Dealer Capacity and US Treasury Market Functionality*

Darrell Duffie[a], Michael Fleming[b], Frank Keane[b], Claire Nelson[c], Or Shachar[b], and Peter Van Tassel[d]

[a]Stanford University
[b]Federal Reserve Bank of New York
[c]Princeton University
[d]Independent

October 2023

Abstract

We show a significant loss in US Treasury market functionality when intensive use of dealer balance sheets is needed to intermediate bond markets, as in March 2020. Although yield volatility explains most of the variation in Treasury market liquidity over time, when dealer balance sheet utilization reaches sufficiently high levels, liquidity is much worse than predicted by yield volatility alone. This is consistent with the existence of occasionally binding constraints on the intermediation capacity of bond markets.

Keywords: Treasury market, liquidity, volatility, dealer intermediation, value-at-risk.
JEL codes: G01, G1, G12, G18, E58

*The views expressed here are the authors' and are not necessarily the views of the Federal Reserve Bank of New York or the Federal Reserve System. The authors are grateful for research assistance from Isabel Krogh and for the comments of Daron Acemoglu, Wenxin Du, Terrence Hendershott, Edith Hotchkiss, Tiff Macklem, Andrew Metrick, Zach Modig, Sungjin Park, Imène Rahmouni-Rousseau, Jeremy Stein, Josh Younger, and the participants of Women in Market Microstructure 2023 conference and BIS Annual Meeting Research Conference. We are also grateful to Damir Filipović, Markus Pelger, and Ye Ye for help with accessing their results on fitted Treasury discounts. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2039656. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Emails: duffie@stanford.edu; michael.fleming@ny.frb.org; frank.keane@ny.frb.org; cn7429@princeton.edu; or.shachar@ny.frb.org; peter.l.vantassel@gmail.com.

1 Introduction

We show that there is a significant loss in US Treasury market functionality when intensive use of dealer balance sheets is needed to intermediate bond markets, as in March 2020. Although yield volatility explains most of the variation in Treasury market liquidity over time, when dealer balance sheet utilization reaches sufficiently high levels, liquidity is much worse than predicted by yield volatility alone. This is consistent with the existence of occasionally binding constraints on the intermediation capacity of bond markets.

The US Treasury market is broadly viewed as the world's deepest and most liquid security market. However, nearly all investor purchases and sales of Treasury securities are handled by dealers, implying that limits on the marginal ability or willingness of dealers to allocate space on their balance sheets to intermediating this market could lead to a loss of market functionality. By many accounts, this occurred in March 2020, spurring the Federal Reserve to initiate asset purchases to support market functioning.[1] This episode also led to extensive and ongoing policy discussions of how Treasury market resilience might be improved and to various policy initiatives to promote a more resilient market.[2]

We estimate a battery of illiquidity metrics, some novel, for the dealer-to-customer (D2C) and interdealer segments of the market for US Treasury securities. We show that volatility is always a key driver of illiquidity. But we also show that as dealer balance sheet utilization reaches moderately high levels, balance sheet utilization metrics begin to add significantly to explaining illiquidity. This is consistent with the often suggested concept that bond market illiquidity is caused in part by limited dealer intermediation capacity, especially relative to the now-much-larger US Treasury market.[3]

The illiquidity metrics that we incorporate from the interdealer market for on-the-run

[1]For example, see Duffie (2020); Goldberg (2020); Brainard (2021) and He, Nagel, and Song (2022).

[2]See Group of Thirty (2021, 2022) and US Department of the Treasury, Board of Governors of the Federal Reserve System, Federal Reserve Bank of New York, US Securities and Exchange Commission, and US Commodity Futures Trading Commission (2021, 2022).

[3]See Comerton-Forde et al. (2010); Adrian et al. (2013); Bessembinder et al. (2018); Duffie (2020); Goldberg (2020); Breckenfelder and Ivashina (2021); Fontaine et al. (2021) and He, Nagel, and Song (2022).

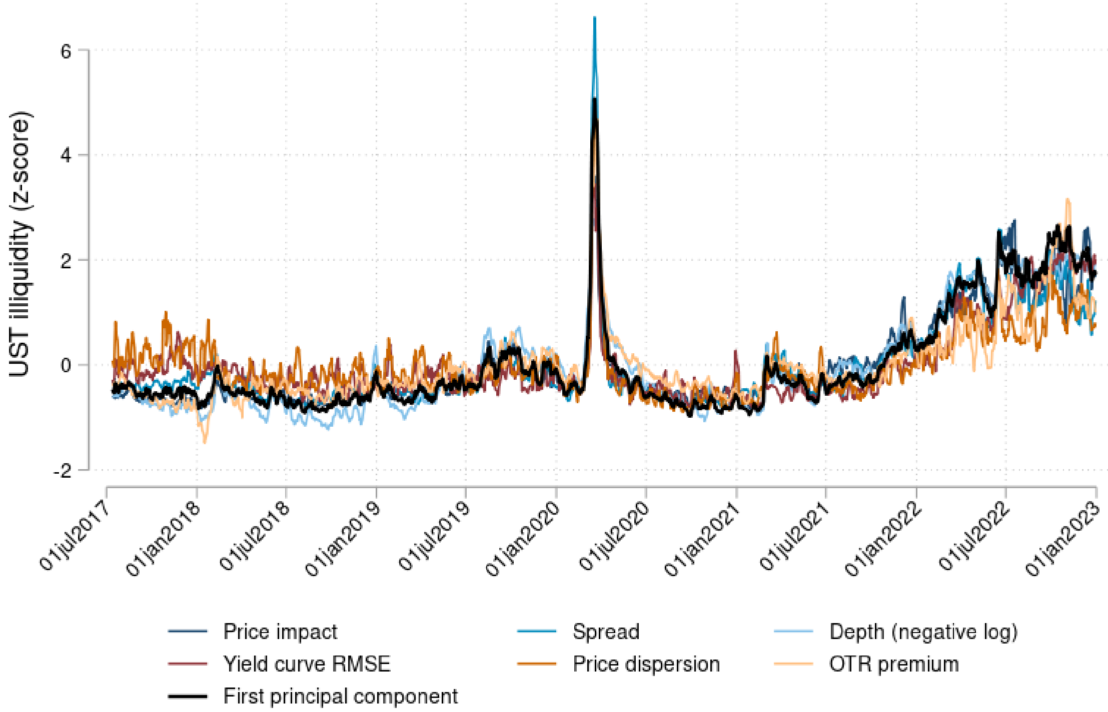


Figure 1. US Treasury market illiquidity measures. Average z-scores across the 2-, 5-, and 10-year maturity sectors of price impact, bid-ask spread, and depth from the interdealer market and on-the-run premium, price dispersion, and yield curve root-mean-squared-error (RMSE) from the dealer-to-customer market. The first principal component of the 18 z-scores is plotted in bold. All variables are shown, for clarity, in the form of five-day moving averages. The sample period is July 10, 2017 to December 31, 2022 excluding holidays and early closes (T=1,336 trading days).

Treasury securities are price impact, depth, and bid-ask spreads. As additional measures from the D2C market, we examine the yield spread between off-the-run and on-the-run Treasuries, the within-security dispersion of off-the-run transaction prices, and a new implementation of the yield-curve root-mean-squared-error (RMSE) measure of Hu, Pan, and Wang (2013), also known as the "noise" measure. Our interdealer data are from BrokerTec and our D2C data are from TRACE. Figure 1 shows these six illiquidity metrics for our main July 2017 - December 2022 sample period. Each of the six time series is the average of the corresponding z-scores across the 2-, 5-, and 10-year maturity sectors. Also shown is the first principal component of the resulting 18 measures of illiquidity.

We construct measures of fixed-income dealer balance sheet capacity utilization using dealer-level weekly FR 2004 reports of primary dealers' net and gross positions in US Treasury securities, agency mortgage-backed securities (MBS), and corporate bonds. Additional measures of intermediation intensity are based on dealer-level purchases from customers over the past three trading days, obtained from TRACE, for Treasuries, agency MBS, and corporate bonds. These D2C flow data offer a different perspective on dealer balance sheet utilization. Firstly, they are observed at a higher frequency. Secondly, freshly arriving flows of assets from customers place a different sort of pressure on dealer balance sheets, because these flows are temporarily warehoused on dealer balance sheets in anticipation of laying them off in the interdealer market or with other customers. All of our position and trade-flow measures are adjusted for risk based on the price-to-yield sensitivity (DV01, or dollar value of a basis point) of each security and on the market-implied yield volatilities of one-month options on maturity-matched interest-rate swaps.

We also estimate dealer capacity utilization based on value-at-risk (VaR) measures of the fixed-income divisions of the largest bank holding companies (BHCs). We include VaR data reported by the dealer divisions of reporting BHCs and we estimate VaRs from quantile regressions of daily dealer profits and losses (PnLs). Relative to our position and trade-flow data, VaR estimates have the advantage that they reflect dealer positions more comprehensively, including the impact of derivatives hedges. We also incorporate estimates of capacity utilization based on the overall holding-company VaR measures, both as reported and as estimated by us from daily holding-company PnLs.

The actual capacity of a dealer to offer liquidity to bond markets is difficult to measure directly because the amount of liquidity provided is limited by the dealer's willingness to take economic risk, by the agency-based concerns of its desk managers and traders, and by the complex processes by which regulatory and economic capital are internally allocated to a firm's divisions and trading desks. We take a revealed-preference approach. On any day t, the degree to which a given dealer i is utilizing its capacity for position levels or trade flows is proxied by the ratio $U_{it} = X_{it}/X_i^*$ of the current level X_{it} of that intermediation measure to its corresponding maximum $X_i^* = \max\{X_{i1}, \ldots, X_{iT}\}$ over the T days in the sample. For the set of observed dealers, the overall level of capacity utilization is estimated as the capacity-weighted average of the capacity utilizations of the individual dealers, which is $\sum_i w_i U_{it}$, where

$$w_i = \frac{X_i^*}{\sum_i X_i^*}.$$

Even in a stationary setting, with a finite sample, X_i^* is a downward biased estimator of the maximum position, trade flow, or VaR that the dealer might achieve with an unlimited number of observations. Moreover, the actual data are not generated from a stationary distribution because of structural changes over time, such as changes in market structure and in the capitalization of dealers.

Figure 2 shows estimates of dealers' capacity utilization based on position and flow data. The corresponding estimates based on VaR are found in Section 8. Supporting the concepts of this paper and the presumption in prior commentaries that we cite, dealer capacity utilization peaks sharply in March 2020, just when the Treasury illiquidity metrics shown in Figure 1 also peak. For example, on March 12, 2020, the World Health Organization declared COVID-19 to be a global pandemic. On that day, the first principal component of Treasury illiquidity reached 5.4 standard deviations above its mean, while dealer balance-sheet utilization, as measured by risk-adjusted gross fixed-income positions, reached its sample record high of 96%.[4]

Figure 3 reveals that most of the variation in Treasury market illiquidity is explained by yield volatility. Figure 4 shows a nearly linear contemporaneous relationship. A high correlation between illiquidity and interest-rate volatility is expected: higher volatility is likely to cause both an increase in the demand for intermediation services at any given trading cost, and a reduction in the supply of intermediation services at any given level of dealer compensation, thus driving up the equilibrium cost of intermediation. In theory,

[4]See Appendix Table C.8, which shows the cross-sectional variation of dealer capacity utilization measures, on peak days and across the entire sample.

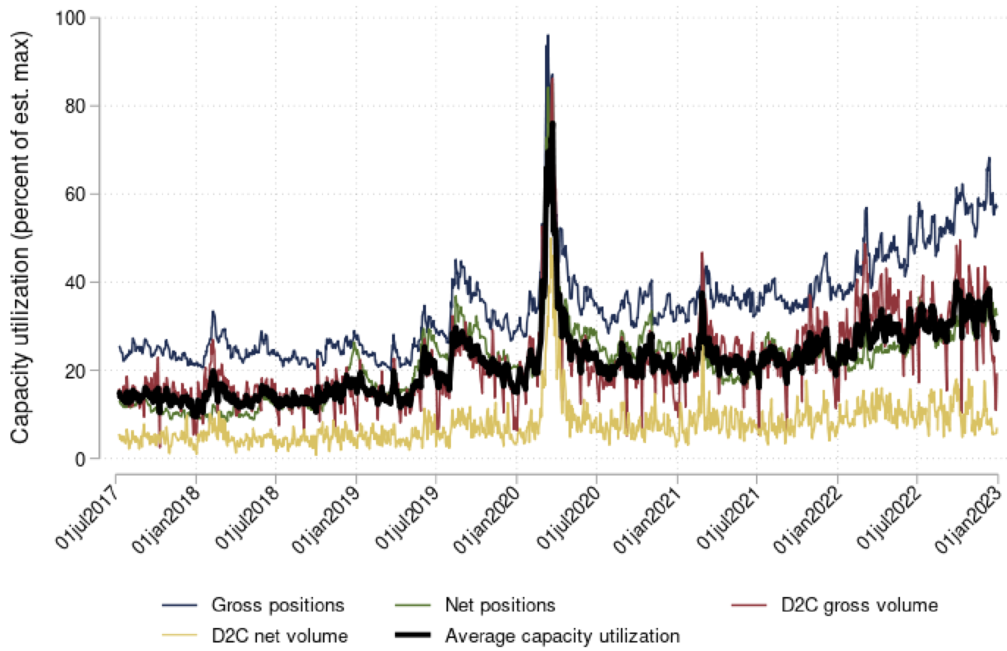


Figure 2. Dealer capacity utilization measures. Dealer capacity utilization measures based on risk-adjusted gross positions, net positions, gross dealer-to-customer (D2C) volume, and net D2C volume. Average capacity utilization is the simple average of the four measures.

the resulting equilibrium quantity of intermediation, for example the volume of trade, could either rise or fall with interest-rate volatility.

With respect to the supply of intermediation, dealers make general risk-return tradeoffs when they provide liquidity to the market. Dealers also consider the association between yield volatility and the cost of adverse selection associated with asymmetric counterparty information related to changes in yields. As for the demand for intermediation services, both dealers and non-dealer investors are likely to demand more liquidity provision from dealers when volatility rises because this increases risk-reduction incentives for some investors more than others. Further, higher volatility is associated with higher macro-monetary uncertainty, which also generates higher demands to trade. The combined effects of a higher demand for liquidity and reduced dealer incentives to supply liquidity at any given level of dealer compensation imply a positive relationship between volatility and illiquidity, as depicted in Figure 3.

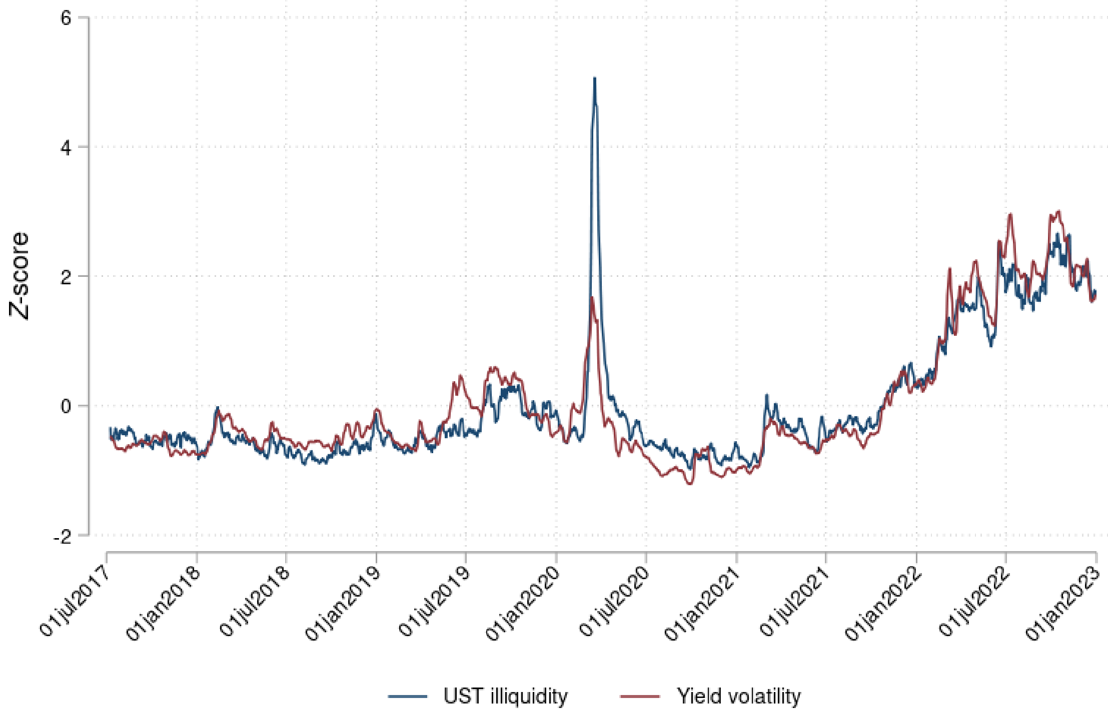


Figure 3. **US Treasury market illiquidity and yield volatility.** Treasury illiquidity is the first principal component of six illiquidity measures across 2-, 5-, and 10-year maturity sectors as shown in Figure 1. Yield volatility is the average volatility implied by swaptions with one-month expirations on 2-, 5-, and 10-year interest-rate swaps. The series are plotted as five-day moving averages of z-scores. The sample correlation of the underlying two daily time-series is 89%.

Figure 4 also reveals clear outliers to this relationship during March 2020. These outliers are partially explained by the peak levels of dealer capacity utilization that were reached at that time. Indeed, Figure 5 shows that, after taking out the component of illiquidity that is predicted by volatility, the remaining component of illiquidity is related to dealer balance-sheet utilization, but only when illiquidity and balance-sheet utilization are both at high levels. This is to be expected from theory, as motivated in Appendix A. Marginal changes in dealer balance sheet utilization should not affect market-making behavior much when dealers have plenty of "space" left to accommodate more positions and trade flows. But, as their balance sheets become intensively loaded, dealers become more guarded about providing liquidity to the market. In addition to this reduction in the supply of intermediation services when capacity utilization gets sufficiently high, there should also be an increased demand by

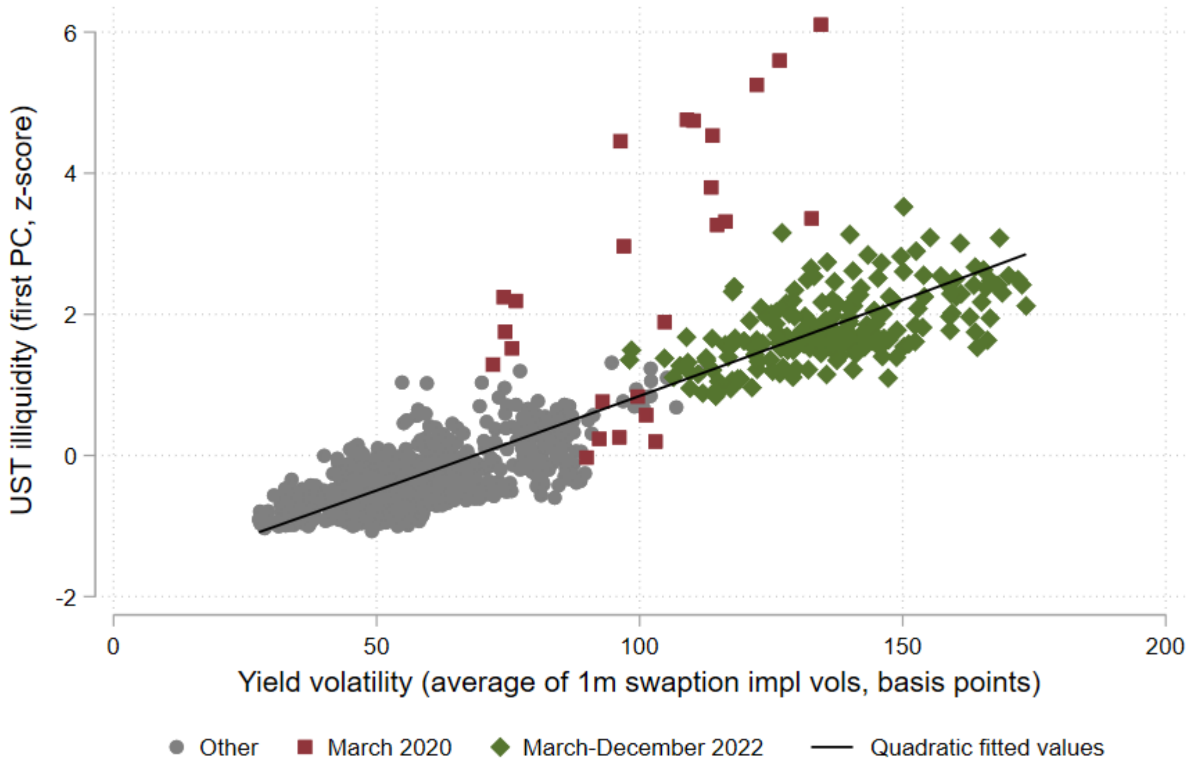


Figure 4. Relationship between US Treasury market illiquidity and yield volatility. A scatter plot and estimated relationship between the principal-component composite measure of Treasury illiquidity and a composite measure of implied volatility, as measured by the average of the standard deviations of benchmark swap rates, in basis points, implied by swaptions on 2-, 5-, and 10-year swaps with one-month expirations. The plotted ordinary-least-squares fit, for July 10, 2017 to December 31, 2022 ($T = 1,336$), is the second-order polynomial $y = -1.81 + 0.026x + 0.000005x^2$, where volatility x is in basis points, $R^2 = 79.5\%$. The constant and linear coefficient estimates have p-values of less than 1% under standard assumptions.

dealers for intermediation services from *other* dealers. The demand by all investors for bond intermediation services could also be accelerated by perceptions that dealer intermediation capacity is increasingly likely to be exhausted.

From Figure 5, when the estimated capacity utilization of dealers is around 20%, there is little estimated marginal impact of increases in capacity utilization on Treasury market illiquidity. However, when dealer capacity utilization rises from 40% to 80%, Treasury illiquidity is estimated to increase by roughly three standard deviations beyond the level of illiquidity predicted by volatility. Although this relationship is noisy, it has high statistical and economic significance. The scatter plot reveals a striking nonlinear relationship between

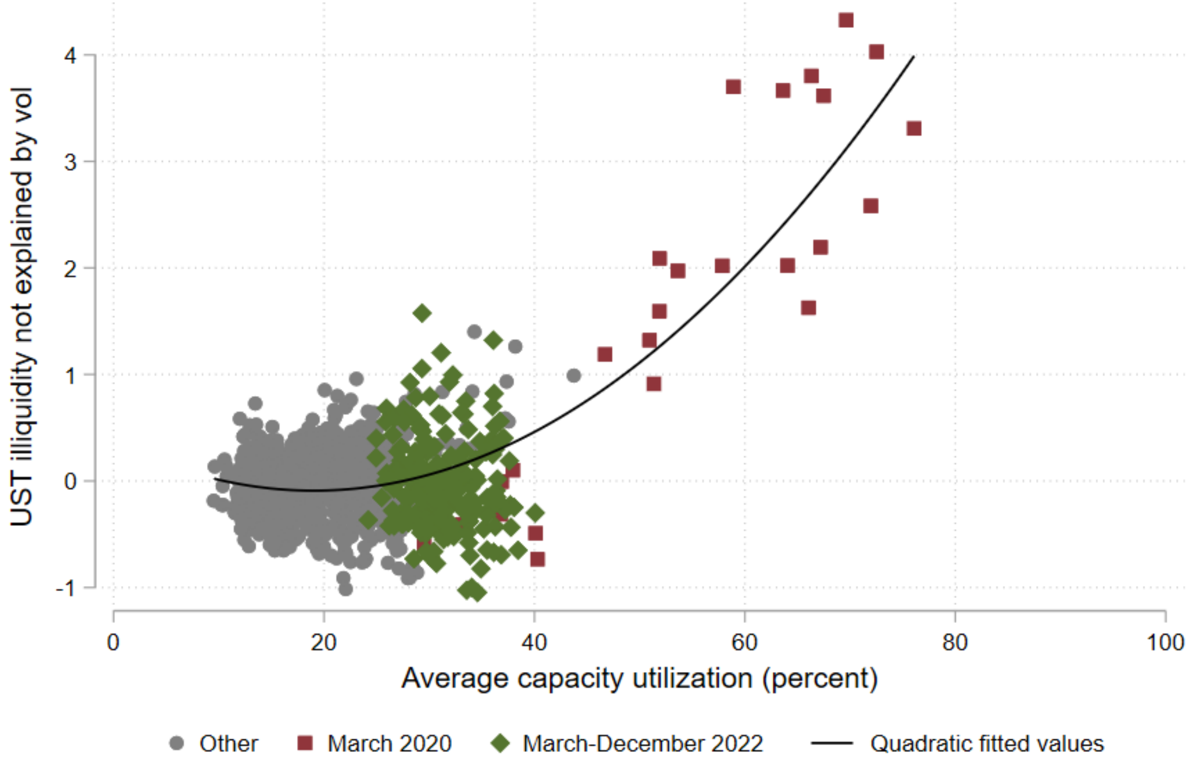


Figure 5. Relationship between US Treasury market illiquidity not explained by yield volatility and average dealer capacity utilization. A scatter plot of the residual illiquidity that remains after controlling for average swaption-implied volatility (the residuals associated with the fitted relationship in Figure 4) and average dealer capacity utilization. The average capacity utilization is the average of the dealer capacity utilization measures based on dealer gross positions, dealer net positions, gross dealer-to-customer volume, and net dealer-to-customer volume. The plotted ordinary-least-squares fit, for July 10, 2017 to December 31, 2022, is the second-order polynomial $y = 0.363 - 0.048x + 0.0013x^2$, with $R^2 = 43.6\%$. All three coefficient estimates have p-values of less than 1% using Newey-West standard errors.

balance sheet utilization and market liquidity.

Equilibrium volatility may be endogenously magnified by dealer balance sheet utilization, especially when utilization is near its limits, creating a feedback effect that may elevate the importance of balance sheet utilization relative to its predicted marginal impact on liquidity after controlling for volatility. In this paper, however, we are unable to disentangle these equilibrium effects. Because of this, our estimates of the impact of dealer balance sheet limitations on Treasury market liquidity are likely to be understated.

We use quantile regression models to explore in more detail how the distribution of Treasury market illiquidity changes in character at tail levels of illiquidity, conditional on yield volatility and dealer capacity utilization. Consistent with our evidence based on ordinary least squares (OLS), at median levels, illiquidity shows no economically important dependence on dealer capacity utilization. However at sufficiently high quantiles of illiquidity, dealer capacity utilization becomes an increasingly important predictor of illiquidity. For example, the 99th percentile of the conditional distribution of composite Treasury illiquidity is estimated to rise over 0.40 standard deviations for each one-standard-deviation increase in dealer capacity utilization, whether measured by gross or net risk-adjusted dealer purchases from customers, even after residualizing for the effect of volatility. Even larger effects are predicted for volatility-residualized dealer capacity utilization measures that are based on gross or net risk-adjusted dealer positions, or based on our estimates of VaR from dealer PnL data.

Our results support the idea that in a fully dealer-intermediated market like the US Treasury market, the cost to market participants of obtaining immediacy—including the cost of delays caused by lack of market depth—can be viewed as having two components. The first component is the direct cost to dealers per unit of asset absorbed from counterparties, including funding costs through the effects of debt overhang, risk-bearing agency costs to dealer traders, and other costs that apply whether dealer inventories are high or low. We find that this "everyday" component of market illiquidity is roughly proportional to yield volatility, which explains about 80% of the daily variation in illiquidity

The second cost component of intermediation is the opportunity cost to dealers of using up some of the remaining available space on their balance sheets. When the amount of balance sheet space remaining is large, the likelihood that taking an additional position from a counterparty will contribute to eventually exhausting the available space is small, and the associated illiquidity costs (including delays) in accessing dealer balance sheets are small. However, when dealer inventories or trade flows are very high, the opportunity cost of giving up some of the remaining available space on dealer balance sheets can be economically significant. Appendix A provides a simple theoretical framework that formalizes these concepts.

As an example, when dealer capacity utilization is between zero and roughly 30%, we find that the first principal component of Treasury illiquidity is predicted to be nearly invariant to dealer capacity utilization, after controlling for yield volatility. However when dealer capacity utilization is 60%, Treasury illiquidity is predicted to be about two standard deviations above the level predicted by contemporaneous volatility. Although the fit of this relationship is quite noisy, the non-linear dependence on dealer capacity utilization is statistically and economically significant.

To further understand how Treasury market functionality has evolved over time, especially with respect to utilization of dealer intermediation capacity, Appendix D extends our analysis to the 2005-2022 sample period. The start of this sample period precedes the existence of TRACE for Treasuries, so for this longer time span we rely exclusively on order-book data from BrokerTec and dealer-position data from the FR 2004. Because of important changes in the capital regulation of BHCs introduced in 2014 that affect the flexibility of bank-affiliated dealers' balance sheets, we break the extended sample period into two sub-periods, one ending December 31, 2013 and the other starting January 1, 2014. For both sub-periods, the results support the thrust of our main findings.

The remainder of the paper is organized as follows. Section 2 describes the relationships of our results with those of prior work on dealer liquidity provision. Section 3 discusses the institutional features of the US Treasury market that underpin market liquidity. Section 4 outlines our data sources. Sections 5 and 6 define our metrics for liquidity and dealer balance sheet utilization, respectively. Section 7 develops the empirical relationships underlying our main findings regarding the connections between Treasury market liquidity, yield volatility, and dealer balance sheet utilization, using measures of capacity utilization based on positions and trade volumes. Section 8 extends our analysis using measures of dealer capacity utilization based on VaR. Section 9 discusses policy implications and Section 10 offers final conclusions.

2 Relation to prior work on dealer liquidity provision

In dealer-intermediated over-the-counter markets, investors typically trade only with dealers. Details for the US Treasury market are outlined in the next section. Dealers temporarily warehouse customer sales and purchases on their balance sheets until laying off excess inventories with other customers or dealers (Wang, 2017). Brunnermeier and Pedersen (2009) emphasize the dependence of market liquidity on intermediary funding liquidity. Because of debt overhang, equity and debt financing of market-making inventories imposes costs on dealer shareholders that reduce the aggressiveness of dealer liquidity provision (Andersen, Duffie, and Song, 2019; Benos and Žikeš, 2018).

In order to safeguard financial stability, regulators typically require more dealer capital for a given amount of inventory than dealers would choose on their own, which further limits liquidity provision (He and Krishnamurthy, 2013; Gromb and Vayanos, 2010; Breckenfelder and Ivashina, 2021).[5] The post-global financial crisis (GFC) tightening of capital regulations therefore increased the balance-sheet costs of dealers (Adrian et al., 2017; Du, Hebert, and Li, 2022). Dealers also bear internal agency costs for intermediation because of the risk aversion and career concerns of traders and their managers. The literature on dealer market liquidity provision has often modeled the effects of dealer inventory costs by assuming risk aversion at the level of the dealer firm itself (Ho and Stoll, 1983; He and Krishnamurthy, 2013; He, Nagel, and Song, 2022).

There is an obvious connection between asset pricing and the balance-sheet costs of dealers, but this connection is not our focus in this paper.[6]

In addition to the effect of direct dealer inventory costs that rise with the level of market-

[5]In the primary market for reverse auctions of UK gilts, Boneva, Kastl, and Zikes (2020) find that dealers sell gilts more aggressively when they have unwanted inventory, or when they took additional gilt positions just before the reverse auctions, or when they are more constrained by the leverage-ratio rule.

[6]See, for example, Adrian, Etula, and Muir (2014); Etula (2009); He, Kelly, and Manela (2017); He and Krishnamurthy (2012); He, Nagel, and Song (2022); Baron and Muir (2022); Du, Hebert, and Li (2022); Fleming, Rosenberg, and Nguyen (2022); Klingler and Sundaresan (2023); Boyarchenko et al. (2018); Jermann (2020); Financial Stability Board (2020); He, Khorrami, and Song (2022); He and Krishnamurthy (2018).

making inventories, there are indirect "shadow costs" implied by an effective upper bound, at least in the short run, on the total amount of inventory that a dealer is able or willing to commit to a given market. In practice, this implicit bound is often called "balance sheet space." The most obvious bound is that implied by the internal allocation of regulatory capital to the relevant dealing activity.

Dealers also impose their own internal risk limits, often based on VaR, in order to protect their firm's franchise value. Dealers are typically reluctant to expand their balance sheet space by issuing new equity, given the associated adverse market signaling as well as the debt-overhang costs to their shareholders. Thus, at least in the short run, when inventory levels are nearing the upper bound associated with allocated capital, the opportunity to capture additional profits by issuing new equity may be forgone unless the associated profit margins are sufficiently high and persistent. An example is the foregone quarter-end profits associated with arbitraging cross-currency bases in the foreign exchange market (Du, Tepper, and Verdelhan, 2018; Wallen, 2022).

These concepts suggest that the marginal shadow cost associated with the balance sheet capacity constraint rises with the degree of utilization of that capacity. This implies in turn that market illiquidity rises nonlinearly with the utilization of dealer intermediation capacity, as motivated by the theoretical discussion in Appendix A. Identifying and quantifying this non-linear effect in the US Treasury market is the main focus of our paper.

In similarly motivated work, Adrian, Boyarchenko, and Shachar (2017) find that in the years leading up to the GFC, corporate bonds that were more heavily traded by more highly levered dealers had better liquidity, but that this relationship reversed after the financial crisis. Breckenfelder and Ivashina (2021) examine the impact of post-2014 leverage-ratio requirements on European coporate bond intermediation. They estimate that for countries in which bank dealers are one percentage point closer to the leverage-ratio-rule requirement, bid-ask spreads are 8 basis points higher. On the other hand, Bicu-Lieb, Chen, and Elliott (2020) find no notable impacts of the leverage-ratio rule on liquidity in the UK gilt market.

As a proxy for balance sheet constraints, they use whether or not the leverage-ratio regulatory capital requirement was in force, and whether or not a dealer had excess capital under the leverage-ratio rule before the rule was imposed.

Huang et al. (2023) find that transactions costs in the foreign exchange market rise when variables that are correlated with the cost of dealer balance sheet space rise, after controlling for dealer-provided volume. For variables that are related to dealer balance sheet costs, Huang et al. (2023) use funding cost spreads and estimates of dealer VaR.[7]

In order to analyze Treasury market liquidity dynamics during March and April of 2020, Goldberg (2020) assumes that a shift in the supply of liquidity by dealers causes opposite-sign changes in prices and quantities. Quantity changes are proxied by weekly changes in dealer gross gross positions from FR2004. A shift in demand by investors is assumed to cause a same-sign change in these two variables. The increase in demand for liquidity in March 2020 is estimated at about 26%, the largest such shift in the sample period, covering 1990 to 2020. He estimates a 17% reduction in the supply of liquidity by dealers, which he ascribes to various factors.

He, Nagel, and Song (2022) estimate significant "inconvenience yields" associated with the SLR regulatory capital constraint on dealer balance sheets and "find that during the two weeks of turmoil, Treasury yields rose substantially above maturity-matched OIS rates, reflecting the inconvenience yield." Huang et al. (2023) find that transactions costs in the foreign exchange market rise when variables that are correlated with the cost of dealer balance sheet space rise, after controlling for dealer-provided volume.

In our work, identifying the impact of balance sheet utilization on liquidity is aided by

[7] They model the cost to dealers of using balance sheet space as $\eta(q + \text{VaR})$, where q is dealer inventory and the cost coefficient η is an input parameter. Huang et al. (2023) assume that VaR is proportional to the variance of position value, $q^2\sigma^2$, where σ^2 is the variance of the asset price. The marginal cost of balance sheet usage is then $\eta + 2\eta\sigma^2 q$. In practice, VaR is defined in a way that is roughly proportional to the standard deviation of the position value, $q\sigma$. With that, the cost of liquidity provision in the sense of Huang et al. (2023) would be of the form $p^S = \eta(1 + \sigma)$, which does not depend on the position size q. Empirically, we find that at any level of dealer capacity utilization, illiquidity rises at a relatively constant rate with volatility, whereas balance-sheet utilization plays a clear additional role only when it reaches a sufficiently high level.

the broad panel of liquidity and dealer capacity utilization metrics that we bring to bear on the US Treasury market and also by a sample period that includes extreme illiquidity to the point of being widely described as "market dysfunctionality." Our results are correspondingly probative. We find statistically and economically substantial non-linear impacts of dealer intermediation capacity utilization on market liquidity. For example, as shown in Figure 5, after controlling for the usual effect of volatility on liquidity, when the capacity utilization of dealers, treated collectively, is around 20%, there is little or no estimated marginal impact of capacity utilization on the first principal component of a panel of Treasury illiquidity metrics. However, again controlling for volatility, when dealer capacity utilization rises from 40% to 80%, market illiquidity is estimated to increase by three to four sample standard deviations. Although these relationships are noisy, having R^2 statistics of only around 20%, the estimated non-linearity is highly statistically significant when tested at traditional confidence levels and the scatter plots plainly reveal striking nonlinear relationships between balance sheet utilization and liquidity.

3 Liquidity provision in the US Treasury market

US Treasury securities trade in a multiple dealer, over-the-counter secondary market. The predominant market makers are the primary government securities dealers, those having a trading relationship with the Federal Reserve Bank of New York. Dealers trade with the Fed, their customers, and one another. Nearly all interdealer trading occurs through interdealer brokers (IDBs), which match buyers and sellers, while ensuring anonymity, even after a trade.

Treasury market structure has evolved considerably since the late 1990s, albeit without substantively affecting the economics of liquidity provision. In the interdealer market, trading of on-the-run notes and bonds (the most recently auctioned of a given original maturity) has migrated from voice-assisted IDBs to fully electronic IDBs. These electronic IDBs opened to non-dealer participants, including principal trading firms, in the mid-2000s.[8] In the D2C market, request-for-quote platforms have developed, on which customers can solicit bids or offers from multiple dealers. Despite these changes, dealers are still counterparties to virtually all customer trades (Chaboud et al., 2022), with the consequence that liquidity provision is dependent on dealer balance sheet capacity. Because some of the largest dealers are within subsidiaries of BHCs, their balance sheets sheets are limited in part by capital and liquidity regulations.

Because of data limitations, studies of Treasury market liquidity have often relied on liquidity proxies, especially for assessing liquidity over long periods of time. The on-the-run/off-the-run spread, for example, gauges the value of a more liquid on-the-run security relative to a less liquid off-the-run security with similar cash flows (Krishnamurthy, 2002; Vayanos and Weill, 2008; Pasquariello and Vega, 2009). Along similar lines, Longstaff (2004) introduces, as a measure of the liquidity premium of Treasuries, the spread between Treasuries and less liquid but equally creditworthy RefCorp bonds. As a measure of arbitrage frictions in the Treasury market, Hu, Pan, and Wang (2013) propose yield-curve "noise," the standard deviation of residuals when fitting market yields with a smooth estimate of the yield curve. We later explain how we enhance this noise measure by exploiting recent improvements in yield-curve fitting (Filipović, Pelger, and Ye, 2022) as well as intraday transactions prices from TRACE.

The increased availability of transactions and order book data have allowed studies to analyze direct measures of liquidity, including bid-ask spreads, depth, and price impact. Several studies analyze price formation and liquidity around macroeconomic announcements (Fleming and Remolona, 1999) or other events (Fleming and Ruela, 2020; Nguyen et al., 2020). Studies assessing liquidity over longer periods of time include Fleming (2003); Adrian, Fleming, and Vogt (2023); Nguyen et al. (2020). In particular, Adrian, Fleming, and Vogt

[8]See US Department of the Treasury, Board of Governors of the Federal Reserve System, Federal Reserve Bank of New York, US Securities and Exchange Commission, and US Commodity Futures Trading Commission (2015).

(2023) assess the evolution of Treasury liquidity using 30 years of limit order book data.

It has long been understood, and is strongly reinforced by our results, that a predominant driver of Treasury market illiquidity is interest rate volatility. Volatility increases the riskiness of intermediating markets, leading to wider bid-ask spreads and reduced depth. For example, Fleming (2003) finds a correlation of 0.82 between realized price volatility and the first principal component of seven liquidity measures for the 2-year note. Adrian, Fleming, and Vogt (2023) find that changes in implied Treasury volatility explain over 20 percent of changes in a Treasury liquidity index composed of bid-ask spreads, depth, and price impact. Nguyen et al. (2020) model the joint dynamics of liquidity, volume, and volatility in the US Treasury market through the GFC and around economic announcements. In contrast, Hu, Pan, and Wang (2013) find no significant relationship between monthly changes in their noise measure and realized volatility, but do find a significant relationship with implied equity volatility.

4 Data sources

Our paper relies on several detailed datasets covering the US Treasury cash and repo markets as well as dealers' balance sheet positions and their riskiness. For the interdealer cash market, we use BrokerTec tick data to capture activity for the on-the-run 2-, 5-, and 10-year notes. For the D2C cash market, we use Treasury TRACE data for on-the-run and off-the-run coupon securities with remaining maturities in the intervals of $(1, 2]$, $(3, 5]$, and $(7, 10]$ years. Our fitted yield curves are from Filipović, Pelger, and Ye (2022). We also use composite bid and offer prices of off-the-run coupons from Tradeweb to help contruct our TRACE-based liquidity measures. Our data on primary dealers' balance sheet positions comes from the Federal Reserve's FR 2004 reports, supplemented with regulatory data for daily VaR and profit-and-loss at bank-affiliated dealers.

Our baseline sample uses the intersection of these datasets spanning the sample period from July 2017 to December 2022 and capturing 23 primary dealers, listed in Table F.13. For robustness, we also construct sub-samples for the periods January 2005 to December 2013 and January 2014 to December 2022, based on implementation in Basel III of the leverage ratio rule in January 2014. By this point, dealers had likely incorporated the implications of this new rule into their balance sheet management. For the extended sample we focus only on interdealer data for our liquidity metrics, given the absence of Treasury TRACE data before July 2017.

BrokerTec. BrokerTec is an electronic trading platform for on-the-run Treasury notes and bonds. In this analysis we rely on the Federal Reserve Bank of New York's reconstructed full limit order book, which is based on a complete record of every order placed on the BrokerTec platform. We also use BrokerTec transaction-level data that include price, quantity, and whether a trade is seller-initiated or buyer-initiated. As mentioned in Section 3, trading activity began migrating from voice-assisted brokers to electronic platforms in the early 2000s, resulting in BrokerTec's coverage significantly improving in 2005 (Adrian, Fleming, and Vogt, 2023). Therefore, January 2005 marks the start of our earlier sub-sample. BrokerTec is also our data source for repurchase agreement ("repo") specials' rates and volumes. These data are used in the adjustment of the on-the-run yield premium for the effect of specialness.

Tradeweb. Tradeweb is an electronic request-for-quote platform in which customers solicit quotes from, and execute trades with, dealers. Tradeweb provides real-time composite bid and offer prices derived from the streaming prices of a cross-section of liquidity providers and a proprietary algorithm. The composite prices are recognized by market participants as being representative of the range in which trades are likely to be filled.

TRACE. Treasury TRACE is the reporting system by which registered broker-dealers report their transactions in Treasuries to the Financial Industry Regulatory Authority (FINRA).[9]

[9] The full reporting requirement are available at FINRA's website.

Reported information includes cusip traded, execution price, trade time, and whether the dealer was a buyer or a seller. TRACE data allow us to calculate liquidity measures for the D2C market and daily gross and net D2C trade flows, as described in Sections 5 and 6. Treasury TRACE reporting began July 10, 2017.

While the focus of our analysis is the US Treasury market, in estimating dealer balance sheet utilization we account for dealer intermediation in other large classes of fixed-income securities. To that end, we use TRACE data for agency MBS and corporate bonds to calculate daily gross and net D2C flows for those securities.

FR 2004A. The Federal Reserve's FR 2004A form collects information on primary dealers' positions in various classes of fixed-income securities, including US Treasuries, agency debt, agency MBS, and corporate debt. These reports covers long and short positions, at fair market values, as of the close of business each Wednesday. Reportable positions reflect outright transactions, including those executed in the when-issued market, as well as securities purchased at auction. While derivatives are not included, the effects of derivatives are captured in regulatory data for daily VaR and profit-and-loss that we analyze in Section 8. Positions are broken down by security types and residual maturity buckets. For Treasuries, in particular, positions are reported for bills, notes and bonds with residual maturities of 0-to-2 years, 2-to-3 years, 3-to-6 years, 6-to-7 years, 7-to-11 years and more than 11 years, Treasury Inflation Protected Securities with residual maturities of 0-to-2 years, 2-to-6 years, 6-to-11 years, and more than 11 years, and Floating Rate Notes. Data are reported for the "legal entity that functions as the primary dealer, including any subsidiaries that it consolidates in its regulatory reports," and thus do not include data from unconsolidated subsidiaries within the same holding company.

5 Liquidity metrics

To capture US Treasury market liquidity we calculate three metrics for the on-the-run, interdealer segment and three for the mostly off-the-run D2C segment. Liquidity in the latter segment has not been studied before due to lack of data. We focus on the 2-, 5-, and 10-year maturity sectors as these are among the most heavily traded securities on Brokertec (Fleming, Mizrach, and Nguyen, 2018) and in the US Treasury market in general (Brain et al., 2018). We calculate our metrics using data from New York trading hours, 7:00 am to 5:00 pm Eastern time.

We follow Adrian, Fleming, and Vogt (2023) and use BrokerTec data to construct our on-the-run, interdealer measures. For each coupon security and day, we calculate bid-ask spread, price impact and quoted depth. The bid-ask spread is the average spread between the best bid and the best ask in the limit order book divided by the bid-ask midpoint. The price impact is the slope coefficient from a regression of one-minute log price changes on one-minute net order flow. Quoted depth is the order book depth at the inside tier, summed across the bid and offer sides. Detailed description of the construction of these measures and the subsequent measures is in Appendix B.

We use Treasury TRACE to calculate our mostly off-the-run D2C measures. Our measure of price dispersion aims to capture the dispersion in trading of the same bond at the same time by different customers in the D2C segment of the US Treasury market. We measure daily price-dispersion measures for off-the-run coupons in each of the three maturity segments of the market as the trade-weighted standard deviation of the signed difference between customer traded prices in TRACE to the most recent mid-quote prior to the trade from Tradeweb.

Our yield curve RMSE measure enhances the noise measure of Hu, Pan, and Wang (2013) by exploiting intraday transactions prices from TRACE as well as recent improvements in yield-curve fitting. We first compute the trade-weighted average yield for each security on each day as compared to the daily fitted yield from a non-parametric estimate of the smoothed curve from Filipović, Pelger, and Ye (2022), derived from CRSP end-of-day quotes. From these differences, we compute the average (issue-weighted) RMSE for each day and maturity segment. By using intraday traded prices, our paper is the first to document the yield curve noise that customers actually experience when they trade. We find that yield curve noise tends to be higher when measured with trades rather than with quotes, including during periods of market stress.[10]

Our sixth and last liquidity metric is the specialness-adjusted on-the-run premium. The on-the-run premium, i.e. the difference between the market yield of an on-the-run nominal security and a hypothetical off-the-run security with identical cash flows, subsumes repo specialness and liquidity effects (Krishnamurthy, 2002). We measure the market yield for the on-the-run benchmark security with the BrokerTec mid-quote at 3:00 pm. The hypothetical yield for the same benchmark security is derived from the off-the-run yield curve from Gürkaynak, Sack, and Wright (2007). The specialness-adjusted on-the-run premium attempts to isolate the liquidity effects by removing repo specialness. Specifically, we regress the on-the-run premium onto a one-week moving average of repo specialness along with fixed effects for weeks since issuance to capture the auction cycle. BrokerTec is our data source for repurchase agreement ("repo") specials' rates and volumes. The specialness-adjusted on-the-run premium is the residual from the regression plus the sample mean of the on-the-run premium. Appendix B.3 provides additional details on the construction of the specialness-adjusted on-the-run premium and reports the results of the regressions.

After calculating the 18 individual metrics (six measures times three maturities), we aggregate them to a single measure by standardizing them and taking the first principal component. This first principal component, the thick black line in Figure 1, aims to capture the overall functioning of the US Treasury market. Figure 6 shows that the first principal

[10]For our July 10, 2017 to December 31, 2022 sample period, we find average yield curve noise of 1.7 basis points using traded prices versus 1.2 basis points using indicative closing quotes from CRSP data. Averages for March 2020 are 3.3 and 2 basis points, respectively.

component has a relatively strong and positive dependence on each of the z-scores, and explains about 61% of their variation, in the usual sense of principal-component analysis. These properties suggest that each of the underlying metrics is relevant and that none of them are close to redundant, given the others. From this point, we will rely on our first principal component as our composite illiquidity measure for much of our predictive analysis of Treasury market illiquidity and its relationship with dealer capacity utilization.

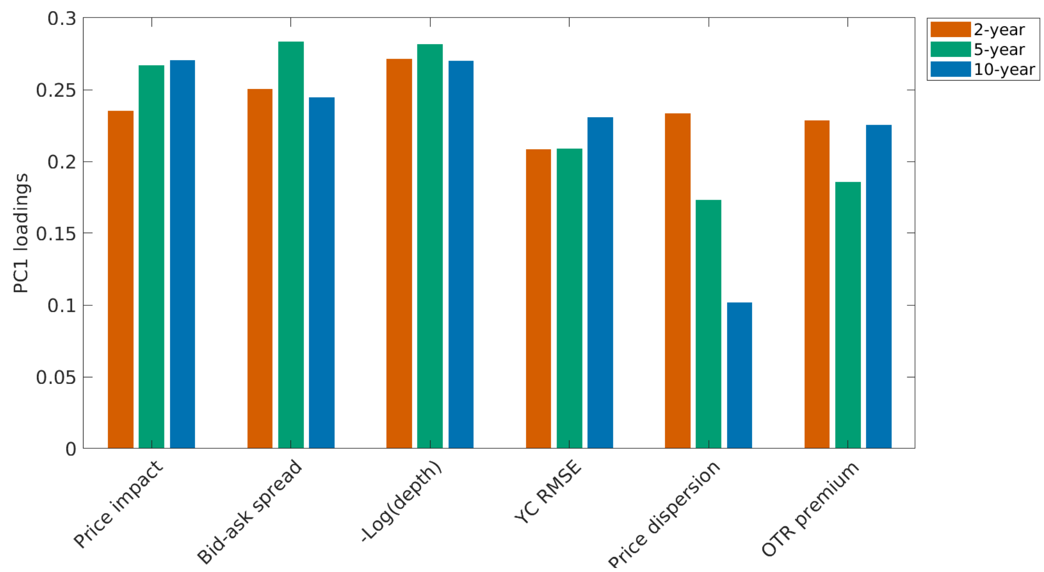


Figure 6. Loadings on first principal component of US Treasury market illiquidity measures. Loadings for 18 individual metrics comprised of six liquidity measures times three maturities.

6 Dealer capacity utilization

Intermediation capacity of dealers is difficult to measure directly. We take an indirect approach and estimate dealers' capacity based on their revealed preference. We estimate the risks associated with (1) gross positions held, (2) net positions held, (3) gross D2C flows, and (4) net D2C flows. For each primary dealer, the capacity for each of these risk measures is proxied by the in-sample historical maximum. Measured capacity is thus likely to be somewhat downward biased. On a given day, a dealer's utilization of its capacity for bearing the given type of risk is estimated as the risk for that day normalized by the historical maximum. The capacity utilization for the collection of dealers is the weighted average of dealer-level utilizations, with weights proportional to historical dealer maximums. Our utilization measures are calculated at a daily frequency.

The gross position for dealer d on day t in security type i is defined as the sum of the dealer's long and short positions, $Q_{dti}^{\text{gross}} = Q_{dti}^{\text{long}} + Q_{dti}^{\text{short}}$. The net position is defined as the difference between the dealer's long and short positions, $Q_{dti}^{\text{net}} = Q_{dti}^{\text{long}} - Q_{dti}^{\text{short}}$. Since FR 2004 positions data are weekly, we measure the gross and net position each day using the data from the most recent Wednesday. We account for positions in Treasuries (bills, coupons, TIPS, and FRNs), agency MBS, and corporate bonds.

The gross D2C flow for dealer d on day t in security type i is defined as the sum of the dealer's buy and sell customer flows, $Q_{dti}^{\text{grossD2C}} = Q_{dti}^{\text{buy}} + Q_{dti}^{\text{sell}}$. The net D2C flow is defined as the difference between the dealer's buy and sell customer flows, $Q_{dti}^{\text{netD2C}} = Q_{dti}^{\text{buy}} - Q_{dti}^{\text{sell}}$. In calculating our flow capacity measures, we use three-day moving averages of the gross and net volumes of dealers' D2C purchases. The three-day lagging window, although somewhat arbitrary, is based on the notion that it typically takes a day or more to settle trades and to find counterparties with whom the acquired positions can be "laid off." As a robustness check, we examined the fit of the empirical relationships between Treasury market illiquidity and dealer capacity utilization based on alternative D2C flow windows of 2 days and 5 days. The quality of fit is slightly improved with a 5-day window, and slightly weaker with a 2-day window.[11]

We adjust for risk within each security type by multiplying the principal amount of the position or flow measure by the price-to-yield sensitivity (DV01) and the market-implied yield volatility from maturity-matched interest-rate swaps.[12] For example, for Treasury coupons

[11]For example, using a 5-day lagging window to measure D2C flows, the pseudo-R^2 measures of fit are slightly higher and the p-values are slightly lower than those for the 3-day window shown Table 1, and the converse applies with a 2-day flow window.

[12]We use an implied volatility measure because implied volatility is forward-looking, reflecting market expectations of volatility going forward, and because it is less likely to be affected by cash market disruptions. Realized volatility, in contrast, reflects past volatility and is likely to be appreciably affected by the market dysfunction we are seeking to explain. That said, as a robustness check, we repeated our analysis using three different measures of realized volatility: 21-day trailing yield volatility, calculated with daily data and

we match the maturity segments of (0,2], (2-3], (3,6], (6,7], (7,11], and (11,30] years from FR 2004 with one-month implied yield volatility for 2-, 3-, 5-, 7-, 10-, and 30-year swaps and on-the-run benchmark DV01 amounts. Table C.6 summarizes the risk-adjustments for each security type. Using these risk-adjusted measures, we aggregate across security types within each dealer-day as follows,

$$Q_{dt}^{\text{gross}} = \sum_i Q_{dti}^{\text{gross}} DV01_{it} Vol_{it} \tag{1}$$

$$Q_{dt}^{\text{net}} = \left| \sum_i Q_{dti}^{\text{net}} DV01_{it} Vol_{it} \right| \tag{2}$$

$$Q_{dt}^{\text{grossD2C}} = \sum_i \left(\frac{1}{3} \sum_{s=0}^{2} Q_{dti-s}^{\text{grossD2C}} \right) DV01_{it} Vol_{it} \tag{3}$$

$$Q_{dt}^{\text{netD2C}} = \left| \sum_i \left(\frac{1}{3} \sum_{s=0}^{2} Q_{dti-s}^{\text{netD2C}} \right) DV01_{it} Vol_{it} \right|, \tag{4}$$

where $DV01_{it}$ is the dollar value of a basis point and Vol_{it} is implied yield volatility. We take the absolute value of the dealer-day, risk-adjusted net positions and flows in order to identify when dealers have large imbalances relative to their historical net positions, irrespective of the sign of the net position. Table C.7 provides summary statistics for the resulting risk-adjusted flow and position measures.

For each measure of risk-adjusted positions or flows, $q_{dt} \in \left\{ Q_{dt}^{\text{gross}}, Q_{dt}^{\text{net}}, Q_{dt}^{\text{grossD2C}}, Q_{dt}^{\text{netD2C}} \right\}$, we measure capacity utilization for each dealer as the fraction $f_{dt} = q_{dt}/m_d \in [0, 1]$ relative to the within-dealer sample maximum $m_d = \max_t q_{dt}$ and then compute aggregate capacity utilization $M_t = \sum_d w_d f_{dt}$ using capacity weights $w_d = m_d / \sum_d m_d$ that are proportional to the within-dealer maximum values. Finally, we adjust for additive, monthly seasonality in the flow-based capacity measures, as described in Section C.3. Figure 2 reports the aggre-

averaged across the 2-, 5-, and 10-year notes, same-day realized volatility, calculated with end-of-minute bid-ask midpoint prices, standardized for each of the 2-, 5-, and 10-year notes, and then averaged, and 21-day trailing realized volatility, calculated in the same manner as same-day realized volatility, but averaged over 21 trading days. Our key findings largely hold with these realized volatility measures. Specifically, the coefficients on our dealer capacity measures reported later in Tables 1 and 2 remain positive and statistically significant in half or more of the eight specifications, depending on the realized volatility measure, with the coefficients statistically insignificant in the remaining specifications.

gate capacity utilization measures for gross positions, net positions, gross D2C flows, and net D2C flows, and Table C.5 summarizes the cross-sectional distribution of the capacity measures.

7 Liquidity when dealer balance sheets are stressed

As illustrated in Section 1, in a period of stressed selling such as March 2020, the US Treasury market is much less liquid than would be predicted by volatility alone. Here, we use quantile regression models to explore in more depth the relationship between Treasury market illiquidity, yield volatility, and dealer balance sheet utilization.[13] We characterize the conditional distribution of illiquidity with a panel of quantile regressions of illiquidity as linear functions of volatility and dealer capacity utilization. This allows us to study the tail behavior of illiquidity more systematically, with an emphasis on understanding the impact of intense use of dealer balance sheet space. We find that measures of dealer capacity utilization provide substantial additional explanation of high-quantile illiquidity. As discussed in Section 2, this finding provides new evidence on the importance of intermediary frictions for asset pricing.

Figure 7 shows the coefficients, sometimes called “loadings,” estimated from quantile regressions of Treasury market illiquidity, over a range of quantiles. We use the composite measure of Treasury illiquidity constructed in Section 5 as the first principal component of 18 underlying standardized measures of illiquidity across three maturity sectors. Panel (a) is based on univariate quantile regressions of illiquidity explained with Treasury yield volatility, measured as the average of the one-month implied volatilities on 2-, 5-, and 10-year interest-rate swaps. As shown, the slope coefficients in these univariate regressions rise sharply at high quantiles of illiquidity. However, Panel (b) shows a much more stable dependence of illiquidity on yield volatility, across quantiles, in bivariate quantile regressions that include dealer capacity utilization as an additional explanatory variable.

[13] See Koenker and Hallock (2001) for background on quantile regression modeling.

In order to clarify the incremental dependence of quantiles of illiquidity on dealer capacity utilization—which depends on volatility—after controlling for volatility itself, we orthogonalize the dealer capacity utilization variable to volatility. That is, we examine the incremental role of the OLS residuals of the capacity utilization variable regressed on volatility.[14]

Panel (c) shows that, at medium quantiles, there is relatively little dependence of predicted illiquidity on dealer capacity utilization. But the dependence on capacity utilization is significantly higher in the right tail of the distribution of illiquidity. At these extremes, the coefficient on dealer capacity utilization is more than two standard errors above the corresponding OLS slope coefficient, also shown in the figure with a 95% confidence band. Appendix Figure G.11 shows similar results when measuring dealer capacity utilization with FR 2004 net positions and with TRACE D2C gross and net purchases from customers. In the next section, we also explore the role of dealer balance sheet utilization when measured based on VaR. In general, the results support an interpretation of the data by which dealer balance sheet capacity is not a major driver of illiquidity during benign periods, but that high levels of dealer balance sheet utilization limit further intermediation.

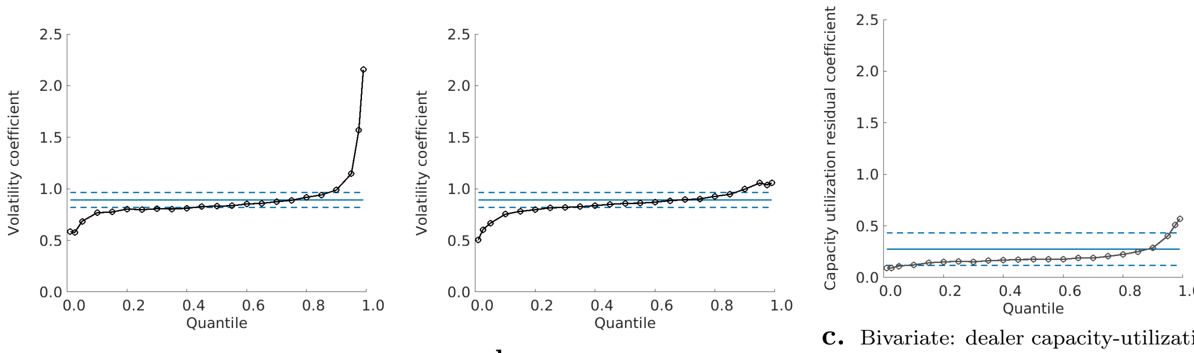
With respect to the economic magnitudes of the effects, each of the variables has been standardized so that the coefficients are comparable across plots. For example, at the 99th percentile, one-standard-deviation increases in volatility and in the capacity utilization residual are associated with 1.06 and 0.57 standard-deviation increases in Treasury market illiquidity, respectively. This contrasts with the estimated conditional median estimation of illiquidity, which shows that illiquidity increases by 0.87 standard deviations for a one-

[14]One can recover the estimated dependence of the quantile of illiquidity on the un-residualized measure of dealer capacity utilization by exploiting the linearity of both the OLS and quantile models. Suppose, for each $t \in \{1, \dots, T\}$, that $x_t = A' z_t$ for x_t and z_t in $\mathbb{R}^n$ and some fixed nonsingular $n \times n$ matrix A. Suppose further that $b^* = \operatorname{argmin}_b \sum_t f(b' z_t)$ for some $f : \mathbb{R} \to \mathbb{R}$. Then

$$\operatorname{argmin}_c \sum_t f(c' x_t) = \operatorname{argmin}_c \sum_t f(c' A' z_t) = A^{-1} b^*.$$

Appendix H applies this result to our quantile regression setting. It reports the first stage regressions of capacity utilization onto volatility from which the implied coefficients can be derived along with quantile regressions using un-residualized capacity utilization for comparison.

standard-deviation increase in volatility, but increases by only 0.17 standard deviations with a one-standard deviation increase in residualized capacity utilization.



a. Univariate: volatility

b. Bivariate: volatility

c. Bivariate: dealer capacity-utilization residuals

Figure 7. US Treasury market illiquidity conditional distribution slope coefficients by percentile based on volatility and dealer capacity utilization measured based on risk-adjusted gross positions. Plots the relationship between Treasury illiquidity, volatility, and residualized capacity utilization from quantile regressions. Panel (a) reports the slope coefficients from univariate quantile regressions with volatility for different percentiles. The ordinary-least-squares (OLS) coefficient is reported as a solid blue line with the dashed lines indicating a 95% confidence interval. If higher levels of volatility resulted in a simple location shift in the distribution of illiquidity we would expect similar coefficients across quantiles to the OLS coefficient. Panels (b) and (c) report the slope coefficients on volatility and the capacity utilization residual from bivariate quantile regressions for different percentiles. Gross position capacity is the risk-adjusted utilization, and residualized relative to volatility. The explanatory variables are standardized to make the slope coefficients comparable.

In order to further investigate the role of dealer balance sheet capacity utilization, Table 1 focuses on the results of quantile regressions for the 99th percentile of illiquidity. Column (1) uses yield volatility as the sole explanatory factor. We see that volatility is statistically and economically significant, with an explanatory power of 54% as measured by pseudo-R^2.[15] At the estimated 99th percentile of illiquidity, a one-standard-deviation increase in volatility is associated with a 2.16 standard deviation increase in illiquidity, which is more than double the 0.89 OLS univariate slope coefficient.[16] The next four columns (2-5) consider bivariate specifications that include the dealer capacity utilization residual as an additional explanatory factor. The dealer capacity utilization residual is significant across specifications

[15]Pseudo-R^2 is defined as $\sum_t \rho_\tau(y_t - x_t'\beta)/\sum_t \rho_\tau(y_t - \hat{y}_\tau)$ where $\hat{y}_\tau$ is the unconditional τ-th quantile and $\rho_\tau(u) = \rho_\tau = u \cdot (\tau - \mathbf{1}_{u<0})$ is the piece-wise linear quantile loss function.

[16]Given the persistence of the dependent and explanatory variables, we use block bootstrapped standard errors following (Gregory, Lahiri, and Nordman, 2018) to measure significance. In unreported results we find similar levels of significance and standard errors using a semiparametric bootstrap procedure following Efron and Tibshirani (1994) that is described in Appendix E

and increases explanatory power to around 71% to 74% as measured by the pseudo-R^2.[17] Moving from the univariate to the bivariate specification causes the loading on volatility to drop to levels closer to that of the univariate OLS coefficient for volatility. These results are consistent with the importance of dealer capacity utilization as a significant determinant of right-tail Treasury illiquidity, alongside volatility.

Without reporting the details, we also estimated 0.99 quantile regressions of illiquidity using dealer capacity utilization as the sole explanatory variable. Of the various capacity utilization measure, that based on risk-adjusted gross positions generates the highest pseudo-R^2, which at 70% is actually larger than that for the model based on volatility alone.

Table 1: Quantile regressions for 99th percentile of US Treasury market illiquidity. Reports quantile regressions for the 99th percentile of Treasury illiquidity onto volatility and capacity utilization. For ease of interpretation all variables are standardized in the regressions. Standard errors are block bootstrapped with a block size of 100 days following the smooth-extended tapered approach from Gregory, Lahiri, and Nordman (2018). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}\ p < 0.01$.

	(1)	(2)	(3)	(4)	(5)
Yield volatility	2.16^{***}	1.06^{***}	1.12^{***}	1.25^{***}	1.14^{***}
	(0.48)	(0.14)	(0.14)	(0.16)	(0.12)
Capacity utilization residual: gross position		0.57^{***}			
		(0.17)			
Capacity utilization residual: net position			0.43^{**}		
			(0.17)		
Capacity utilization residual: gross D2C volume				0.51^{***}	
				(0.17)	
Capacity utilization residual: net D2C volume					0.41^{**}
					(0.16)
Constant	1.80^{***}	1.01^{***}	0.97^{***}	1.16^{***}	0.91^{***}
	(0.50)	(0.15)	(0.15)	(0.21)	(0.13)
N	1336	1336	1336	1336	1336
Pseudo R2	0.54	0.74	0.74	0.71	0.74

[17]In our main results, we aggregate the net measures of dealer capacity utilization using the absolute value of the net trade volume or net position scaled by its historical maximum. The net measures thus identify when dealers have large imbalances relative to their historical net positions, irrespective of the sign of the net position or of the correlation of net positions across dealers. If dealers are less connected during periods of market stress, aggregating dealer level net measures in this manner may be appropriate for highlighting times when dealers have large imbalances or absolute net positions relative to their within-dealer historical levels. We also considered a specification in which we aggregate the signed net position across dealers and then compute the absolute value after aggregating. If dealers have offsetting net positions or flows, this alternative specification of the net measure will be lower than our baseline measure which looks at the imbalance or absolute net measure within dealer. In unreported results we found that this alternative definition resulted in similar statistical significance and explanatory power of net positions. Defining net D2C purchase volumes in this alternative way results in a positive slope coefficient but the relationship is no longer significant.

To further illustrate the results, Figure 8 shows heat maps of fitted quantiles of Treasury market illiquidity. These show that dealer capacity utilization matters more in the right tails of illiquidity. The higher predicted tail quantiles of illiquidity are reflected by the darker red cells in the northeast region of the heat map, when volatility and dealer capacity utilization are both elevated.

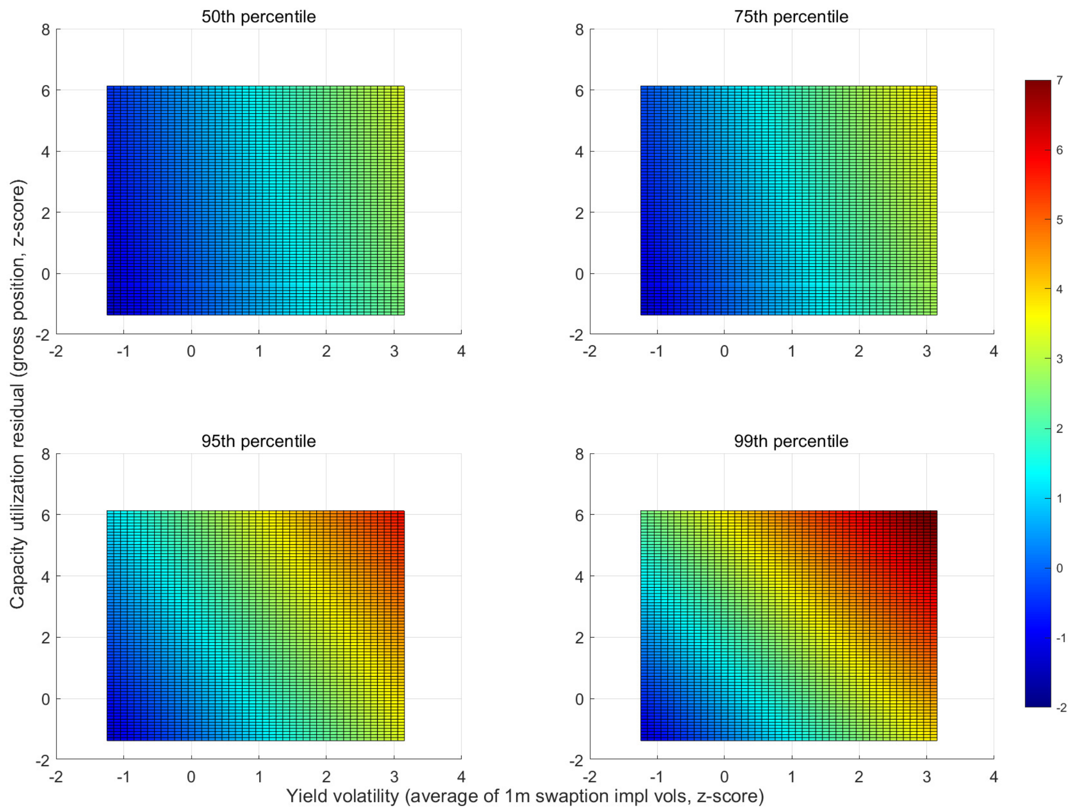


Figure 8. Heatmap of US Treasury market illiquidity. Plots the fitted value for Treasury illiquidity given the level of volatility and residual capacity utilization for the 50th, 75th, 95th, and 99th percentiles of the conditional distribution from quantile regressions. Volatility is measured as the average one-month implied volatility on 2-, 5-, and 10-year swaps. Capacity utilization is measured with dealer gross positions. The residual of capacity utilization captures the component not explained by volatility. Each of the variables is standardized. The range of the variables is determined by the historical minimum and maximum during our sample period.

In order to examine the relationships between Treasury market liquidity, yield volatility, and dealer capacity utilization over a longer time span, Appendix D, extends our analysis back to 2005. Because TRACE data covering the customer-to-dealer segment of the Treasury market become available only in 2017, this extended analysis relies on FR 2004 dealer-positions-based capacity measures and on on-the-run interdealer liquidity measures based on BrokerTec data. We measure on-the-run Treasury illiquidity over the longer horizon using the first principal component of spreads, depth, and price impact for the 2-, 5-, and 10-year benchmark securities–nine liquidity measures in all.

Significant changes in BHC regulations introduced beginning in 2014, including the enhanced Supplementary Leverage Ratio, have important implications for the provision of liquidity to the Treasury market by bank-affiliated dealers (Tarullo, 2023; Group of Thirty, 2021). Because of this important structural change in regulation, we estimate the role of dealer capacity utilization separately for two sub-periods: (i) May 25, 2005 to December 30, 2013 and (ii) January 2, 2014 to December 30, 2022.

Figures D.5b and D.6 in Appendix D support the thrust of our empirical findings for the main sample period. First, yield volatility explains the majority of variation in Treasury market illiquidity. Here again, this relationship is roughly linear. Second, when illiqduity is much higher than would be explained by volatility, dealer capacity utilization is also high. During the earlier subperiod, 2005-2013, this "stressed-market" role of dealer capacity utilization shows up in the fall of 2008, after the fall of Lehman. In the latter sub-period, this effect applied in March 2020, much as we found in our main sample period with a richer set of measures of market liquidity and dealer capacity utilization. Quantile regressions, reported in Table D.11, also support our main-sample finding of significant incremental roles for dealer capacity utilization measures, given volatility, in the conditional tails of Treasury market illiquidity.

8 Illiquidity and VaR capacity utilization

In measuring dealer capacity utilization, we capture the largest fixed-income security positions and flows. Nonetheless, our measures do not capture all types of positions that might contribute to capacity limits. For example, FR 2004 positions and TRACE volume data do not capture exposures related to fixed-income derivatives such as futures and interest-rate swaps. Moreover, the amount of risk capital or regulatory capital allocated to the fixed-income divisions of BHCs may depend on the risk exposures of the overall firm. To account for these sources of risk, we derive additional capacity utilization measures based on confidential supervisory data for daily VaR and PnL from the Market Risk Rule (MRR).[18] The MRR data are used for determining market risk capital requirements and for monitoring compliance with restrictions on proprietary trading as required by Section 13 of the Bank Holding Company Act (the Volcker Rule). The VaR data are measures of one-day holding period 99th percentile VaR and the PnL data are based on the net change in the price of positions held as of the previous business day. Our sample of regulatory data runs from July 2017 to December 2022, which overlaps with our main sample period. The data are available for 18 bank-affilitaed dealers at the BHC level, and for 15 dealers include a breakdown focusing on trading desks with large exposures to interest rate risk.

In addition to dealer-reported VaRs, we use a simple quantile regression model to estimate the corresponding VaRs from PnL data as described in Appendix E. Our estimated VaRs could be less accurate than those reported by dealers because dealers can exploit detailed position data and sophisticated risk management systems. However, the dealer estimates are reported for regulatory purposes and are not necessarily used for internal risk management. Dealer regulatory capital requirements rise with reported VaR, and are higher to the extent that PnL exceeds reported VaR estimates.[19] The dealer-reported VaRs also suffer from VaR model heterogeneity across dealers, which likely reduces the reliability of cross-dealer aggregate measures such as capacity utilization. For example, one dealer might report realized

[18]The MRR data is collected pursuant to Basel III Subpart F (Market Risk) Section 205 as described here. The MRR applies to Board-regulated institutions with aggregate trading assets and liabilities equal to $1 billion or more or that are equal to 10% or more of quarter-end total assets as reported on the most recent quarterly Call Report (FR Y-9C). The MRR data includes a daily measure of VaR for each subportfolio calibrated to a one-tail, 99.0 percent confidence level and the daily PnL of the subportfolio based on the net change in the price of the positions held as of the end of the previous business day, excluding any commissions, fees, or bid-ask spreads that are generated by new positions.

[19]Dealers must include multipliers and add-on factors in their required regulatory capital when PnL exceeds reported VaR thresholds. Abboud et al. (2021) describes a large number of PnL "exceptions" during March 2020.

VaR from a trailing window, whereas another might use a forward-looking approach that takes into account current positions and current market-implied volatility.

Using dealer capacity utilization measures based on reported and estimated VaR, for interest-rate related positions and at the firm-wide level, Table 2 repeats our quantile regression analysis of the role of dealer balance sheet capacity utilization in the tail behavior of US Treasury market illiquidity. Column (1) repeats the univariate quantile regression based on yield volatility. Columns (2-5) report bivariate specifications using the capacity utilization residual measured with the reported and estimated VaR at the BHC level and for desks with large exposures to interest rate risk. After residualizing the VaR measures to yield volatility, each of the four VaR measures has a positive slope coefficient. As before, the variables are standardized so that magnitudes can be interpreted in standard-deviation units. The coefficients for the estimated VaR measures (columns and 4 and 5) are statistically and economically significant, with somewhat higher explanatory power (pseudo-R^2) than for the respective dealer-reported VaR measures (columns 2 and 3).

The capacity utilization measures based on estimated VaR are highly correlated with those based on position and trade-flow data while the correlation is somewhat lower for reported VaR, as reported in Appendix Table C.10. The explanatory power for high-quantile Treasury market illiquidity achieved with the estimated VaR-based models is similar to those based on positions and trade-flow data, reported in Table 1, a robustness check of our finding that dealer capacity utilization is an important explanatory factor for the tail behavior of Treasury market liquidity.

Table 2: Quantile regressions for dealer capacity utilization measured with dealer Value-at-Risk. Reports quantile regressions for the 99th percentile of US Treasury market illiquidity onto yield volatility and capacity utilization. Capacity utilization is measured with value-at-risk (VaR) at the bank holding company (BHC) level and for desks with exposure to interest rate risk (IR). VaR is a one-day 99-percentile measure that is reported by the dealer as required by the Market Risk Rule or that is estimated from daily profit-and-loss data. For ease of interpretation all variables are standardized in the regressions. Standard errors are block bootstrapped. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}\ p < 0.01$.

	(1)	(2)	(3)	(4)	(5)
Yield volatility	2.16***	2.11***	2.20***	1.32***	1.26***
	(0.47)	(0.47)	(0.52)	(0.23)	(0.16)
Capacity utilization residual: BHC VaR Dealer		0.26*			
		(0.15)			
Capacity utilization residual: IR VaR Dealer			0.25*		
			(0.14)		
Capacity utilization residual: BHC VaR Estimated				0.52***	
				(0.17)	
Capacity utilization residual: IR VaR Estimated					0.44***
					(0.17)
Constant	1.80***	1.51***	1.61***	0.96***	0.85***
	(0.50)	(0.40)	(0.45)	(0.21)	(0.13)
N	1336	1336	1336	1336	1336
Pseudo R2	0.54	0.60	0.57	0.72	0.75

9 Policy implications

Central banks and other market authorities can benefit from being able to identify when government securities markets are not functioning adequately—and whether dealers' intermediation capacity limits are playing a role. The knowledge can be valuable to designing regulations, considering changes in market structure, or gauging whether official-sector financing or purchasing programs would be useful in supporting market functionality (Duffie and Keane, 2023).

To illustrate, suppose that the fixed-income divisions of most dealers have significant unused balance-sheet capacity, but are quoting wide bid-offer spreads because fundamental macroeconomic uncertainty and hence yield volatility are high. This may describe the situation in the second half of 2022 and in March 2023. Treasury yield volatility was elevated by heightened monetary policy uncertainty and disruptions in the banking sector, and illiquidity was proportionately high, but dealers were not near their intermediation capacity limits, at least according to the measures provided in this paper. In this situation, our results suggest relatively low effectiveness for central-bank purchases as a way to restore market liquidity, as shown in Figures 4 and 5. Indeed, the Fed did not intervene with market-function purchases during late 2002 and March 2023, even though Treasury market liquidity was poor—and worse by some metrics in the 2-year sector than in March 2020.

Figure 3 shows that the typical close relationship between Treasury market illiquidity and yield volatility broke down in March 2020 for a period of roughly four weeks.[20] The Fed indeed intervened during this period with significant purchases, buying nearly $1 trillion of Treasuries in the three weeks following March 12, 2020. The Fed continued to purchase significant quantities of Treasuries, however, after market functionality had largely returned.

Our findings suggest that one way that policymakers may be able to identify when markets are increasingly vulnerable to an episode of dysfunctionality is by monitoring dealer capacity utilization. While some episodes of market dysfunction may be triggered without warning, other instances may become more likely as dealers approach their intermediation capacity limits, and be mitigated by market intervention. Conversely, when dealer capacity measures do not appear to be stretched by investor demands for liquidity, monitoring dealer capacity utilization may help officials avoid unnecessary market intervention. A focus on dealer capacity is natural for the US Treasury market given the rapid growth in debt outstanding and the much less robust growth of dealer balance sheets (Duffie, 2020). Moreover, as we show, fixed-income-market dealers also devote a substantial fraction of their balance-sheets to other key bond markets, particularly the agency MBS market, contributing significantly to peak pressures on the space available for liquidity provision in the Treasury market.

While it was suspected that central bank purchases in March 2020 were useful for restoring market liquidity because the purchases freed space on dealer balance sheets for further

[20]From the end of the first week of March 2020 to the end of the first week of April 2020, the residuals of the regression shown Figure 3 were at least two standard deviations above the standard error of the regression, which is 0.45.

intermediation,[21] this paper may be the first to support that conjecture with empirical research. Prior empirical research (Bernardini and De Nicola, 2020; Fleming et al., 2022; Vissing-Jorgensen, 2021) on the effectiveness of central bank "pandemic" purchases of government securities does not identify a specific channel for the impact of the purchases on market functionality, such as the role of dealer intermediation capacity.

The US Treasury market is commonly viewed as the world's most important safe haven, given its large size, its liquidity, its active repo market, and the extremely low default risk of its securities. Yet, there have been several episodes of abrupt deterioration in market liquidity in recent years, most notably in March of 2020, when liquidity became so impaired that official asset purchases were viewed as necessary to restore and sustain market function (Logan, 2020). Changes that might improve the resilience of the Treasury market are being explored by the Inter-Agency Working Group for Treasury Market Surveillance.[22]

Expanded central clearing is one of the proposals for improving the resilience of the Treasury market.[23] Broader central clearing, among other benefits, could expand dealer balance-sheet capacity for handling trade flows through the netting of purchases against sales. Fleming and Keane (2021) find, in a counterfactual empirical analysis of trade settlement in the US Treasury market, that central clearing reduces total dealer settlement exposures significantly, especially when trade volumes are higher, and by up to 70% on peak days in March 2020. They also find that broad central clearing leads to a significant reduction in settlement failures, which increase counterparty risk, tie up additional dealer balance sheet space, and involve obvious matching inefficiencies.

Another approach for improving the resilience of the US Treasury market is to expand the use of all-to-all trading, which is currently almost non-existent in the market. Although

[21]See, for example, Duffie (2020); Logan (2020); Liang and Parkinson (2020); Busetto et al. (2022); Fleming et al. (2022).

[22]See US Department of the Treasury, Board of Governors of the Federal Reserve System, Federal Reserve Bank of New York, US Securities and Exchange Commission, and US Commodity Futures Trading Commission (2021, 2022). See also Group of Thirty (2021, 2022).

[23]See "SEC Proposes Rules to Improve Risk Management in Clearance and Settlement and to Facilitate Additional Central Clearing for the US Treasury Market," press release, September 14.

dealers play a crucial liquidity provision role on D2C trade venues, augmenting the market structure through the ability of non-dealers to trade directly with each other could conserve dealer balance sheet space, improve allocative efficiency, and increase market resilience under stress (Chaboud et al., 2022; Allen and Wittwer, 2023; Kutai, Nathan, and Wittwer, 2022). The SEC has proposed a rule change that, among other measures, would expand "fair access" to US Treasury trading platforms, where quotes are currently provided only by dealers and some specialized high-frequency trading firms.

While changes to market structure may help make the Treasury market more resilient, the importance of dealer balance sheet capacity for market liquidity in certain situations is likely to remain salient given the rapid growth in debt outstanding relative to the growth in dealer balance sheets. Efforts to closely track Treasury market liquidity and its drivers, and to monitor dealer balance sheet capacity, will likely remain of value to policymakers in the years ahead.

10 Conclusion

This study combines highly relevant data on dealer-level balance-sheet positions and comprehensive transaction-level Treasury security trades, among other data sets, to show that there is a significant loss in US Treasury market functionality when intensive use of dealer balance sheets is needed to intermediate bond markets, as in March 2020. While yield volatility explains most of the variation in Treasury market liquidity over time, when dealer balance-sheet utilization reaches sufficiently high levels, liquidity is much worse than predicted by yield volatility alone. This is consistent with the existence of occasionally binding constraints on the intermediation capacity of bond markets.

A limitation of our analysis is that we cannot directly observe the amount of a dealer's balance sheet that is available for intermediating bond markets. These limits depend on multifaceted internal assignments based on risk, funding, and various regulatory measures of capital and liquidity within large complex financial institutions. We resort to an approach that relies on various measures of the use of dealer balance sheets, relative to sample-record maxima. Our results are similar in character whether the measures are based on gross or net dealer positions, gross or net customer-to-dealer trade flows, or value-at-risk (whether as reported by dealers or as estimated by us, and whether firm-wide or specific to fixed-income positions).

Our quantification of the role of capacity utilization on illiquidity, after controlling for the everyday role of yield volatility, is also limited by the fact that high capacity utilization is likely to cause increases in yield volatility. Given this endogeneity, our estimated impacts of capacity utilization are likely understated.

The status of US Treasury securities as the world's premier safe haven rests in part on the depth and liquidity of the market in which they are traded. Our results shed new light on the dependence of market liquidity on asset volatility and dealer intermediation capacity, and add focus to ongoing policy efforts to improve the resilience of the US Treasury market, an anchor of global capital markets.

References

Abboud, A., C. Anderson, A. Game, D. Iercosan, H. Inanoglu, and D. Lynch (2021): "Banks' Backtesting Exceptions during the COVID-19 Crash: Causes and Consequences," FEDS Notes. Board of Governors of the Federal Reserve System.

Adrian, T., N. Boyarchenko, and O. Shachar (2017): "Dealer Balance Sheets and Bond Liquidity Provision," *Journal of Monetary Economics*, 89, 92–109.

Adrian, T., E. Etula, and T. Muir (2014): "Financial intermediaries and the cross-section of asset returns," *Journal of Finance*, 69, 2557–2596.

Adrian, T., M. Fleming, J. Goldberg, M. Lewis, F. M. Natalucci, and J. J. Wu (2013): "Dealer Balance Sheet Capacity and Market Liquidity during the 2013 Selloff in Fixed Income Markets," Federal Reserve Bank of New York *Liberty Street Economics*.

Adrian, T., M. Fleming, O. Shachar, and E. Vogt (2017): "Market liquidity after the financial crisis," *Annual Review of Financial Economics*, 9, 43–83.

Adrian, T., M. Fleming, and E. Vogt (2023): "The Evolution of Treasury Market Liquidity: Evidence from 30 Years of Limit Order Book Data," Federal Reserve Bank of New York Staff Report No. 827.

Allen, J. and M. Wittwer (2023): "Centralizing Over-The-Counter Markets," *Journal of Political Economy*, forthcoming.

Andersen, L., D. Duffie, and Y. Song (2019): "Funding Value Adjustments," *Journal of Finance*, 74, 145–192.

Baron, M. and T. Muir (2022): "Intermediaries and asset prices: International evidence since 1870," *Review of Financial Studies*, 35, 2144–2189.

Benos, E. and F. Žikeš (2018): "Funding constraints and liquidity in two-tiered OTC markets," *Journal of Financial Markets*, 39, 24–43.

Bernardini, M. and A. De Nicola (2020): "The market stabilization role of central bank asset purchases: high-frequency evidence from the COVID-19 crisis," Bank of Italy Temi di Discussione (Working Paper) No. 1310.

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman (2018): "Capital Commitment and Illiquidity in Corporate Bonds," *Journal of Finance*, 73, 1615–1661.

Bicu-Lieb, A., L. Chen, and D. Elliott (2020): "The leverage ratio and liquidity in the gilt and gilt repo markets," *Journal of Financial Markets*, 48, 100510.

Boneva, L., J. Kastl, and F. Zikes (2020): "Dealer balance sheets and bidding behavior in the UK QE reverse auctions," Working paper.

Boyarchenko, N., T. M. Eisenbach, P. Gupta, O. Shachar, and P. Van Tassel (2018): "Bank-intermediated arbitrage," Federal Reserve Bank of New York Staff Report No. 858.

Brain, D., M. D. Pooter, D. Dobrev, M. Fleming, P. Johansson, F. Keane, M. Puglia, A. Rodrigues, and O. Shachar (2018): "Breaking Down TRACE Volumes Further," Federal Reserve Bank of New York *Liberty Street Economics*.

Brainard, L. (2021): "Some Preliminary Financial Stability Lessons from the COVID-19 Shock," Speech at the 2021 Annual Washington Conference Institute of International Bankers, Board of Governors of the Federal Reserve System.

Breckenfelder, J. and V. Ivashina (2021): "Bank leverage constraints and bond market illiquidity during the COVID-19 crisis," European Central Bank Research Bulletin No. 89.

Brunnermeier, M. K. and L. H. Pedersen (2009): "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201–2238.

Busetto, F., M. Chavaz, M. Froemel, M. Joyce, I. Kaminska, and J. Worlidge (2022): "QE at the bank of England: a perspective on its functioning and effectiveness," *Bank of England Quarterly Bulletin*, Q1.

Chaboud, A., E. Correia-Golay, C. Cox, M. J. Fleming, Y. Huh, F. M. Keane, K. Lee, K. Schwarz, C. Vega, and C. Windover (2022): "All-to-all trading in the US Treasury market," Federal Reserve Bank of New York Staff Report No. 1036.

Comerton-Forde, C., T. Hendershott, C. Jones, P. Moulton, and M. Seasholes (2010): "Time Variation in Liquidity: The Role of Market-Maker Inventories and Revenues," *Journal of Finance*, 65, 295–331.

Du, W., B. Hebert, and W. Li (2022): "Intermediary Balance Sheets and the Treasury Yield Curve," Federal Reserve Bank of New York Staff Report No. 1023, forthcoming, *Journal of Financial Economics*.

Du, W., A. Tepper, and A. Verdelhan (2018): "Deviations from covered interest rate parity," *Journal of Finance*, 73, 915–957.

Duffie, D. (2020): "Still the World's Safe Haven? – Redesigning the US Treasury Market After the COVID-19 Crisis," Hutchins Center Working Paper No. 62, Brookings Institution.

——— (2023): "Resilience Redux in the US Treasury Market," Jackson Hole Symposium paper, Stanford University.

Duffie, D. and F. M. Keane (2023): "Market-Function Asset Purchases," Federal Reserve Bank of New York Staff Report No. 1054.

Efron, B. and R. J. Tibshirani (1994): *An introduction to the bootstrap*, CRC press.

Etula, E. (2009): "Broker-Dealer Risk Appetite and Commodity Returns," Federal Reserve Bank of New York Staff Report No. 406.

Filipović, D., M. Pelger, and Y. Ye (2022): "Stripping the Discount Curve: A Robust Machine Learning Approach," Swiss Finance Institute Research Paper No. 22-24.

Financial Stability Board (2020): "Holistic Review of the March Market Turmoil," Financial Stability Board Reports to the G20.

Fleming, M. and F. Keane (2021): "The Netting Efficiencies of Marketwide Central Clearing," Federal Reserve Bank of New York Staff Report No. 964.

Fleming, M., H. Liu, R. Podjasek, and J. Schurmeier (2022): "The Federal Reserve's Market Functioning Purchases," *Federal Bank of New York Economic Policy Review*, 9, 210–241.

Fleming, M. and F. Ruela (2020): "Treasury Market Liquidity during the COVID-19 Crisis," Federal Reserve Bank of New York *Liberty Street Economics*.

Fleming, M. J. (2003): "Measuring Treasury market liquidity," *Federal Reserve Bank of New York Economic Policy Review*, 9, 83–108.

Fleming, M. J., B. Mizrach, and G. Nguyen (2018): "The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform," *Journal of Financial Markets*, 40, 2–22.

Fleming, M. J., G. H. Nguyen, and F. Ruela (2023): "Tick Size, Competition for Liquidity Provision, and Price Discovery: Evidence from the US Treasury Market," *Management Science*, forthcoming.

Fleming, M. J. and E. M. Remolona (1999): "Price formation and liquidity in the US Treasury market: The response to public information," *Journal of Finance*, 54, 1901–1915.

Fleming, M. J., J. V. Rosenberg, and G. Nguyen (2022): "How do Treasury dealers manage their positions?" Federal Reserve Bank of New York Staff Report No. 299.

Fontaine, J.-S., C. Garriott, J. Johal, J. Lee, and A. Uthemann (2021): "COVID-19 Crisis: Lessons Learned for Future Policy Research," Staff Discussion Paper 2021-2, Bank of Canada.

Garbade, K. D. and W. L. Silber (1976): "Price dispersion in the government securities market," *Journal of Political Economy*, 84, 721–740.

Giglio, S., B. Kelly, and S. Pruitt (2016): "Systemic risk and the macroeconomy: An empirical evaluation," *Journal of Financial Economics*, 119, 457–471.

Goldberg, J. (2020): "Dealer Inventory Constraints during the COVID-19 Pandemic: Evidence from the Treasury Market and Broader Implications," FEDS Notes. Board of Governors of the Federal Reserve System.

Gregory, K. B., S. N. Lahiri, and D. J. Nordman (2018): "A smooth block bootstrap for quantile regression with time series," *Annals of Statistics*, 46, 1138–1166.

Gromb, D. and D. Vayanos (2010): "A model of financial market liquidity based on intermediary capital," *Journal of the European Economic Association*, 8, 456–466.

Group of Thirty (2021): "US Treasury Markets: Steps Toward Increased Resilience," G30 Working Group on Treasury Market Liquidity, Group of 30, Washington, DC.

——— (2022): "US Treasury Markets: Steps Toward Increased Resilience: Status Update 2022," G30 Working Group on Treasury Market Liquidity, Group of 30, Washington, DC.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007): "The US Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54, 2291–2304.

He, Z., B. Kelly, and A. Manela (2017): "Intermediary Asset Pricing: New Evidence from Many Asset Classes," *Journal of Financial Economics*, 126, 1–35.

He, Z., P. Khorrami, and Z. Song (2022): "Commonality in credit spread changes: Dealer inventory and intermediary distress," *Review of Financial Studies*, 35, 4630–4673.

He, Z. and A. Krishnamurthy (2012): "A model of capital and crises," *Review of Economic Studies*, 79, 735–777.

——— (2013): "Intermediary Asset Pricing," *American Economic Review*, 103, 732–70.

——— (2018): "Intermediary asset pricing and the financial crisis," *Annual Review of Financial Economics*, 10, 173–197.

He, Z., S. Nagel, and Z. Song (2022): "Treasury Inconvenience Yields during the COVID-19 Crisis," *Journal of Financial Economics*, 43, 57–79.

Ho, T. S. and H. R. Stoll (1983): "The dynamics of dealer markets under competition," *Journal of Finance*, 38, 1053–1074.

Hu, G. X., J. Pan, and J. Wang (2013): "Noise as information for illiquidity," *Journal of Finance*, 68, 2341–2382.

Huang, W., A. Ranaldo, A. Schrimpf, and F. Somogyi (2023): "Constrained Liquidity Provision in Currency Markets," Working paper, Bank for International Settlements.

Jermann, U. (2020): "Negative swap spreads and limited arbitrage," *Review of Financial Studies*, 33, 212–238.

Klingler, S. and S. Sundaresan (2023): "Diminishing Treasury convenience premiums: Effects of dealers' excess demand and balance sheet constraints," *Journal of Monetary Economics*, 135, 55–69.

Koenker, R. and K. F. Hallock (2001): "Quantile regression," *Journal of Economic Perspectives*, 15, 143–156.

Krishnamurthy, A. (2002): "The bond/old-bond spread," *Journal of Financial Economics*, 66, 463–506.

Kutai, A., D. Nathan, and M. Wittwer (2022): "Exchanges for Government Bonds? Evidence During COVID-19," Working paper, Bank of Israel.

Liang, N. and P. Parkinson (2020): "Enhancing Liquidity of the US Treasury Market Under Stress," Hutchins Center Working Paper No. 72, Brookings Institution.

Logan, L. (2020): "Treasury market liquidity and early lessons from the pandemic shock," Remarks at Brookings-Chicago Booth Task Force on Financial Stability meeting.

Longstaff, F. A. (2004): "The Flight-to-Liquidity Premium in US Treasury Bond Prices," *Journal of Business*, 77, 511–526.

Nguyen, G., R. Engle, M. Fleming, and E. Ghysels (2020): "Liquidity and volatility in the US Treasury market," *Journal of Econometrics*, 217, 207–229.

Pasquariello, P. and C. Vega (2009): "The on-the-run liquidity phenomenon," *Journal of Financial Economics*, 92, 1–24.

Tarullo, D. (2023): "Capital Regulation and the Treasury Market," Hutchins Center, Brookings Institution.

US Department of the Treasury, Board of Governors of the Federal Reserve System, Federal Reserve Bank of New York, US Securities and Exchange Commission, and US Commodity Futures Trading Commission (2015): "Joint Staff Report: The US Treasury Market on October 15, 2014," Washington DC.

——— (2021): "Recent Disruptions and Potential Reforms in the US Treasury Market: A Staff Progress Report," Washington DC.

——— (2022): "Enhancing the Resiliency of the US Treasury Market: 2022 Staff Progress Report," Washington DC.

Vayanos, D. and P.-O. Weill (2008): "A search-based theory of the on-the-run phenomenon," *Journal of Finance*, 63, 1361–1398.

Vissing-Jorgensen, A. (2021): "The Treasury Market in Spring 2020 and the Response of the Federal Reserve," *Journal of Monetary Economics*, 124, 19–47.

Wallen, J. (2022): "Markups to financial intermediation in foreign exchange markets," Working paper, Harvard University.

Wang, C. (2017): "Core-Periphery Trading Networks," Working paper, Stanford University.

Weill, P.-O. (2007): "Leaning Against the Wind," *Review of Economic Studies*, 74, 1329–1354.

A Theoretical role of dealer capacity utilization

This appendix outlines a theoretical motivation for the relevance for the provision of market liquidity of varying the amount of space available on a dealer's balance sheet. A model of dealer balance sheet constraints with only one trading period is nearly degenerate from this perspective because either *(i)* the balance sheet constraint is binding, in which case one cannot characterize how dealer pricing depends on variation in capacity utilization in typical ranges below 100%, or *(ii)* the balance-sheet constraint is not binding, in which case there is no shadow cost associated with the capacity limit. This is merely restating the complementary slackness condition for constrained optimization problems. A dynamic model of dealer pricing and balance-sheet constraints is therefore needed to obtain realistic implications for dealer pricing of varying the remaining balance sheet space currently available to dealers. We present an extremely simple model of this type to illustrate that point.

Suppose a monopolistic dealer quotes a bid price b_t and an ask price a_t at each integer time t between 0 and T. The asset pays some uncertain amount v at time T, after trade at time T. For technical simplicity, the set of states of the world is finite. At time t and at any bid price of b, investors buy $B(b, X_t) \geq 0$ from the dealer, where X_t is a stochastic process describing the state of the market at time t. We assume that X_t is exogenous to dealer behavior. Likewise, at any ask price a, investors sell $A(a, X_t) \geq 0$ to the dealer. At each time t the dealer must meet the capacity constraint $q_t \leq C(X_t)$, where q_t is the dealer's inventory at time t and $C(X_t)$ is an inventory limit that incorporates the effects of market conditions such as volatility. Examples could include a regulatory capital constraint or an internal VaR constraint. For simplicity, we ignore any constraint on the dealer's short position size, but that could be analyzed with the same approach.

The dealer has some initial inventory y and, to take a simple objective, solves the expected profit maximization problem

$$\sup_{a,b} E\left[\sum_{t=0}^{T} A(a_t, X_t)(a_t - v) + B(b_t, X_t)(v - b_t)\right]$$

subject to $q_t \leq C(X_t)$ for $0 \leq t \leq T$, where

$$q_t = y + \sum_{s=0}^{t} A(a_s, X_s) - B(b_s, X_s).$$

We suppose that the investor demand and supply functions, $B(\,\cdot\,, X_t)$ and $A(\,\cdot\,, X_t)$, satisfy enough technical regularity that Kuhn-Tucker "first-order" optimality conditions apply. The first-order conditions reveal in a natural way how the dealer's current liquidity-provision incentives depend on the tightness of current and expected future balance-sheet constraints.

By collecting all terms in the Lagrangian that involve the initial ask price a_0, we have the optimality condition

$$A(a_0^*, X_0) = A_a(a_0^*, X_0)\left[E\left(v - \sum_{t=0}^{T} \lambda_t\right) - a_0^*\right],$$

where $\lambda_t(\omega)$ is the Lagrange multiplier associated with the capacity constraint and A_a denotes the derivative of the supply function A with respect to the ask price (determining the elasticity of supply). This is like the usual monopolistic first-order condition for unconstrained profit maximization, except that the dealer bids for the asset as though its expected payoff $E(v)$ is reduced by the sum of expected future shadow costs of balance sheet space. Similarly, for the initial bid price b_0, we have

$$B(b_0^*, X_0) = B_b(b_0^*, X_0)\left[E\left(v - \sum_{t=0}^{T}\lambda_t\right) - b_0^*\right].$$

If time zero is typical, the dealer will not initially be at its balance-sheet limit, given the opportunity value of reserving at least a little bit of balance sheet space in anticipation of the possibility of heavy and relatively inelastic customer selling at a future time at which the dealer could get significant price concessions. When unconstrained at time zero, the dealer bids less aggressively to extent that future balance-sheet constraints are expected to be more binding.

The expected shadow cost $E(\lambda_t)$ of balance sheet space at a future time t is increasing in the initial inventory y, decreasing in upward (stochastic dominating) shifts in the probability distribution of the balance-sheet capacity $C(X_t)$, increasing in upward shifts in future investor supply functions, and decreasing in upward shifts in future investor demand functions. From a welfare perspective, increases in the expected future shadow cost $E(\lambda_t)$ of balance sheet space reduce the incentives of counterparties to sell at time zero, and increase the incentives of counterparties to buy. (If we were to incorporate non-negative or short position constraints into the model, these would dominate at sufficiently low levels of inventory.)

For a simple parametric example, suppose that $A(a, X_0) = k_a e^{-\alpha a}$ and $B(b, X_0) = k_b e^{\beta b}$, for positive parameters k_a, k_b, α, and β. With this parameterization, the first-order conditions for optimality are both necessary and sufficient, and are solved by ask price

$$a_0^* = E\left(v - \sum_{t=0}^{T}\lambda_t\right) + \frac{1}{\alpha}$$

and bid price

$$b_0^* = E\left(v - \sum_{t=0}^{T}\lambda_t\right) - \frac{1}{\beta}.$$

The bid-ask spread, $\alpha^{-1}+\beta^{-1}$, is constant, and in particular does not depend on the expected shadow cost of future balance-sheet space because both the bid price and the ask price are shifted down equally by this shadow cost. Balance sheet constraints nevertheless cause a distortion in prices.

In a related continuous-time Markovian model that considers the impact of dealer quotes on inventories, Duffie (2023) finds the bid-ask spread $\alpha^{-1} + \beta^{-1} - V''(y)$, where $V(x)$ is the optimal expected value to the dealer of discounted future dealing profits, considering the effect of inventory constraints. The component $\alpha^{-1}+\beta^{-1}$ of the bid-ask spread is a monopoly rent taking from immediate counterparties, which would apply with or without inventory constraints. The second component $-V''(y)$ increases as y approaches the dealer's inventory constraint because of the declining marginal value to the dealer of adding to inventory as the amount of space available nears exhaustion. This is consistent with our empirical results.

Taking a different theoretical approach, Weill (2007) considers the impact of a capacity limit on the dealer's rate of transactions. One can also consider the direct costs to dealers of holding inventory, such as risk-bearing costs or shareholder debt-overhang costs for financing inventories, sometimes called "funding value adjustments" by practitioners. For this, we can extend the original setting of this appendix to consider the modified dealer objective

$$\sup_{a,b} E\left[\sum_{t=0}^{T} A(a_t, X_t)(v - a_t) - B(b_t, X_t)(v - b_t) - K(q_t, X_t)\right],$$

where K is a per-period inventory cost function that is convex and differentiable in q. In this case, the first-order condition for the optimal bid and ask prices are, respectively,

$$A(a_0^*, X_0) = A_a(a_0^*, X_0)\left[E\left(v - \sum_{t=0}^{T} \lambda_t - K_q(q_t, X_t)\right) - a_0^*\right],$$

$$B(b_0^*, X_0) = B_b(a_0^*, X_0)\left[E\left(v - \sum_{t=0}^{T} \lambda_t - K_q(q_t, X_t)\right) - b_0^*\right].$$

With this, the dealer bids and offers at time zero as it would in a market with no constraints or direct balance sheet costs, for an asset whose payoff is $E\left(v - \sum_{t=0}^{T} \lambda_t - K_q(q_t, X_t)\right)$. Thus, our empirical results are roughly consistent with an absence of inventory constraints and a marginal inventory cost function $K_q(q_t, X_t)$ that is relatively linear in volatility and highly convex in inventory at high levels of inventory.

Our cost-based measures of Treasury market illiquidity vary roughly linearly with volatility, suggesting that $K(q)$ may vary roughly linearly with position size, implying that the marginal direct frictional cost $K'(q_t)$ to dealers of holding inventory may not vary markedly with the level of dealer inventory, but we can't be confident of this. The linearity of K is reasonable for the cost of inventory funding, which is roughly the product of the quantity of required funding and the dealer's funding credit spread (Andersen, Duffie, and Song, 2019).

While an individual monopolistic dealer's expected profits decline as its balance-sheet becomes more constrained, an oligopolistic dealer is better off to the extent that *other* dealers are more balance-sheet constrained. Depending on specific market conditions, it can be either less or more profitable for dealers, overall, to make markets in a world in which all dealer balance sheets are more constrained. At one extreme, one can take the trivial case in which all dealer balance sheet space is reduced to virtually zero, with at least some non-zero lower bound on the price elasticity of investor demand and supply. Clearly, dealer profitability would suffer from such severe constraints. However, consider the tightening of a regulation that causes dealer balance sheet space to be reduced. Investor trade could then be more constrained, causing investors to have a higher willingness to pay for intermediation from dealers. In an oligopolistic setting, this could cause all dealers to earn greater rents. Those dealers that are relatively less constrained would earn relatively greater increases in profits than dealers that are relatively more constrained. The tightening of a risk-based balance-sheet regulation could easily be welfare improving overall, however, because of a sufficiently large improvement in financial stability.

One could extend the Cournot setting of Wallen (2022), who models an oligopolistic market in which dealers without quarter-end capital regulations benefit if other dealers have binding quarter-end capital constraints, and calibrates the model to the cross-currency basis. A simpler example is a cartel that is able to enforce capacity limits on its members. While each member of the cartel would, if able, benefit from a relaxation of only its own capacity constraint, all members earn higher profits by cartelizing, with an optimally chosen capacity constraint that binds all members to some fraction of their uncartelized capacities.

B Additional details about the illiquidity measures

This appendix provides additional details on the liquidity metrics in Section 5.

B.1 On-the-run interdealer measures

Bid-ask spread is calculated for each security and day as the average spread between the best bid and the best ask in the limit order book as reported by BrokerTec divided by the bid-ask midpoint. In calculating the average we weight all ticks (changes in the order book) equally, implicitly giving greater weight to periods with more active updates to the order book. The bid-ask spread is measured in basis points (in return space).

Price impact is calculated for each security and day as the slope coefficient from a regression of the one-minute log price changes onto that security's one-minute net order flow. Net order flow is calculated as buyer-initiated trading volume minus seller-initiated trading volume, where the signing of trades is based on BrokerTec trade data. Price impact is measured in basis points (in return space) per $100 million net trading volume and is inflation adjusted to the 2022 price level.

Quoted depth is the order book depth at the inside tier, summed across the bid and offer sides. Depth is measured in millions of US dollars par, and is inflation adjusted to the 2022 price level. We then take the negative log of the depth so a higher number indicates worse liquidity, similar to the other measures we calculate.

2-year tick size change, and our adjustment US Treasury securities are quoted and traded in 32nds of a point, where a point equals one percent of par, with the 32nds split into fractions. On November 19, 2018, BrokerTec halved the 2-year note tick size from $1/4$ to $1/8$ of a 32nd, while tick sizes of other securities remained unchanged. Fleming, Nguyen, and Ruela (2023) find that the transition to a smaller tick size have led to an immediate narrowing of the inside bid-ask spread by about 50% and increase in trading activity, whereas the depth near the top of the limit order book when measured at a fixed distance from the bid-ask midpoint changed modestly (as is also observed in Figure B.1).

To quantify the effects of the tick size reduction on spreads and depths, we run the following differences-in-differences regression for each security i on day t:

$$Y_{i,t} \quad = \quad \alpha_i + \gamma_t + \beta \, \text{Post}_i \times \text{Treatments}_i + \varepsilon_{i,t} \tag{5}$$

where $Y_{i,t}$ is either bid-ask spread or the negative log depth, $Treatment$ is an indicator variable equal to 1 for the 2-year note and 0 otherwise, $Post$ is an indicator variable equal to 1 for the period following the tick size change, and α_i and γ_t and security and day fixed effects, respectively. The coefficient β captures the effect of the tick size reduction on the liquidity series. For the bid-ask spread, we run the regression using *raw* spreads and then we convert to proportional spreads, as the latter is also affected by the structural break. Table B.1 reports the regression results for bid-ask spreads and negative log depth.

Table B.1: Effects of tick size reduction on spreads and depths. Reports the results of the differences-in-differences regression in Equation (6) estimated for the 2-, 5-, and 10-year notes from September 24, 2018 through January 11, 2019. t statistics in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

	Spread	Negative log depth
Post tick-size	-0.119***	1.108***
change * **1**[2-Year]	(-114.65)	(25.95)
FEs	Term, trading date	Term, trading date
N	222	222
R2	1.000	0.976

To adjust the liquidity series for the effect of the tick size reduction, we subtract the coefficient β from the liquidity series: $\hat{Y}_{i,t} = Y_{i,t} - \beta$. As Figure B.1 shows, the resulting adjusted series no longer exhibit a structural break at the time of the tick size reduction.

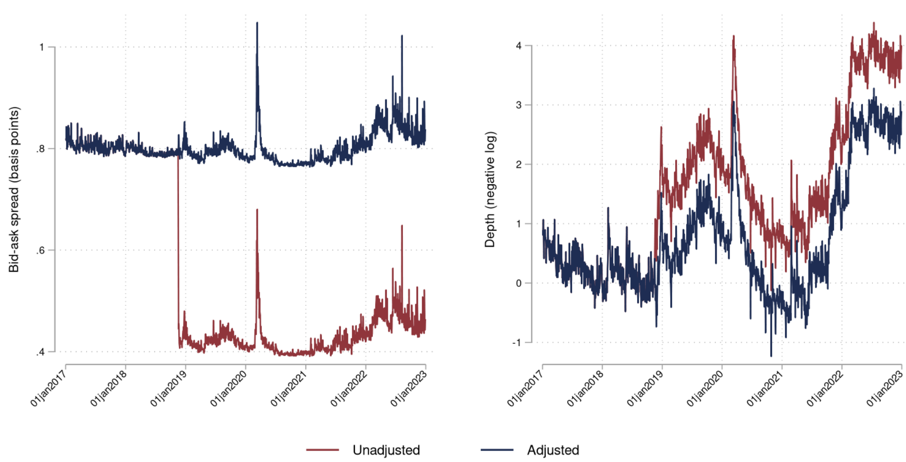


Figure B.1. Tick-size adjustment for 2-year spread and depth. Time-series plot of adjusted and unadjusted 2-year proportional spreads and negative log depth for January 2, 2017 through December 30, 2022.

B.2 Off-the-run D2C measures

Price dispersion In the absence of market frictions, one expects that traded prices would be equal to the expected market valuation. However, in the presence of market frictions, such as dealer inventory risk and client search costs, traded prices can be either higher or lower than the market expected value. Frictions can also result in situations where the security is traded at significantly different prices at approximately the same time.

Garbade and Silber (1976) is one of the earliest works to analyze price dispersion models in the context of the US Treasury market. Using daily bid and ask quote sheets from government securities dealers for small trades, they compute dispersion as the 90 percent range or root-mean-squared distance of bid and ask quotes from the average quote. Dispersion is found to be larger on days when an issue has a large absolute price change.

Our measure of price dispersion for the Treasury market compares customer traded prices in TRACE to the most recent mid-quote prior to the trade from Tradeweb.[24] We measure daily price- dispersion measures for each of the three maturity segments of the market as the trade-weighted standard deviation of the signed difference between customer traded prices and mid-quotes expressed in basis points of yield-to-maturity.

Yield curve root mean square error (RMSE) Following Hu, Pan, and Wang (2013), our yield curve RMSE (or noise) measure is intended to capture the deviation between Treasury prices and those implied by a smooth yield curve. Our approach differs from Hu, Pan, and Wang (2013) in its use of intraday transactions prices from TRACE and by incorporating recent improvements in yield-curve fitting. Using intraday traded prices allows us to measure the yield curve noise that customers actually experience when they trade. The smoothed yield curve we use is a non-parametric estimate from Filipović, Pelger, and Ye (2022) that is derived from CRSP end-of-day quotes.

As with price dispersion, when measuring yield curve noise we need to account for the fact that customer trades occur asynchronously over the day. To do this, we adjust each traded price by adding the change between the CRSP end-of-day quote and the mid-quote just prior to the trade. This adjustment synchronizes all trades to match the end-of-day timing from which the smoothed yield curve is derived. After synchronizing the trades, we compute the trade-weighted average yield for each security on each day and compare this to the daily fitted yield from the smoothed curve. From these differences, we compute the average (issue-weighted) RMSE for each day and maturity segment.

B.3 Specialness-adjusted on-the-run premium

One contribution of our paper lies in the calculation of the on-the-run premium controlling for repo specialness. For security i on day t, we calculate the specialness-adjusted on-the-run

[24] The median amount of time between the trade and quote is less than 2 seconds for the off-the-run institutional D2C trades in our sample. We require that each trade have a quote for the same security within at least one hour before the trade to be included in the sample. We winsorize the trade-level dispersion measure at 25 basis points in yield space to mitigate the impact of outliers. This is meant to be a conservative filter in that for our sample of trades, the 99th percentile is 8.5 basis points.

premium as the residual from the following regression:

$$p_{it} \quad = \quad a + b \cdot s_{it} + \gamma_w + \varepsilon_{it}, \tag{6}$$

where p_{it} is the on-the-run premium, s_{it} is the one-week moving average repo specialness of the same security, γ_w is a fixed effect for the number of weeks since issuance capturing auction cycle patterns, and ε_{it} is the residual. Repo specialness is the difference between the daily general collateral repo rate from Bloomberg and the volume-weighted average specials repo rate for the benchmark security from BrokerTec. For the 10-year benchmark security, we include an additional explanatory variable in the regression, $s_{it} \cdot \mathbf{1}_{\text{FirstMonth}_{\text{it}}}$, an interaction term of repo specialness and an indicator for the first month since a security's issuance. This interaction term captures the effect of differing levels of specialness for 10-year note before and after it is re-opened one month from its original issuance.

The term $b \cdot s_{it}$ (plus some part of the intercept a and fixed effects γ_w) captures the component of the on-the-run premium that comes from repo specialness. The residual ε_{it} (plus the remaining intercept component) captures the on-the-run liquidity premium. On a stress day, when the off-the-run segment of the market becomes more impaired relative to the on-the-run market segment, we expect the residual ε_{it} to be bigger than normal because the present value of specialness is arguably about the same as usual, and has about the same level of noise as usual in its contribution to the residual. Therefore, an elevated level of ε_{it} is a noisy signal that the segment of off-the-run market near this benchmark maturity is more impaired than normal relative to the market for this benchmark.

Table B.2 reports the results of the specialness-adjustment regressions for the 2-, 5-, and 10-year securities. Table B.3 reports summary statistics of on-the-run premium, fitted value from the regression, and specialness-adjusted on-the-run premium which equals the residual from the regression plus the sample mean of the on-the-run premium.

Table B.2: Specialness-adjustment regressions. This table reports the specialness-adjustment regressions results for the the 2-, 5-, and 10-year on-the-run yield premiums. Fixed effects for the number of weeks since original issuance of the security. The t-statistics are reported in parentheses using HAC standard errors with 100 lags. Sample period is January 3, 2005 to May 31, 2023. $^{*}p < 0.1$, $^{**}p < 0.05$, *** $p < 0.01$.

	2-year (bps)	5-year (bps)	10-year (bps)
Specialness, one-week MA (bps)	0.0158***	-0.000671	0.0253
	(2.81)	(-0.17)	(1.51)
Specialness one-week MA (bps) $\times \mathbf{1}_{\text{within one month of original issuance}}$			-0.0209**
			(-2.01)
FEs	✓	✓	✓
R2 Adj	0.025	0.003	0.014
Observations	4,473	4,473	4,467

B.4 lliquidity measures summary statistics

Table B.4 reports summary statistics of the illiquidity measures used throughout the paper.

Table B.3: Specialness-adjusted on-the-run premium summary statistics. Statistics of the sample distribution of the on-the-run premium, fitted value from the specialness regression, and the residual plus the sample mean for 2-, 5-, and 10-year benchmark securities, in basis points. Sample period is January 3, 2005 to May 31, 2023. The residuals are mean zero.

		Mean	Std. Dev.	p5	p25	p50	p75	p95
2-Year	OTR Premium	0.99	3.60	-3.18	-0.31	1.09	2.67	5.62
	Fitted Value	0.99	0.59	0.44	0.64	0.90	1.06	2.04
	Residual plus sample mean	0.99	3.55	-3.10	-0.19	1.12	2.59	5.40
5-Year	OTR Premium	1.68	2.71	-2.60	0.15	1.72	3.18	5.60
	Fitted Value	1.68	0.17	1.29	1.51	1.75	1.76	1.85
	Residual plus sample mean	1.68	2.70	-2.58	0.15	1.70	3.18	5.59
10-Year	OTR Premium	7.66	9.78	-0.78	2.08	5.32	8.80	27.11
	Fitted Value	7.65	1.29	6.02	6.75	7.61	8.32	9.52
	Residual plus sample mean	7.66	9.70	-0.65	2.19	5.32	8.53	27.36

Table B.4: Summary statistics of illiquidity measures. This table reports the summary statistics of illiquidity measures for the 2-, 5-, and 10-year maturity sectors. Price impact is in basis points (in return space) per $100 million net trading volume adjusted for inflation, bid-ask spread is in basis points (in return space), negative log depth where depth is in billions of US dollars par adjusted for inflation, yield curve RMSE is in basis points (in yield space), price dispersion is in basis points (in yield space), and the specialness-adjusted OTR premium in basis points (in yield space). Bid-ask spread and negative log depth for the 2-year sector have been adjusted for the tick size change, and the OTR premium has been adjusted for the auction cycle and repo specialness as described in Section B.3. The sample period is July 10, 2017 to December 30, 2022.

Measure	Maturity	Mean	SD	p5	p25	p50	p75	p95
Price impact	2y	0.26	0.24	0.07	0.12	0.18	0.31	0.76
	5y	0.64	0.48	0.23	0.32	0.46	0.80	1.68
	10y	1.77	1.34	0.62	0.86	1.24	2.18	4.59
Bid-ask spread	2y	0.80	0.03	0.77	0.78	0.79	0.81	0.85
	5y	0.84	0.04	0.80	0.81	0.83	0.85	0.92
	10y	1.70	0.s12	1.59	1.63	1.67	1.73	1.90
Negative log depth	2y	0.78	1.01	-0.38	0.03	0.44	1.29	2.79
	5y	2.37	0.70	1.58	1.81	2.21	2.64	3.79
	10y	2.50	0.56	1.81	2.07	2.38	2.80	3.55
YC RMSE	2y	2.25	1.21	0.96	1.40	1.89	2.79	4.72
	5y	1.25	0.69	0.57	0.79	1.04	1.48	2.61
	10y	1.39	1.04	0.55	0.78	1.03	1.55	3.95
Price dispersion	2y	1.41	0.99	0.36	0.67	1.11	1.84	3.57
	5y	0.60	0.38	0.19	0.33	0.52	0.75	1.28
	10y	0.46	0.44	0.11	0.19	0.32	0.57	1.21
Specialness adjusted	2y	1.87	2.27	-0.40	0.48	1.28	2.54	6.38
OTR premium	5y	2.23	1.97	0.05	0.97	1.95	2.97	5.62
	10y	3.18	3.59	-0.82	0.67	2.05	5.40	10.20

C Additional details about the capacity measures

C.1 Risk adjustment

Table C.5 summarizes the dealer-level positions and flows underlying the capacity utilization measures. The table reports the gross and net measures for the FR 2004 positions and TRACE D2C volume data for Treasuries, agency MBS, and corporate bonds at the dealer-day level in billions of dollars of par. Table C.6 reports the DV01 and swaption-implied yield volatility used to risk adjust each security type and maturity bucket. Table C.7 summarizes the risk-adjusted measures expressed in millions of dollars of 95% monthly VaR for each dealer-day. For each measure Q_{it} for dealer i on day t, we compute the risk-adjusted measure as,

$$Q_{it}^{adj} = Q_{it} \cdot DV01_{it} \cdot Vol_{it} \cdot \sqrt{\frac{1}{12}} \cdot 1.96,$$

where Q_{it} is expressed in millions of par value, $DV01_{it}$ is the dollar value of a basis point for one million of par value, and Vol_{it} is swaption implied volatility expressed in basis points per annum.[25]

Table C.5: Summary statistics of dealer-level, primary dealer positions and D2C flows for Treasuries, MBS and corporate bonds. Summary statistics for the gross and net measures using FR 2004 positions and TRACE D2C volume data for Treasuries, agency MBS, and corporate bonds at the dealer-day level in billions of dollars of par. The sample is primary dealer, dealer-day observations from July 10, 2017 to December 30, 2022.

	Mean	Std. Dev.	p1	p5	p25	p50	p75	p95	p99
Gross Treasury	24.8	25.6	1.4	2.5	7.2	15.4	34.5	75.5	122.9
Net Treasury	7.0	12.0	-2.1	-0.9	0.7	2.7	7.4	28.9	63.2
Gross MBS	31.6	36.0	0.0	0.0	2.4	11.9	62.9	95.8	133.1
Net MBS	3.0	4.1	-0.5	0.0	0.0	1.3	4.5	11.2	18.3
Gross Corporate	3.3	4.0	0.0	0.0	0.2	1.2	6.2	10.9	14.9
Net Corporate	0.4	0.9	-1.2	-0.3	0.0	0.1	0.6	2.3	3.3
Gross D2C Treasury Volume	8.6	8.8	0.0	0.1	2.3	5.0	12.8	27.0	37.0
Net D2C Treasury Volume	-0.9	1.3	-5.1	-3.4	-1.4	-0.6	-0.1	0.4	1.4
Gross D2C MBS Volume	6.4	10.7	0.0	0.0	0.0	1.0	8.8	28.7	48.4
Net D2C MBS Volume	0.1	0.5	-1.4	-0.6	-0.0	0.0	0.1	0.8	1.8
Gross D2C Corporate Volume	1.0	1.4	0.0	0.0	0.1	0.5	1.5	3.6	6.4
Net D2C Corporate Volume	-0.2	0.8	-3.9	-1.1	-0.0	-0.0	0.0	0.1	0.3

[25] For a 3-month Treasury bill we use $\$25 = 1\text{mn} \times 1\text{bp} \times 3/12$ as the DV01. For the on-the-run benchmarks we compute the DV01 each day from the change in price associated with a parallel shift up or down in the yield curve by 10 basis points, e.g. $DV01 = 1\text{mn} \cdot (P(r_t + dr) - P(r_t - dr))/(2 \cdot dr) \cdot 1\text{bp}$ where $P(\cdot)$ is the price of the bond, equal to \$1 at maturity, as a function of the zero-coupon discount curve r_t from Filipović, Pelger, and Ye (2022). The average DV01 amounts are \$194, \$290, \$476, \$652, \$892, \$2051 for 2-, 3-, 5-, 7-, 10-, and 30-year on-the-run Treasuries during our sample period.

Table C.6: Risk adjustment of dealer positions. Reports the DV01 measure and swaption implied volatility for the risk adjustment of each security type and maturity bucket.

Security type i	$DV01_{it}$	Vol_{it}	Security type i	$DV01_{it}$	Vol_{it}
Coupons (0,2]	2-year OTR	2-year 1-month	TIPS (0,2]	2-year OTR	2-year 1-month
Coupons (2,3]	3-year OTR	3-year 1-month	TIPS (2,6]	5-year OTR	5-year 1-month
Coupons (3,6]	5-year OTR	5-year 1-month	TIPS (6,11]	10-year OTR	10-year 1-month
Coupons (6,7]	7-year OTR	7-year 1-month	TIPS (11,30]	30-year OTR	30-year 1-month
Coupons (7,11]	10-year OTR	10-year 1-month	Bills	3-month bill	2-year 1-month
Coupons (11,30]	30-year OTR	30-year 1-month	FRNs	0	2-year 1-month
Agency MBS	7-year OTR	10-year 3-month			
Corporates	5-year OTR	5-year 1-month			

Table C.7: Summary statistics of dealer-level, primary dealer-level risk-adjusted positions and D2C flows for Treasuries, agency MBS and corporate bonds. Summary statistics for the gross and net measures using FR 2004 positions and TRACE D2C volume data for Treasuries, agency MBS, and corporate securities at the dealer-day level expressed in millions of dollars of monthly 95% VaR. The sample is primary dealer, dealer-day observations from July 10, 2017 to December 30, 2022.

	Mean	Std. Dev.	p1	p5	p25	p50	p75	p95	p99
Gross Treasury Adj.	680.8	854.4	16.0	43.2	139.7	310.8	929.3	2578.0	3999.8
Net Treasury Adj.	208.6	436.1	-68.7	-30.7	-0.8	42.4	212.8	1045.5	2374.3
Gross MBS Adj.	883.4	1090.6	0.0	0.0	56.4	320.6	1541.9	2939.7	4278.5
Net MBS Adj.	83.9	118.4	-12.1	0.0	0.7	34.7	121.3	350.4	507.9
Risk-Ad Gross Corporate Adj.	64.9	95.4	0.0	0.0	2.5	19.2	99.4	255.8	430.0
Net Corporate Adj.	7.3	15.9	-27.7	-6.4	0.1	2.1	10.8	39.2	60.6
Gross D2C Treasury Volume Adj.	161.7	220.2	0.0	2.6	23.7	83.1	211.5	596.3	1012.4
Net D2C Treasury Volume Adj.	-3.3	20.8	-66.9	-32.5	-8.1	-1.3	2.4	21.4	53.8
Gross D2C MBS Volume Adj.	176.2	298.8	0.0	0.0	0.3	32.0	237.2	790.9	1354.7
Net D2C MBS Volume Adj.	1.5	10.1	-23.3	-9.1	-0.3	0.0	2.4	15.7	36.1
Gross D2C Corporate Volume Adj.	18.7	26.7	0.0	0.0	1.8	8.1	25.3	72.3	126.3
Net D2C Corporate Volume Adj.	-0.0	0.0	-0.1	-0.0	-0.0	-0.0	0.0	0.0	0.0

C.2 Dealer-level distribution of capacity utilization

Figure C.2 illustrates the cross-sectional distribution of the four capacity utilization measures, weighted by the within-dealer maximum of the position or flow measure. This figure shows that, while many dealers reached maximum capacity utilization during March 2020, several dealers were still far below their historical maximums. Throughout 2022, amidst market volatility and monetary policy uncertainty, several dealers approached capacity utilization levels close to those observed in March 2020. By the end of 2022, there was considerable heterogeneity in capacity utilization across dealers, with the top quartile of dealers utilizing 80% of their capacities and the bottom quartile utilizing just 30%, in terms of gross positions. Table C.8 summarizes the dealer-level capacity utilization measures for all days and the sample as well as the peak days on which the respective mean-across-dealers capacity utilization measure reached its maximum.

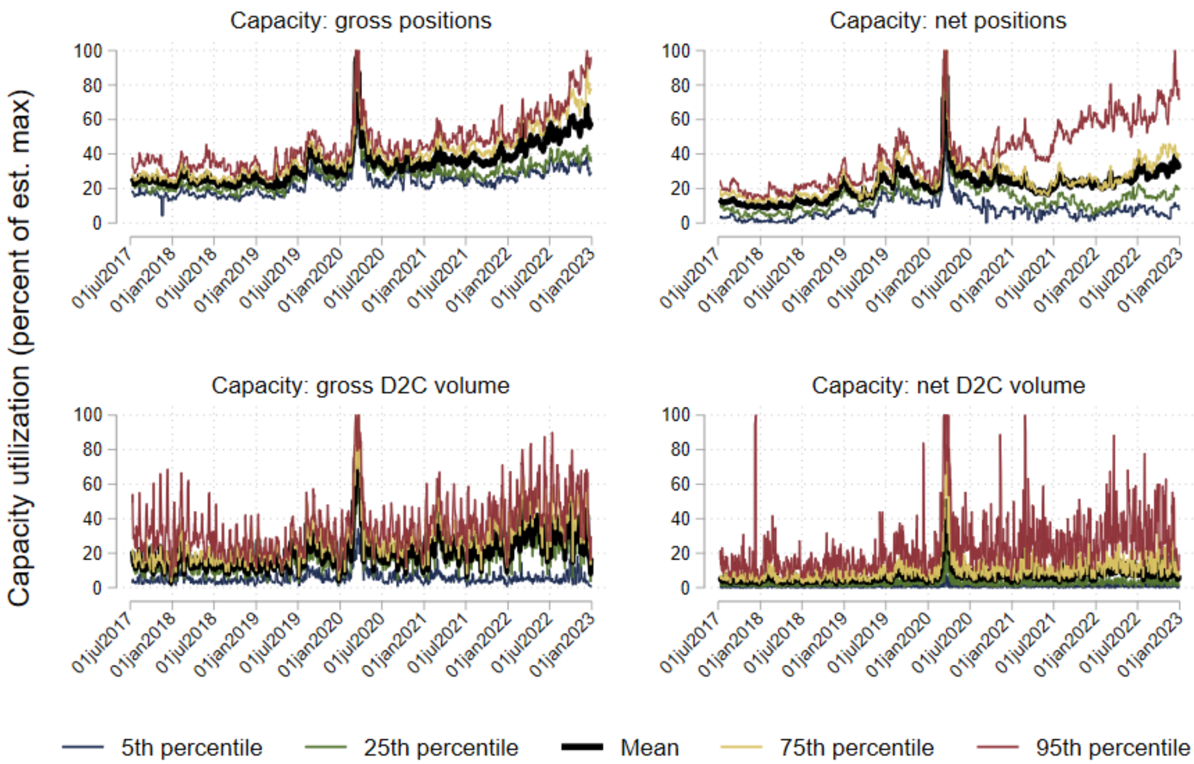


Figure C.2. Cross-sectional distribution of capacity utilization. The average and 5th, 25th, 75th, and 95th percentiles of dealer-level capacity utilization based on gross positions, net positions, gross D2C volume, and net D2C volume. Percentiles and averages are weighted according to the within-dealer maximum of the respective risk-adjusted position or flow measures.

Table C.8: Summary statistics of the cross-sectional distribution of capacity utilization. The 25th, 75th, 95th percentiles and the mean on peak mean-dealer capacity utilization days and on average across all days. The peak days are defined as the maximums of each of the respective mean-across-dealers capacity utilization measure. For gross positions the peak day is March 12, 2020, for D2C net volume the peak day is March 19, 2020, and for net positions and D2C gross volume the peak day is March 23, 2020. Percentiles and averages are weighted according to the within-dealer maximum of the respective risk-adjusted position or flow measure.

	All p25	All mean	All p75	All p95	Peak p25	Peak mean	Peak p75	Peak p95
D2C gross volume	17.94	22.58	28.26	36.64	76.10	83.83	100.00	100.00
D2C net volume	2.74	7.84	10.64	23.35	26.49	50.34	82.65	100.00
Gross positions	28.12	34.88	40.69	46.34	95.19	96.05	100.00	100.00
Net positions	14.33	21.78	26.00	39.95	82.68	85.04	100.00	100.00

C.3 Seasonality adjustment of flow-based capacity utilization

We adjust the flow-based capacity utilization measures for monthly seasonality as follows:

$$M_t \;=\; \alpha_t + \sum_{i=1}^{31} \beta_i D_i + \epsilon_t, \tag{7}$$

where D_i equals 1 on the ith day of the month and 0 otherwise. Figure C.3 reports the results of Equation (7) for gross D2C volume and net D2C volume capacity measures. To compute the seasonally adjusted capacity utilization measure, we then subtract the estimated day-of-month effects: $\hat{M}_t = M_t - \sum_{i=1}^{31} \hat{\beta}_i D_i$. Figure C.4 plots the adjusted and unadjusted flow-based capacity utilization measures.

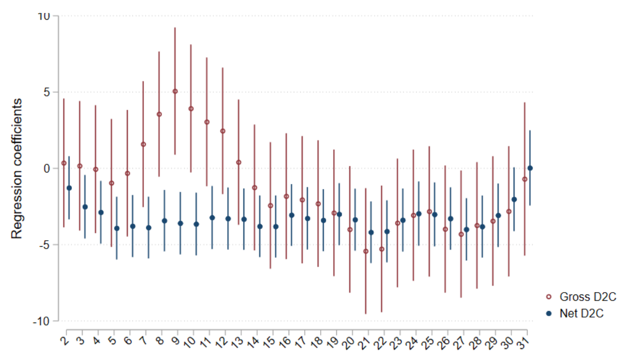


Figure C.3. Seasonal adjustment Point estimates and 95% confidence bands from regression of flow-based capacity utilization measures onto day-of-month dummies relative to the first day of the month which is captured by the constant term. For gross dealer-to-customer volumes, the regression constant is 23.6 and $R^2 = 0.073$. For net dealer-to-customer volumes, the regression constant is 10.9 and $R^2 = 0.036$.

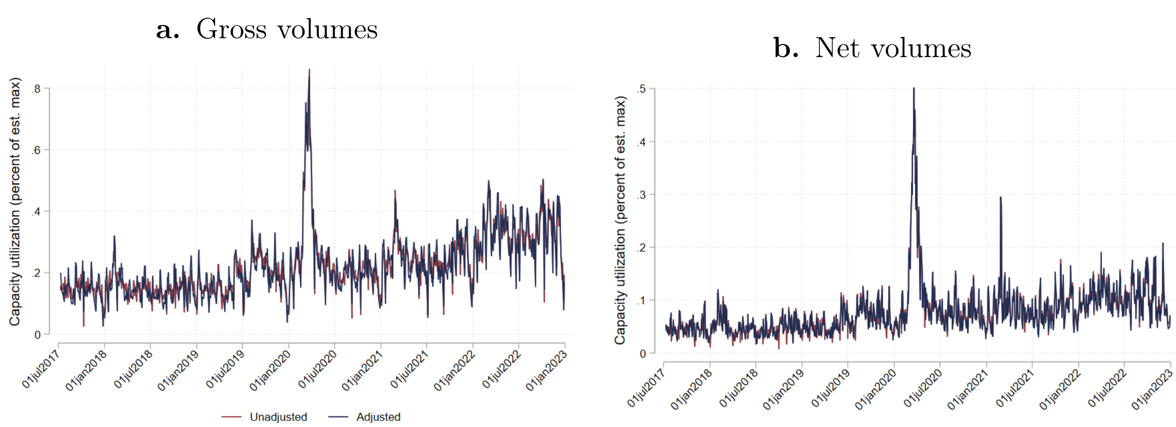


Figure C.4. Seasonal adjustment of capacity utilization based on D2C volumes measure. Unadjusted and seasonally adjusted gross dealer-to-customer volume-based capacity utilization.

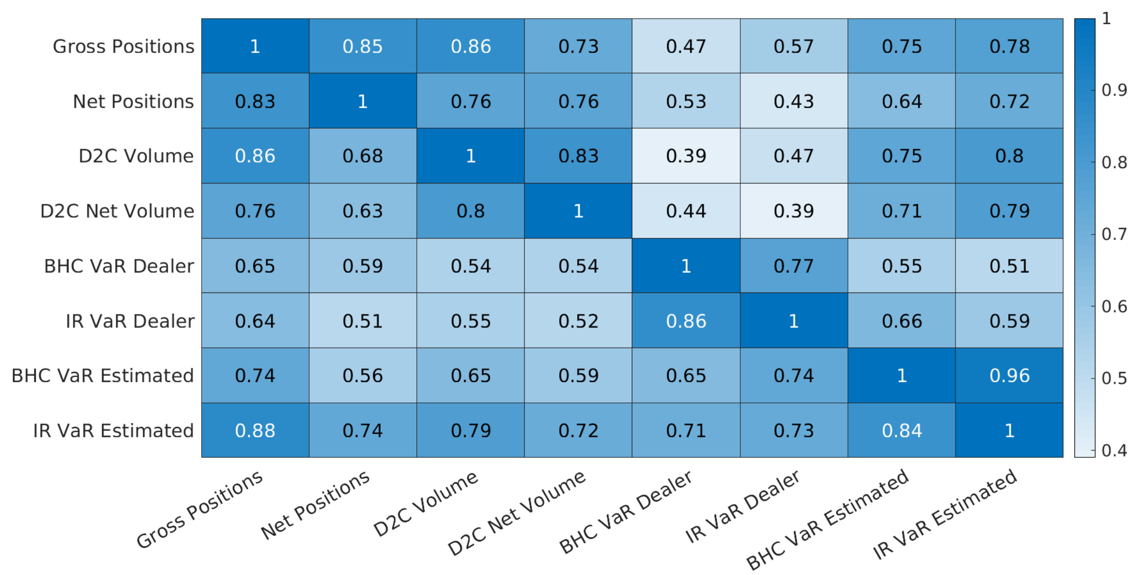
C.4 Capacity utilization measures summary statistics

Table C.9 reports summary statistics of the capacity utilization measures. Table C.10 reports the correlation matrix of the capacity utilization measures. The capacity utilization measures broadly exhibit strong positive correlations despite their different specifications.

Table C.9: Summary statistics of capacity utilization measures. Capacity utilization measures are in percent of estimated maximums. The sample period is July 10, 2017 to December 30, 2022.

Measure	Source	Mean	SD	p5	p25	p50	p75	p95
Gross Positions	FR2004	34.8	11.2	21.6	25.4	33.7	39.6	57.0
Net Positions	FR2004	21.8	9.0	9.9	14.7	22.4	26.4	33.2
D2C Volume	TRACE	22.5	9.8	11.6	15.4	20.5	27.1	39.9
D2C Net Volume	TRACE	7.8	4.8	3.1	4.9	6.9	9.4	14.6
Dealer BHC VaR Reported	Market Risk Rule	20.7	9.1	12.5	13.5	18.0	24.4	38.9
Dealer IR VaR Reported	Market Risk Rule	12.7	4.1	8.5	9.6	11.2	15.6	18.7
Dealer BHC VaR Estimated	Market Risk Rule	18.7	11.2	10.2	11.6	13.6	22.9	41.3
Dealer IR VaR Estimated	Market Risk Rule	12.6	9.5	5.8	6.8	8.5	16.0	28.2

Table C.10: Correlation of capacity utilization measures. Correlations of the capacity utilization measures from July 2017 to December 2022. The upper diagonal is the Pearson-linear correlation. The lower diagonal is the Spearman-rank correlation.



D Full-horizon analysis

In this section we present the results of our full-horizon analysis, as described at the end of Section 7.

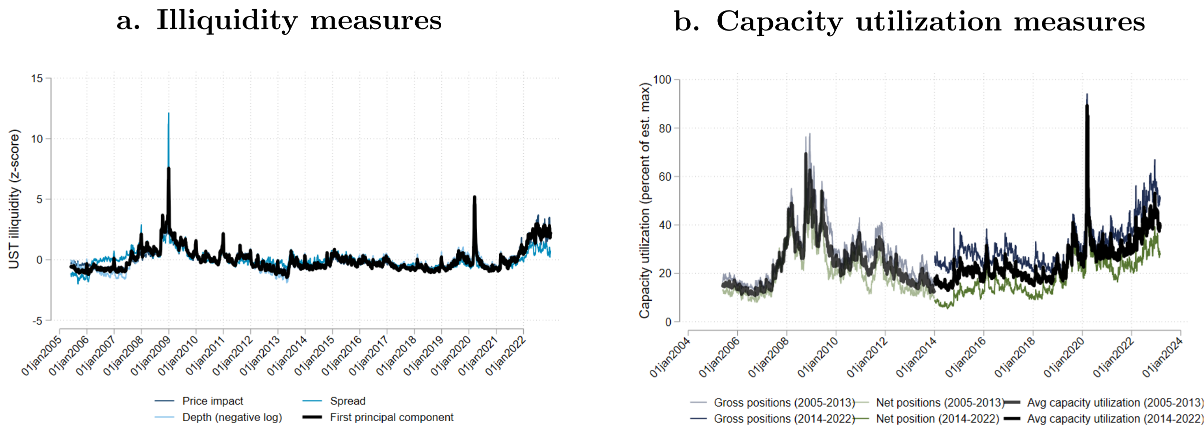


Figure D.5. Full horizon on-the-run US Treasury market illiquidity and composite capacity utilization. Panel (a) shows the the z-scores of three measures of interdealer on-the-run illiquidity – price impact, bid-ask spread, and depth – averaged across the 2-, 5-, and 10-year maturity sectors. The first principal component of the nine z-scores is plotted in bold. All variables are shown, for clarity, as five-day moving averages. Panel (b) plots the time-series of dealer capacity utilization measures based on risk-adjusted gross positions and net positions. Average capacity utilization is the simple average of the two measures. Capacity utilization is estimated separately for sub-periods: 2005-2013 and 2014-2022.

Table D.11: Quantile regressions for 99th percentile of US Treasury illiquidity. Quantile regressions for the 99th percentile of US Treasury illiquidity predicted by volatility and capacity utilization. Columns (1)-(3) covers the period from May 25, 2005 through December 30, 2013, and columns (4)-(6) covers the period from January 2, 2014 through December 30, 2022. The dependent variable is the first principal component of the illiquidity measures for the on-the-run 2-, 5-, and 10-year notes. For ease of interpretation all variables are standardized in the regressions. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}\ p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
Yield volatility	1.39***	1.23***	1.17***	1.63***	1.29***	1.30***
	(0.40)	(0.30)	(0.35)	(0.31)	(0.12)	(0.12)
Capacity utilization residual: gross position		0.68***			0.34***	
		(0.23)			(0.08)	
Capacity utilization residual: net position			0.66***			0.32***
			(0.23)			(0.09)
Constant	1.67***	1.59***	1.54***	1.68***	1.14***	1.19***
	(0.32)	(0.29)	(0.33)	(0.34)	(0.08)	(0.09)
N	2135	2135	2135	2249	2249	2249
Pseudo R2	0.44	0.50	0.50	0.54	0.69	0.69

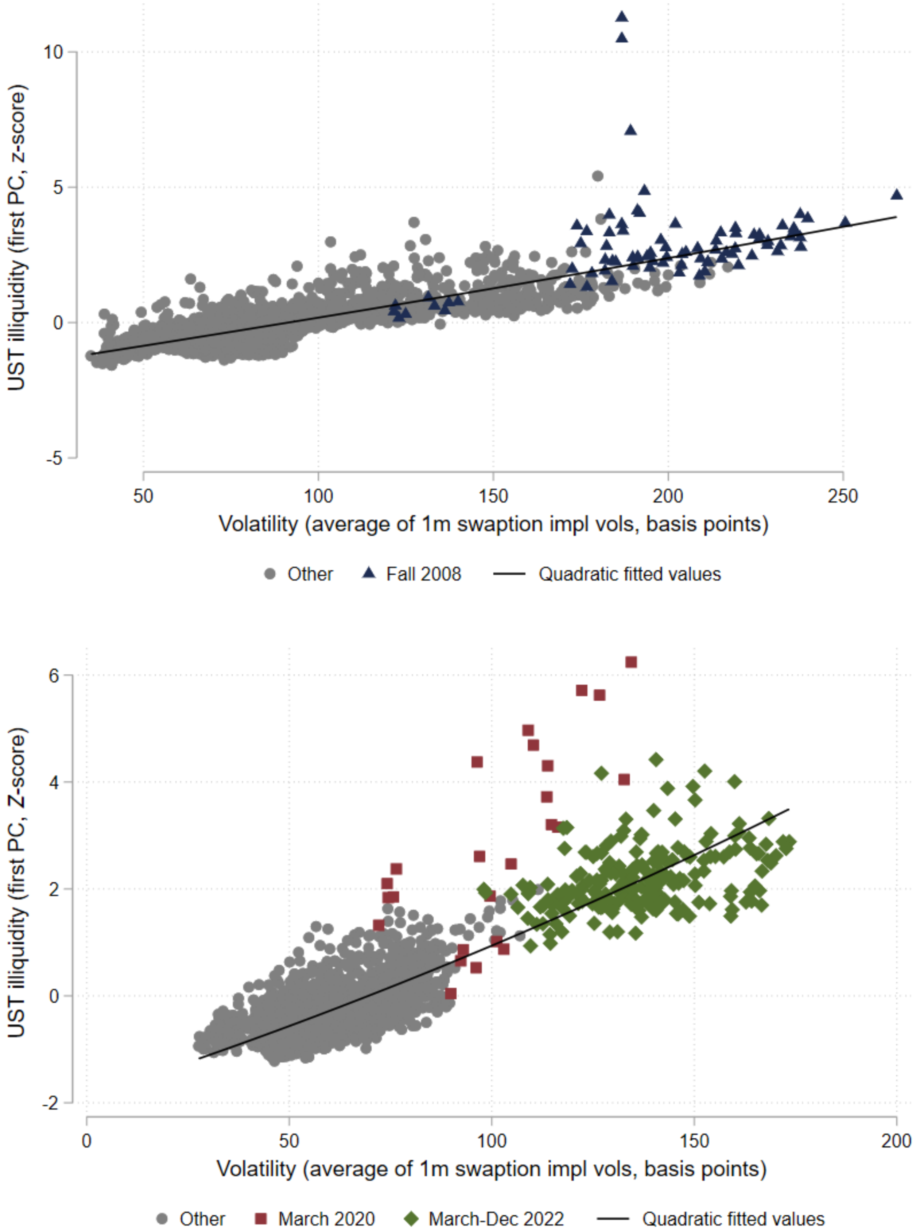


Figure D.6. Full-horizon prediction of illiquidity given yield volatility, 2005-2013 and 2014-2022. A scatter plot and estimated relationship between the principal-component composite measures of Treasury market illiquidity and a composite measure of implied volatility, as measured by the average of the standard deviations of benchmark swap rates, in basis points, implied by swaptions on 2-, 5-, and 10-year swaps with one month expirations. The top panel plots the scatter and estimated relationship for May 25, 2005 through December 30, 2013. The plotted OLS fit, for 2135 observations, is the second-order polynomial $y = -1.86 + 0.020x + 0.000008x^2$, where volatility x is in basis points, with $R^2 = 67.2\%$. All constant and linear term coefficient estimates have p-values of less than 1% under standard assumptions. The bottom panel plots the scatter plot and estimated relationship for January 2, 2014 through December 30, 2022. The plotted OLS fit, for 2249 observations, is the second-order polynomial $y = -1.88 + 0.024x + 0.000038x^2$, where volatility x is in basis points, with $R^2 = 75.4\%$ All three coefficient estimates have p-values of less than 1% under standard assumptions.

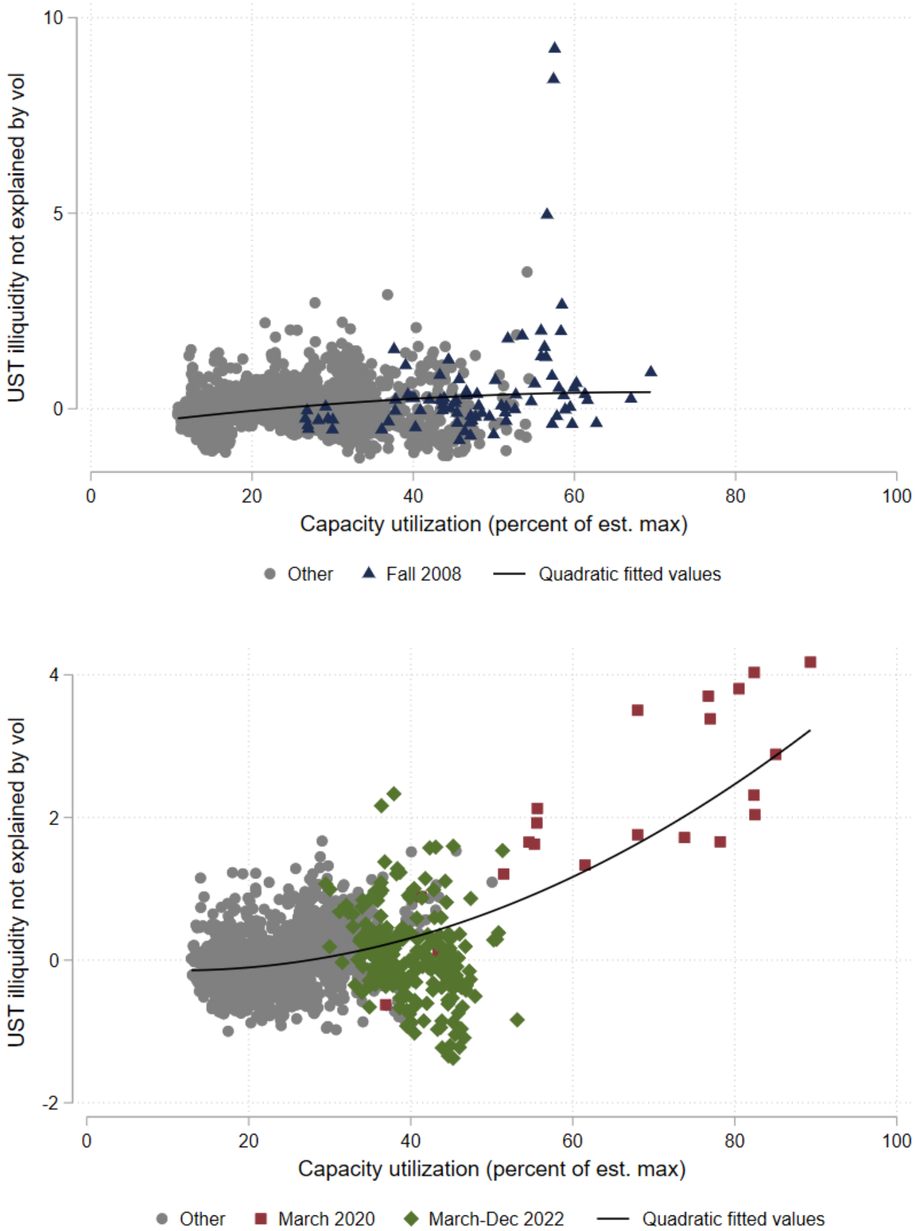


Figure D.7. Full-horizon composite capacity utilization, 2005-2013 and 2014-2022. The relationship between the average dealer capacity utilization and the residual component of Treasury market illiquidity that remains after controlling for average swaption-implied volatility (the residuals associated with the fitted relationship in Figure D.6). The average capacity-utilization measure is based on dealer gross positions and dealer net positions. The top panel plots the scatter and estimated relationship for May 25, 2005 through December 30, 2013. The plotted OLS fit, for 2135 observations, is the second-order polynomial $y = -0.52 + 0.027x - 0.000196x^2$, with $R^2 = 7.2\%$. The constant and linear term coefficients have p-values of less than 1%, and the quadratic term coefficient has a p-value of less than 5% under standard assumptions. The bottom panel plots the scatter plot and estimated relationship for January 2, 2014 through December 30, 2022. The plotted OLS fit, for 2249 observations, is the second-order polynomial $y = -0.07 - 0.013x + 0.000554^2$, with $R^2 = 24.8\%$. All three coefficient estimates have p-values of less than 1% under standard assumptions.

E Estimating dealer VaR

We estimate dealer VaR from rolling quantile regressions. The advantage of this approach is that we can apply a transparent and common modeling approach to all dealers. The disadvantage is that we abstract from positions data. If a dealer takes on a large position that impacts the true (unobserved) VaR, in our estimation approach we would only see this position at a lag to the extent that it affects PnL. In particular, we estimate VaR from predictive quantile regressions of daily PnL onto the VIX index, MOVE index, and a measure of realized PnL volatility lagged by a day,

$$\mathrm{PnL}_t = \beta_0 + \beta_{\mathrm{VIX}}\mathrm{VIX}_{t-1} + \beta_{\mathrm{MOVE}}\,\mathrm{MOVE}_{t-1} + \beta_{\mathrm{RV}}\mathrm{RV}^{\mathrm{PnL}}_{t-1} + u_t,$$

where realized PnL volatility is an exponentially weighted moving average,

$$\mathrm{RV}^{\mathrm{PnL}}_t = \sqrt{\sum_{d=0}^{t} \omega_d\,\mathrm{PnL}^2_{t-d}}, \quad \text{with } \omega_d = \frac{e^{-\beta d}}{\sum_{d=0}^{t} e^{-\beta d}},$$

and $\beta = -\log(1/2)/10$ for a half-life of 10 trading days.[26] For comparison, we also run predictive quantile regressions of daily PnL onto dealer reported VaR,

$$\mathrm{PnL}_t = \beta_0 + \beta_{\mathrm{VaR}}\mathrm{VaR}_{t-1} + u_t,$$

which we call the reported VaR model.

Table E.12 compares the performance of the estimated and reported VaR models. We run the quantile regressions at the 1st percentile to estimate a one-day, 99th percentile VaR for each dealer. The table reports summary statistics across dealers for the regression coefficients and explanatory power. Comparing the models, we see that the simple model based on market implied volatility and dealer realized PnL volatility has much higher in-sample and out-of-sample pseudo-R^2 compared to the model using the reported VaR as an explanatory variable in the quantile regressions. For the estimated model the in-sample pseudo-R^2 is 38% on average with a 75th percentile of 49%. For the model based on dealer reported VaR the in-sample pseudo-R^2 is 9% on average with a 75th percentile of 15%. The expanding window out-of-sample results are similar, illustrating a robustness for the findings despite zooming in on the tail of the distribution.[27] In addition to its higher explanatory power, the slope coefficients on the VIX, MOVE, and PnL volatility have the expected negative sign on average across dealers in the estimated model.[28] This means that VaR

[26] Similar results were found using GARCH (1,1) and EGARCH(1,1,1) models.

[27] The out-of-sample pseudo-R^2 is computed following Giglio, Kelly, and Pruitt (2016) using an expanding window at a daily frequency with a one-year burn in period. For the reported VaR model we also report the null pseudo-R^2 (R^2_{H0}) imposing the null hypothesis that $\beta_0 = 0$ and $\beta_{VaR} = 1$. The explanatory power imposing the null (R^2_{H0}) is similar to that of the expanding window out-of-sample estimates (R^2_{OOS}), with slightly lower performance on average but higher performance for the median dealer.

[28] Since dealers have different levels of PnL volatility, it is not straightforward to compare the magnitudes of the slope coefficients β_{VIX}, β_{MOVE}, and β_{RV} in the estimated model across firms. To account for this, the table reports summary statistics for the t-statistics for the slope coefficients to indicate the signs of the factors and to show which factors are significant. For the dealer reported VaR there the null hypothesis that

becomes larger in magnitude and more negative when market implied volatility increases and when lagged PnL volatility increases. The most significant factor as measured by the average t-statistics is dealer PnL volatility (-2.94) which is followed by the VIX (-1.70) and the MOVE (-0.48).[29]

Table E.12: Dealer VaR quantile regression summary statistics. Summary statistics across dealers from the dealer-level VaR quantile regressions. The dependent variable is the daily profit and loss (PnL). We run the quantile regressions at the 1st percentile to estimate a one-day, 99th percentile VaR. The estimated VaR model includes the VIX, MOVE, and PnL realized variance lagged by a day as explanatory factors. The reported VaR model includes dealer reported VaR as the sole explanatory factor. R^2_{IS} is the in-sample pseudo-R^2. R^2_{OOS} is the daily expanding window pseudo-R^2 with a one-year burn-in period. R^2_{H0} is the pseudo-R^2 imposing the null that $\beta_0 = 0$ and $\beta_{VaR} = 1$ in the reported VaR model with a one-year burn-in period for comparability to R^2_{OOS}. Panel A reports results at the BHC level. Panel B reports results for desks with exposure to interest rate risk. The sample period is July 2017 to December 2022.

Panel A: Bank holding company VaR

Estimated	Mean	SD	p25	p50	p75	Reported	Mean	SD	p25	p50	p75
R^2_{IS}	0.38	0.15	0.32	0.38	0.49	R^2_{IS}	0.09	0.09	0.02	0.07	0.15
R^2_{OOS}	0.32	0.19	0.22	0.34	0.48	R^2_{OOS}	0.07	0.10	0.00	0.06	0.14
$t[\beta_0]$	0.77	1.30	0.57	1.07	1.52	R^2_{H0}	0.06	0.18	-0.05	0.11	0.19
$t[\beta_{VIX}]$	-1.70	0.93	-2.32	-1.57	-0.99	β_{VaR}	0.49	0.62	-0.08	0.65	0.85
$t[\beta_{MOVE}]$	-0.48	1.12	-0.90	-0.48	0.06	$SE[\beta_{VaR}]$	0.37	0.19	0.23	0.36	0.45
$t[\beta_{RV}]$	-2.94	2.15	-4.60	-3.29	-1.19	$t[\beta_{VaR}]$	1.61	2.06	-0.16	1.71	2.89

Panel B: Interest rate desks VaR

Estimated	Mean	SD	p25	p50	p75	Reported	Mean	SD	p25	p50	p75
R^2_{IS}	0.32	0.11	0.29	0.33	0.39	R^2_{IS}	0.10	0.08	0.03	0.10	0.18
R^2_{OOS}	0.24	0.16	0.22	0.28	0.32	R^2_{OOS}	0.08	0.09	0.01	0.10	0.16
$t[\beta_0]$	0.39	1.53	-0.19	0.58	1.43	R^2_{H0}	-0.15	0.84	-0.07	0.06	0.19
$t[\beta_{VIX}]$	-1.42	0.98	-1.70	-1.08	-0.83	β_{VaR}	0.41	0.61	0.16	0.44	0.73
$t[\beta_{MOVE}]$	-0.18	1.09	-0.65	-0.20	0.47	$SE[\beta_{VaR}]$	0.34	0.24	0.19	0.24	0.48
$t[\beta_{RV}]$	-3.73	1.84	-5.10	-4.51	-2.18	$t[\beta_{VaR}]$	1.41	1.97	0.71	2.10	2.80

For the reported model the point estimate β_{VaR} is below 1 on average and is close to significant at the 10% level on average. However, the 25th and 75th percentiles reveal that there is heterogeneity across dealers in the significance of the reported VaR. For some dealers the reported VaR results in a point estimate that is close to 0 that is insignificant. For other dealers the reported VaR results in a point estimate that is close to 1 that is strongly significant, making it difficult to reject the null hypothesis of an unbiased model. The heterogeneity in the performance of the reported VaR estimates further motivates our approach using a common methodology for all dealers. The results are broadly similar for

$\beta_{VaR} = 1$ makes it more straightforward to compare the point estimates across firms, so the table reports summary statistics for the point estimate, standard error, and t-statistic.

[29] We measure statistical significance using a semiparametric bootstrap procedure following Efron and Tibshirani (1994). In the semiparametric bootstrap, we model the explanatory variables x_t with a first order vector autoregression (VAR) and the quantile regression residuals $\hat{u}_t = y_t - \hat{\beta}' x_t$ with a first-order autoregression AR(1). To obtain a bootstrap sample we resample from the VAR and AR(1) errors with a moving blocks bootstrap to generate (x^*_t, u^*_t) and then obtain $y^*_t = \hat{\beta}' x^*_t + u^*_t$. The t-statistic is computed as the quantile regression point estimate divided by the bootstrapped standard error.

VaR from desks with exposure to interest rate risk where we again observe the expected sign of the coefficients on market implied volatility and dealer realized PnL volatility along with higher explanatory power for the estimated model in comparison to the dealer reported VaR.

Figure E.8 reports the time-series of the capacity utilization measures based on reported and estimated VaR at the BHC level and for desks with exposure to IR risk. Each of the measures spikes in March 2020 after a more quiescent period at the start of the sample. The measures based on estimated VaR decline relatively quickly after the COVID-19 shock and then increase more significantly again in 2022. In contrast, the measures based on dealer reported VaR decline somewhat initially but then remain elevated for an entire year until early 2021. This pattern might be explained by some dealers reporting realized VaR estimates based on a rolling one-year window, which is also consistent with the lower explanatory power of the reported measures in the quantile regressions. Capacity utilization based on the reported measures then increases in 2022, but to a lesser extent in comparison to the measures based on the estimated VaR. Finally, we can observe some spikes in the reported measures that may reflect the presence of large positions that dealers are temporarily warehousing, which is a feature of the reported measures that our estimated measures will not capture. Despite these differences, Table C.10 shows that all of the capacity utilization measures are positively correlated with each other, including the capacity utilization measures based on FR 2004 positions and TRACE volumes.

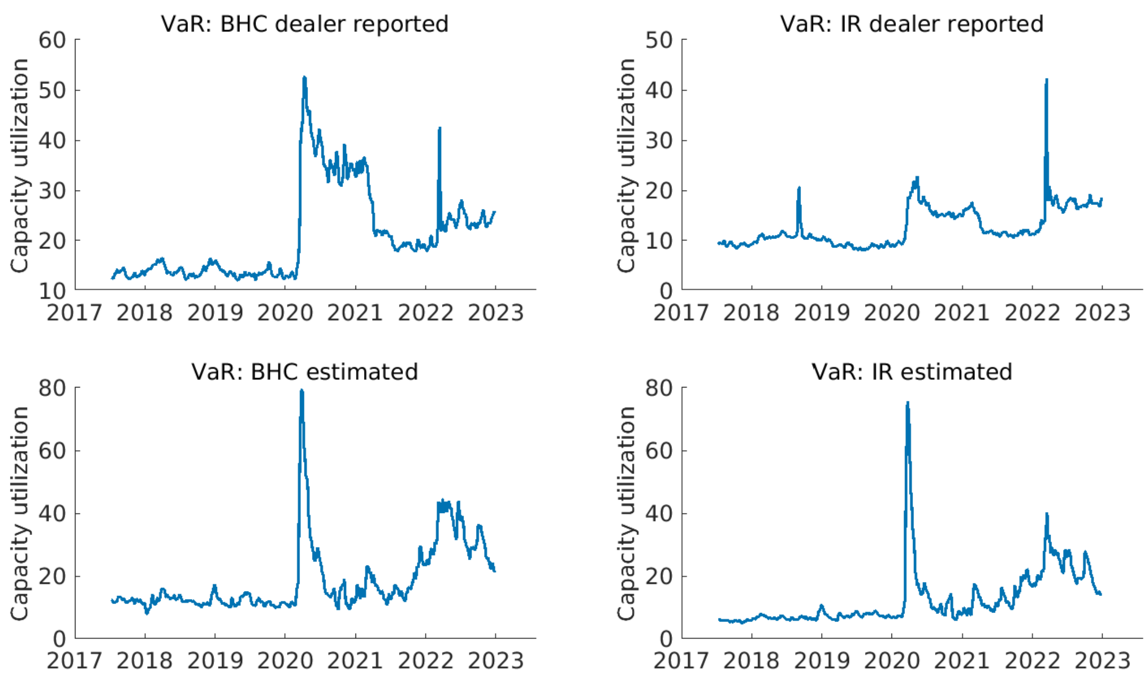


Figure E.8. Capacity utilization measured with dealer reported and estimated VaR. Plots of capacity utilization as measured by dealer reported VaR and estimated VaR models at the BHC level and for desks with exposure to interest-rate risk (IR) as a five-day moving average. The estimated models reflect the in-sample fit for the first year of data and then an expanding-window estimate that is updated daily.

F Additional details about the Data

F.1 Treasury TRACE

FINRA member firms started reporting transactions in US Treasury securities via the Trade Reporting and Compliance Engine (TRACE) on July 10, 2017. US Treasury securities include all marketable Treasuries, including bills, notes, bonds, TIPS, and FRNs. All secondary market transactions are reported, except for purchases of a US Treasury security from the Treasury Department as part of an auction; and, repurchase and reverse repurchase transactions. When-issued transactions, which can take place after the Treasury Department's announcement of an auction but before the auction and issuance of the securities, are reportable. Reopening transactions in a US Treasury security that is the subject of an auction are treated as "when-issued transactions."

Other than adjusting trades that were corrected or cancelled after they were reported, we apply the following conditions: (1) keep trades during New York trading hours (7:30 am – 5:00 pm) with standard settlement and at least 365 days-to-maturity; (2) remove remove trades that are reported in yield terms, or whose entered price is below 25 or above 250, to avoid influence of price outliers; (3) remove trades that are flagged with a special modifier code for being part of a series of transactions where one or more transactions are executed at a fixed and pre-determined price or that are part of a series of transactions involving a futures contract, e.g., Treasury-futures basis trades; (4) exclude trades at off-market prices that are flagged as being part of a package trade for price-based calculations, but collapse their volume across "matched" customer trades with the same reporting firm, security, timestamp, price, trade size, and trade direction for volume-based measures. For the volume and flow measures we include all trade sizes. When calculating RMSE and price dispersion measures, however, we restrict the sample to trades with a par-value of at least $1 million dollars. D2C illiquidity measures are calculated based on D2C transactions data, excluding interdealer trades and non-member affiliate trades.[30]

Identifying primary dealers in Treasury TRACE Treasury TRACE reporting is required by registered broker-dealers that are FINRA members and, since September 2022, it is also required by depository institutions. The level of reporting is at the Market Participant Identifier (MPID). Depending on a firm's organizational structure, a single firm could report under multiple MPIDs. Therefore, we aggregate multiple MPIDs based on their Central Registration Depository (CRD) number, which links same-firm MPIDs, and then use manual mapping based on firms' names.

To allow some comparison between utilization measures that are based on FR 2004 form and utilization measures that are based on TRACE data, we focus on flows and volumes of MPIDs that are linked to primary dealer entities. The legal entities under the "primary dealer" entity do not necessarily align with the TRACE MPIDs. Hence, other than the aggregation above, when handling primary dealers' mergers, we sum TRACE volumes of

[30]Certain type of affiliate trades, usually with same price that are executed at or around the same time, are not economically distinct and, as such, do not provide useful information for pricing. For example, transfer of a trade across affiliates for book-keeping purposes. Other types of affiliate trades are at arms-length.

the later-to-be-merged entities throughout the sample, except for Santander.[31] Overall, we match 26 primary dealers to 291 MPIDs. Finally, we omit primary dealers with little to no coverage in TRACE, resulting in a sample of 23 primary dealers, listed in the table below.

Table F.13: The sample of primary dealers

ASL Capital Markets Inc.	Jefferies LLC
Bank of Montreal, Chicago Branch / BMO	J.P. Morgan Securities LLC
BNP Paribas Securities Corp.	Mizuho Securities USA LLC
Barclays Capital Inc.	Morgan Stanley & Co. LLC
BofA Securities, Inc. (includes Merrill Lynch)	NatWest Markets Securities Inc. / RBS
Cantor Fitzgerald & Co.	Nomura Securities International, Inc.
Citigroup Global Markets Inc.	RBC Capital Markets, LLC
Credit Suisse	Santander US Capital Markets LLC / Amherst Pierpoint
Daiwa Capital Markets America Inc.	TD Securities (USA) LLC
Deutsche Bank Securities Inc.	UBS Securities LLC.
Goldman Sachs & Co. LLC	Wells Fargo Securities, LLC
HSBC Securities (USA) Inc.	

F.2 FR 2004 granularity

The FR 2004 form is periodically revised, typically to increase the granularity of reporting. One such change impacting our calculations concerns the reporting of corporate securities. Prior to January 31, 2013, FR 2004 data combined commercial paper, non-agency MBS, and corporate bonds under "corporate securities." We use this broader definition of corporate securities when we extend our analysis back to 2005 (see Appendix D). In contrast, since January 31, 2013, our analysis of corporate securities is limited to corporate bonds. This January 2013 break in our corporate series might impact statistical inference.

[31] On July 15, 2021, Santander Holdings USA reached an agreement to acquire Amherst Pierpont Securities. Both Santander and Amherst Pierpont have been TRACE reporters, but Santander was not a primary dealer prior to the merger with Amherst Pierpont. When determining the weights of dealers based on their D2C trading volume, we consider Amherst Pierpont separately up to the merger, and only then adding Santander and Pierpont.

G Additional quantile modeling of illiquidity

Figure G.9 shows the fitted distribution of illiquidity conditional on only Treasury yield volatility, in the form of a family of quantile regressions of US Treasury illiquidity. For this purpose, yield volatility is the average of the one-month swaption-implied volatilities of 2-, 5-, and 10-year swaps. Here and throughout the remainder, we use the US Treasury illiquidity measure constructed in Section 5. This is the first principal component of 18 underlying measures of illiquidity. Yield volatility is measured as the average of the one-month swaption-implied volatilities of 2-, 5-, and 10-year swaps, in basis points. As the plot makes clear, the slope coefficients in these univariate regressions increase with the quantile, implying that the conditional distribution of illiquidity becomes wider as volatility increases and that as the quantile to be predicted rises, predicted illiquidity is more sensitive to changes in volatility.

Figure G.10 shows the time-series of the conditional distribution for the bivariate specification based on dealer gross positions. According to this specification, illiquidity was elevated during March 2020, reaching levels between the 95th and 99th percentile of the conditional distribution. In contrast, amidst the high levels of volatility in 2022 that coincided with the most recent interest rate hiking cycle, illiquidity was elevated but was generally closer to the 75th or 50th percentile of the conditional distribution, alongside lower levels of dealer capacity utilization than those realized in March 2020.

Figure G.11 shows how the quantile regression coefficients vary by percentile when measuring dealer capacity utilization with FR 2004 net positions and TRACE D2C gross and net volumes. This analysis complements Figure 7 in the main text for FR 2004 gross positions. Across dealer capacity utilization measures, we see the coefficient or loading on the capacity utilization residual become more significant in the right-tail quantiles. We also see that the relationship between illiquidity and volatility is more stable in the right-tail in the bivariate specifications compared to the univariate specification, as before.

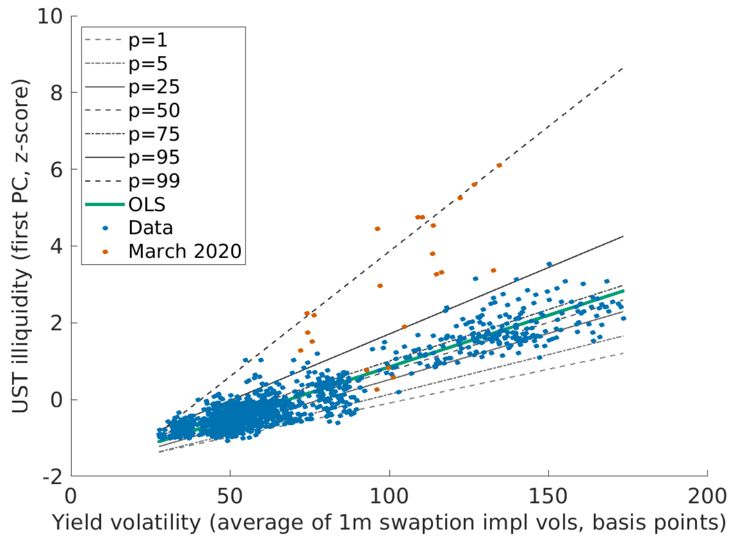


Figure G.9. US Treasury market illiquidity conditional distribution given volatility. A scatter plot of US Treasury illiquidity versus volatility alongside fitted values from quantile regressions showing the conditional distribution, and an OLS regression showing the conditional mean. High levels of illiquidity in March 2020 are large positive residuals in the OLS regression that are fit by a much higher slope for the 99th percentile quantile regression in comparison to the lower percentiles. US Treasury illiquidity is the standardized first principal component of the standardized illiquidity measures. Yield volatility is the average one-month implied volatility on 2-, 5-, and 10-year swaps in basis points. Daily data from July 10, 2017 to December 30, 2022.

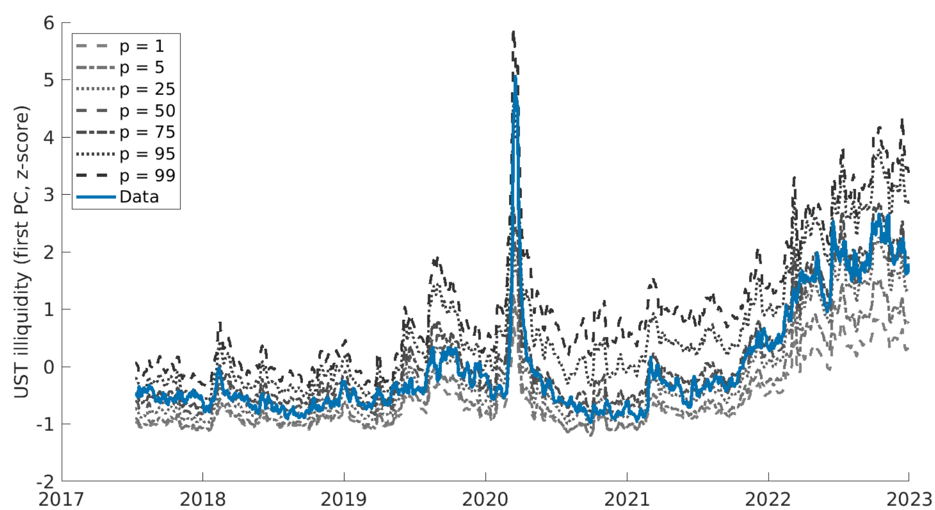


Figure G.10. US Treasury illiquidity conditional distribution over time. The conditional distribution of US Treasury illiquidity over time from quantile regressions that include volatility and residual capacity utilization as explanatory variables. Volatility is measured as the average one-month implied volatility on 2-, 5-, and 10-year swaps. Capacity utilization is measured with dealer gross positions. The residual of capacity utilization is the component not explained by volatility, using an OLS regression. US Treasury illiquidity and the conditional quantiles are plotted as five-day moving averages.

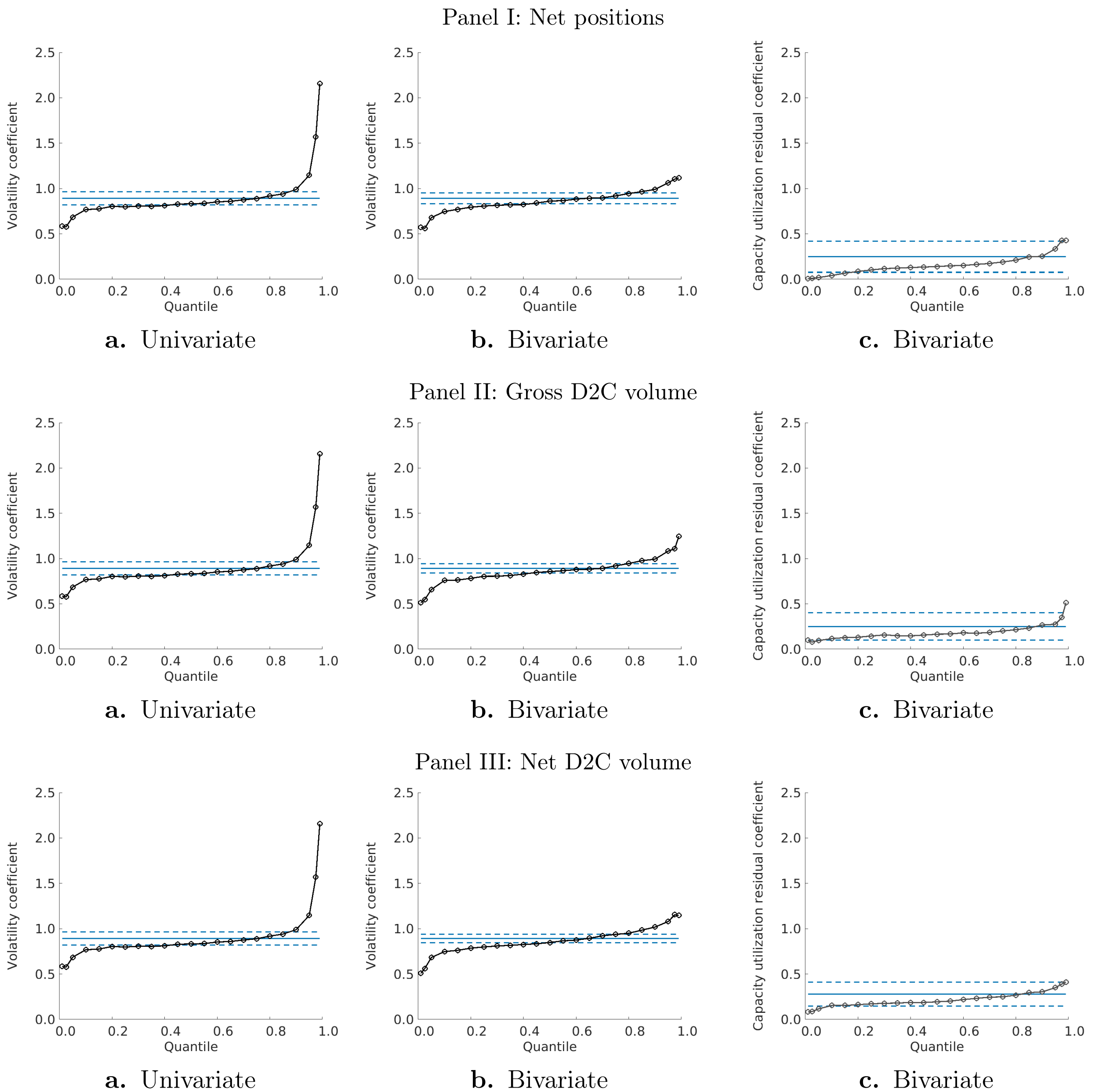


Figure G.11. US Treasury market illiquidity conditional distribution slope coefficients by percentile. Reproduces analysis from Figure 7 in the main text using dealer capacity measures based on FR 2004 net positions and TRACE D2C gross and net volumes to complement the analysis based on FR 2004 gross positions.

H Quantile regression with un-residualized capacity utilization

In the main body of the paper we orthogonalize dealer capacity utilization to volatility when studying the conditional distribution of Treasury market illiquidity. This allows us to study the incremental role of the OLS residuals of capacity utilization regressed onto volatility. Table H.14 reports the corresponding first-stage regressions. For several of the specifications of capacity utilization we see there is a statistically and economically significant relationship with volatility. For example, volatility has explanatory power of around 60% for capacity utilization based on gross positions.

As discussed in the main text, it is possible to exploit the linearity of the OLS and quantile regressions to map between quantile regression specifications using either residualized or un-residualized capacity utilization as an explanatory variable alongside volatility and a constant term. To show this result explicitly, let cu, vol, and e denote capacity utilization, yield volatility, and the first stage OLS residual. Let β_0 and β_1 be the OLS estimates from the first stage regression of cu onto vol where $cu = \beta_0 + \beta_1 vol + e$. Denote the z-scored variables as

$$\tilde{vol} = \frac{vol - \mu_{vol}}{\sigma_{vol}}, \quad \tilde{cu} = \frac{cu - \mu_{cu}}{\sigma_{cu}}, \quad \tilde{e} = \frac{e}{\sigma_e}.$$

The main body of the paper includes quantile regressions of Treasury market illiquidity y predicted by $\tilde{vol}$ and $\tilde{e}$. If we denote the point estimates from this regression as γ_0, γ_1, and γ_2, the fitted values may be rewritten as,

$$\begin{aligned}
\hat{y} &= \gamma_0 + \gamma_1 \tilde{vol} + \gamma_2 \tilde{e} \\
&= \gamma_0 + \gamma_1 \tilde{vol} + \tfrac{\gamma_2}{\sigma_e}(cu - \beta_0 - \beta_1 vol) \\
&= \gamma_0 + \gamma_1 \tilde{vol} + \tfrac{\gamma_2}{\sigma_e}((\tfrac{cu - \mu_{cu} + \mu_{cu}}{\sigma_{cu}})\sigma_{cu} - \beta_0 - \beta_1(\tfrac{vol - \mu_{vol} + \mu_{vol}}{\sigma_{vol}})\sigma_{vol}) \\
&= \gamma_0 + \gamma_1 \tilde{vol} + \tfrac{\gamma_2}{\sigma_e}(\tilde{cu}\sigma_{cu} + \mu_{cu} - \beta_0 - \beta_1(\tilde{vol}\sigma_{vol} + \mu_{vol})) \\
&= (\gamma_0 + \tfrac{\gamma_2}{\sigma_e}(\mu_{cu} - \beta_0 - \beta_1 \mu_{vol})) + (\gamma_1 - \gamma_2 \tfrac{\sigma_{vol}}{\sigma_e}\beta_1)\tilde{vol} + \gamma_2 \tfrac{\sigma_{cu}}{\sigma_e}\tilde{cu} \\
&\equiv \lambda_0 + \lambda_1 \tilde{vol} + \lambda_2 \tilde{cu}.
\end{aligned}$$

A quantile regression of Treasury market illiquidity y onto $\tilde{vol}$ and $\tilde{cu}$ thus has the point estimates λ_0, λ_1, and λ_2 which can be derived from,

$$\Theta = (\beta_0, \beta_1, \mu_{vol}, \sigma_{vol}, \mu_{cu}, \sigma_{cu}, \sigma_e, \gamma_0, \gamma_1, \gamma_2).$$

That λ_0, λ_1, and λ_2 are the point estimates follows from the observation that the fitted values $\hat{y}$ and residuals $u = y - \hat{y}$ are the same in a quantile regression of y onto $\tilde{vol}$ and $\tilde{e}$ using the parameters $(\gamma_0, \gamma_1, \gamma_2)$ or of y onto $\tilde{vol}$ and $\tilde{cu}$ using the parameters $(\lambda_0, \lambda_1, \lambda_2)$. If $\lambda \neq \arg\min_\theta \sum_t \rho_\tau(y_t - \theta_0 - \theta_1 \tilde{vol}_t - \theta_2 \tilde{cu}_t)$, it would be possible to achieve a better fit in the original quantile regression of y onto $\tilde{vol}$ and $\tilde{e}$, contradicting $\gamma = \arg\min_\theta \sum_t \rho_\tau(y_t - \theta_0 - \theta_1 \tilde{vol}_t - \theta_2 \tilde{e}_t)$. Table H.15 shows this result empirically by reporting the quantile regression results using un-residualized capacity utilization as an explanatory variable. One can derive the point estimates λ_0, λ_1, and λ_2 in this table by applying the formulas above combined with the first stage regression results, z-score parameters, and the quantile regressions in the paper using residualized capacity utilization.

Table H.14: First stage OLS regression of capacity utilization onto yield volatility. First-stage OLS regressions of capacity utilization onto yield volatility for the different specifications of capacity utilization considered in the paper. The z-score parameters for yield volatility used in the second-stage quantile regressions are $\mu_{vol} = 68.65$ and $\sigma_{vol} = 33.21$. Newey-West standard errors using 100 lags in parentheses, $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

Panel A: Capacity utilization based on FR2004 positions and TRACE D2C volumes

	Gross positions	Net positions	Gross D2C volume	Net D2C volume
Yield volatility	0.26***	0.14***	0.19***	0.06***
	(0.03)	(0.04)	(0.03)	(0.02)
Constant	16.88***	12.52***	9.29***	3.65***
	(3.33)	(3.35)	(2.32)	(1.06)
N	1,336	1,336	1,336	1,336
μ_{cu}	34.80	21.76	22.47	7.80
σ_{cu}	11.23	9.04	9.80	4.80
σ_{e}	7.15	7.86	7.45	4.37
R2 adj.	0.595	0.244	0.423	0.174

Panel B: Capacity utilization based on reported and estimated dealer VaR

	BHC VaR	IR VaR	Est. BHC VaR	Est. IR VaR
Yield volatility	0.01	0.05***	0.21***	0.17***
	(0.05)	(0.02)	(0.04)	(0.03)
Constant	20.16***	9.33***	4.14	1.09
	(5.16)	(1.67)	(2.75)	(2.08)
N	1,336	1,336	1,336	1,336
μ_{cu}	20.72	12.67	18.66	12.57
σ_{cu}	9.06	4.10	11.25	9.53
σ_{e}	9.06	3.76	8.79	7.75
R2 adj.	0.000	0.155	0.390	0.339

Table H.15: Quantile regressions with un-residualized capacity utilization Reports quantile regressions for the 99th percentile of US Treasury illiquidity onto yield volatility and un-residualized capacity utilization. For ease of interpretation all variables are standardized in the regressions. Standard errors are block bootstrapped.

Panel A: FR 2004 positions and TRACE D2C volumes

	(1)	(2)	(3)	(4)
Yield volatility	0.37	0.87***	0.81***	0.96***
	(0.25)	(0.14)	(0.15)	(0.11)
Capacity utilization: gross position	0.90***			
	(0.26)			
Capacity utilization residual: net position		0.49**		
		(0.19)		
Capacity utilization residual: gross D2C volume			0.67***	
			(0.22)	
Capacity utilization residual: net D2C volume				0.45**
				(0.18)
Constant	1.01***	0.97***	1.16***	0.91***
	(0.15)	(0.14)	(0.20)	(0.13)
N	1336	1336	1336	1336
Pseudo R2	0.74	0.74	0.71	0.74

Panel B: Reported and estimated dealer VaR

	(1)	(2)	(3)	(4)
Yield volatility	2.10***	2.10***	0.91***	0.94***
	(0.48)	(0.48)	(0.17)	(0.11)
Capacity utilization: BHC VaR Dealer	0.26			
	(0.16)			
Capacity utilization: IR VaR Dealer		0.27*		
		(0.17)		
Capacity utilization: BHC VaR Estimated			0.67***	
			(0.21)	
Capacity utilization: IR VaR Estimated				0.54***
				(0.20)
Constant	1.51***	1.61***	0.96***	0.85***
	(0.40)	(0.44)	(0.19)	(0.13)
N	1336	1336	1336	1336
Pseudo R2	0.60	0.57	0.72	0.75