Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Trading Costs v. Indicative Liquidity in the Off-the-Run Treasury Market

Oleg Sokolinskiy

2025-049

Please cite this paper as:
Sokolinskiy, Oleg (2025). "Trading Costs v. Indicative Liquidity in the Off-the-Run Treasury Market," Finance and Economics Discussion Series 2025-049. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2025.049.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Trading Costs v. Indicative Liquidity in the Off-the-Run Treasury Market [*]

Oleg Sokolinskiy[†]

July 7, 2025

Abstract

This paper estimates trading costs in the off-the-run Treasury market using comprehensive transactions data and machine learning techniques. The analysis reveals several key findings that enhance the understanding of the off-the-run Treasury market liquidity. First, the indicative bid-ask spread is shown to be a biased measure of liquidity, even when not considering transaction volume. Specifically, bid-ask spreads systematically overstate trading costs of more seasoned Treasuries, and the liquidity of benchmark, on-the-run securities affects how off-the-run bid-ask spreads map to trading costs. Second, the paper demonstrates that trading costs may scale non-monotonically with transaction volume, which suggests selective, opportunistic liquidity-taking. Additionally, transaction size has greater effect on off-the-run securities' trading costs when benchmark, on-the-run liquidity is lower. Finally, indicative bid-ask spreads may notably overstate trading costs for larger orders of relatively less liquid securities. These findings contribute to our understanding of actual liquidity in the off-the-run Treasury market, while highlighting the limitations of a traditional liquidity measure. By providing a more nuanced view of trading costs, this study contributes valuable insights for supporting financial stability and optimal asset allocation.

Keywords: liquidity, Treasury market, off-the-run, effective bid-ask spread

JEL codes: G10, G12

[*] This article represents the views of the author, and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff. I am grateful for helpful comments from and discussions with David Bowman, Dobrislav Dobrev, Maximilian Dunn, Sebastian Infante Bilbao, Peter Johansson, Don Kim, Edith Liu, Andrew Meldrum, Marius Rodriguez, and Min Wei.

[†] Board of Governors of the Federal Reserve System. Email: oleg.v.sokolinskiy@frb.gov

1 Introduction

The U.S. Treasury market is the deepest fixed income market in the world. Market liquidity – ability to transact significant amounts without temporarily dislocating prices – allows the Treasury market to serve as a benchmark for valuation and a source of near risk-free collateral. However, Treasury market functioning became a concern at the height of the COVID-19 crisis (e.g., see Logan, 2020; Fleming and Ruela, 2020; Duffie et al., 2023).[1] The March 2023 SVB collapse and the April 2025 trade tensions led to two more episodes of severely strained Treasury market liquidity. Furthermore, the supply of Treasury securities has been increasing relative to the primary dealers' ability to intermediate in this market (see Duffie, 2023, 2025). As a result, the functioning of the Treasury market is likely to remain a key focus for both regulators and investors.

The Treasury market has two major segments that differ in their microstructure and, consequently, liquidity. First, there are on-the-run Treasuries – most recently-issued securities for a given maturity date. On-the-run Treasuries have much higher trading volumes than other Treasuries, with a significant proportion of trades being on centralized markets with firm quotes.[2] However, on-the-run Treasuries are only a small fraction of the notional of all outstanding marketable Treasury debt. Second, there are off-the-run Treasuries – all Treasury securities that are not on-the-run. In contrast to on-the-run securities, off-the-run Treasuries trade infrequently in a decentralized, bilateral market providing indicative quotes.[3] Due to this microstructure of the off-the-run Treasury market, its liquidity is difficult to gauge.

In this paper, I suggest an estimate of the effective the bid-ask spread measure (see Dem-

[1] Adrian et al. (2025) considers Treasury market liquidity over a longer timeline starting from the GFC.

[2] A firm quote with an associated volume is a commitment of a liquidity provider to trade in that size if requested to do so.

[3] Indicative quotes are not firm commitments to trade, and the price of a potential trade is may be set upon further negotiations.

setz, 1968) for the off-the-run Treasury market liquidity largely based on actual trades.[4] The core components of the suggested method for the assessment of actual trading costs in the off-the-run Treasury market are (i) reliance on off-the-run Treasuries' transaction data along with firm quotes for on-the-run Treasuries and (ii) state-of-the-art machine learning approach to fitting the term structure of benchmark, on-the-run rates. The method relies on the on-the-run Treasuries' quotes as the source of most recent information on changes in the benchmark interest rate term structure. In turn, periodic snapshots of indicative quotes enable the estimation of idiosyncratic price components of particular off-the-run securities. The machine learning-based method of Filipović et al. (2022) for fitting the term structure of interest rates – the second critical component of the suggested method – delivers sufficient accuracy and stability to estimate the term structure of benchmark, on-the-run rates at each moment when a trade in an off-the-run security occurs.[5]

Having constructed a measure of trading costs for off-the-run Treasuries, I explore how it deviates from indicative liquidity. First, I document biases in indicative bid-ask spreads that are unconditional on transaction volume. In particular, indicative bid-ask spreads tend to systemically overstate trading costs for more seasoned securities. Another bias of indicative bid-ask spreads results from their failure to reflect benchmark, on-the-run Treasury market liquidity. When benchmark, on-the-run liquidity is degraded, indicative bid-ask spreads for off-the-run securities translate into greater trading costs. This bias is of particular significance for assessing financial stability – a market dysfunction resulting from prohibitive trading costs in the off-the-run Treasury market may occur at lower off-the-run bid-ask spread levels than a simpler linear regression model may suggest.

Second, overcoming the inherent limitation of indicative quotes, I explore the effect of transaction size on price. While indicative bid-ask spreads do not provide information on

[4] Bid-ask spread is the difference between the lowest price at which a security may be bought and the highest price at which the security may be sold.

[5] Subject to considering trades that occur during the BrokerTec platform's active hours.

how trading costs scale with the size of an order, effective bid-ask spreads are associated with particular transaction volumes. Within a non-parametric framework for modeling how trading costs vary with transaction volume, I estimate the corresponding scaling laws conditional on either benchmark liquidity or relative liquidity of the specific off-the-run security. I show how the effect of volume on trading costs in the off-the-run market varies with the liquidity conditions in the benchmark, on-the-run Treasury market: worse benchmark liquidity translates into higher trading costs in the off-the-run market, even controlling for the correspondingly wider off-the-run securities' indicative bid-ask spreads. Finally, I demonstrate that indicative bid-ask spreads may notably overstate expected trading costs for larger transaction volumes of relatively less liquid off-the-run securities. More generally, I obtain evidence in favor of selective liquidity taking – endogeneity of trading in response to available liquidity.

Literature

In the remainder of this section, I place this paper within the framework of the existent research. In particular, there are three strands of research that this paper contributes to: (i) liquidity measurement in the decentralized, bilateral Treasury off-the-run market; (ii) differences between indicative and actual liquidity; and (iii) scaling of trading costs with transaction volume.

First, the suggested method for the calculation of the effective bid-ask spread contributes to the literature on liquidity measurement in the off-the-run Treasury market. The opacity of the off-the-run Treasury market shifted research focus to the liquidity in the on-the-run Treasury market, for which there is ample excellent research by Fleming (2001) among others. In contrast, liquidity in the off-the-run market is the subject of far fewer studies. Due to sparseness of trades in specific off-the-run Treasuries, the indicative bid-ask spread (or some transform thereof) is a prominent measure of Treasury market liquidity, as in Pasquariello and Vega (2009) and Goyenko et al. (2011). Another strand of literature relies on liquidity measures that are more similar to price impact. Among these papers, Babbel et al. (2004) is conceptually closest to the approach I adopt in this paper. Babbel et al. introduces a price pressure measure in the off-the-run Treasury market, which is the difference between a synthetic 'on-the-run price' – price of the security obtained by discounting cash flows using the interest rate term structure from on-the-run Treasuries – and the actual trade price for relatively low volume transactions. Similar to Babbel et al., I also use on-the-run Treasuries as the benchmark for most up-to-date pricing, but fit the term structure using the more flexible model of Filipović et al. (2022) instead of the Nelson and Siegel (1987) parametric model.[6] The approach of Babbel et al. also relies on the identification of whether the trade was initiated by the buy or sell side based on the assumption that $1mm trades have a negligible effective bid-ask spread. In contrast, using indicative quotes I circumvent the necessity to identify the trade-initiating side and, instead, directly estimate the spread between cash flow equivalent on- and off-the-run securities. Among other papers in the strand of literature going beyond indicative bid-ask spreads are Furfine and Remolona (2002) and Diaz and Escribano (2017). Furfine and Remolona uses the Hasbrouck (1991) price impact measure to estimate liquidity over longer periods based on daily returns. Diaz and Escribano uses a set of measures including the Amihud (2002) measure and their version of the Jankowitsch et al. (2011) dispersion measure. Finally, the spread between on- and off-the-run Treasuries is commonly referred to as the liquidity premium, as in Diaz and Escribano. However, this spread captures not only liquidity but differing financing costs of on- and off-the-run securities, (Krishnamurthy, 2002), as well as repo counterparty risk, (Liu and Wu, 2017).

Second, this paper contributes to the literature on differences in indicative and actual liquidity. In equity markets, Blume and Goldstein (1992) demonstrates notable differences between the effective and displayed spread, with the differences being affected by the posted

[6] Andersen (2007) argues that parametric models do not allow for sufficient flexibility and are not in common use by market makers – a critical requirement for obtaining a benchmark, on-the-run interest rate term structure.

volume available at the quotes. In the market for collateralized loan obligations, Hendershott et al. (2024) shows that indicative quotes in over-the-counter markets may lead to an overestimate of available liquidity when failed attempts to trade are ignored. In this paper, I uncover several biases in indicative bid-ask spreads. In particular, indicative bid-ask spreads overestimate trading costs for seasoned and relatively less liquid securities. Moreover, I show that reliance on the regular relationship between indicative and effective bid-ask spreads leads to an underestimate of the liquidity deterioration in the off-the-run Treasury market during periods of low benchmark, on-the-run Treasury market liquidity. This evidence augments the results in Furfine and Remolona (2002) that off-the-run securities' liquidity – as measured by price impact – deteriorates proportionately more than on-the-run securities' liquidity during periods of market stress.

Finally, this paper contributes to research on how trading costs scale with transacted volume in decentralized, bilateral markets. The majority of existent empirical results are for CLOB-driven markets, with the notable exception of Babbel et al. (2004). Bouchaud et al. (2009), Farmer et al. (2006), and Lillo and Farmer (2004) among others find that trading costs are non-linear in transaction volume. A notable distinction of this paper from most empirical studies in the field is that I allow for the scaling law to vary with market conditions and relative liquidity of the security. Therefore, I consider conditional scaling laws. Furthermore, I find evidence in support of selective liquidity taking – large trades get executed when liquidity is high. This finding complements Hendershott et al. (2024) that draws attention to unobserved failed trades when available liquidity is low.

The rest of the paper is organized as follows. In Section 2, I suggest a method for calculating the effective bid-ask spread for off-the-run Treasuries. In Section 3, I explore the relationship between actual and indicative liquidity by comparing the statistical properties of effective bid-ask spreads and price improvements relative to indicative bid-ask spreads.

In conclusion of that section, I identify various biases in indicative bid-ask spreads using the linear mixed effects modeling framework. In Section 4, I extend the linear mixed effects model to obtain scaling laws for the conditional dependence of the effective bid-ask spread on transacted volume. I allow the scaling with volume to depend on benchmark, on-the-run Treasury market liquidity in Section 4.1 and on relative liquidity characteristics of securities in Section 4.2. In Section 5, I reiterate the main results of the paper and suggest topics for further research. Finally, in a series of appendices, I provide some background details, as well as explore the robustness and extensions of results to various closely related Treasury market segments. In Appendix A I provide a brief overview of on- and off-the-run Treasury market microstructure. Then, I consider (i) the effect of Alternative Trading Systems (ATS) intermediation on liquidity in Appendix B, (ii) interconnectedness of liquidity in the dealer-to-client and dealer-to-dealer segments in Appendix C, and (iii) execution quality of retail clients' trades in Appendix D.

2 Effective Bid-Ask Spread for Off-the-Run Treasuries

The methodological framework of this paper has two components – (i) an algorithm for calculating the effective bid-ask spread for off-the-run Treasuries, and (ii) a linear mixed model framework that allows for random effects associated with each security, while enabling non-parametric exploration of how the effective bid-ask spread scales with order size. I cover the first component in this section and defer the description of the mixed modeling framework until the time it is applied in Section 3.3.

2.1 Advantages over Competing Measure of Liquidity

The effective bid-ask spread (see Demsetz, 1968) is the difference between the trade price of an asset and its fair value. The wider the absolute value of the effective bid-ask spread, the greater the trading costs are.[7] The effective bid-ask spread is a measure of actual, experienced liquidity for each individual transaction. Consequently, transaction characteristics – like traded volume – are linked to each observation of the effective bid-ask spread.

The effective bid-ask spread has three main advantages over indicative bid-ask spreads. First, the effective bid-ask spread is based on actual transactions and not mere indications of where a generic trade could occur (absent possible negotiations). Second, the effective bid-ask spread allows the exploration of how trading costs scale with volume, which is entirely latent in indicative bid-ask spreads. There may not be a precise order size associated with the quote. This is particularly the case for streamed quotes.[8] In general, a 'normal market size' may serve as the implied potential transaction volume. However, what is 'normal' may change depending on market conditions. Finally, due to the bilateral nature of trading in the off-the-run Treasury market and multiple competing intermediaries, obtaining reliable best indicative bid and ask prices for all off-the-run Treasuries may be prohibitively expensive even for some market participants. In summary, market microstructure considerations favor the effective bid-ask spread over the indicative bid-ask spread as a measure of liquidity in the off-the-run Treasury market.

The effective bid-ask spread is also a more fitting measure for off-the-run Treasury liquidity than price impact. Price impact – how much trade or order flow of a given magnitude over a set execution horizon affects the market price – is a prominent metric in Central Limit Order Book (CLOB) markets.[9] On CLOB-based platforms, market participants split their total desired transaction into multiple smaller child orders that they submit over time while targeting a particular average execution price, with possible opportunistic deviations in response to the CLOB dynamics. This manner of execution is supported, if not required, by

[7] Under the assumption that positive (negative) difference between the trade price of an asset and its fundamental value corresponds to buyer-initiated (seller-initiated) transactions.

[8] As noted above, market makers may stream quotes to clients via their in-house platforms or external Alternative Trading Systems (ATS). In contrast to the request for quote (RFQ) protocol, clients do not contact dealers to obtain quotes and, thus, hide their interest in a potential trade.

[9] BrokerTec platform is a prominent example of a CLOB-based market for on-the-run Treasuries.

the limited immediately available liquidity on the CLOB. In stark contrast, such execution would be considered a breach of expected behavior in the bilateral market for off-the-run Treasuries (see Wood, 2018). If a client splits a large order among multiple market makers, the price of the asset is likely to move against the market makers as the information subsequently sips into the market. If a client were to engage in such order splitting, it may expect to receive considerably less favorable quotes in its future interactions with market makers.[10] Thus, the atomic nature of the effective bid-ask spread that does not incorporate flows over a period of time is more suitable for the off-the-run Treasuries market. In addition, as noted in Hendershott et al. (2024), accurate estimation of price impact at a daily frequency requires more transactions than commonly occur in many individual off-the-run Treasuries.

2.2 Estimation Algorithm

The analytical definition of the *effective bid-ask spread*, adapted from Hagströmer (2021), is:

$$2D_i\left(P_{i,t}^{trade} - P_{i,t}^{fair}\right), \tag{1}$$

where $P_{i,t}^{trade}$ is the transaction price and $P_{i,t}^{fair}$ is the fair value of security i at time t; D_i is equal to 1 for buyer-initiated and -1 for seller-initiated trades. The scaling by 2 expresses the effective bid-ask spread on the same scale as the indicative bid-ask spread. For transaction prices I rely on the comprehensive Treasury Trade Reporting and Compliance Engine (TRACE) data from the Financial Industry Regulatory Authority (FINRA). Since Treasury TRACE data do not contain an explicit identifier of the side initiating a transaction, I make the assumption that aggressive buying (selling) leads to positive (negative) deviations of the

[10] The same logic does not apply in centralized, CLOB-based markets. While the gradual execution of a large order would also move the price against market makers that provided liquidity for the earlier child orders, the anonymity on CLOB-based markets does not allow market makers to retaliate against clients in future interactions based on their past behavior.

trade price from the fair value, that is, I set:

$$D_i := \mathbf{sign}\left(P_{i,t}^{trade} - P_{i,t}^{fair}\right).$$

In order to make the effective bid-ask spread comparable for securities of different maturities, I introduce the *duration-normalized effective bid-ask spread*:

$$\mathbf{EBA}_{i,t} = 2 \cdot \frac{\left|P_{i,t}^{trade} - P_{i,t}^{fair}\right|}{\mathbf{duration}_{i,t}}, \tag{2}$$

where $\mathbf{duration}_{i,t}$ measures the sensitivity of security i's price to changes in interest rates at time t (its modified duration multiplied by its price). Normalization by duration expresses the effective bid-ask spread in the yield space, rather than in the price space – duration-normalized effective bid-ask spread is the parallel shift of the interest rate term structure that would cause twice the dollar difference between the transaction price and the fair price of a security. Such normalization of the effective bid-ask spread by duration enables comparison and unified modeling of liquidity across the curve. For brevity, I henceforth refer to the duration-normalized effective bid-ask spread in Eq. (2) as the effective bid-ask spread.

While the definition of the effective bid-ask spread is straightforward, its estimation is complex. The trade price is observed, but the fair value has to be estimated. The most common estimate of the fair value is the mid-point between the best bid and ask quotes at the time of the trade. However, Hagströmer (2021) shows that this fair value estimate biases the effective bid-ask spread measure. Moreover, unlike in the equity market, best bid and ask prices are opaque for off-the-run Treasury securities. Trading in off-the-run Treasuries is bilateral and relatively reliable quotes from a specific dealer are available only via the request for quote protocol; streamed quotes tend to be a notably rougher indication of price.[11] Thus,

[11] Clients cannot routinely request quotes from the entire universe of dealers to assess the true best bid and ask quotes. Requesting quotes without a certain amount of trading would violate the market etiquette – in many cases, generation of off-the-run Treasury quotes is not fully automated and is, thus, costly to the dealer.

some indicative quotes may differ from the best available bids and offers. Also, conversations with market participants indicate that market makers differ in the relative firmness of their indicative quotes – with some being more reliable indications of the price at which a trade can occur. Furthermore, in bilateral relationships the identities of counterparties are known and market makers may shade their quotes based on their assessment of how informed the client is (see Wood, 2018). Consequently, there is no single price that would be available to all market participants.

Guided by the Treasury market microstructure, I estimate off-the-run Treasuries' fair values based on transparent firm quotes for on-the-run Treasuries from the leading BrokerTec ATS and indicative NPQS quotes from the Federal Reserve Bank of New York. In greater detail, to calculate the fair value of an off-the-run Treasury, $P_{i,t}^{fair}$, I deconstruct it into the benchmark and idiosyncratic price components, as follows.

The *benchmark component* of the off-the-run Treasury's price is the price of the hypothetical on-the-run security with equivalent cash flows – it reflects the interest rate term structure under conditions of high liquidity and low financing costs. An *idiosyncratic component* of the off-the-run Treasury's price captures the effects of lower liquidity and less favorable financing of a specific security relative to the on-the-run benchmark. The benchmark component reflects the information in news announcements and flows of funds. Consequently, the benchmark component may be highly variable intraday. In contrast, the financing cost and liquidity premia are relatively stable intraday – these premia primarily depend on the financing and liquidity conditions over the remaining life of the bond.[12]

Due to higher liquidity and trading volumes in the on-the-run Treasury market, it is the locus of Treasury price discovery. Thus, to estimate the benchmark price component I use

[12] Cheap financing and high liquidity benefit the investor over the holding period, and their values over the remaining life of the security will likewise affect future holders of the security. Consequently, financing and liquidity characteristics over the remaining life of the security determine the corresponding premia. That said, value of liquidity can spike during market stress. In Section 3.1, I introduce a forecast accuracy filter that mitigates the effect of such spikes on liquidity measurement, albeit at the cost of retaining fewer observations.

firm on-the-run Treasuries' quotes from one of the leading inter-dealer broker (IDB) BrokerTec Alternative Trading System (ATS).[13] Crucially, firm on-the-run Treasuries' quotes are available at a high frequency, which precludes pricing based on stale information.Another argument in favor of the on-the-run Treasury market as the source of the benchmark price component is that off-the-run Treasuries can be quoted by dealers in terms of their spreads to the corresponding on-the-run securities. Dealers also engage in *swap box* trading where they effectively hedge their off-the-run Treasuries with on-the-run securities within a single transaction. Finally, the differences between NPQS and BrokerTec ATS quotes for on-the-run securities are generally negligible, suggesting that market makers refer to prominent CLOB-driven ATS when quoting prices for on-the-run securities to reduce the probability of being exploited by arbitrageurs.

A term structure model ingests firm on-the-run Treasury quotes to generate the benchmark price component for any Treasury characterized by its remaining cash flows. At each trade time t in Treasury TRACE data, I fit the state-of-the-art machine learning term structure model of Filipović et al. (2022).[14] To reiterate, the benchmark price component is the value of a security's cash flows when discounted using time t fitted on-the-run term structure.

Next, to estimate the idiosyncratic price component – capturing the security's financing and liquidity premia – I use indicative NPQS quotes at 08:40am, 11:30am, 2:15pm, and 3:30pm. For each NPQS quote snapshot, I fit a Filipović et al. (2022) term structure model to the NPQS quotes for on-the-run Treasuries. For each off-the-run Treasury security, the difference between the NPQS quote and the value of its cash flows when discounted using the

[13] Other prominent ATS include Dealerweb and Fenics UST. Due to principal trading firms being active on all major ATS, the mid points between best bid and ask quotes on each major ATS must be sufficiently close to prevent straightforward arbitrage.

[14] The term structure is fitted with the smoothness parameter $\lambda = 1$ (before scale normalization), maturity weight of $\alpha = 0.05$, and tension parameter $\delta = 0.001$ (see Filipović et al., 2022, for details). I found that the method of Filipović et al. (2022) delivers superior bond yield forecasting accuracy relative to popular alternatives, including Fama and Bliss (1987), Nelson and Siegel (1987), Tanggaard (1997) and Andersen (2007).

fitted on-the-run term structure is the idiosyncratic price component. I take these idiosyncratic price components to be constant between consecutive NPQS quote snapshots within the trading day.

Finally, the fair value of an off-the-run Treasury i at time t, $P_{i,t}^{fair}$, is the time t benchmark price component adjusted for the security's idiosyncratic price component estimated at the latest NPQS quote snapshot time that precedes time t.

3 Actual v. Indicative Liquidity

3.1 Sample Construction

The sample comprises dealer-to-client trades in nominal Treasury Notes and Bonds, covering the period from January 2018 to June 2024. ATS-intermediated dealer-to-client trades and direct dealer-to-dealer trades are the subjects of Appendices B and C, respectively. I also remove likely retail trades by considering transactions of at least $10mm in notional value; I consider retail trades in Appendix D.[15]

Only transactions during the active trading hours between 8:40am and 5:00pm are retained, which tempers the liquidity seasonality at the market opening. With most of the repo trades done early in the trading day, the financing conditions are largely determined before 8:40am. Thus, the assumption of stable intraday financing premiums in Section 2 is more likely to hold for transactions that occur post 8:40am.

Since an accurate determination of the fair value is critical to the measurement of the effective bid-ask spread, I retain transactions in securities for which the model achieves accurate forecasts. Specifically, I consider intraday forecasts of mid-quotes at three snapshot times – 11:30am, 2:15pm, and 3:30pm.[16] I retain security i on day d in the sample when the

[15] This threshold is somewhat discretionary – it may also lead to the omission of very small institutional trades. Also, wealthy individuals may transact in more than $10mm notional amount of Treasuries, but such trades are likely to be relatively rare.

[16] The price forecasts are model estimates of fair values based on the most recent benchmark, on-the-run

maximum absolute value of the forecast error normalized by the indicative bid-ask spread for security i on day d is less than one half. Analytic expression for this requirement is:

$$\mathbf{max}\left\{\frac{\left|P_{i,t}^{fair}-\frac{P_{i,t}^{ask}+P_{i,t}^{bid}}{2}\right|}{P_{i,t}^{ask}-P_{i,t}^{bid}}\right\}_{t\in\mathcal{T}}<\frac{1}{2},$$

where $\mathcal{T}=\{d\ 11:30am, d\ 2:15pm, d\ 3:30pm\}$ are indicative quote snapshot times on day d.

As many off-the-run Treasuries trade rather infrequently, making it difficult to reliably isolate the security-level effect from that of explanatory variables, I construct a sample of actively traded securities. First, I select 250 off-the-run Treasury securities with the largest number of trades in the full sample. Second, on any given trading day, I retain trades in a security only if it traded at least 50 times during that day. These sample selection criteria justify a caveat that the current analysis applies to the more liquid segment of the off-the-run Treasury market. However, it is important to note that the sample is not dominated by cheapest-to-deliver securities – securities that market participants prefer to deliver into Treasury futures.

Finally, to further reduce the effect of outliers and non-obvious data-entry errors in Treasury TRACE data, I aggregate effective bid-ask spreads over a set of volume bins that reflect common trade sizes. Specifically, the effective bid-ask spread measure is set to its median within each aggregation group defined by the tuple of the security identifier, trading day, and volume bin, $\{i,d,v\}$:

$$\mathbf{EBA}_{i,d,v}^{med}:=\mathbf{med}\left(\{\mathbf{EBA}_{i,t}\}_{t\in d,v\in V}\right), \tag{3}$$

where i is the security index, t is the time of the trade, d is the trading day ($t\in d$), **med**

quotes and prior estimates of idiosyncratic components. On each trading day, the quote snapshot at 08:40am is necessary for estimating initial idiosyncratic price components and, thus, belongs to the in-sample period.

is the median operator, and V is the set of volume bins defined by the cutoff values of $\{10, 15, ..., 50, 60, ..., 100, 125, ...250\}$, in millions of dollars. Due to their rarity, I exclude trades with notionals greater than $250mm.

In summary, such filtering and aggregation of Treasury TRACE data results in a sample of 101,978 observations of trades in 118 different Treasury securities over 1,589 trading days. Other data sources include BrokerTec inter-dealer-broker ATS for on-the-run Treasury quotes and the Federal Reserve Bank of New York NPQS for lower-frequency intraday snapshots of all Treasury securities' quotes.

3.2 Distributions of Effective Bid-Ask Spreads and Price Improvements

In this section, I describe the empirical distributions of the effective and indicative bid-ask spreads. Panel A of Figure 1 depicts the empirical density of the effective bid-ask spread. While the majority of the effective bid-ask spreads fall in the zero to two basis points range, the heavy right tail is evident even when the graph is truncated at 10 basis points. The natural conjecture is that market participants obtain good execution for most transactions, while paying substantially larger trading costs for a not insignificant number of transactions. Panel A of Table 1 contains summary statistics for the effective and indicative bid-ask spreads. The median of the effective bid-ask spread is below that of the indicative bid-ask spread – suggesting better-than-expected execution for most trades. On the contrary, the right tail is notably more pronounced for the effective relative to the indicative bid-ask spread, as evidenced by the comparison of their 95^{th} percentiles. Kurtosis of the indicative bid-ask spread is above 21, while that of the effective bid-ask spread is an order of magnitude larger still. The positive skew of the effective bid-ask spread distribution also far exceeds the still notable skew of the indicative bid-ask spread.

To quantify the deviation between actual and expected execution quality, I suggest a *price improvement metric*:

$$\mathbf{IMPRVT}_{i,d,v} = 1 - \mathbf{med}\left(\left\{\frac{\mathbf{EBA}_{i,t}}{\mathbf{IBA}_{i,t-}}\right\}_{t \in d, v \in V}\right), \tag{4}$$

where $\mathbf{EBA}_{i,t}$ is the effective and $\mathbf{IBA}_{i,t-}$ is the indicative bid-ask spread for security i, expressed in basis points, at times t and $t- := max\,(s \in \{quote\ snapshot\ times\} : s \leq t)$, respectively. Thus, the price improvement is measured relative to the indicative bid-ask spread – it is positive only when the effective bid-ask spread is less than the indicative bid-ask spread, with the maximal improvement being unity. Negative price improvement values correspond to trades where the effective bid-ask spread exceeds the corresponding indicative bid-ask spread.

Panel B of Figure 1 depicts the empirical probability density of price improvements. The shape of the distribution validates the above conjecture that the majority of trades occur at prices that are more favorable to the initiating side than indicative bid-ask spreads suggest. One explanation for the apparent ubiquity of price improvements is that clients conduct mini-auctions among market makers – a client obtains quotes from multiple market makers and selects the quote corresponding to the smallest effective bid-ask spread. Another possibility is that market makers see their quotes as a first step in negotiations and want to have some leeway to provide their clients with a 'good deal' relative to their initial indication. Finally, indicative quotes may reflect quotes available for a generic client, while actual client relationships allow for better prices available to preferred and less informed clients. Since the majority of clients likely fall into the category of less informed market participants – participants that do not have superior ability to forecast Treasury yields – we observe price improvements for the majority of trades. There are also many trades that occur at prices that are significantly worse to the initiating side than indicative bid-ask spreads suggest. This may reflect either trading by informed clients or larger transaction volumes – I discuss the latter possibility next.

Figure 2 depicts empirical probability densities of effective bid-ask spreads and price improvements separately for trades of less than and greater than \$50mm in notional value. The probability of a lower effective bid-ask spread and a corresponding price improvement are considerably higher for lower volume transactions. Panels B and C of Table 1 contain summary statistics for effective bid-ask spreads and price improvements conditional on transaction volume of below and above \$50mm, respectively. For smaller-volume transactions, the 95^{th} percentiles of the effective and indicative bid-ask spreads are nearly identical. On the other hand, the right tail of the effective bid-ask spread, as measured by its 95^{th} percentile, is considerably greater for larger-volume transactions. Finally, the median effective and indicative bid-ask spreads are nearly identical for larger-volume transactions, with the corresponding price improvements being rather modest for half the trades in this category.

3.3 Biases in Indicative Bid-Ask Spreads

In this section, I construct a series of linear mixed models to document biases in indicative bid-ask spreads. The linear mixed model framework naturally fits the data structure – as the same security may be traded multiple times during a given day, there is natural variance error clustering associated with the grouping of observations by traded security. A linear mixed model allows for random effects associated with each security:

$$\mathbf{EBA}_i = \alpha_0 + \alpha_1 \cdot \mathbf{IBA}_i + \mathbf{X}_i \gamma + Z_i \beta + \epsilon_i, \tag{5}$$

where i is the trade index, **EBA** and **IBA** are the effective and indicative bid-ask spreads, respectively; $\mathbf{X}_i$ is a vector of variables that may help detect biases in indicative bid-ask spreads; $\{\alpha_0, \alpha_1, \gamma\}$ are fixed effect coefficients; Z_i is the i^{th} row of the $n \times s$ random-effects model matrix – a sparse indicator matrix that captures the grouping of n observations by s traded securities; β is the vector of random effects that have a multivariate normal distribution, $\mathcal{N}(\mathbf{0}, \Sigma)$; ϵ_i is the noise term.

To detect potential biases in indicative bid-ask spreads, I consider four model specifications that differ in explanatory variables, $\mathbf{X}_i$. Within the specification of Eq. (5), the hypothesis of unbiased indicative bid-ask spreads corresponds to the joint restriction of $\alpha_0 = 0$, $\alpha_1 = 1$, and $\gamma = \mathbf{0}$. Table 2 contains maximum likelihood parameter estimates of these models.[17]

Model I is the benchmark specification corresponding to an empty set of control variables, X. The intercept is positive and significant, suggesting a constant bid-ask spread component that is not represented by the variation in indicative bid-ask spreads. At the same time, the coefficient in front of the indicative bid-ask spread is significantly below unity, which aligns well with commonly observed price improvements noted in Section 3.2. Thus, already a simple benchmark model suggests that indicative bid-ask spreads do not translate nearly one-to-one to effective bid-ask spreads.

Model II introduces an interaction term between the indicative bid-ask spread and security's relative age, $IBA \cdot RA$. I define a security's relative age as the ratio of the remaining time to maturity to its original time to maturity. The significant negative coefficient in front of $IBA \cdot RA$ suggests that market makers tend to quote overly conservative indicative spreads for more seasoned securities. Consequently, measures of off-the-run v. on-the-run liquidity premiums are inflated if they are based on indicative bid-ask spreads, especially for more seasoned securities. In Model II, the coefficient in front of the indicative bid-ask spread is notably higher than in Model I, suggesting a closer correspondence between indicative and effective bid-ask spreads for more recently issued securities; however, it still remains significantly below unity. For a security that has just become off-the-run, a one basis point change in its indicative bid-ask spread translates into an expected 0.73 basis points change in its effective bid-ask spread. As a result, a widening of indicative bid-ask spreads

[17] Maximum likelihood estimation follows Bates et al. (2015) – it involves repeated applications of penalized least squares, which allows for expressions of various probability densities required for the calculation of the log-likelihood.

may exaggerate the magnitude of an actual liquidity deterioration.

Model III extends Model II by introducing an interaction term between the indicative bid-ask spread and an indicator of whether the security is the cheapest-to-deliver (CTD) into a Treasury futures contract, $IBA \cdot CTD$.[18] CTD securities are Treasury securities that parties holding short Treasury futures positions find most profitable to deliver into the futures, if they were to make a delivery. Given the demand for CTD securities from the Treasury futures market participants (e.g., from Treasury cash-futures basis traders), CTD securities are more likely to trade special in the repo market (resulting in cheaper financing) and exhibit greater liquidity. The coefficient in front of the interaction term, $IBA \cdot CTD$, is insignificant. Thus, indicative bid-ask spreads properly reflect any effects induced by the CTD status of Treasuries.

Model IV extends Model III by introducing an interaction term between the indicative bid-ask spread in the off-the-run market and a measure of liquidity in the corresponding maturity sector of the benchmark, on-the-run Treasury market. To some extent, Model IV complements the conditional volume scaling analysis of Section 4.1, but in a simpler framework. Next, I describe the construction of control variables for the on-the-run Treasury market liquidity, which must be tailored to corresponding off-the-run securities.

Liquidity conditions can differ notably with a security's maturity, as exemplified by notably worse deterioration in the liquidity of shorter maturity Treasuries during the most recent monetary policy tightening cycle. Recognizing maturity-dependent liquidity conditions, I assign off-the-run Treasury securities to a set of duration groups.[19] Then, I assess benchmark, on-the-run Treasury market liquidity conditions separately for each considered duration group. Specifically, I first match each duration group to an appropriate on-the-run

[18] CTD securities' CUSIPs are identified based on data from J.P. Morgan Chase & Co., MorganMarkets and DataQuery, https://markets.jpmorgan.com.

[19] Duration is a measure of interest sensitivity that depends on the security's remaining time to maturity. I consider four duration groups $\{(1,2], (2,5], (5,10], (10,30]\}$.

Treasury security.[20] Combining the above two maps, each off-the-run security, s may be associated with an on-the-run security $\xi(s)$. Then, for each trading day, I use a transform of the time-weighted average bid-ask spread from the BrokerTec ATS as a measure of overall benchmark, on-the-run market liquidity in the corresponding sector of the curve.[21] A bid-ask spread on the BrokerTec ATS is bounded from below by the corresponding tick size – the minimal price change increment – prescribed by the trading platform. In general, tick sizes differ for on-the-run Treasury securities of different maturities. To capture the liquidity conditions in different sectors of the market on a common scale, I consider a security-specific empirical cumulative probability of the bid-ask spread as the measure of the benchmark, on-the-run Treasury market liquidity in a duration sector:

$$\theta_i = \mathbf{ecdf}_{\xi(s)}\left(\mathbf{BBA}_d^{\xi(s)}\right), \tag{6}$$

where i is the transaction index for a trade in security s, belonging to a duration group that maps to the on-the-run security $\xi(s)$, during trading day d; $\mathbf{BBA}_d^{\xi(s)}$ is the time-weighted average – with weights corresponding to lengths of time during which each spread level prevailed – bid-ask spread of the on-the-run benchmark, on-the-run Treasury security $\xi(s)$ during trading day d; $\mathbf{ecdf}_{\xi(s)}$ is the empirical cumulative distribution function of the bid-ask spread for on-the-run security $\xi(s)$. High values of θ correspond to higher bid-ask spreads for benchmark, on-the-run securities and, thus, lower benchmark liquidity.

Estimates of Model IV suggest that benchmark, on-the-run Treasury market liquidity is a significant scaling factor for indicative bid-ask spreads. When benchmark, on-the-run liquidity is low – θ is high – changes in indicative bid-ask spreads have a greater impact on the corresponding effective bid-ask spreads. This effect is sufficiently strong to make

[20] In detail, I use the 2-, 5-, 10-, and 30-year on-the-run Treasury securities as the corresponding benchmarks for $\{(1,2], (2,5], (5,10], (10,30]\}$ duration groups.

[21] Time-weighted average bid-ask spread on day d is the weighted average of bid-ask spreads where weights reflect the lengths of time that a given spread was observed during the active hours of trading day d.

the coefficient in front of the unscaled indicative bid-ask spread, $\mathbf{IBA}_i$, insignificant. Thus, the relationship between indicative and effective bid-ask spreads for off-the-run Treasuries depends on liquidity in the benchmark, on-the-run Treasuries market. Section 4 develops these insights further by conditioning the dependence between the effective bid-ask spread and transaction volume on liquidity in the on-the-run Treasuries market.

Since clients in the off-the-run Treasuries market do not commonly split their volumes among multiple dealers to obtain their best prices, transaction volume is particularly relevant for the effective bid-ask spread. Models of this section served to document biases in the simplest framework, with the non-linear effects of volume left for exploration in the following section. In Section 4, I focus on the relationship between the effective bid-ask spread and transaction volume. However, already at this stage, I note that the bias of indicative bid-ask spreads against more seasoned Treasury securities remains significant when controlling for volume. Furthermore, Section 4.1 extends the results on the effect of benchmark, on-the-run Treasury market liquidity. Beyond illuminating the dependence between the effective bid-ask spread and order volume, Section 4.1 may be seen as an extension of Model II that controls for order volume within a flexible, non-parametric specification.

4 Scaling of Trading Costs with Order Volume

Several channels exist for why transaction volume affects the trade price and, thus, the effective bid-ask spread. First, the informational content of a trade may vary with volume, as in Barclay and Warner (1993). Second, to support larger transaction sizes, market makers need to either carry a larger inventory on their balance sheet – the channel explored in Cohen et al. (2023) – or be able to source a large amount of the security in the inter-dealer market. The third channel is the endogeneity in liquidity-taking – large trades are entered into when there is sufficient liquidity.[22] Thus, selective liquidity-taking would suggest that clients in the off-the-run Treasury market are sufficiently sophisticated and opportunistic to seek to avoid large price dislocations. Either one or, more likely, a combination of these channels gives rise to complex scaling of the effective bid-ask spread with transaction volume.

Power law scaling is a common choice for modeling price impact's dependence on volume (see Plerou et al., 2004; Zhang, 1999, among others). In this section I find that one-dimensional power law models are overly restrictive for two reasons. First, selective liquidity taking may cause a non-monotone relationship between the effective bid-ask spread and transaction volume. Second, other explanatory variables – like benchmark, on-the-run Treasury liquidity – can affect the dependence of the effective bid-ask spread on volume. Ignoring such explanatory variables is tantamount to estimating an average scaling law, as opposed to the scaling law conditional on relevant market and security-specific variables.

The linear mixed model framework is well-suited for modeling how the effective bid-ask spread scales with transaction volume. First, the task requires retention of multiple transactions for each security on any given trading day. Such data structure is supported by a linear mixed model that accounts for random effects associated with each security. Second, the shape of the relationship between trading costs and transaction volume may be too complex for a simple polynomial model. The linear mixed model framework has the advantage of allowing for non-parametric models of multivariate surfaces through the tensor spline methodology of Wood et al. (2013). In particular, bivariate tensor splines enable the relationship between trading costs and transaction volume to be conditional on another variable. I identify two conditioning variables that significantly affect the scaling relationship – benchmark, on-the-run Treasury market liquidity conditions and the relative liquidity of the off-the-run security.

To accommodate the dependence of the scaling law on another market or security-specific

[22] Farmer et al. (2004) suggests that large price moves occur due to market participants taking liquidity when it is scarce.

variable, I model the effect of volume via the two-dimensional tensor product smoothing method of Wood et al. (2013). Following the results of Section 3.3, in all model specifications, I control for the indicative bid-ask spread and its interaction with the relative age of the security. So, I extend the model specification in Eq.(5) to the non-parametric framework of Wood et al. (2013):

$$\mathbf{EBA}_i = \alpha_0 + \alpha_0 \cdot \mathbf{IBA}_i + \alpha_1 \cdot \mathbf{IBA}_i \times \mathbf{RA}_i + Z_i\beta + \left[\sum_j L_{ij}\mathcal{T}_j\left(\mathbf{V}_i, \theta_i\right)\right] + \epsilon_i, \tag{7}$$

where $\mathbf{EBA}_i$ and $\mathbf{IBA}_i$ are the effective and indicative bid-ask spreads for transaction i in security s, $\mathbf{IBA}_i \times \mathbf{RA}_i$ is the interaction term between the indicative bid-ask spread and security's s relative age. $\mathcal{T}_j\left(\mathbf{V}_i, \theta_i\right)$ are unknown smooth functions of transaction volume, $\mathbf{V}_i$, and of the explanatory variable, θ_i, that affects the scaling relationship. The degree of smoothness of functions $\mathcal{T}_j\left(\mathbf{V}_i, \theta_i\right)$ is not known in advance but there is an associated penalty functional, $J_j\left(\mathcal{T}\right)$, for each function. L_{ij} are, in general, known linear functionals.

In Section 4.1, I show how the effective bid-ask spread scales with volume under different benchmark, on-the-run Treasury market liquidity conditions. Then, in Section 4.2, I explore whether the scaling is different for relatively less liquid securities among the off-the-runs.

4.1 Scaling Law Conditional on Benchmark, On-the-Run Treasury Liquidity

In this section, I consider how the effective bid-ask spread scales with transaction volume under different benchmark, on-the-run Treasury market liquidity conditions. The construction of a maturity sector-specific measure of benchmark liquidity conditions follows the approach in Section 3.3 – specifically, I set θ_i in Eq. (7) in accordance with the definition in Eq. (6).

Figure 3 shows the conditional effect of volume on the expected effective bid-ask spread, as well as the associated 95 percent confidence intervals. I consider the scaling law conditional on either median or bad benchmark, on-the-run Treasury liquidity conditions, corresponding to $\theta = 0.5$ and $\theta = 0.75$, respectively. Controlling for indicative bid-ask spreads, when benchmark liquidity is low, transaction volume has a uniformly greater positive effect on the effective bid-ask spread – indicating higher trading costs. This finding extends the results of Model IV from Section 3.3 – indicative bid-ask spreads do not adequately capture how off-the-run Treasury market liquidity changes with benchmark, on-the-run Treasury liquidity conditions.

The relationship between the effective bid-ask spread and transaction volume is concave, irrespective of benchmark, on-the-run liquidity conditions. A novel empirical finding is that the effect of selective liquidity-taking – clients transacting in large volumes only when market makers offer sufficiently attractive quotes – can make incremental order volume decrease the expected effective bid-ask spread for sufficiently large transactions. The resulting non-monotone dependence cannot be adequately described by classic power laws. The point at which the effect of incremental volume on the effective bid-ask spread becomes negative is at a lower volume level under bad benchmark, on-the-run liquidity conditions. This observation suggests that selective liquidity-taking motive is stronger when benchmark liquidity is lower.

4.2 Scaling Law for Securities of Varying Liquidity

In this section, I consider whether relatively less liquid securities exhibit different dependence of the effective bid-ask spread on transaction volume. To determine the relative liquidity of securities on a specific trading day, I rank securities within their duration groups in accordance with their indicative bid-ask spreads. Then, I measure the relative liquidity of the security by an empirical cumulative probability of the indicative bid-ask spread – a fraction of securities with lower indicative bid-ask spreads within the duration group on a given day. The greater the measure, the less the relative liquidity of the security. Since the ranking of securities is specific to both trading day and duration group, overall Treasury market liquidity does not affect this measure of relative liquidity.

Analytically, the relative liquidity variable, θ_i is:

$$\theta_i = \mathbf{ecdf}_{\psi,d}\left(\mathbf{IBA}_d^s\right), \tag{8}$$

where i is the transaction index for a trade in security s, belonging to duration group ψ, during trading day d; $\mathbf{IBA}_d^s$ is the simple average indicative bid-ask spread of security s during trading day d, estimated by an average of the corresponding bid-ask spreads over NPQS quote snapshots on day d; $\mathbf{ecdf}_{\psi,d}$ is the empirical cumulative distribution function of average indicative bid-ask spreads on day d for the duration group ψ. High values of θ correspond to comparatively less liquid securities within their duration groups for the particular trading day.

Figure 4 depicts the conditional effect of volume on the expected effective bid-ask spread, as well as associated 95 percent confidence intervals. I consider scaling laws conditional on median and lower liquidity securities, corresponding to $\theta = 0.5$ and $\theta = 0.75$, respectively. The level and shape of the scaling law for median liquidity securities is essentially similar to the scaling law obtained when conditioning on benchmark, on-the-run market liquidity (see Section 4.1 above). For less liquid securities, the effect of volume is somewhat lessened for transactions of modest volumes. The results are notably more pronounced for larger transaction volumes, albeit confidence intervals are also wider: indicative bid-ask spreads may notably overstate trading costs for larger volumes of relatively less liquid off-the-run securities. Moreover, the threshold defining the region where selective liquidity taking prevails occurs for notably smaller volumes. These results suggest that higher indicative bid-ask spreads tend to be overly conservative relative to the median indicative bid-ask spreads. In other words, indicative bid-ask spreads tend to present a somewhat distorted picture of relative liquidity across different off-the-run Treasuries. One explanation is that relatively higher indicative bid-ask spreads may be biased by some dealers that prefer not to trade in the particular security, while clients can still locate dealers that have no such preference against the potential trade.

5 Conclusion

Actual liquidity experienced by clients in the off-the-run Treasury market can differ systemically and notably from the levels reflected in indicative bid-ask spreads. In particular, I show that indicative bid-ask spreads are biased for seasoned Treasury securities, Moreover, indicative bid-ask spreads have a greater effect on actual trading costs when benchmark, on-the-run Treasury market liquidity is low.

While better-then-indicated execution is common for moderate transaction volumes, the effective bid-ask spread of large trades can be substantially wider. Large order sizes incur notable trading costs that are often far in excess of what indicative bid-ask spreads suggest. And it is the capacity to transact in significant volume without causing price dislocations that is necessary for securities to act as benchmarks and near risk-free collateral. Thus, the dependence of the effective bid-ask spread on transaction volume – the scaling law – is critical for assessing actual liquidity.

In agreement with earlier studies, I also find a non-linear relationship between trading costs and transaction volume. In this paper, I introduce two key features that sharpen the results. First, I consider the relationship between the effective bid-ask spread and transaction volume conditional on either benchmark, on-the-run Treasury market liquidity or the relative liquidity of a particular security – this is in contrast to unconditional modeling that is prevalent in the existent literature. I find that both conditioning variables have notable effects on the relationship between the effective bid-ask spread and transaction volume. Second, I obtain evidence of selective liquidity-taking – market participants time their large trades to coincide with the willingness of market makers to quote large volumes at favorable prices. The inflection point where the effective bid-ask spread starts to decrease with marginal transaction volume marks the start of the region where selective liquidity-taking becomes the dominant factor. The resultant non-monotonicity of the effective bid-ask spread's dependence on transaction volume cannot be captured by power laws.

The empirical framework of this paper can support a number of directions for future research into the off-the-run Treasury market liquidity. First, future research can model liquidity heterogeneity – market makers may offer different quality of execution under various market conditions. If clients have strong ties to specific market makers – perhaps, due to other lines of business – clients will be exposed to heterogeneous liquidity conditions in the market. Furthermore, clients vary in their information advantage and negotiating power. Resultant price discrimination by market makers may lead to vastly different liquidity conditions for different clients. Second, future research can investigate the time series aspects of aggregate liquidity in the off-the-run Treasury market. Specifically, it can explore the association between aggregate liquidity in the off-the-run Treasury market and interest rate uncertainty under various market conditions (possibly, along the lines of Meldrum and Sokolinskiy, 2025). The off-the-run market liquidity's sensitivity to interest rate uncertainty may increase with the level of market stress. Such non-linearity may have notable financial stability implications as it would magnify the probability of a market dysfunction.

References

Adrian, Tobias, Michael J Fleming, and Kleopatra Nikolaou, "US Treasury Market Functioning from the GFC to the Pandemic," *FRB of New York Staff Report*, 2025, (1146).

Amihud, Yakov, "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of financial markets*, 2002, *5* (1), 31–56.

Andersen, Leif, "Discount Curve Construction with Tension Splines," *Review of Derivatives Research*, 2007, *10* (3), 227–267.

Babbel, David F, Craig B Merrill, Mark F Meyer, and Meiring De Villiers, "The Effect of Transaction Size on Off-the-Run Treasury Prices," *Journal of Financial and Quantitative Analysis*, 2004, *39* (3), 595–611.

Barclay, Michael J and Jerold B Warner, "Stealth Trading and Volatility: Which Trades Move Prices?," *Journal of financial Economics*, 1993, *34* (3), 281–305.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 2015, *67* (1), 1–48.

Blume, Marshall E and Michael A Goldstein, "Displayed and Effective Spreads by Market," *Rodney L. White Center for Financial Research Working Paper*, 1992, (27-92).

Bouchaud, J.-P., J. D. Farmer, and F. Lillo, "How Markets Slowly Digest Changes in Supply and Demand," in T. Hens and K. R. Schenk-Hoppé, eds., *Handbook of Financial Markets: Dynamics and Evolution*, Elsevier, 2009, pp. 57–160.

Cohen, Assa, Mahyar Kargar, Benjamin R Lester, and Pierre-Olivier Weill, "Inventory, Market Making, and Liquidity in OTC Markets," *Available at SSRN 4650063*, 2023.

Demsetz, Harold, "The Cost of Transacting," *The quarterly journal of economics*, 1968, *82* (1), 33–53.

Diaz, Antonio and Ana Escribano, "Liquidity Measures throughout the Lifetime of the US Treasury Bond," *Journal of Financial Markets*, 2017, *33*, 42–74.

Duffie, Darrell, "Resilience redux in the US Treasury market," in "Jackson Hole Symposium, Federal Reserve Bank of Kansas City, August" 2023.

_ , “How US Treasuries Can Remain the Worldâs Safe Haven,” *Journal of Economic Perspectives*, 2025, *39* (2), 195–214.

_ **, Michael J Fleming, Frank M Keane, Claire Nelson, Or Shachar, and Peter Van Tassel**, “Dealer Capacity and US Treasury Market Functionality,” *FRB of New York Staff Report*, 2023, (1070).

Fama, Eugene F and Robert R Bliss, “The Information in Long-Maturity Forward Rates,” *The American Economic Review*, 1987, pp. 680–692.

Farmer, Doyne J, Laszlo Gillemot, Fabrizio Lillo, Szabolcs Mike, and Anindya Sen, “What Really Causes Large Price Changes?,” *Quantitative finance*, 2004, *4* (4), 383–397.

Farmer, J. D., A. Gerig, F. Lillo, and S. Mike, “Market Efficiency and the Long-Memory of Supply and Demand: Is Price Impact Variable and Permanent or Fixed and Temporary?,” *Quantitative Finance*, 2006, *6* (02), 107–112.

Filipović, Damir, Markus Pelger, and Ye Ye, “Stripping the Discount Curve – a Robust Machine Learning Approach,” *Swiss Finance Institute Research Paper*, 2022, (22-24).

Financial Industry Regulatory Authority (FINRA), *Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE).*

Fleming, Michael J, “Measuring Treasury Market Liquidity,” *FRB of New York Staff Report*, 2001, (133).

_ **and Francisco Ruela**, “Treasury Market Liquidity during the COVID-19 Crisis,” Technical Report, Federal Reserve Bank of New York 2020.

Furfine, C and Eli Remolona, “Whatâs behind the Liquidity Spread? On-the-Run and Off-the-Run US Treasuries in Autumn 1998,” *BIS Quarterly Review*, 2002, *6*, 51–58.

Goyenko, Ruslan, Avanidhar Subrahmanyam, and Andrey Ukhov, "The Term Structure of Bond Market Liquidity and its Implications for Expected Bond Returns," *Journal of Financial and Quantitative Analysis*, 2011, *46* (1), 111–139.

Hagströmer, Björn, "Bias in the Effective Bid-Ask Spread," *Journal of Financial Economics*, 2021, *142* (1), 314–337.

Hasbrouck, Joel, "Measuring the Information Content of Stock Trades," *The Journal of Finance*, 1991, *46* (1), 179–207.

Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff, "When Failure is an Option: Fragile Liquidity in Over-the-Counter Markets," *Journal of Financial Economics*, 2024, *157*, 103859.

Jankowitsch, Rainer, Amrut Nashikkar, and Marti G Subrahmanyam, "Price Dispersion in OTC Markets: A New Measure of Liquidity," *Journal of Banking & Finance*, 2011, *35* (2), 343–357.

Krishnamurthy, Arvind, "The Bond/Old-Bond Spread," *Journal of Financial Economics*, 2002, *66* (2-3), 463–506.

Lillo, F. and J. D. Farmer, "The Long Memory of the Efficient Market," *Studies in Nonlinear Dynamics and Econometrics*, 2004, *8* (3).

Liu, Sheen and Chunchi Wu, "Repo Counterparty Risk and On-/Off-the-Run Treasury Spreads," *The Review of Asset Pricing Studies*, 2017, *7* (1), 81–143.

Logan, Lorie, "Treasury Market Liquidity and Early Lessons from the Pandemic Shock," in "Remarks at Brookings-Chicago Booth Task Force on Financial Stability Meeting (via videoconference), October," Vol. 23 2020.

Meldrum, Andrew C and Oleg Sokolinskiy, "The Relationship between Market Depth and Liquidity Fragility in the Treasury Market," Technical Report, Board of Governors of the Federal Reserve System (US) 2025.

Nelson, Charles R and Andrew F Siegel, "Parsimonious Modeling of Yield Curves," *Journal of business*, 1987, pp. 473–489.

Pasquariello, Paolo and Clara Vega, "The On-the-Run Liquidity Phenomenon," *Journal of Financial Economics*, 2009, *92* (1), 1–24.

Plerou, Vasiliki, Parameswaran Gopikrishnan, Xavier Gabaix, and H Eugene Stanley, "On the Origin of Power-Law Fluctuations in Stock Prices," *Quantitative Finance*, 2004, *4* (1), C11.

Tanggaard, Carsten, "Nonparametric Smoothing of Yield Curves," *Review of Quantitative Finance and Accounting*, 1997, *9*, 251–267.

Warga, Arthur, "Bond Returns, Liquidity, and Missing Data," *Journal of Financial and Quantitative Analysis*, 1992, *27* (4), 605–617.

Wood, Duncan, "Hunt for Toxic Flow Hits One of Banking's Old Problems," 2018.

Wood, Simon N, Fabian Scheipl, and Julian J Faraway, "Straightforward Intermediate Rank Tensor Product Smoothing in Mixed Models," *Statistics and Computing*, 2013, *23*, 341–360.

Zhang, Yi-Cheng, "Toward a Theory of Marginally Efficient Markets," *Physica A: Statistical Mechanics and its Applications*, 1999, *269* (1), 30–44.

Table 1: Biases in Indicative Bid-Ask Spreads

Panel A. Entire sample

	EBA	IBA	IMPRVT
5% quantile	0.03	0.18	-6.57
10% quantile	0.07	0.23	-1.65
50% quantile	0.37	0.68	0.50
90% quantile	2.79	2.72	0.90
95% quantile	6.22	4.16	0.95
mean	1.51	1.22	-1.30
sd	6.74	1.53	13.10
skew	36.93	3.53	-50.10
kurt	2188.18	21.40	4803.63

Panel B. Trades in less than $50mm notional

	EBA	IBA	IMPRVT
5% quantile	0.03	0.19	-2.60
10% quantile	0.06	0.24	-0.71
50% quantile	0.32	0.71	0.55
90% quantile	2.09	2.75	0.91
95% quantile	4.29	4.16	0.95
mean	1.16	1.24	-0.50
sd	5.90	1.52	10.05
skew	49.79	3.44	-56.67
kurt	3704.79	20.77	4939.37

Panel C. Trades in more than $50mm notional

	EBA	IBA	IMPRVT
5% quantile	0.05	0.18	-24.14
10% quantile	0.09	0.21	-8.80
50% quantile	0.62	0.62	0.20
90% quantile	6.27	2.62	0.85
95% quantile	12.41	4.15	0.93
mean	2.59	1.17	-3.76
sd	8.77	1.58	19.57
skew	21.53	3.79	-37.59
kurt	773.63	23.08	2879.57

Notes: This table reports the summary statistics for the effective bid-ask spreads, EBA defined in Eq. (3), NPQS indicative bid-ask spreads IBA, and the price improvement, **IMPRVT**, defined in Eq. (5). The sample consists of direct dealer-to-client trades from January 2018 to June 2024, filtered and aggregated as described in Section 3.1. Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

Table 2: Biases in Quoted Bid-Ask Spreads

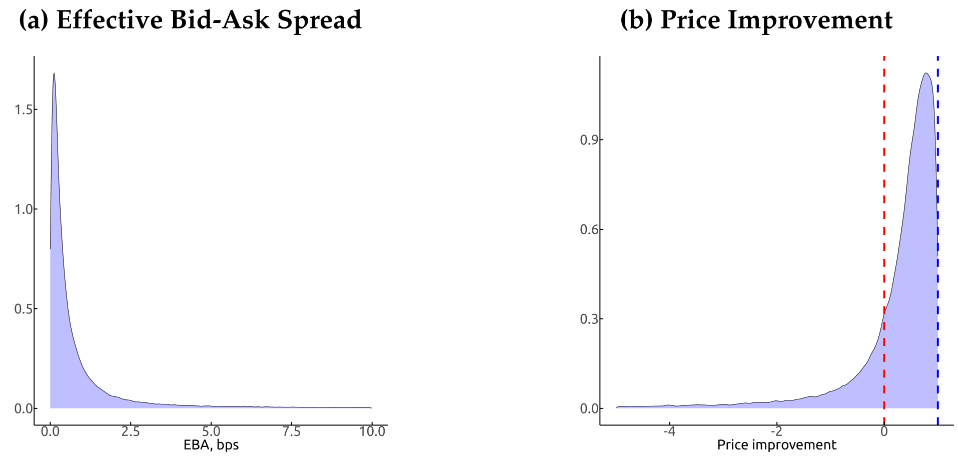
	Model I	**Model II**	**Model III**	**Model IV**
α_0	0.91^{***}	0.81^{***}	0.81^{***}	0.99^{***}
	(0.04)	(0.05)	(0.05)	(0.05)
IBA	0.48^{***}	0.73^{***}	0.73^{***}	-0.07
	(0.01)	(0.03)	(0.03)	(0.07)
$IBA \cdot RA$		-0.51^{***}	-0.51^{***}	-0.15^{*}
		(0.05)	(0.05)	(0.06)
$IBA \cdot CTD$			-0.04	-0.02
			(0.10)	(0.10)
$IBA \cdot \theta$				0.80^{***}
				(0.07)
BIC	677253.40	677180.44	677194.64	677073.21
Num. obs.	101,978			
Num. cusips	118			
Random effects variance	0.13	0.13	0.13	0.14

Notes: This table reports estimates of linear mixed models following the general specification in Eq. (5), reproduced here for convenience:

$$\mathbf{EBA}_i = \alpha_0 + \alpha_1 \cdot \mathbf{IBA}_i + \mathbf{X}_i\gamma + Z_i\beta + \epsilon_i,$$

where **EBA** is the effective bid-ask spreads for trade i; $\mathbf{IBA}_i$ is the indicative bid ask spread; $\mathbf{X}_i$ is a vector of control variables that identifies potential biases; $\{\alpha_0, \alpha_1, \gamma\}$ are fixed effect coefficients; Z_i is the i^{th} row of the $n \times s$ random-effects model matrix – a sparse indicator matrix that captures the grouping of n observations by s traded securities; β is the vector of random effects that have a multivariate normal distribution, $\mathcal{N}(\mathbf{0}, \Sigma)$; ϵ_i is the noise term. Control variables include the interactions of indicative bid-ask spreads with the following variables. RA is the relative age of the security, defined as the ratio of remaining time to maturity to the original time to maturity. CTD is an indicator of whether the security is cheapest-to-deliver into a front-month Treasury futures. θ a security-specific empirical cumulative probability of the bid-ask spread in the on-the-run Treasury market. The sample consists of direct dealer-to-client trades from January 2018 to June 2024, filtered and aggregated as described in Section 3.1. ***; **; * denote significance at the 0.1, 1 and 5 percent levels respectively. Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS; J.P. Morgan Chase & Co., MorganMarkets and DataQuery, https://markets.jpmorgan.com.

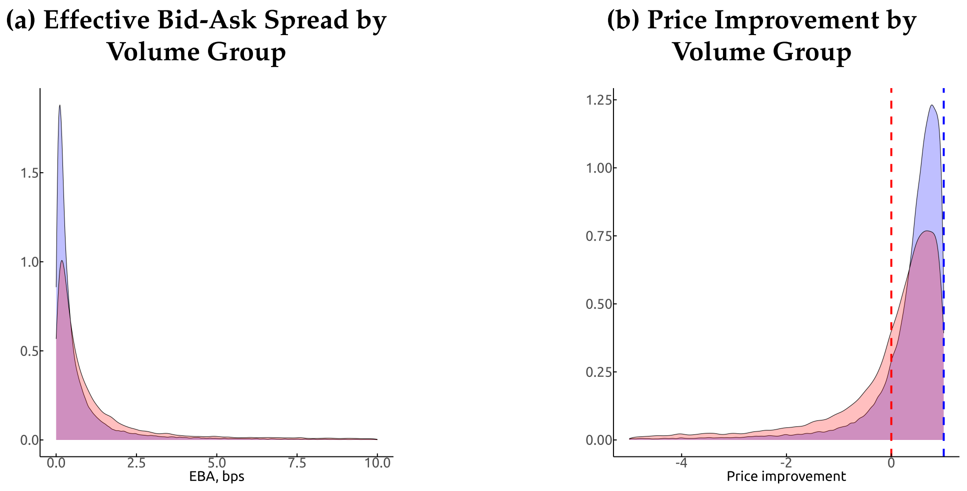
Figure 1: Distributions of Effective Liquidity and Execution Quality


(a) Effective Bid-Ask Spread

(b) Price Improvement

NOTE: the effective bid-ask spreads, EBA, is defined in Eq. (3) and the price improvement is defined in Eq. (4); the graph is truncated at 10 basis points. The sample consists of direct dealer-to-client trades from January 2018 to June 2024, filtered and aggregated as described in Section 3.1. Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

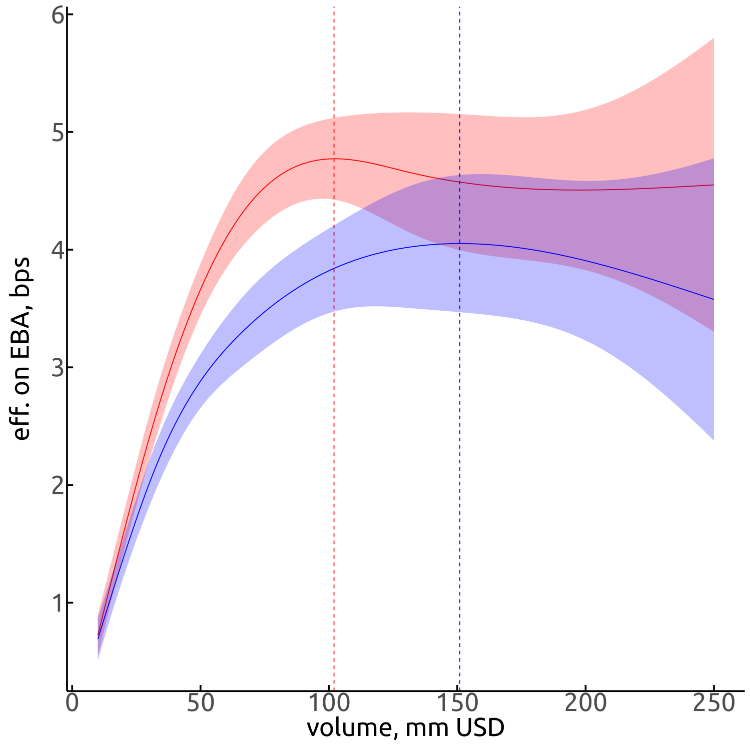
Figure 2: Distributions of Effective Liquidity and Execution Quality Conditional on Volume


(a) Effective Bid-Ask Spread by Volume Group

(b) Price Improvement by Volume Group

NOTE: the effective bid-ask spreads, EBA, is defined in Eq. (3) and the price improvement is defined in Eq. (4); the graph is truncated at 10 basis points. The distributions are conditional on transaction volume: trades of less than \$50mm – blue-shaded distribution, trades of \$50mm or more – red-shaded distribution. The sample consists of direct dealer-to-client trades from January 2018 to June 2024, filtered and aggregated as described in Section 3.1. Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

Figure 3: Effect of Volume on Effective Bid-Ask Spread Conditional on Benchmark, On-the-Run Liquidity



NOTE: The expected effects of volume on the effective bid-ask spread conditional on benchmark, on-the-run liquidity, as estimated based on specification in Eq. (7), reproduced here for convenience:

$$\mathbf{EBA}_i = \alpha_0 + \alpha_0 \cdot \mathbf{IBA}_i + \alpha_1 \cdot \mathbf{IBA}_i \times \mathbf{RA}_i + Z_i\beta + \left[\sum_j L_{ij}\mathcal{T}_j\left(\mathbf{V}_i, \theta_i\right)\right] + \epsilon_i,$$

where $\mathbf{EBA}_i$ and $\mathbf{IBA}_i$ are the effective and indicative bid-ask spreads for transaction i in security s, $\mathbf{IBA}_i \times \mathbf{RA}_i$ is the interaction term between the indicative bid-ask spread and security s relative age; $\{\alpha_0, \alpha_1, \gamma\}$ are fixed effect coefficients; Z_i is the i^{th} row of the $n \times s$ random-effects model matrix – a sparse indicator matrix that captures the grouping of n observations by s traded securities; β is the vector of random effects that have a multivariate normal distribution, $\mathcal{N}\left(\mathbf{0}, \Sigma\right)$; ϵ_i is the noise term; $\mathcal{T}_j\left(\mathbf{V}_i, \theta_i\right)$ are unknown smooth functions of volume, $\mathbf{V}_i$, and of the explanatory variable. θ_i is the security-specific empirical cumulative probability of the bid-ask spread, BBA, as the measure of benchmark, on-the-run Treasury market liquidity in a duration sector:

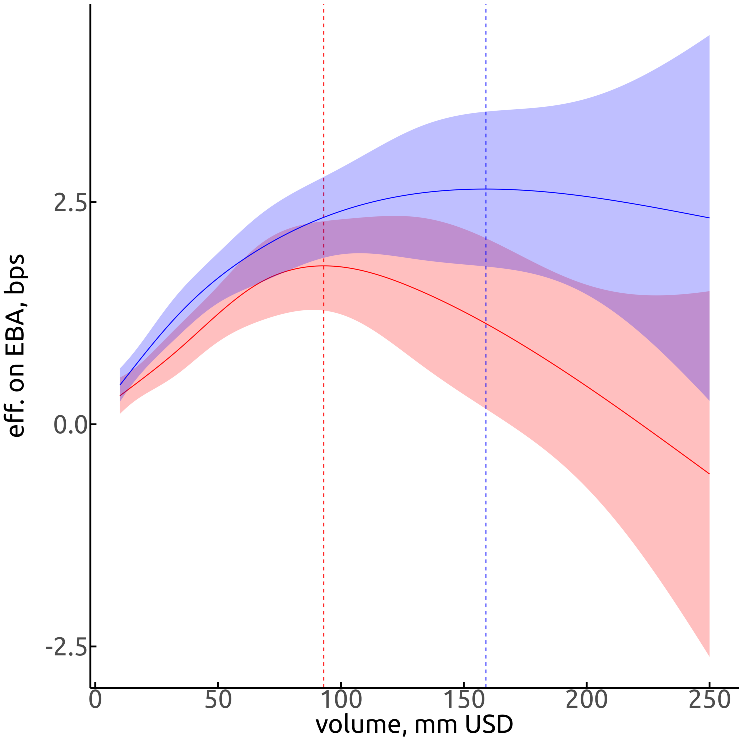
$$\theta_i = \mathbf{ecdf}_{\xi(s)}\left(\mathbf{BBA}_d^{\xi(s)}\right),$$

where i is the transaction index for a trade in security s, belonging to a duration group that maps to the on-the-run security $\xi\left(s\right)$, during trading day d; $\mathbf{BBA}_d^{\xi(s)}$ is the time-weighted average – with weights corresponding to lengths of time during which each spread level prevailed – bid-ask spread of the on-the-run benchmark, on-the-run Treasury security $\xi\left(s\right)$ during trading day d; $\mathbf{ecdf}_{\xi(s)}$ is the empirical cumulative distribution function of the bid-ask spread for on-the-run security $\xi\left(s\right)$. High values of θ correspond to higher bid-ask spreads for benchmark, on-the-run securities and, thus, lower benchmark liquidity.

The scaling law conditional on median benchmark, on-the-run Treasury liquidity conditions, $\theta = 0.5$, is in blue. The scaling law conditional on bad benchmark, on-the-run Treasury liquidity conditions, $\theta = 0.75$, is in red. The shaded regions correspond to the 95 percent confidence intervals. The dashed vertical lines mark the local maxima of the expected effects of volume on the effective bid-ask spread.

Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

Figure 4: Effect of Volume on Effective Bid-Ask Spread Conditional on the Security's Relative Liquidity



NOTE: The expected effects of volume on the effective bid-ask spread conditional on the security's relative liquidity, as estimated based on specification 7, reproduced here for convenience:

$$\mathbf{EBA}_i = \alpha_0 + \alpha_0 \cdot \mathbf{IBA}_i + \alpha_1 \cdot \mathbf{IBA}_i \times \mathbf{RA}_i + Z_i\beta + \left[\sum_j L_{ij}\mathcal{T}_j\left(V_i, \theta_i\right)\right] + \epsilon_i,$$

where $\mathbf{EBA}_i$ and $\mathbf{IBA}_i$ are the effective and indicative bid-ask spreads for transaction i in security s, $\mathbf{IBA}_i \times \mathbf{RA}_i$ is the interaction term between the indicative bid-ask spread and security s relative age; $\{\alpha_0, \alpha_1, \gamma\}$ are fixed effect coefficients; Z_i is the i^{th} row of the $n \times s$ random-effects model matrix – a sparse indicator matrix that captures the grouping of n observations by s traded securities; β is the vector of random effects that have a multivariate normal distribution, $\mathcal{N}\left(\mathbf{0}, \Sigma\right)$; ϵ_i is the noise term; $\mathcal{T}_j\left(V_i, \theta_i\right)$ are unknown smooth functions of volume, V_i, and of the explanatory variable. θ_i is the security- and trading day-specific empirical cumulative probability of indicative bid-ask spreads:

$$\theta_i = \mathbf{ecdf}_{\psi,d}\left(\mathbf{IBA}_d^s\right),$$

where i is the transaction index for a trade in security s, belonging to duration group ψ, during trading day d; $\mathbf{IBA}_d^s$ is the simple average indicative bid-ask spread of security s during trading day d, estimated by an average of the corresponding bid-ask spreads over NPQS quote snapshots on day d; $\mathbf{ecdf}_{\psi,d}$ is the empirical cumulative distribution function of average indicative bid-ask spreads on day d for the duration group ψ. High values of θ correspond to comparatively less liquid securities within their duration groups for the particular trading day.
The scaling law for a median liquidity security, $\theta = 0.5$, is in blue. The scaling law for a relatively illiquid security, $\theta = 0.75$, is in red. The shaded regions correspond to the 95 percent confidence intervals. The dashed vertical lines mark the local maxima of the expected effects of volume on the effective bid-ask spread.

Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

Appendices

A Treasury Market Microstructure – On- v. Off-the-Run

With existent ample research into on-the-run Treasury market liquidity, it is important to detail the causes making on-the-run and off-the-run Treasury market liquidity notably different, and, thus, requiring a separate study of the off-the-run Treasury market.

Most trades in on-the-run Treasuries occur on alternative trading systems (ATS) that offer the central limit order book (CLOB) protocol for matching buy and sell orders. The CLOB contains time-prioritized queues of buy and sell orders at various price levels that may be observed by all market participants. Thus, the CLOB provides a good indication of immediately available liquidity.[23]

In contrast, off-the-run Treasury market liquidity is more opaque because trading occurs bilaterally, even when intermediated by an ATS. A client willing to trade can ask several market makers for quotes for a specific transaction volume and receive non-firm, indicative quotes that could form the basis of further negotiations. Alternatively, some market makers stream quotes to clients via their in-house platforms or external ATS allowing potential clients to hide their intention to trade. Generally available streamed quotes are, however, even less reliable indicators of where actual trades can occur.[24]

Complete anonymity of participants in CLOB-based markets is another key difference between trading benchmark, on-the-run Treasuries via an inter-dealer broker ATS and trading off-the-run Treasuries in the dealer-to-client Treasury market segment. Lack of anonymity in direct bilateral trading allows market makers to discriminate against 'toxic flow' – trades by well-informed clients with a track record of either predicting price changes or executing in a manner that would move the price against the market maker (see Wood, 2018).

In summary, on-the-run and off-the-run Treasury markets have vastly different mi-

[23] While immediately available liquidity at the best prices is relatively modest and orders trading through multiple levels of the book are rare under normal market conditions, market participants may observe the CLOB's response to trades to infer the willingness of market makers to replenish the CLOB.

[24] This is natural given that market makers would prefer to discriminate among clients based on their perceived ability to forecast future prices.

crostructure that makes liquidity of even similar securities notably different.

Besides the economic importance of the off-the-run Treasury market, one more argument that recommends research into its liquidity is the opportunity to study a critical decentralized market where bilateral relationships and search costs matter. While on-the-run Treasuries often serve as barometers of liquidity, there is no simple relationship between liquidity of on-the-run and off-the-run securities. Particularly during periods of market stress, when there is a flight to quality, off-the-run Treasuries' liquidity tends to suffer more than on-the-run Treasuries' liquidity (see Furfine and Remolona, 2002).[25]

B Trades Intermediated by Alternative Trading Systems

Market microstructure in the Alternative Trading System (ATS) intermediated sector differs from that of the direct dealer-to-client sector. Crucially, ATS may offer anonymity to their participants, thereby removing the possibility of price discrimination by market makers based on a client's perceived sophistication. That said, curated markets that seek to exclude 'toxic' flows from most informed participants have been a popular innovation. In addition, anonymity removes the consideration of a client-dealer relationship. Finally, ATS may offer a variety of trading protocols that suit different groups of market participants.[26]

In this section, I show how the effective-bid ask spread depends on transaction volume in the ATS-intermediated sector of the dealer-to-client market. First, I consider the model specification in Eq. (7) with θ of Eq. (6) to allow for the effect of benchmark, on-the-run Treasury market liquidity. Panel (A) of Figure A-1 depicts the impact of transaction volume on the expected effective bid-ask spread for client trades intermediated by ATS, conditional on median and poor benchmark, on-the-run Treasury market liquidity. Relative to the direct dealer-to-client trades, the dependence of the effective-bid ask spread on transaction volume has a more complex shape for the ATS-intermediated

[25] The spreads between off-the-run and on-the-run Treasury yields compensate for liquidity differential between the two types of securities. These spreads tend to widen notably during periods of market stress (see Warga, 1992; Furfine and Remolona, 2002).

[26] A single ATS may offer multiple trading protocols. Unfortunately, Treasury TRACE data do not contain a variable indicating the protocol. I leave less direct ways of identifying the trading protocol for future research.

trades – the mean effects function has both convex and concave regions. Similar to the direct dealer-to-client market, I obtain evidence for selective liquidity taking – a region where the expected effective bid-ask spread decreases with incremental volume. However, in the ATS-intermediated market, for even greater transaction volumes, selective liquidity-taking loses its dominance. This may be an indication that orders where clients do not have time discretion are more likely to be executed via an ATS. As in the case of the direct dealer-to-client market, the effect of transaction volume on the effective bid-ask spread is greater when overall Treasury market liquidity is low.

Second, I consider the model specification in Eq. (7) with θ of Eq. (8) to allow for systemic differences between securities of varying liquidity. Panel (B) of Figure A-1 depicts the impact of transaction volume on the expected effective bid-ask spread for client trades intermediated by ATS, for securities of median and low relative liquidity levels. Relative to the direct dealer-to-client trades, the differences in the scaling laws for securities of varying relative liquidity are much more modest. One notable attribute of the scaling law for ATS-intermediated trades is a greater region of convexity, especially for lower relative liquidity securities. Once again, this is evidence that selective liquidity-taking is more prevalent in the direct dealer-to-client segment.

C Liquidity in the Dealer-to-Dealer Market

Dealer-to-dealer off-the-run Treasury market is also a decentralized, bilateral market – just like the dealer-to-client market. However, there are important differences. First, trading norms differ with respect to typical counterparty search patterns; in particular, explicit mini-auctions may be less likely. Second, there is, arguably, less information heterogeneity among leading dealers relative to that among clients.

While the microstructure differences between the dealer-to-dealer and dealer-to-client markets are important, the two markets are closely tied. First, when a market maker does not have the security requested by a client in its inventory, it can act as a (i) broker – sourcing the security in the inter-dealer market on an agency basis, or (ii) dealer – selling the security that it does not possess with the intention of obtaining it in the inter-dealer market later that business day – all done on a principal basis.[27] In summary, there are apparent differences in the relationships between the participants in the dealer-to-client and dealer-to-dealer markets, but also close links between the two exist.

Panels A and B of Figure A-2 depict the probability density plots of the effective bid-ask spreads and price improvements, respectively, for transactions in the dealer-to-client and dealer-to-dealer segments. The close links between the dealer-to-client and dealer-to-dealer segments make the distributions of effective bid-ask spreads and price improvements near identical in the two segments.

D Retail Trades

Retail investors are active in the Treasury market, likely constituting the majority of the trade count, but not volume. Systemic differences in trading costs of retail and institutional investors is due to information and relationship effects. First, retail clients are commonly seen as less ‘toxic’, that is, less informed.[28] Second, there is the offsetting factor in the form of less value assigned to each retail client-dealer relationship. Third, dealers may see their retail clients as less likely to shop around for better execution, as retail investors are unlikely to be able to accurately assess execution quality.

While retail trades are not explicitly identified in Treasury TRACE data, transactions of less than $1mm in notional may only be reasonably attributed to retail investors.[29] When such small notional transactions occur between large dealers, it must mean that one dealer sold Treasuries to a retail investor that the dealer did not have at the moment of receiving the client’s order and had to immediately source the securities from another dealer.

Panels A and B of Figure A-3 depict the probability density plots of the effective bid-

[27] Thus, a dealer is effectively a broker-dealer, but the common practice is to refer to the market segment as ‘dealer-to-dealer’ – and not ‘broker-dealer-to-broker-dealer’ – for brevity. A market participant can sell a security that it does not possess at the moment the transaction is entered into, with the intention of sourcing it later, because Treasury markets settle on the next business day. The security can also be borrowed in the specials repo market where a specific security is used as collateral. Finally, if a dealer cannot source the security in time, a fail occurs and corresponding penalties are levied.

[28] Principal trading firms in the equity market are entering in special arrangements with brokerages to gain access to such less-toxic trade flows.

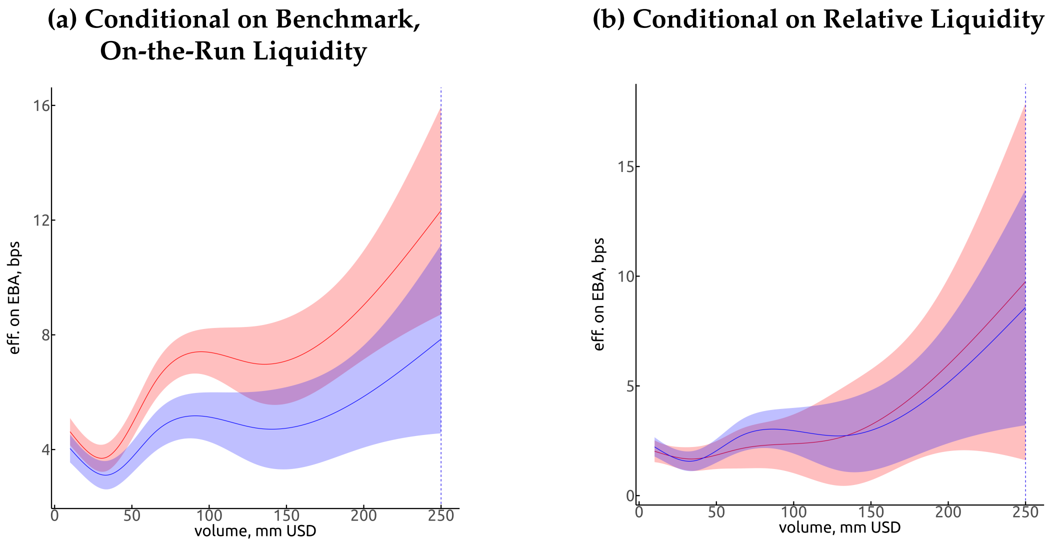
[29] Retail investors may also trade in larger notional amounts, but this would likely constitute a negligible fraction of all trades.

ask spreads and price improvements, respectively, for transactions of retail vis-à-vis institutional investors. There is clear evidence that retail investors get worse execution – they pay higher effective bid-ask spreads and do not experience the same degree of price improvement as institutional investors.[30] Thus, retail investors do not benefit from their credible non-toxicity, while their inability to assess execution quality and the lack of sufficient client-dealer relationship value result in wider effective bid-ask spreads.

[30] Since retail trades are all in small volumes, analysis of the dependence of the effective bid-ask spread on transaction volume is irrelevant for this market segment. Accounting for the effect of volume on the effective bid-ask spread of institutional investors' trades would make the results even more pronounced.

Appendix Figures

Figure A-1: Conditional Effect of Volume on Effective Bid-Ask Spread for Client Trades Intermediated by an ATS


(a) Conditional on Benchmark, On-the-Run Liquidity

(b) Conditional on Relative Liquidity

NOTE: The expected effects of volume on the effective bid-ask spread, as estimated based on specification in Eq. (7), reproduced here for convenience:

$$\mathbf{EBA}_i = \alpha_0 + \alpha_0 \cdot \mathbf{IBA}_i + \alpha_1 \cdot \mathbf{IBA}_i \times \mathbf{RA}_i + Z_i \beta + \left[\sum_j L_{ij} \mathcal{T}_j \left(\mathbf{V}_i, \theta_i \right) \right] + \epsilon_i,$$

where $\mathbf{EBA}_i$ and $\mathbf{IBA}_i$ are the effective and indicative bid-ask spreads for transaction i in security s, $\mathbf{IBA}_i \times \mathbf{RA}_i$ is the interaction term between the indicative bid-ask spread and security s relative age; $\{\alpha_0, \alpha_1, \gamma\}$ are fixed effect coefficients; Z_i is the i^{th} row of the $n \times s$ random-effects model matrix – a sparse indicator matrix that captures the grouping of n observations by s traded securities; β is the vector of random effects that have a multivariate normal distribution, $\mathcal{N}\left(\mathbf{0}, \Sigma\right)$; ϵ_i is the noise term; $\mathcal{T}_j\left(\mathbf{V}_i, \theta_i\right)$ are unknown smooth functions of volume, $\mathbf{V}_i$, and of the explanatory variable.

In Panel (A), I condition the dependence between the effective bid-ask spread and order volume on the benchmark, on-the-run Treasury market liquidity in a duration sector, with θ_i – the security-specific empirical cumulative probability of the bid-ask spread:

$$\theta_i = \mathbf{ecdf}_{\xi(s)} \left(\mathbf{BBA}_d^{\xi(s)} \right),$$

where i is the transaction index for a trade in security s, belonging to a duration group that maps to the on-the-run security $\xi(s)$, during trading day d; $\mathbf{BBA}_d^{\xi(s)}$ is the time-weighted average – with weights corresponding to lengths of time during which each spread level prevailed – bid-ask spread of the on-the-run benchmark Treasury security $\xi(s)$ during trading day d; $\mathbf{ecdf}_{\xi(s)}$ is the empirical cumulative distribution function of the bid-ask spread for on-the-run security $\xi(s)$. High values of θ correspond to higher bid-ask spreads for benchmark, on-the-run securities and, thus, lower benchmark liquidity.

In Panel (B), I condition the dependence between the effective bid-ask spread and order volume on the relative liquidity of a security within its duration group on the corresponding day, with θ_i – the security- and trading day-specific empirical cumulative probability of indicative bid-ask spreads:

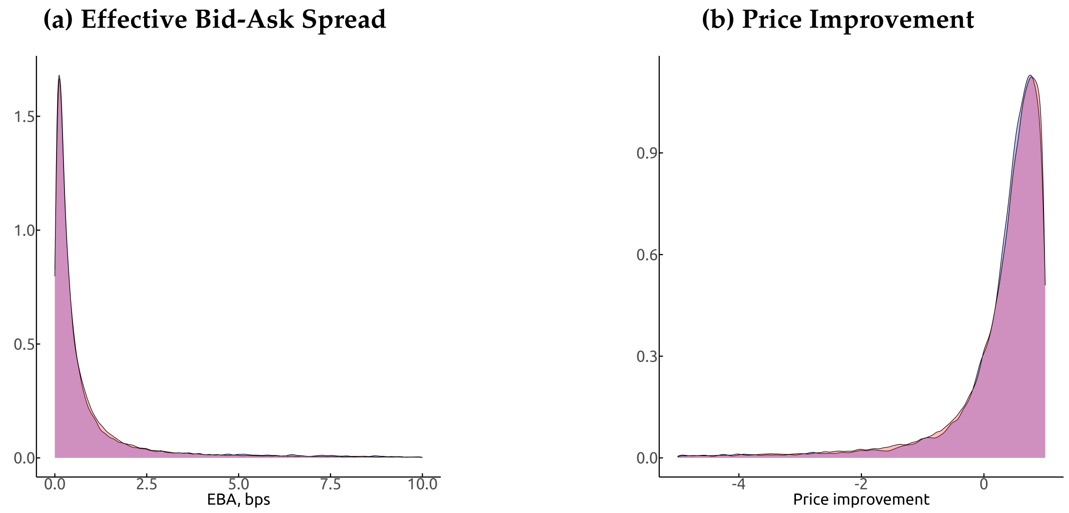
$$\theta_i = \mathbf{ecdf}_{\psi,d} \left(\mathbf{IBA}_d^s \right),$$

where i is the transaction index for a trade in security s, belonging to duration group ψ, during trading day d; $\mathbf{IBA}_d^s$ is the simple average indicative bid-ask spread of security s during trading day d, estimated by an average of the corresponding bid-ask spreads over NPQS quote snapshots on day d; $\mathbf{ecdf}_{\psi,d}$ is the empirical cumulative distribution function of average indicative bid-ask spreads on day d for the duration group ψ. High values of θ correspond to comparatively less liquid securities within their duration groups for the particular trading day.

Panel (A) depicts the scaling law conditional on median benchmark, on-the-run Treasury liquidity conditions, $\theta = 0.5$, in blue, and the scaling law conditional on bad benchmark, on-the-run Treasury liquidity conditions, $\theta = 0.75$, in red. Panel (B) depicts the scaling law for a median liquidity security, $\theta = 0.5$, in blue, and the scaling law for a relatively illiquid security, $\theta = 0.75$, in red. The shaded regions correspond to the 95 percent confidence intervals. The dashed vertical lines mark the local maxima of the expected effects of volume on the effective bid-ask spread.

Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

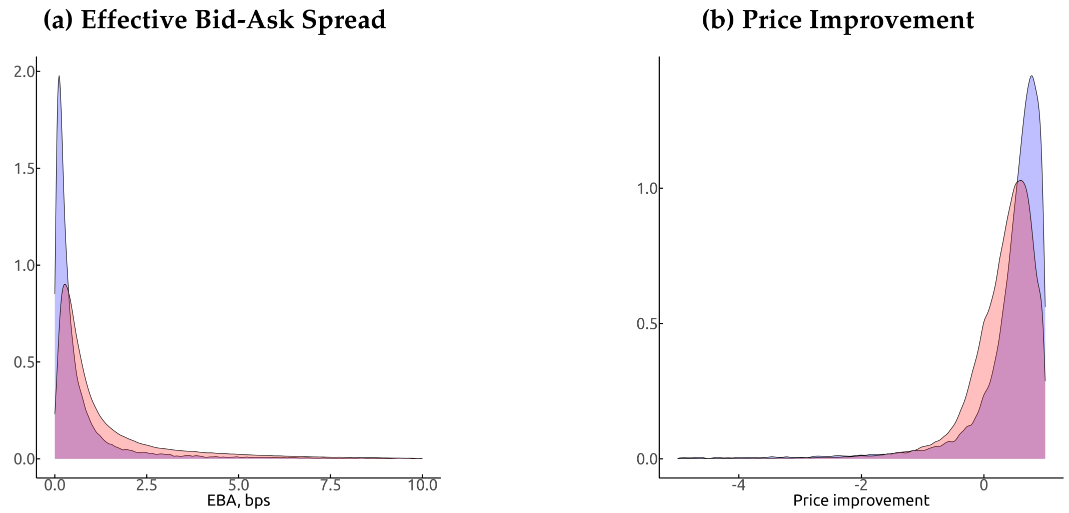
Figure A-2: Dealer-to-Dealer Trades: Distributions of Effective Liquidity and Execution Quality


(a) Effective Bid-Ask Spread

(b) Price Improvement

NOTE: the effective bid-ask spreads, EBA, is defined in Eq. (3) and the price improvement is defined in Eq. (4). The sample consists of direct dealer-to-dealer trades and dealer-to-client trades from January 2018 to June 2024, filtered and aggregated as described in Section 3.1. The empirical probability densities of metrics for dealer-to-dealer trades are in blue, while the empirical probability densities of metrics for dealer-to-client trades are in red. The empirical probability densities for dealer-to-dealer trades and dealer-to-client trades are nearly identical. Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.

Figure A-3: Retail Trades: Distributions of Effective Liquidity and Execution Quality


(a) Effective Bid-Ask Spread

(b) Price Improvement

NOTE: the effective bid-ask spreads, EBA, is defined in Eq. (3) and the price improvement is defined in Eq. (4). The sample consists of retail dealer-to-client trades, identified as trades of less than \$10mm notional, and direct institutional dealer-to-client trades, identified as trades of \$10mm notional or greater, from January 2018 to June 2024, filtered and aggregated as described in Section 3.1. The empirical probability densities of metrics for institutional trades are in blue, while the empirical probability densities of metrics for retail trades are in red. Data sources: Financial Industry Regulatory Authority (FINRA), Bond Trade Dissemination System (BTDS) and Trade Reporting and Compliance Engine (TRACE); information obtained from the repo Inter Dealer Broker community; Federal Reserve Bank of New York, NPQS.