secured rates are those derived from transactions that require collateral. On one hand, the absence of collateral introduces heterogeneity to transactions and can thus increase noise because the different transaction counterparties may have different credit risks. On the other hand, the presence of collateral can introduce noise from the underlying collateral markets themselves. Our estimation results illustrate the latter effect — SOFR is a very noisy benchmark with noise shares that remain high since its launch, reflecting substantial spikes in SOFR that subsequently revert back. This dynamic is likely a consequence of supply and demand in the market for investment-grade securities that are used as collateral in secured repo transactions (such as the "SOFR surge event" discussed in Section 3.3). In the second half of 2021, high noise share estimates for SOFR also reflect a relatively stable USD LIBOR (many days with the same LIBOR rate at the zero bound).

Aggregation window. The period of time over which transactions are aggregated to determine the benchmark rate determines the breadth and heterogeneity of those transactions. A longer aggregation window means more volume and greater cost of manipulation, therefore, lower noise. On the other hand, a shorter window could have the benefit of less heterogeneity in the transactions. Across the ARRs we examine, most have a similar window of 7-9 hours, while the reformed SONIA rate aggregates transactions over 18 hours (see Table 3).

Statistical stabilization. Statistical stabilization techniques, such as rate trimming, should be noise-reducing because they eliminate outliers. The positive effect is especially pronounced if applied to a broad heterogeneous market but could be detrimental if applied to a narrow market with only a few transactions. Two rates that apply the statistical trimming techniques in our sample are SONIA and ESTR. SONIA was successfully reformed, with one of the key new features being a switch to the volume-weighted trimmed mean methodology. As shown in Table 2, the noise share of SONIA went down from 54.7% to 11.7% post-reform. In the time series of SONIA noise shares (Panel B Figure 1), we also observe that outliers in the data became less common after the reform, leading to lower noise shares in SONIA relative to GBP LIBOR. ESTR, being a newly launched rate, does not have a pre-transition NS estimate, but post-transition, ESTR is also less noisy than EUR LIBOR, with 19.6% noise share.

Weighting or discarding small trades. To discard small trades is to use only the substantial transactions for price formation, taking to the limit the argument in Duffie and Dworczak (2021), who suggest a volume-weighting scheme that places small weight on small transactions. However, discarding small transactions altogether reduces the breadth of the underlying market. So if a given market experiences little manipulation because of non-market forms of deterrence (e.g., legal threat, monitoring), then discarding the small trades could be detrimental to market quality. Duffie and Dworczak (2021) conclude that, absent manipulation, equal weighting is optimal in that it is the statistically most efficient way of computing the benchmark rate.

Contract value fixing on the rate. Contract value, or the contract-value-to-underlying-market-volume ratio, captures the incentive to manipulate a benchmark. This incentive depends on the value or volume of benchmark-linked derivatives. When benchmark-linked derivative volume exceeds the underlying benchmark-forming volume of transactions, the benefit-to-cost of manipulation is high. Among the benchmarks we study, the newly launched benchmarks like SOFR and ESTR tend to have lower contract-to-market ratios because the derivative markets have not accumulated sufficient volume.

Multiplicity of benchmarks. In several currencies (USD, EUR, JPY), two alternative benchmarks co-exist. The two-benchmark approach, as in the case of USD AMERIBOR co-existing with SOFR, EURIBOR with ESTR, and TIBOR with TONAR, could be beneficial by decreasing the contract-to-market ratio, and therefore reducing incentives for manipulation. On the other hand, the multiplicity of benchmarks could lead to segmentation and reduce the underlying market liquidity of the benchmark. Empirically, we observe high noise shares in USD SOFR, but relatively low noise shares in TONAR and ESTR, suggesting that the multiplicity of benchmarks in these rates is not a major factor affecting ARR noisiness.

Regulatory clarity. The time series patterns support the notion that transparent regulatory action around the reference rate transition tends to increase the legitimacy of ARRs and bring in extra liquidity in both the underlying and derivatives markets. For example, the noise share in CHF SARON (Panel A Figure 1) shows a clear downward trend in 2020–2021. That trend coincides with the Swiss National Bank (SNB) making a transition from LIBOR to SARON in managing liquidity in the money market (i.e., targeting SARON instead of LIBOR to be in line with the SNB policy rate). The Bank of Japan (BOJ), on the other hand, has not provided the same degree of clarity on the JPY LIBOR transition. For example, Bloomberg (2020) reports that regulators have only started publishing data on the term structure of TONAR in 2021 and a functioning curve in mid-2021. This difference may explain the upwards-trending noise levels in JPY TONAR (Panel B Figure 2).

4 Wealth transfers due to noise in LIBOR and ARRs

We have shown in the previous section that the LIBOR replacements (ARRs) are rather noisy in USD and significantly less noisy in GBP, CHF, JPY, and EUR. ARR noise shares also vary over time, driven by factors such as methodology changes (e.g., adopting the trimmed mean calculation as in the case of SONIA).

One of the reasons noise matters is because reference rates determine payments on trillions of dollars of financial instruments such as interest rate swaps, forward rate agreements, interest rate options, cross-currency swaps, interest rate futures, business loans, bonds, and securitized products.[10] Given the large number and total value of contracts using the benchmark rates, even a small distortion in the benchmark rates due to noise can result in substantial noise-driven wealth transfers between contract counterparties. For example, on a day when $10 trillion of value fixes on a benchmark rate, a mere one basis point error in the benchmark due to noise can result in $1 billion of incorrect wealth transfers.

4.1 Data

Our empirical method allows us to estimate the daily series of pricing errors and calculate the value of wealth transfers that occur because of noise in reference rates. To arrive at the value of noise-driven wealth transfers, we use (i) model-derived noise estimates and (ii) data on entity-netted notional (ENN) value of contracts tied to the OIS rates. The total notional amounts are publicly

[10] According to the New York Fed, the US LIBOR market footprint (as of Q4 2020) is $223 trillion in outstanding notional exposure. By asset class, the notional exposures include $171 trillion in outstanding OTC derivatives (interest rate swaps, forward rate agreements, interest rate options, and cross currency swaps), $43 trillion in outstanding exchange-traded derivatives (interest rate options and futures), $4.8 trillion in outstanding business loans, $1.4 trillion in consumer loans, $1.1 trillion in bonds, and $1.6 trillion in securitized products.

available from the Commodity Futures Trading Commission (CFTC) and the ENN estimates are from Baker, Haynes, Roberts, Sharma and Tuckman (2021).[11]

We correct for netting between long and short positions in the interest rate swap market using CFTC estimates of Entity-Netted Notionals (ENN) Baker et al. (2021), which are approximately 6% of total notional outstanding.[12] We use ENN in our calculations because many intermediaries offset outstanding positions by entering a new offsetting position rather than canceling the initial position.

Using only OIS-linked contracts rather than all maturities of interest rate benchmarks means that we obtain a lower-bound estimate of the total value of wealth transfers. OIS-linked contracts account for about 14% of the total interest rate swaps notional according to CFTC data for 2020. By currency, OIS-linked notional value represents 9% of USD swaps, 14% of EUR, 31% of GBP, and 4% of JPY. The data on CHF are not available from CFTC.

Limiting our estimation to OIS-linked contracts has several advantages. First, OIS contracts share the same tenor (O/N) as the LIBOR-ARR pairs used in the benchmark quality estimation. This means the noise in O/N rates directly affects floating leg payments in OIS contracts.[13] Second, OIS payments are simpler to estimate given they occur daily and not on specific roll dates. Finally, unlike other instruments (e.g., interest rate options, futures, syndicated loans, etc), swaps have weekly data available from CFTC.

Among the various interest rate contracts that reference LIBOR or ARRs, interest rate swaps have by far the most substantial outstanding notional value. For example, in USD, they account for $99 trillion out of $152 trillion of total USD notional outstanding. In EUR, $92 trillion out of $132 trillion. In GBP, $44 trillion out of $54 trillion. In JPY, $35 trillion out of $37 trillion. And in CHF, $3.2 trillion out of $3.6 trillion.[14] We use the CFTC Swaps Report to obtain data on overnight index swaps (OIS), which allow market participants to exchange a floating rate (e.g.,

[11]The data are available from CFTC (2023).

[12]This estimate comes from the following calculation: ENN as a fraction of total notional is 13.9/231 = 0.06. According to Baker et al. (2021), the size of the interest rate swaps market measured by notional amount is $231 trillion, but, measured by ENNs, is only $13.9 trillion. Therefore, ENN constitutes 6% of total notional and represents a better proxy for the quantity of risk transfers than total notional.

[13]If we were to do the estimation for other tenors (e.g., 3-month or 6-month LIBOR vs. ARR), the compounding conventions in ARRs of longer tenors would complicate the estimation.

[14]These data are as of 2020 Q2, as referenced in BIS Statistics on interest rate derivatives notional amounts outstanding (Bank for International Settlements, 2020).

LIBOR or an ARR) for an agreed fixed rate.[15]

4.2 Estimation procedure

We estimate the wealth transfers as follows. First, from the empirical price discovery model, we obtain the time series of daily pricing (rate) errors in overnight LIBOR and ARR rates.[16] Next, we match these daily benchmark error estimates with the ENN value of OIS contracts fixing on the reference rate for each currency. We then compute the value of swap payments that are due to noise, assuming in one case that the currency's OIS ENN is referencing LIBOR, and in a second case assuming that it is referencing the corresponding ARR. The difference between these two cases gives the "excess" noise-related wealth transfers.

We provide the estimates for the year 2020, as that is the only year with data available on all rates and all OIS contracts. The data on CHF notional are not available from CFTC, so we omit CHF from our calculations. That is not a substantial omission, as the CHF OIS market is significantly smaller compared to the rest of the currencies. Although we do not have the data on CHF OIS notional outstanding, one indication of a much smaller CHF market is that notional traded,as reported on ISDA's website (International Swaps and Derivatives Association, 2020), in CHF OIS is only about \$11.49 billion weekly, on average over 2019–2020, compared to \$2,120.76 billion in USD, \$454.14 billion in GBP, and \$73.61 billion in JPY.

4.3 Results

Table 4 presents the results. The noise in LIBOR (ARRs) creates an estimated \$345 billion p.a. (\$1.1 trillion p.a.) in wealth transfers between OIS contract counterparties in 2020. This number is substantial — it represents 18% (60%) of the average ENN in 2020, or 1.1% (3.5%) of the total notional. These estimates are a lower bound of the actual noise-related transfers, as we only consider swap contracts directly referencing overnight rates, as opposed to all types of derivatives, across all tenors. Across all rates, SOFR generates the highest dollar value of noise-related wealth

[15] The interest rate payments are exchanged on reset dates (daily). The floating rate fixing dates determine which floating rate applies. For EUR and GBP OIS contracts, the standard floating leg fixing lag is 0 (the rate is fixed on the reset date). For CHF, the fixing date is one business day following the reset date. For USD, the fixing date is one day prior to the reset date. The OIS contract duration can range from seven days to two years. IMM/ EOM Roll Dates do not apply to OIS. For detailed OIS contract specifications, see the Internet Appendix Section 2.

[16] The pricing errors used to compute wealth transfers are obtained from the smoothed disturbance vector of the main state-space model that includes the central bank policy rate as a control variable.

transfers (\$1.05 trillion p.a.), driven by the high notional value of USD OIS and substantial noise in SOFR.

[Table 4 about here.]

Due to their lower levels of noise, the ARRs for EUR, GBP, and JPY generate lower noise-related transfers relative to LIBOR. In contrast, USD SOFR creates a higher value of noise-related transfers compared to USD LIBOR and compared to all the other ARRs. Noise-related transfers between USD OIS counterparties would have amounted to approximately \$767 billion less in 2020 if all USD OIS contracts were based on USD LIBOR compared to basing all such contracts on SOFR. In the other currencies, noise-related transfers would have been \$6.69 billion higher in EUR LIBOR compared to ESTR, \$274 billion higher in GBP LIBOR compared to SONIA, and \$3.81 billion higher in JPY LIBOR compared to TONAR.

To put these estimates into perspective, we express them relative to the notional dollar value of outstanding OIS. In LIBOR, GBP has the largest value of noise-related wealth transfers per dollar of notional (3.72%). In ARRs, USD has the largest value of noise-related wealth transfers per dollar of notional (9.32%). USD LIBOR, on the other hand, generates the least noise-related transfers per dollar of notional (3 bps). In terms of relative noisiness of LIBOR vs ARRs, EUR LIBOR generates 0.06% more noise-related transfers per dollar of OIS compared to ESTR, GBP LIBOR 3.09% more compared to SONIA, JPY LIBOR 0.82% more compared to TONAR. In USD, the opposite is true: 9.31% more noise-related transfers in SOFR compared to USD LIBOR.

The overall conclusion is that noise in interest rate benchmarks is economically meaningful given the large wealth transfers that it creates between contract counterparties. Furthermore, the choice of benchmark has a material effect on the magnitude of these noise-related wealth transfers.

Figures 4 and 5 show the time series dynamics of noise-related wealth transfers in OIS. In the time series, we observe an already familiar contrast between USD SOFR and the rest of the rates. For SOFR (Figure 4), noise-related transfers tend to be higher than for corresponding LIBOR. For ESTR, SONIA, and TONAR (Figures 4, 5), on the other hand, we observe relatively lower noise-related transfers in ARRs relative to LIBOR post-transition.

[Figure 4 about here.]

[Figure 5 about here.]

5 Conclusions

We use a state-space model to separate information from noise in interest rate benchmarks and assess their quality. We then apply the method to assess the quality of LIBOR and alternative interest rate benchmarks (ARRs). We find that most ARRs are less noisy than their LIBOR counterparts, in line with the predictions in (Duffie and Dworczak, 2021), who propose trade-based benchmarks to replace submission-based LIBOR. The results also show that the noise shares of newly established ARRs tend to decrease through time as the new benchmarks gain adoption.

However, the replacement for USD LIBOR, SOFR, stands out as a much noisier interest rate benchmark compared to USD LIBOR. In contrast to the other benchmarks, SOFR is based on collateralized repo transactions, which makes it vulnerable to extreme spikes based on the supply and demand of the underlying collateral. As a result, there are instances in the data where SOFR spikes more than a full percentage point in a single day before reverting to its prior level.

We find that the noise in interest rate benchmarks creates economically meaningful wealth transfers between contract counterparties that would not occur but for the noise. For example, if during 2020, all overnight interest rate swaps (OIS) were based on LIBOR, \$345 billion p.a. would change hands between contract counterparties as a direct result of temporary pricing errors (noise) in the benchmark. In contrast, if during that year all OIS contracts were based on the ARRs, the equivalent number would be smaller in most currencies, but much larger overall (\$1.1 trillion) because of the high level of noise in SOFR. These estimates are adjusted for the netting of long and short positions using entity-netted notional values. The annual value of these noise-related wealth transfers using ARRs equates to approximately 2.5% of the total outstanding notional value of OIS contracts.

Finally, we find that well-designed reforms can significantly reduce noise in benchmarks. For example, GBP SONIA experienced a decline in noise share from 54.7% to 11.7% following the SONIA reform of April 23, 2018. In this reform, the reference market was expanded so that the benchmark would be based on a higher volume of transactions and mean trimming was introduced. However, many benchmark design choices involve tradeoffs. For example, a broad reference market can benefit a benchmark by being more liquid, but can be detrimental if it increases the heterogeneity of transactions.

Overall our findings point to the importance of benchmark design, in particular for systemically important benchmarks that underpin large volumes of contracts. The methods in this paper provide a means for market designers to empirically test the outcomes of benchmark design variations or learn from benchmark designs that have been implemented in other markets.

References

Bai, Y. et al. (2022), 'Term risk-free rates: Methodologies, challenges, and the future', *Journal of Derivatives* **29**(3), 30–45.

Baillie, R. T., Booth, G. G., Tse, Y. and Zabotina, T. (2002), 'Price discovery and common factor models', *Journal of Financial Markets* **5**(3), 309–321.

Baker, L., Haynes, R., Roberts, J., Sharma, R. and Tuckman, B. (2021), 'Risk transfer with interest rate swaps', *Financial Markets, Institutions & Instruments* **30**(1), 3–28.

Bank for International Settlements (2020), 'BIS statistics'.
URL: *https://stats.bis.org/statx/srs/table/d7*

Bank of England (2018), 'What you need to know about LIBOR transition'.
URL: *https://www.bankofengland.co.uk/-/media/boe/files/markets/benchmarks/what-you-need-to-know-about-libor-transition*

Black, F. (1986), 'Noise', *Journal of Finance* **41**(3), 528–543.

Bloomberg (2020), 'A Guide to the world's new benchmarks after LIBOR'.
URL: *https://www.bloomberg.com/news/articles/2020-08-11/a-global-guide-to-the-new-benchmarks-that-will-replace-libor*

Bloomberg (2021), 'The world's most important number'.
URL: *https://www.bloomberg.com/opinion/articles/2018-04-03/the-world-s-most-important-number*

Brogaard, J., Nguyen, T. H., Putnins, T. J. and Wu, E. (2021), 'What moves stock prices? The role of news, noise, and information', *Review of Financial Studies* .

Campbell, J. Y. and Shiller, R. J. (1988), 'Interpreting cointegrated models', *Journal of Economic Dynamics and Control* **12**(2-3), 505–522.

CFTC (2023), 'CFTC Swaps Report'.
URL: *https://www.cftc.gov/MarketReports/SwapsReports/Archive/index.htm*

Duffie, D. and Dworczak, P. (2021), ‘Robust benchmark design’, *Journal of Financial Economics* **142**(2), 775–802.

Duffie, D., Dworczak, P. and Zhu, H. (2017), ‘Benchmarks in search markets’, *Journal of Finance* **72**(5), 1983–2044.

Duffie, D., Harry, C., Luck, S., Wang, Z., Yang, Y. et al. (2022), Bank funding risk, reference rates, and credit supply, Technical report.

Duffie, D. and Stein, J. C. (2015), ‘Reforming LIBOR and other financial market benchmarks’, *Journal of Economic Perspectives* **29**(2), 191–212.

Durbin, J. and Koopman, S. J. (2012), *Time series analysis by state space methods*, Vol. 38, Oxford University Press.

Eisl, A., Jankowitsch, R. and Subrahmanyam, M. G. (2017), ‘The manipulation potential of Libor and Euribor’, *European Financial Management* **23**(4), 604–647.

Fassas, A. P. (2021), ‘Price discovery in US money market benchmarks: LIBOR vs. SOFR’, *Economics Letters* **204**, 109882.

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

Hasbrouck, J. (1993), ‘Assessing the quality of a security market: A new approach to transaction-cost measurement’, *Review of Financial Studies* **6**(1), 191–212.

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

Hasbrouck, J. (1999), ‘The dynamics of discrete bid and ask quotes’, *Journal of Finance* **54**(6), 2109–2142.

Hasbrouck, J. (2007), *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*, Oxford University Press.

Hasbrouck, J. (2021), ‘Price discovery in high resolution’, *Journal of Financial Econometrics* **19**(3), 395–430.

Heitfield, E. and Park, Y.-H. (2019), 'Inferring term rates from SOFR futures prices'.

Hendershott, T. and Menkveld, A. J. (2014), 'Price pressures', *Journal of Financial Economics* **114**(3), 405–423.

ICE (2018), 'ICE LIBOR'.
URL: *https://www.theice.com/iba/libor*

Indriawan, I., Jiao, F. and Tse, Y. (2021), 'The SOFR and the Fed's influence over market interest rates', *Economics Letters* **209**, 110095.

International Swaps and Derivatives Association (2020), 'ISDA Swaps Info'.
URL: *http://swapsinfo.org/*

Kirti, D. (2022), 'What are reference rates for?', *Journal of Banking & Finance* **144**, 106635.

Klingler, S. and Syrstad, O. (2021), 'Life after LIBOR', *Journal of Financial Economics* **141**(2), 783–801.

Menkveld, A. J., Koopman, S. J. and Lucas, A. (2007), 'Modeling around-the-clock price discovery for cross-listed stocks using state space methods', *Journal of Business & Economic Statistics* **25**(2), 213–225.

Morck, R., Yeung, B. and Yu, W. (2000), 'The information content of stock markets: Why do emerging markets have synchronous stock price movements?', *Journal of Financial Economics* **58**(1-2), 215–260.

New York Fed (2021), 'Progress Report: The Transition from U.S. Dollar LIBOR'.
URL: *https://www.newyorkfed.org/medialibrary/Microsites/arrc/files/2021/USD-LIBOR-transition-progress-report-mar-21.pdf*

Putnins, T. J. (2013), 'What do price discovery metrics really measure?', *Journal of Empirical Finance* **23**, 68–83.

Schrimpf, A. and Sushko, V. (2019), 'Beyond LIBOR: a primer on the new benchmark rates', *BIS Quarterly Review (March)* .

Skov, J. B. and Skovmand, D. (2021), 'Dynamic term structure models for SOFR futures', *Journal of Futures Markets* **41**(10), 1520–1544.

Watson, M. W. (1986), 'Univariate detrending methods with stochastic trends', *Journal of Monetary Economics* **18**(1), 49–75.

Yan, B. and Zivot, E. (2010), 'A structural analysis of price discovery measures', *Journal of Financial Markets* **13**(1), 1–19.

Youle, T. (2014), How much did manipulation distort the Libor, Technical report, Working Paper, University of Minnesota.

Table 1: Details of LIBOR and ARRs in five major currencies

This table summarizes the key details about the LIBORs and alternative reference rates (ARRs) in five major currencies. The table relies on information from the Bank of England. The data on transition dates are sourced from the web pages of ARR administrators in the respective currencies.

LIBOR pub-n time (London)	ARR pub-n time (London)	Transition date	ARR Administrator	ARR Collateralization
CHF LIBOR 11:55 am	SARON 5:00 pm	Oct 31, 2018	SIX Swiss Exchange	Secured
EUR LIBOR 11:55 am	ESTR 7:00 am	Oct 1, 2019	European Central Bank	Unsecured
GBP LIBOR 11:55 am	SONIA 9:00 am	Apr 23, 2018	Bank of England	Unsecured
JPY LIBOR 11:55 am	TONAR 2:00 am	Aug 31, 2018	Bank of Japan	Unsecured
USD LIBOR 11:55 am	SOFR 1:00 pm	Apr 2, 2018	New York Fed	Secured

Notes on transition dates.

Oct 31, 2018 – SARON Futures start trading at Eurex. Source: https://www.snb.ch/en/ifor/finmkt/fnmkt_benchm/id/finmkt_NWG_milestones

Oct 1, 2019 – ECB starts publishing ESTR. Source: https://www.ecb.europa.eu/paym/interest_rate_benchmarks/WG_euro_risk-free_rates/html/milestones.en.html

Apr 23, 2018 – SONIA reforms implemented, firms develop SONIA futures trading infrastructure. Source:https://www.bankofengland.co.uk/markets/transition-to-sterling-risk-free-rates-from-libor

Aug 31, 2018 – Cross-Industry Committee on Japanese Yen Interest Rate Benchmarks established. Source: https://www.boj.or.jp/en/announcements/press/koen_2020/data/ko200219a1.pdf

Apr 2, 2018 – The New York Fed starts publishing SOFR. Source: https://www.newyorkfed.org/medialibrary/microsites/arrc/files/libor-timeline.pdf

Table 2: Benchmark quality estimates for LIBOR – ARR pairs

This table reports information shares (IS), noise shares (NS), and information-to-noise ratios (IN) for five pairs of LIBOR rates and alternative reference rates (ARRs). The sequential price state-space model and estimation details are described in Section 2. The estimation windows are in the "Period" column. In Panel A, the pre-transition period covers two years prior to the transition from LIBOR to ARR in each respective currency, and the post-transition period corresponds to the time from the transition milestone until December 31, 2021 (the date of cessation of LIBOR). Table 1 notes explain the transition dates. In Panel B, the pre- and post-transition periods are the same for all currencies. The pre-transition period is from July 1, 2017 (when the FCA announced the commitment to phasing out LIBOR) to April 25, 2018 (the date of LIBOR methodology change for GBP). The post-transition period starts on October 1, 2019 (the date of ESTR launch) and finishes on December 31, 2021 (the end of our sample). We do not estimate the pre-transition price discovery shares for SOFR and ESTR, as they are newly established ARRs and therefore have not existed prior to transition.

Panel A. Benchmark quality estimates around the major transition milestones

Currency	ARR	Period	IS (LIBOR)	NS (LIBOR)	IN (LIBOR)	IS (ARR)	NS (ARR)	IN (ARR)
CHF	SARON	Oct 31, 2016 - Oct 31, 2018	55.2%	80.9%	1.33%	44.8%	19.1%	4.44%
CHF	SARON	Oct 31, 2018 - Dec 31, 2021	20.5%	98.9%	0.78%	79.5%	1.13%	72.6%
EUR	ESTR	n/a	-	-	-	-	-	-
EUR	ESTR	Oct 1, 2019 - Dec 31, 2021	26.1%	80.5%	0.64%	73.9%	19.6%	7.00%
GBP	SONIA	Apr 23, 2016 - Apr 23, 2018	25.9%	45.3%	33.1%	74.1%	54.7%	54.0%
GBP	SONIA	Apr 23, 2018 - Dec 31, 2021	69.7%	88.3%	0.12%	30.3%	11.7%	0.38%
JPY	TONAR	Aug 31, 2016 - Aug 31, 2018	43.6%	100.0%	2.90%	56.4%	0.05%	98.8%
JPY	TONAR	Aug 31, 2018 - Dec 31, 2021	54.4%	99.2%	3.61%	45.6%	0.85%	78.4%
USD	SOFR	n/a	-	-	-	-	-	-
USD	SOFR	Apr 2, 2018 - - Dec 31, 2021	0.13%	0.00%	50.0%	99.9%	100%	0.61%

Panel B. Robustness test using alternative transition milestones

Currency	ARR	Period	IS (LIBOR)	NS (LIBOR)	IN (LIBOR)	IS (ARR)	NS (ARR)	IN (ARR)
CHF	SARON	Jul 1, 2017 - Apr 23, 2018	27.1%	83.0%	1.13%	72.9%	17.0%	13.1%
CHF	SARON	Oct 1, 2019 - Dec 31, 2021	21.0%	98.8%	1.08%	79.0%	1.25%	76.4%
EUR	ESTR	n/a	-	-	-	-	-	-
EUR	ESTR	Oct 1, 2019 - Dec 31, 2021	26.1%	80.5%	0.64%	73.9%	19.6%	7.00%
GBP	SONIA	Jul 1, 2017 - Apr 23, 2018	46.7%	50.8%	94.1%	53.3%	49.2%	94.9%
GBP	SONIA	Oct 1, 2019 - Dec 31, 2021	99.8%	90.6%	61.9%	0.16%	9.38%	2.41%
JPY	TONAR	Jul 1, 2017 - Apr 23, 2018	63.6%	99.0%	4.45%	36.4%	1.02%	72.1%
JPY	TONAR	Oct 1, 2019 - Dec 31, 2021	53.9%	96.9%	4.95%	46.1%	3.14%	57.9%
USD	SOFR	n/a	-	-	-	-	-	-
USD	SOFR	Oct 1, 2019 - Dec 31, 2021	0.11%	0.01%	50.0%	99.9%	100.0%	6.91%

Table 3: Summary of ARR methodology

This table summarizes the key design features of alternative reference rates (ARRs), using the data from regulatory reports and the websites of rate administrators.

	SOFR	**ESTR**	**SONIA (pre-reform)**	**SONIA (post-reform)**	**SARON**	**TONAR**
Trade-based?	Trade	Trade	Trade	Trade	Transactions and binding quotes	Trade
Underlying market						
Counterparties	banks and wholesale	banks and wholesale	banks and wholesale	banks and wholesale	banks	banks and wholesale
Secured?	yes	no	no	no	yes	no
Underlying volume*	$1 trn	$45 bn	$14 bn	$57 bn	$3.5 bn	$40 bn
Tenor	overnight	overnight	overnight	overnight	overnight	overnight
Aggregation window	7 hours (7:00-14:00 CT)	8 hours (8:30-17:30 GMT+2)	n/a	18 hours (00:00-18:00 GMT+1)	9.5 hours (8:30-18:00 GMT+2)	n/a
Weighting	volume-weighted median	volume-weighted trimmed mean	n/a	volume-weighted trimmed mean	volume-weighted mean	volume-weighted mean
Price trimming	no	yes[1]	n/a	yes[1]	yes[2]	n/a
Volume trimming	no	yes[3]	no	yes[4]	no[5]	n/a
Contracts fixing on the rate						
$ Notional derivatives **	$200 bn	$4.69 bn	n/a	$7,913 bn	$27.67 bn	$248 bn
Contract-to-market ratio ***	0.20	0.10	n/a	138.82	7.91	6.20
Other						
Other benchmarks?	yes[6]	yes[7]	n/a	no	no	yes[8]
ARR start	Apr 2018 new	Oct 2019 new	March 1997 existing	Apr 2018 reformed	Aug 2009 existing	2016 existing

Notes on ARR methodology choices.

*Underlying volume data is for the year 2019, collected from the web-sites of rate administrators. Data availability is limited for other years.

**We present the notional $ value traded in derivatives fixing on the rate, in USD billion, for 2019. We use notional traded because ISDA does not provide dollar volume numbers. We use 2019 for comparability because underlying market ARR volumes are for 2019 only.

*** Contract-to-market ratio is $ notional derivatives divided by underlying volume.

1.In ESTR and SONIA, they remove the top and bottom 25% of volume-weighted rates, after sorting rates from lowest rate to highest rate.

2. In SARON, they apply a quote filter: eliminate quotes from outside +/- 3bps from midpoint; trade filter: eliminate prices outside +/-50 bps of average trade price on completed trades.

3. In ESTR, a pro-rata calculation is applied to volumes that span the thresholds for trimming to ensure that exactly 50% of the total eligible volume is used in the calculation of the volume-weighted mean.

4. In reformed SONIA, the trimmed mean is calculated as the volume-weighted mean rate, based on the central 50% of the volume-weighted distribution of rates.

5. In SARON, the volume of quotes is limited to CHF 100 million.

6. In USD, AMERIBOR is an unsecured rate calculated by AFX, since 2015, now with $2bn underlying transactions.

7. In EUR, EURIBOR is a daily reference rate, published by the European Money Markets Institute, based on the averaged interest rates at which Eurozone banks borrow unsecured funds from counterparties in the euro wholesale money market (or interbank market). It has existed since 1999.

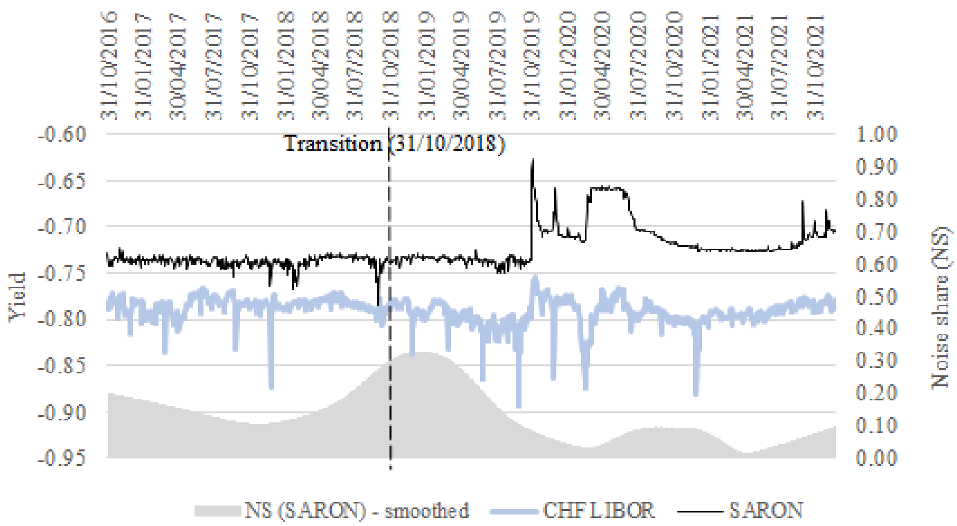
8. In JPY, TIBOR is the daily reference rate derived from the interest rates that banks charge to lend funds to other banks in the Japanese interbank market. There are two types of TIBOR rates—the European TIBOR rate (since 1998) and the Japanese Yen TIBOR rate (since 1995).

Table 4: Estimates of noise-related wealth transfers

This table presents the annual noise-related wealth transfers during the year 2020 in overnight index swaps (OIS). We calculate noise-related wealth transfers for four currencies: EUR, GBP, JPY, and USD (CHF data is missing from CFTC reports). We estimate the time series of pricing errors using the sequential state-space model. We calculate weekly wealth transfers by multiplying the entity-netted notional outstanding by the estimated noise term at the end of that week and then sum the weekly wealth transfers for the year 2020. For comparison, the table also reports the average weekly notional outstanding and entity-netted notional outstanding in OIS for each currency in 2020.

Currency	(1) Wealth transfers, LIBOR ($bn)	(2) Wealth transfers, ARR ($bn)	Difference (2-1), ARR-LIBOR ($bn)	Notional outstanding ($bn)	ENN Notional outstanding ($bn)	(3) Wealth transfers, LIBOR(%) % of notional	(4) Wealth transfers, ARR(%) % of notional	Difference (4-3) ARR - LIBOR % of notional
EUR	9.70	3.01	-6.69	10,407.82	624.47	0.0932%	0.0289%	-0.0643%
GBP	331.00	56.15	-274.85	8,881.41	532.88	3.7269%	0.6322%	-3.0947%
JPY	3.84	0.03	-3.81	462.32	27.74	0.8304%	0.0062%	-0.8242%
USD	0.34	1,053.21	1,052.87	11,300.05	678.00	0.0030%	9.3204%	9.3174%
Total	344.88	1,112.40	767.52	31,051.60	1863.10	1.16%	2.50%	1.33%

Panel A. CHF (SARON)



Panel B. GBP (SONIA)

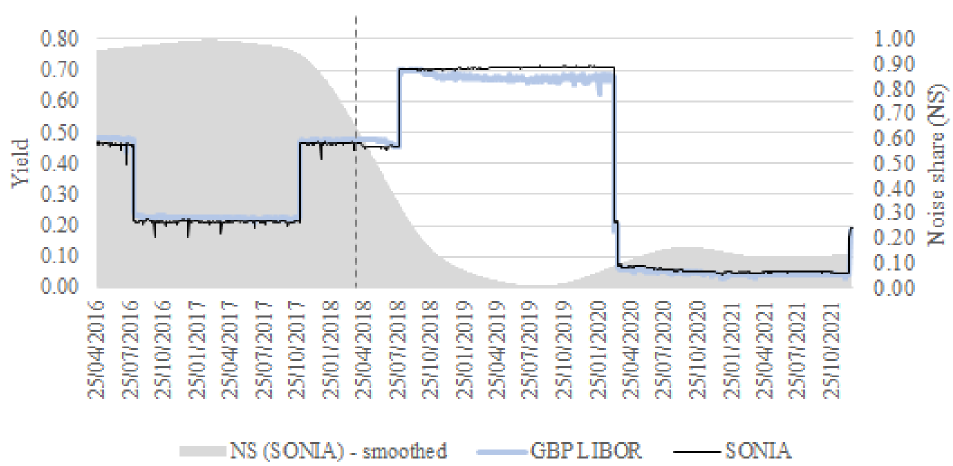
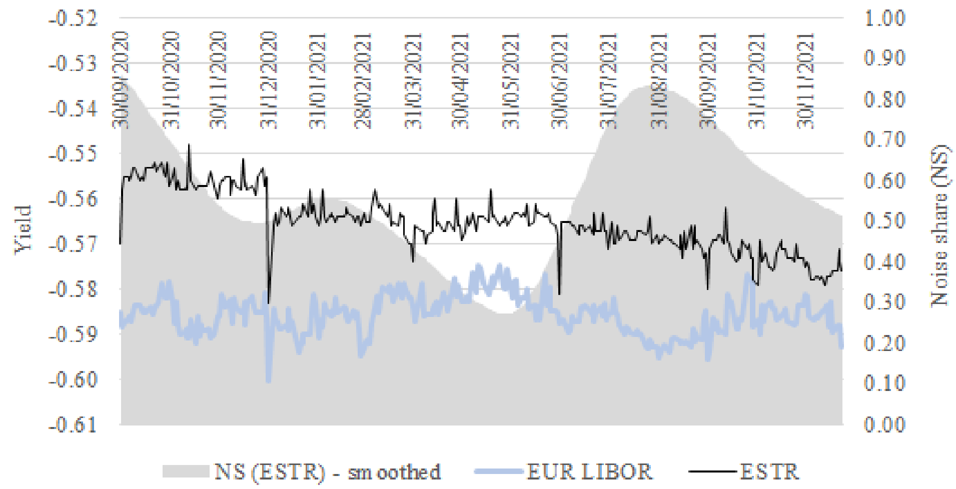


Figure 1: LIBORs, ARRs, and noise shares in ARRs for CHF and GBP

This figure plots the time series of LIBORs and alternative reference rates (ARRs), overlayed with the noise shares in ARRs for SARON (CHF) and SONIA (GBP). We estimate the noise shares using the sequential price state-space model described in Section 2 with a one-year rolling window shifted daily to generate the estimates of IS and NS as the daily time series. The lines in the plot are smoothed using Locally Weighted Scatter-plot Smoothing (LOWESS) with tuning parameter of one-fourth.

Panel A. EUR (ESTR)



Panel B. JPY (TONAR)

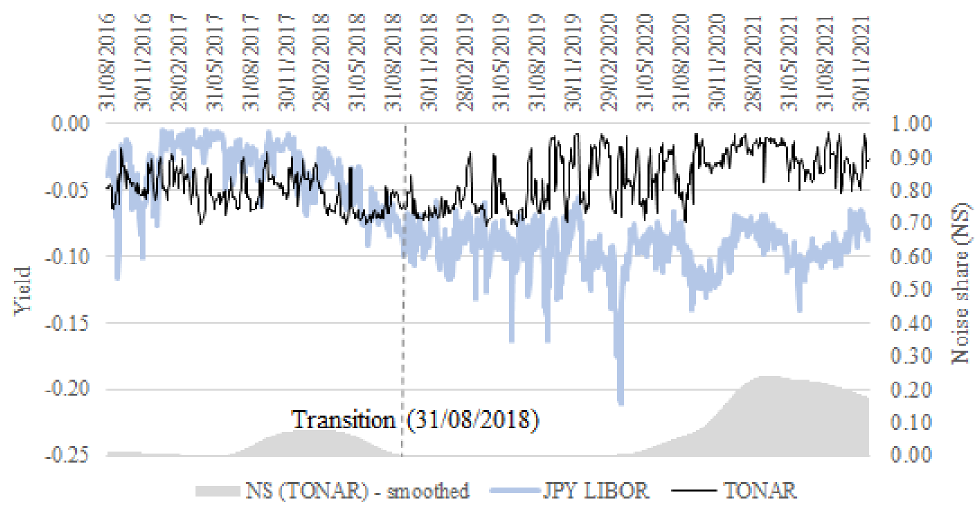


Figure 2: LIBORs, ARRs, and noise shares in ARRs for EUR and JPY

This figure plots the time series of LIBORs and alternative reference rates (ARRS), overlayed with the noise shares in ARRs for ESTR (EUR) and TONAR (JPY). We estimate the noise shares using the sequential price state-space model described in Section 2 with a one-year rolling window shifted daily to generate the estimates of IS and NS as the daily time series. The lines in the plot are smoothed using Locally Weighted Scatter-plot Smoothing (LOWESS) with tuning parameter of one-fourth.

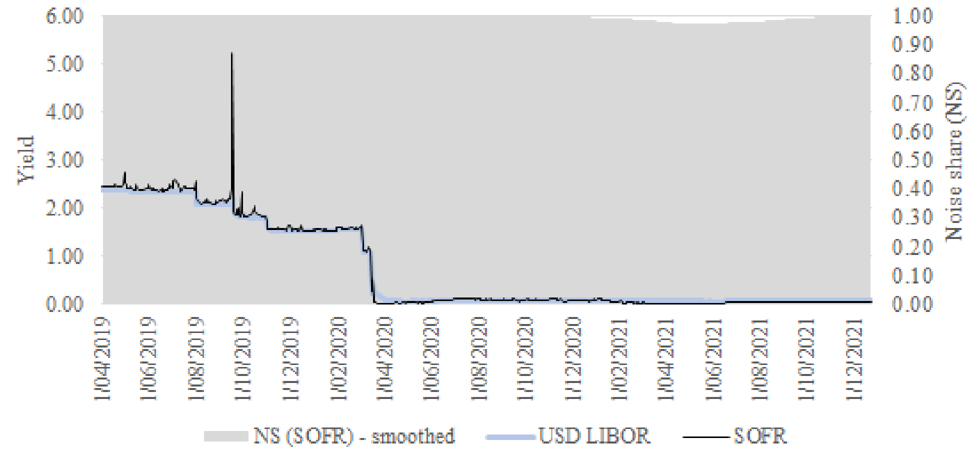
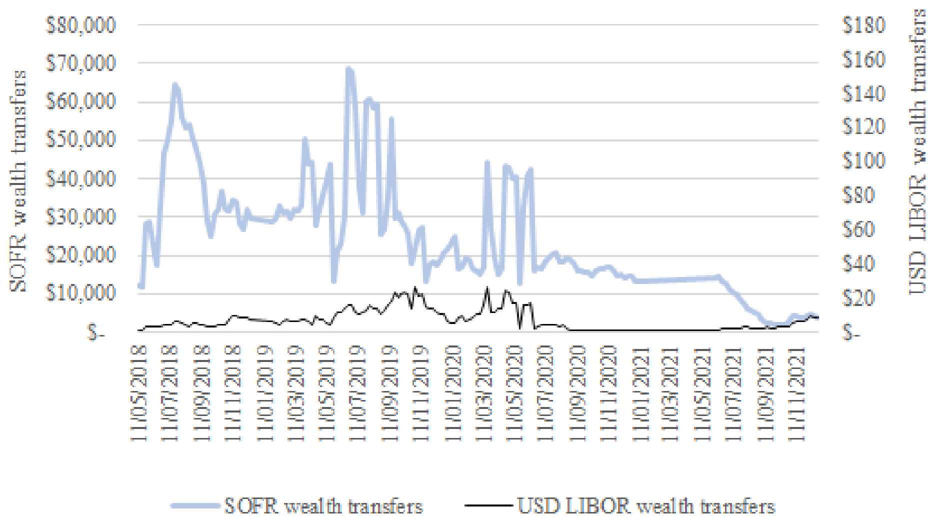


Figure 3: LIBORs, ARRs, and noise shares in ARRs for USD

This figure plots the time series of LIBORs and alternative reference rates (ARRs), overlayed with the noise shares in ARRs for SOFR (USD). We estimate the noise shares using the sequential price state-space model described in Section 2 with a one-year rolling window shifted daily to generate the estimates of IS and NS as the daily time series. The lines in the plot are smoothed using Locally Weighted Scatter-plot Smoothing (LOWESS) with tuning parameter of one-fourth.

Panel A. Wealth transfers in USD LIBOR and SOFR



Panel B. Wealth transfers in EUR LIBOR and ESTR

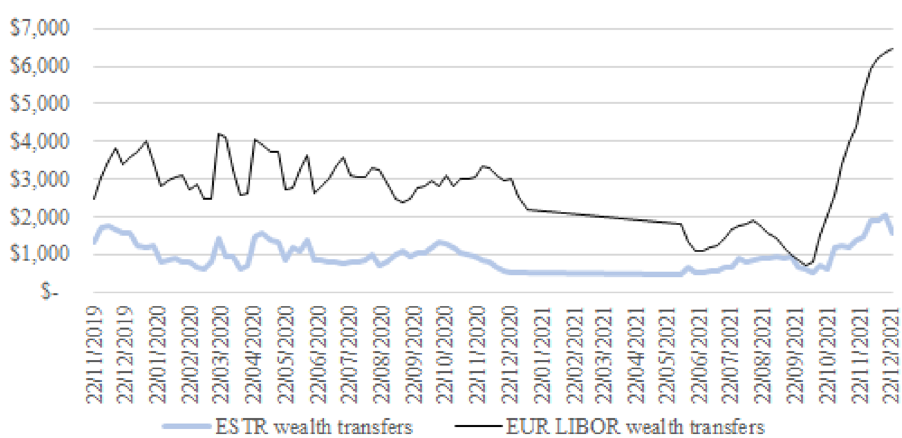
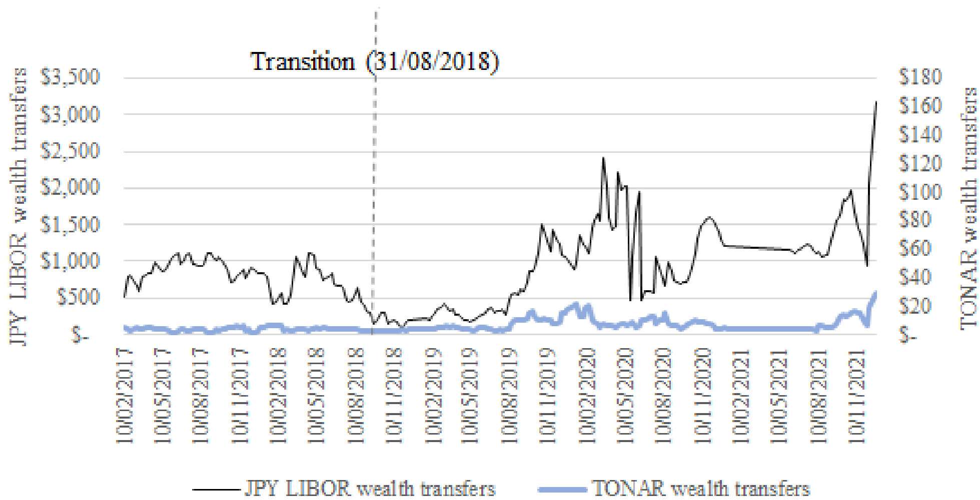


Figure 4: The time series of estimated wealth transfers in USD and EUR

This figure plots the 8-week moving average of estimated wealth transfers in overnight interest swaps (OIS) due to noise in reference rates. Panel A (B) plots noise-related transfers in USD (EUR) OIS, calculated by multiplying the entity-netted notional outstanding by the estimate of noise for the end of that week. The data on OIS notional outstanding in each currency are from CFTC swaps reports. Estimation periods are as indicated in Table 2.

Panel A. Wealth transfers in JPY LIBOR and TONAR



Panel B. Wealth transfers in GBP LIBOR and SONIA

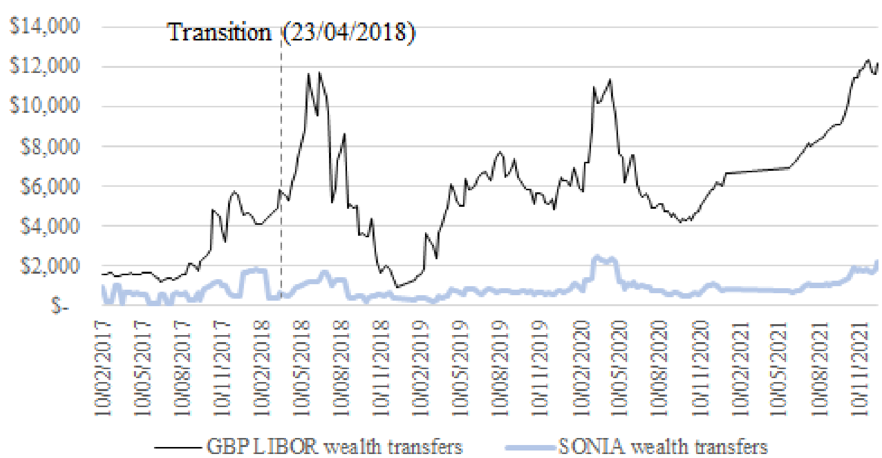


Figure 5: The time series of estimated wealth transfers in JPY and GBP

This figure plots the 8-week moving average of the estimated wealth transfers in overnight interest swaps (OIS) due to noise in the reference rates. Panel A (B) plots noise-related transfers in JPY (GBP) OIS, calculated by multiplying the entity-netted notional outstanding by the estimate of noise for the end of that week. The data on OIS notional outstanding in each currency are from CFTC swaps reports. Estimation periods are as indicated in Table 2.