Intermediary Balance Sheets and the Treasury Yield Curve *

Wenxin Du (Chicago and FRBNY)

Benjamin Hébert (Stanford)

Wenhao Li (USC)

December 22, 2022

Abstract

We document a regime change in the U.S. Treasury market post-Global Financial Crisis (GFC): dealers switched from net short to net long Treasury bonds. Consistent with this change, we derive "net-long" and "net-short" Treasury curves that account for dealers' balance sheet costs, and show that actual Treasury yields moved from the net short curve pre-GFC to the net long curve post-GFC. This regime change helps explain negative swap spreads post-GFC and the co-movement among swap spreads, dealer Treasury positions, yield curve slope, and covered-interest-parity violations, and implies changing effects for a wide range of monetary and regulatory policy interventions.

*Du: University of Chicago Booth School of Business, Federal Reserve Bank of New York, NBER, and CEPR. Email: wenxin.du@chicagobooth.edu. Hébert: Stanford University Graduate School of Business and NBER. Email: bhebert@stanford.edu. Li: University of Southern California, Marshall School of Business. Email: liwenhao@marshall.usc.edu.

A portion of this research was conducted at the Becker-Friedman Institute, whose hospitality B. Hébert gratefully acknowledges. We would like to thank Adrien Auclert, Patrick Augustin, Bruno Biais, Rohan Kekre, Darrell Duffie, Michael Fleming, Thierry Foucault, Valentin Haddad, Terrence Hendershott, Sebastian Infante, Arvind Krishnamurthy, Ralph Koijen, Tim Landvoigt, Stefan Nagel, Vania Stavrakeva, and Fabrice Tourre for insightful comments, as well as seminar participants at UC Berkeley, Chicago Fed, the University of British Columbia, Washington University in St. Louis, HEC Paris, USC, and Cornell. We in particular would like to thank Ken Singleton for both his insightful feedback and for sharing his term structure estimation code. We thank Filippo Cavaleri for excellent research assistance. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of New York or the Federal Reserve System. All remaining errors are our own.

Introduction

Prior to the GFC, the yield on the long-term Treasury bonds was below the fixed rate on interest rate swaps of the same maturity. This positive swap-Treasury spread ("swap spread," for short) pre-GFC was in part due to the "convenience yield" of the Treasury bonds (à la Krishnamurthy and Vissing-Jorgensen (2012)). However, post-GFC, yields on long-maturity U.S. Treasury bonds have been consistently above the corresponding swap rates, suggesting that Treasury bonds are now in some sense "inconvenient." Likewise, prior to the GFC, CIP deviations were almost non-existent, but have emerged as significant and persistent following the crisis. The emergence of CIP deviations and the change in the sign of long-maturity swap spreads are both well-documented in the literature.[1]

We provide new empirical evidence to link swap spreads and CIP deviations with *regimes* in which primary dealers are net long or net short with respect to Treasury bonds. Figure 1 illustrates a significant change in the Treasury market over the past twenty years. Pre-GFC, primary dealers on net maintained a short position in Treasury bonds, and the swap spread was positive. Post-GFC, primary dealers switched to a net long position in Treasury bonds, and the swap spread became negative. That is, dealers were short Treasury bonds when those bonds had yields lower than swap rates, and became long Treasury bonds once those bonds had higher yields than swap rates. The sign switch in the swap spread exactly coincided with the sign switch in dealers' net position.

We explain the collection of evidence in Figure 1 via a unified framework that incorporates the convenience (or lack thereof) of Treasury bonds, covered interest rate parity (CIP) deviations, and primary dealers' net Treasury positions. Our framework emphasizes the role of balance-sheet constrained dealers in the Treasury market. This framework naturally leads to regimes in which dealers are net long, net short, or flat with respect to Treasury bonds. Intuitively, in the long regime, Treasury yield should be above the swap rate of the same maturity to compensate the balance sheet cost, while in the short regime, the opposite is true. We show that the effects of policies such as quantitative easing/tightening (QE/QT) are regime-dependent, and can use our framework to explain why Treasury yields fell relative to other interest rates during the Global Financial Crisis (GFC) but rose during the March 2020 COVID-19 crisis (Duffie (2020)).

Our formal analysis proceeds in the following steps. We first derive what we call the "net-long"

[1] On swap spreads, see Feldhütter and Lando (2008), Klingler and Sundaresan (2019), Jermann (2020), Augustin et al. (2021), and Fleckenstein and Longstaff (2021). On the CIP deviations, see Du, Tepper, and Verdelhan (2018b).

and "net-short" curves, which can be viewed as arbitrage bounds on Treasury yields that account for funding costs, balance sheet costs, and interest rate risk. We use CIP deviations as a measure of balance sheet costs (i.e. the shadow value on the balance sheet constraint).[2] Quantifying these curves generates predictions of dealer positions consistent with the data. We next build a two-period, two-market model of dealers-as-arbitrageurs interacting with return-seeking clients, and in which CIP deviations, Treasury yields, and dealer Treasury positions are endogenous. We use this model to illustrate the idea of regimes and explain the facts in Figure 1. Lastly, we consider a variety of conventional and unconventional policy interventions in our model, and show that the effects of these policies are regime-dependent.

We begin by constructing estimates of the yields at which dealers would be willing to take a net long or net short position in Treasury bonds. The net long curve describes a Treasury yield above which a dealer would always want to be net long Treasury bonds, regardless of its beliefs about future Treasury yields. Similarly, the net short curve provides a yield below which a dealer would always be willing to be net short the bond. We construct these curves taking the interest rate swap curve and the structure of CIP violations as given. Our construction uses interest-rate swaps[3] to measure discount rates, CIP violations to measure the cost of balance sheet, and takes into account the spreads between financing rates and discount rates. The use of swap rates and CIP violations is motivated by a single assumption: that all zero-cost, zero-balance-sheet trading strategies are at least weakly unprofitable under a common stochastic discount factor (SDF).[4] This assumption allows us to compare the strategy of financing the purchase of a Treasury bond with the strategy of borrowing unsecured and engaging in CIP arbitrage (the difference between these is a zero-cost, zero-balance-sheet strategy). We use this comparison to derive our net long curve, and construct our net short curve in a similar manner.

The net long and net short curves differ for two reasons: financing rates and balance sheet costs. In particular, because taking either a net short or net long position increases the size of a dealer's balance sheet, higher balance sheet costs will increase the net long yield but decrease the net short yield. We find that actual Treasury yields are quite close to the estimated net short curve

[2] The interpretation of these costs is consistent with the view of Jermann (2020) that swap spreads are affected by "bond holding costs," but more specific in that these costs affect both swap spread and CIP deviations.

[3] Specifically, we use overnight index swaps (OIS), which are swaps based on the effective federal funds rate.

[4] This assumption is implicit in the models of Jermann (2020) and Du et al. (2022). Under this assumption, the SDF prices zero-cost and zero-balance-sheet strategies (e.g. derivatives including swaps with negligible balance sheet cost and zero initial cost), but does not directly determine the price of cash securities such as Treasuries.

pre-GFC and close to the estimated net long curve post-GFC, across a variety of maturities. This finding is consistent with the data on dealer net positions (i.e. dealers were net short pre-GFC and net long post-GFC), providing evidence in support of our view of dealers as arbitrageurs and our notion of Treasury market regimes.

We next construct a two-period, two-market supply-demand model for Treasury bonds and synthetic dollar lending in the foreign exchange (FX) swap market. Dealers arbitrage the spread between synthetic dollar lending rate and the swap rate (i.e., the CIP deviation), and the spread between swap rate and Treasury yield (i.e., the swap spread). In the model, the CIP deviation (endogenously) reflects the shadow cost of dealer balance sheet. Apart from dealers, we model two types of clients for Treasury bonds. One type is "real-money" investors such as domestic pension funds who do not rely on dealers' balance sheets to fund their positions. Real-money investors decide between holding Treasury bills versus long-term Treasury bonds, and their demand for bonds is an increasing function of the expected excess returns on the Treasury bonds. Second, we model foreign investors who invest in U.S. Treasury bonds financed by "synthetic dollars" (foreign currency converted to dollars using short-dated FX swaps).[5] We assume that foreign investors' demand for bonds increases in the expected return on the Treasury bond net the synthetic dollar rate.

We close the model by imposing market clearing conditions for Treasury bonds and synthetic dollars, as well as an intermediary balance sheet constraint. The model has a unique equilibrium, which falls into one of the three regimes: intermediaries are either long, short, or flat with respect to Treasury bonds. The regime itself is determined by the difference between Treasury bond supply and client demand; that is, dealers act as "buyer of last resort" in the long regime and a "short-seller of last resort" in the short regime.

The model helps explain a new empirical fact we document in the paper: dealer's net long position in Treasury bonds is negatively correlated with the term spread between the long-term Treasury bond yield and short rate in the post-GFC period. That is, dealers hold more bonds when the price is high relative to the short rate (i.e. a small term spread) and fewer bonds when the price is low relative to the short rate (i.e. a large term spread). This behavior at first appears puzzling, on the grounds that a larger term spread predicts larger expected returns of bonds vs.

[5]FX-hedged foreign investors do not use leverage but nevertheless rely on dealer balance sheets to obtain FX swap hedging. A canonical example of an FX-hedged foreign investor is a Japanese life insurance company, with liabilities denominated in yen and substantial U.S. dollar fixed-income assets.

the short rate (Campbell and Shiller (1991)). Our view of dealers as arbitrageurs absorbing the excess supply/demand for Treasury bonds can rationalize this fact: during the post-GFC period, when the term spread is low, client demand for Treasury is low. To clear the market, dealers must increase their holdings, which requires that Treasury bonds offer an attractive yield relative to interest rate swaps (after accounting for financing rates and balance sheet costs). In contrast, the existing literature has often emphasized instead returns-seeking behavior (as in the intermediary asset pricing literature, He and Krishnamurthy (2013)) and the role of dealers as over-the-counter market makers (as in Duffie et al. (2005)). These perspectives are not mutually exclusive with our view; however, they do not naturally lead to the prediction that dealers should buy high and sell low in the Treasury market, as suggested by the data.

In the last section of the paper, we demonstrate that the effects of a variety policy interventions are regime-dependent, using our two-period model. We also show that the effects of dealer distress differ across regimes, offering an explanation for why Treasury yields went down relative to swap rates in the GFC and up during March 2020. These differences arise from differences in the comparative statics of our model across regimes, which ultimately are a consequence of the difference in the way balance sheet costs affect Treasury yields. The model, with minimal modification, is capable of speaking to the effects of quantitative easing and tightening, inter-central-bank swap lines, regulatory exemptions to the supplementary leverage ratio (SLR), and interest rate policy. We discuss the effects of each of these policies, and emphasize how they differ in the long and short regimes. Finally, we discuss the implications of the framework for the ongoing tightening cycle, and draw a parallel with the experience of 2017-2019 tightening cycle.

Our paper is most closely related to Jermann (2020) and Du, Hébert, and Huber (2022), in that we model swap spreads (Jermann (2020)) and CIP violations (Du, Hébert, and Huber (2022)) as arising from constraints on intermediaries. Our paper combines these perspectives and shows that they can jointly explain data on dealers' positions.[6] Relatedly, Favara, Infante, and Rezende (2022) show evidence that SLR shocks have reduced large banks' participation in the U.S. Treasury market. Du, Tepper, and Verdelhan (2018b), Hébert (2020), and Du, Hébert, and Huber (2022) argue that CIP deviations can proxy for the shadow cost of the these constraints. The strong co-movement of CIP violations and swap spreads post-GFC we document is consistent with these

[6] Another difference between our paper and Jermann (2020) is that Jermann (2020) treats Treasury yields as exogenous, and as a result predicts that dealer positions are increasing in the term spread, in contrast to our framework and the data.

perspectives.[7]

Considering swap spreads and CIP deviations together helps address a puzzle: why would long-maturity (e.g. 30 year) swap spreads be affected by fluctuations in current balance sheet costs? The answer suggested by Du et al. (2022) is that there is a substantial risk premium associated with the risk that balance sheet costs increase; our quantitative analysis confirms that this risk premium plays a significant role in long maturity swap spreads. Our quantitative term-structure framework is substantially more general than the models employed in these papers, and our two-market equilibrium model is able to address a broader range of policy questions.

The key limitation of our static model is that it treats swap rates as exogenous, and we do not attempt to explain the behavior of dealers' swap counter-parties. In contemporaneous work, Hanson, Malkhozov, and Venter (2022) adopt an approach broadly similar to our two-period model to explain the way in which shocks to the demand for interest rate swaps affect swap spreads (see also Klingler and Sundaresan (2019)). Hanson, Malkhozov, and Venter (2022) focus on the swap market, treating the Treasury market as exogenous; combining the two approaches is an interesting direction for future research. Their interest is in separating supply and demand shocks, whereas our focus is on providing estimates of the dealer net long and short Treasury curves, validating the notion of Treasury market regimes, and analyzing policy interventions.

More recently, the dislocation of the Treasury bond market during the height of the COVID-19 pandemic in March 2020 has led some authors to question whether U.S. Treasury bonds remain convenient (such as in Duffie (2020) and He, Nagel, and Song (2022)). Relative to this literature, we highlight the importance of the regime change in the dealers' net position and the interaction between dealer balance sheet constraints and client demand for Treasury bonds.

Complementary to our analysis on the role of dealers, the role of hedge funds in the Treasury market has been examined in Barth and Kahn (2021) and Kruttli et al. (2021), particularly regarding their activities in the Treasury cash-futures basis arbitrage funded by dealers' balance sheets. We develop the net long and net short curve from the perspective of a securities dealer, but argue in Appendix Section E.1 that dealers will in effect transmit their balance sheet costs to hedge funds and other levered clients, consistent with Boyarchenko et al. (2018). As a result, the curves we develop are applicable for these levered clients as well.

[7] Recent work by Siriwardane et al. (2021) examines market segmentation across different near-arbitrages. The correlation between CIP deviations and the swap spread stands out being among the highest, which supports the use of CIP deviations as a balance sheet cost proxy for Treasury trading activities.

The negative correlation between dealers' net position and Treasury yield curve slope we document is the opposite to the pattern observed in typical (excluding the largest) commercial bank portfolios (Haddad and Sraer (2020)). This contrast emphasizes the importance of distinguishing between the Treasury activities of securities dealers and commercial bank subsidiaries. Fluctuations in dealers' inventory have also been linked to overall liquidity conditions (Goldberg (2020)).

The structure of the paper is as follows. We provide institutional background on Treasury trading by dealers in Section 1. We derive and estimate the net long and the net short curves in Section 2. We introduce the demand from real-money investors and build an equilibrium model for Treasury market dynamics in Section 3. We analyze policy implications in Section 4, and in Section 5 we conclude.

1 Institutional Background

In this section, we outline the mechanics about how dealers go long or short Treasury securities and hedge with swaps. We then discuss when these strategies are arbitrages, and compare them with CIP arbitrage. We denote Treasury yields as y, financing rates as i, and swap rates as r.

1.1 The Long Treasury vs. Swap Arbitrage

Suppose that the dealer goes long (buys) a zero-coupon Treasury bond of maturity n at time t. At the onset of the trade, the dealer buys the Treasury bond at yield $y_{n,t}$, and finances the entire position period-by-period at the time $t+j$ (for $j = 0, ..., n-1$) financing rate i^l_{t+j}. This trade is zero-cost, in the sense that the entire purchase cost of the bond is financed through a mix of secured (repo) and unsecured debt financing (i^l_{t+j} is the effective one-period financing rate of this mix at time $t+j$).[8]

A dealer can hedge the interest rate risk of this trade by entering into a swap contract.[9] Interest rate swaps exchange a fixed interest payment ($r_{n,t}$) for a sequence of floating interest payments

[8] See Internet Appendix Section A for more details.

[9] Dealers have an incentive to hedge their net interest rate risk, due to risk-based capital requirements, but will typically do so at the trading desk level or the whole book level as opposed to trade-by-trade.

(r_{t+j}).[10] To hedge a net long bond position, the dealer pays the fixed rate and receives the floating rate in a swap, whose maturity is set to match the bond (n in our example).

After hedging, the dealer will receive the spread between the floating rate r_{t+j} and the financing rate i^l_{t+j} each period. If the dealer holds the bond and swap to maturity, the dealer will earn the yield $y_{n,t}$ on the bond and pay the fixed rate $r_{n,t}$ on the swap; the difference of these two is the negative of the "swap spread" $r_{n,t} - y_{n,t}$. Thus, if the floating rate is guaranteed to exceed the financing rate ($r_{t+j} \geq i^l_{t+j}$ for all j), and the swap spread is negative, the dealer is guaranteed to earn a profit on this trade if held to maturity. If in addition the spread $r_{t+j} - i^l_{t+j}$ is constant, the dealer's profit is certain ex-ante (because $r_{n,t} - y_{n,t}$ is known at the trade's inception). In this sense, the swap hedges the bond.[11] During this trade, the dealer will have a larger balance sheet, approximately equal to the value of the Treasury net long position.[12]

1.2 The Short Treasury vs. Swap Arbitrage

Short-selling a Treasury bond and hedging with a swap works in an essentially identical way, with the directions of the trades reversed. When a dealer short-sells a Treasury bond, the dealer borrows the bond from its owner, and offers the owner cash as collateral. The dealer raises that cash by selling the borrowed bond, and each period receives an effective interest rate of i^s_{t+j} on this cash.[13]

The dealer can hedge with an interest rate swap by receiving the fixed rate $r_{n,t}$ and paying the floating rate r_{t+j}. If the spread between r_{t+j} and i^s_{t+j} is bounded above, and the swap spread $r_{n,t} - y_{n,t}$ exceeds this bound, the dealer is guaranteed to make a profit, assuming the position is held to maturity. This arbitrage also increases the size of the dealer's balance sheet. The asset in

[10] Our analysis will focus on overnight index swaps (OIS), in which one party pays a fixed rate of interest in exchange for a series of floating payments indexed to the overnight interbank federal funds rate. Prior to the GFC, swaps indexed to LIBOR were more commonly used, and recently swaps indexed to SOFR (Secured Overnight Financing Rate) have been introduced. OIS rates are available for our entire sample and are similar to LIBOR swap rates pre-GFC and to SOFR swap rates in the recent period.

[11] If $r_{t+j} - i^l_{t+j}$ is not constant, there is residual "basis risk," but negative swap spreads are nevertheless an arbitrage opportunity if $r_{t+j} - i^l_{t+j}$ is guaranteed to be positive.

[12] To a first approximation, the interest rate swap is entirely off-balance-sheet. More precisely, trading interest rate swaps can increase the size of the balance sheet slightly. The total exposure includes initial and variation margins (typically a couple percent of total notional), and an additional 0-1.5% of the swap notional calculated for off-balance sheet interest rate derivative exposure using the Current Exposure Method, depending on the maturity of the interest rate swaps. We ignore the additional balance sheet costs of trading derivatives to simplify our analysis.

[13] See Internet Appendix Section A.2 for more details.

this case is a "loan" to the original owner of the bond (the dealer has given the original owner cash in exchange for Treasury collateral), and the liability is a security to be delivered.

Thus, the long Treasury-swap and short Treasury-swap arbitrage strategies are both zero-cost, n-period strategies that use balance sheet and, under certain assumptions, deliver a guaranteed profit if held to maturity. We will compare these strategies to another n-period, zero-cost, balance-sheet-intensive arbitrage opportunity: CIP arbitrage.

1.3 CIP Arbitrage

One-period USD-EUR CIP arbitrage involves borrowing dollars in the cash market, converting the dollars to euros in the spot foreign exchange market, lending out the euros, and using a forward contract to lock in the exchange rate at which the euro proceeds of the loan are converted back to dollars. Suppose the dealer borrows dollars in unsecured funding markets at the rate r_t (the same rate used in the interest rate swap), and define the "synthetic" dollar rate r_t^{syn} as the synthetic dollar lending rate in the FX swap market obtained by converting the euro lending rate into dollars. The profits of the one-period CIP arbitrage are the difference between these two rates, $r_t^{cip} = r_t^{syn} - r_t$, and are positive for most major USD-pairs in the post-GFC period (Du, Tepper, and Verdelhan (2018b)). This trading strategy is also zero-cost, in that the synthetic dollar lending is entirely financed by unsecured dollar borrowing.

A dealer who plans to engage in the CIP arbitrage for n periods can lock in the profits of this trade using interest rate and cross-currency basis swaps (the details of which are unimportant at this stage of our discussion). However, engaging in CIP arbitrage will increase the size of the dealer's balance sheet; in this case, the asset is the euro loan and the liability is the dollar borrowing.

1.4 Comparing Arbitrages

Each of these three arbitrages lasts for n periods if held to maturity, has zero initial cost, and uses balance sheet.[14] The core of our analysis focuses on perturbations in which a dealer does less CIP arbitrage and more of one of the two Treasury vs. swap arbitrages. These perturbations are zero-cost, zero-balance sheet trading strategies, illustrated in Figures 4 and 5.

[14] There are differences between the strategies with regards to the timing of payments. We develop our term structure model to account for these kinds of issues.

Empirically, the rate at which dealers finance the long Treasury trade (i^l_{t+j}) is higher than the rate at which they finance the short Treasury rate (i^s_{t+j}). We document in Internet Appendix Section A.2 that the rate the original bond owners ("security lenders") pay to dealers is typically roughly 20 basis points below i^l_{t+j}, even for securities that are not "special", using data from Markit Securities Finance.[15] Because of this difference, the long and short Treasury vs. swap arbitrages will never both be appealing to dealers. If swap spreads are sufficiently low/negative, as in the post-GFC period, the long arbitrage will be profitable, while if swap spreads are sufficiently large, as in the pre-GFC period, the short arbitrage will be profitable.

Each of these arbitrages is subject to market-to-market fluctuations. For example, the two-period CIP arbitrage has a risk-premium associated with the possibility that the one-period CIP arbitrage becomes larger next period (Du et al. (2022)). To account for this, we next construct a model, based on a single critical assumption: that all zero-cost, zero-balance-sheet trading strategies are weakly unprofitable under a common SDF. Using this assumption, we carefully compare the Treasury vs. swap and CIP arbitrages.

2 The Long and Short Treasury Yield Curves

In this section, we construct what we call "net long" and "net short" yield curves. These yield curves represent (approximate) arbitrage bounds at which a dealer would be willing to go net long or net short Treasury bonds. In frictionless models (in which arbitrage capacity is unlimited and the financing rates i^l_{t+j} and i^s_{t+j} are equal), there is a single Treasury yield curve. At yields above this yield curve, dealers would want to go net long bonds, while at lower yields, dealers would want to go net short bonds. In our model, two frictions create a wedge between the yield at which dealers would go net long and the yield at which dealers would go net short.

The first friction is balance sheet costs. As emphasized in the previous section, going either net long or net short a Treasury bond increases the size of dealer balance sheets. This has an opportunity cost, and this opportunity cost acts like a tax on both these trades. It raises the yield dealers require when going net long, while lowering the yield they require when going net short. We proxy for balance sheet costs using CIP deviations.

[15] When a dealer makes a repo loan to a hedge fund client (typically at a rate above i^l_{t+j}), the client selects the Treasury collateral. A dealer who wants to short a specific bond cannot rely on this kind of lending to find the bond. Instead, the dealer borrows the bond from a security lender at the lower rate i^s_{t+j}.

The second friction is the difference in financing rates of the long and short Treasury positions. When a dealer goes net long, they finance the position at a relatively high interest rate (i^l_{t+j}); when a dealer goes net short, they receive a relatively low interest rate on their cash (i^s_{t+j}). The difference between these two rates crates another wedge between the net long and net short curves.

Our analysis in this section is partial equilibrium. We take swap rates, financing rates, and Treasury yields as given, and ask if dealers are willing to go net long or net short Treasury bonds. We will then compare our answers to this question, which are based on price data, with our quantity data on dealers' net Treasury positions.

We proceed in four steps. First, we will introduce a very simple model of dealer behavior, to illustrate the trade-off a dealer faces between going net long or short Treasury bonds and other arbitrage activities. Second, we generalize the ideas illustrated in this simple model, and construct the net long and net short curves. Third, we discuss the assumptions under which these curves represent arbitrage bounds. Finally, we build a term structure model to estimate the net long and net short curves.

2.1 One-Period Net-Long and Net-Short Treasury Yields

We first consider the problem of a risk-neutral dealer who can choose between trading a single n-period zero-coupon Treasury bond and a one-period CIP arbitrage, subject to a fixed balance sheet constraint. We make these stark assumptions (only one bond, fixed balance sheet) to illustrate the main idea; they are not imposed in our more general term structure analysis in Section 2.4. The dealer makes this choice at a single date in this simplified example; for this reason, we will omit time subscripts.

Let q^{bond} be the dealer's bond position (in dollars, not notional), with negative values implying short-selling. Let q^{syn} be the dealer's "synthetic dollar" position (in dollars), which we define as the currency hedged investment leg of the CIP arbitrage (the other leg is borrowing dollars). We assume that the synthetic dollar rate (e.g. the currency-hedged euro rate) is above the dollar borrowing rate ($r^{syn} > r$), so that the direction of the arbitrage is to lend in synthetic dollars and finance in actual dollars.

Suppose the dealer faces a fixed balance sheet constraint,

$$q^{syn} + |q^{bond}| = \bar{q}, \tag{1}$$

where $\bar{q} > 0$ is the balance sheet capacity for Treasury and CIP arbitrage. Note that the absolute value of q^{bond} enters this expression, reflecting the fact that both long and short positions require balance sheet.

Let y be the log-yield on the Treasury bond. We assume, following the discussion above, that if the dealer buys the bond, the dealer will finance its position with at the log interest rate i^l. If the dealer instead short-sells the bond, it receives a log return i^s on its cash.

We define $p_{\mathbb{Q}}$ as the risk-neutral expectation of the bond price in the next period, when the trade will be unwound. In this simple example, we will treat this as exogenous. In our more general term structure analysis, this value will be determined primarily by the n-period swap rate and the n-period CIP deviation, in keeping with the idea that we are using swap rates and CIP violations to measure interest rate and balance sheet risk.

The dealer's problem is

$$\begin{aligned} \max_{q^{bond},q^{syn}} & \Big(\underbrace{\frac{p_{\mathbb{Q}}}{e^{-ny}}}_{\text{return of selling after one period}} - \underbrace{e^{i^l}}_{\text{financing of long position}} \Big) \cdot \max\{q^{bond},0\} \\ & + \Big(\underbrace{e^{i^s}}_{\text{return earned on cash collateral}} - \underbrace{\frac{p_{\mathbb{Q}}}{e^{-ny}}}_{\text{return of buying back after one period}} \Big) \cdot \max\{-q^{bond},0\} \\ & + \Big(\underbrace{e^{r^{syn}} - e^{r}}_{\text{CIP arbitrage spread}} \Big) \cdot q^{syn} \end{aligned} \tag{2}$$

subject to the balance sheet constraint in (1). That is, the dealer chooses between the long Treasury trade, the short Treasury trade, and the CIP arbitrage.[16]

We assume that the profit of Treasury trade does not dominate the synthetic lending trade, so that the synthetic lending amount is always non-zero. This assumption eliminates the corner case of $q^{syn} = 0$, which has less empirical relevance, and implies that the shadow cost of balance sheet constraint is measured by CIP deviations. We also assume, consistent with the data, that $i^s < i^l$.

Solving the problem, we find three different regimes. The first regime features $q^{bond} > 0$ and $q^{syn} > 0$. The yield in this regime is equal to the "net long yield" y^l, solved from the first-order

[16]We have omitted the swap part of the Treasury-swap arbitrages from these calculations, as a simplification. This simplification is justified under the assumption that the swaps are fairly priced under the $\mathbb{Q}$ measure.

conditions of q^{bond} and q^{syn},

$$e^{-ny^l} \equiv \frac{p_{\mathbb{Q}}}{e^{i^l} + (e^{r^{syn}} - e^r)}, \tag{3}$$

The second regime features $q^{bond} < 0$ and $q^{syn} > 0$. The yield in this regime is equal to the "net short yield" y^s, solved from the first-order conditions of q^{bond} and q^{syn},

$$e^{-ny^s} \equiv \frac{p_{\mathbb{Q}}}{e^{i^s} - (e^{r^{syn}} - e^r)}. \tag{4}$$

Comparing these two first-order conditions, we can see that the discount rate (denominator) is lower in the short regime than in the long regime, for two reasons. First, the funding rate is lower, and second, the opportunity cost of balance sheet ($e^{r^{syn}} - e^r$) is subtracted from the short funding rate but added to the long funding rate. As a result, the short yield is below the long yield, reflecting (intuitively) that the dealer requires a low yield/high price to justify a net short position and a high yield/low price to justify a net long position.

The third possible regime features $q^{bond} = 0$ and $q^{syn} > 0$. In this regime, the Treasury bond yield has to satisfy $y^s \leq y \leq y^l$. That is, the profit on the long and short Treasury strategies does not justify the balance sheet cost, so the dealer chooses to have a zero net Treasury position.

To gain further intuition, consider the case of a one-period bond that matures next period ($p_{\mathbb{Q}} = 1$). The log-linearized version of (3) and (4) is, defining $r^{cip} = r^{syn} - r$,

$$\begin{aligned} y^l &\approx r - (r - i^l) + r^{cip}, && (5) \\ y^s &\approx r - (r - i^s) - r^{cip}, && (6) \end{aligned}$$

The net long yield can differ from r (which we will think of as the one-period swap rate) for two reasons. First, holding Treasury bonds takes up bank balance sheet, so the yield has to be higher by an amount equal to the opportunity cost of the balance sheet (measured by r^{cip}). Second, if dealers' financing rate is lower than their unsecured funding cost, $i^l < r$, then there is a financing benefit to owning the Treasury bond, which makes the dealer willing to accept a lower yield.

The net short yield can differ from r for similar reasons. The impact of the opportunity cost of balance sheet affects the sell yield with a negative sign. The sell yield has to be lower (the price to be higher) to justify dealer's short position, which also takes balance sheet. The sell yield is further lowered if the return on the cash collateral is lower than the dealer's borrowing cost, $i^s < r$.

2.2 Multi-Period Net Long and Net Short Curves

We next extend the logic of these yields to a more general, multi-period setting, constructing what we will call the "net long curve" and "net short curve."

We make three key assumptions: (i) there is an SDF that prices interest rate swaps and cross-currency basis swaps, (ii) some synthetic dollar lending occurs in equilibrium, and (iii) all zero-cost, zero-balance-sheet trading strategies are weakly unattractive under this SDF (i.e. $0 \geq E[MR]$ in the standard notation). We will interpret this SDF as a dealer's SDF, under the usual intermediary asset pricing assumption that dealers are active in all of these markets. Let $\mathbb{Q}$ denote the risk-neutral measure associated with this SDF.

The first assumption is justified by the observation that essentially all of the arbitrages documented in the literature involve the use of balance sheet.[17] That is, no-arbitrage between derivatives appears to hold in the data. The second is justified by the empirical observation that currency dealers actively engage in CIP arbitrage.

The third assumption generalizes the first-order condition of our simple model. In that model, the strategy of increasing or decreasing q^{bond}, financed at the appropriate financing rate, and offsetting the change in balance sheet by changing q^{syn}, is weakly unattractive under the risk-neutral measure $\mathbb{Q}$. Our generalization of this condition assumes that all feasible zero-cost, zero-balance sheet perturbations are weakly unattractive.

These assumptions are much more general than our simple model. In particular, we have made no assumptions on the pricing of balance-sheet-increasing or balance-sheet-reducing strategies, and hence are agnostic about why dealers find it costly to increase the size of their balance sheet. We have also said nothing about how many maturities of Treasury bonds are available, or what other assets are also traded by dealers.

In what follows, all rates are annual and each period is one month. Consider the strategy of going long the Treasury bond at yield $y_{n,t}$, financing via repo, and offsetting the balance sheet

[17] This is true, for example, of all of the arbitrages considered in and Boyarchenko et al. (2018) and Siriwardane, Sunderam, and Wallen (2021).

effect by reducing CIP activity. We must have

$$\underbrace{e^{-ny_{n,t}}e^{i_t^l}}_{\text{secured financing}} + \underbrace{e^{-ny_{n,t}}(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}r_t})}_{\text{forgone CIP profits}} \geq E_t^{\mathbb{Q}}[e^{-(n-1)y_{n-1,t+1}}]. \tag{7}$$

The left-hand side of this expression represents costs paid at time $t+1$. The financing of the bond purchase must be repaid (the first term), and the profits of the forgone CIP arbitrage are lost (the second term). These must be weighed against the benefits of selling the bond at time $t+1$. This condition is a generalization of (3) from our simple model (with an expectation under $\mathbb{Q}$ playing the role of $p_{\mathbb{Q}}$). Note that because both sides of this equation are defined in terms of $t+1$ payoffs, the discount rate associated with the SDF is irrelevant.

Define the monthly log interest rate

$$x_{1,t} = \ln(e^{\frac{1}{12}i_t^l} - e^{\frac{1}{12}r_t} + e^{\frac{1}{12}r_t^{syn}}), \tag{8}$$

and iterate:

$$e^{-ny_{n,t}} \geq e^{-ny_{n,t}^l} = E_t^{\mathbb{Q}}[\exp(-\sum_{j=0}^{n-1} x_{1,t+j})]. \tag{9}$$

The yield curve $y_{n,t}^l$ defines what we call the "net long curve." This curve represents the point at which dealers would be willing to switch from CIP arbitrage activity to taking a net long position in a Treasury bond. Since zero-cost, zero-balance-sheet strategies are weakly unattractive, this net long curve is, by induction, an upper bound on Treasury yields (under the assumption that dealers engage in CIP arbitrage).

This net long curve can also be viewed as a lower bound on swap spreads (defined as the difference between swap rates and Treasury yields of matching maturity). Let $r_{n,t}^{cip}$ be the n-period CIP violation (the n-period synthetic dollar rate minus the n-period swap rate). Linearizing (9) and recalling that $r_{n,t}$ is the n-period swap rate,

$$r_{n,t} - y_{n,t} \geq r_{n,t} - y_{n,t}^l \approx - \underbrace{r_{n,t}^{cip}}_{\text{n-period CIP violation}} + E_t^{\mathbb{Q}}[\frac{1}{n}\sum_{j=0}^{n-1} \underbrace{(r_{t+j} - i_{t+j}^l)}_{\text{financing benefit}}]. \tag{10}$$

The difference between the short-maturity swap rate r_{t+j} and the financing rate i_{t+j}^l is generally small and stable over time. As a result, equation (10) implies that if yields are close to the net long curve, we should expect swap spreads to be negative and close in absolute value to the matched-maturity CIP violation.

We should also emphasize the important role of risk premia that is hidden in this expression. The n-period CIP violation $r_{n,t}^{cip}$ is well above the physical ($\mathbb{P}$) measure expectation of future short-maturity CIP violations, as emphasized by Du, Hébert, and Huber (2022). That is, net long curve yields are higher than swap rates both because of expectations of non-zero future balance sheet costs and because of the risk associated with the possibility that these costs become larger.

We develop the net short curve via similar logic. Consider the strategy of short-selling the Treasury bond at yield $y_{n,t}$, borrowing the bond against cash collateral, and offsetting the balance sheet effect by reducing CIP activity. We must have

$$\underbrace{e^{-ny_{n,t}}}_{\text{sale price}} \times \underbrace{e^{\frac{1}{12}i_t^s}}_{\text{gross return on cash collateral}} \leq \underbrace{E_t^{\mathbb{Q}}[e^{-(n-1)y_{n-1,t+1}}]}_{\text{repurchase price}} + e^{-ny_{n,t}} \underbrace{(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}r_t})}_{\text{forgone CIP profits}}. \tag{11}$$

The left-hand side of this expression is the cash generated at date $t+1$ by selling the bond and posting the cash as collateral. The right-hand side represents the costs of this trade at date $t+1$, including both the cost of repurchasing the bond and the forgone CIP profits. Define the monthly log interest rate

$$x_{2,t} = \ln(e^{\frac{1}{12}i_t^s} + e^{\frac{1}{12}r_t} - e^{\frac{1}{12}r_t^{syn}}) \tag{12}$$

and iterate:

$$e^{-ny_{n,t}} \leq e^{-ny_{n,t}^s} = E_t^{\mathbb{Q}}[\exp(-\sum_{j=0}^{n-1} x_{2,t+j})]. \tag{13}$$

The yield curve $y_{n,t}^s$ defines what we call the "net short curve." This curve represents the point at which dealers would be willing to switch from CIP arbitrage activity to taking a net short position in a Treasury bond. The net short curve is a lower bound on Treasury yields (again under the assumption that dealers engage in CIP arbitrage).

The net short curve can also be viewed as an upper bound on swap spreads. Linearizing (13),

$$r_{n,t} - y_{n,t} \leq r_{n,t} - y_{n,t}^s \approx \underbrace{r_{n,t}^{cip}}_{\text{n-period CIP violation}} + E_t^{\mathbb{Q}}[\frac{1}{n}\sum_{j=0}^{n-1} \underbrace{(r_{t+j} - r_{t+j}^{sec})}_{\text{security lending spread}}]. \tag{14}$$

The difference between the short-maturity swap rate and the security lending rate is positive, and the n-period CIP violation in our definition is also positive and proxies for the balance sheet cost. Taking these two forces together, equation (14) implies that if yields are close to the net short curve, we should expect positive swap spreads.

Our assumptions are sufficient to determine an upper and lower bound on Treasury yields (or, equivalently, on swap spreads), but are not enough to pin down Treasury yields themselves. In a frictionless world, the net long and net short curves converge to one curve and thus exactly pin down Treasury yields. In the presence of frictions, yields can fall anywhere in between the net short and net long curves. We will show, empirically, that yields were close to the net short curve before the GFC and close to the net long curve after the GFC. We will then construct a model in which dealers interact with non-dealers. In this model, the demands of non-dealers will determine where Treasury yields fall within the net short and net long curve bounds.

2.3 Discussion

Before proceeding, we elaborate on the interpretation of these curves and on the role of interest rate swap hedges in this framework.

The Net Short and Net Long Curves as Arbitrage Bounds. The bound $y_{n,t} \in [y^s_{n,t}, y^l_{n,t}]$ is an arbitrage bound if $y_{n-1,t+1} \in [y^s_{n-1,t+1}, y^l_{n-1,t+1}]$ with probability one, which follows from our key assumptions. This observation gives rise to the following interpretation: the net long yield is a yield at which a dealer would be willing to go long even if the dealer believed all dealers would be net long the bond in the future with probability one. Likewise, the net short yield is the yield at which a dealer would be willing go short, even if the dealer believed all dealers would be net short in the future with probability one. Yields falling in between these bounds can be loosely interpreted as related to the probability that dealers will be either net long or net short in the future.

In Internet Appendix Section E.2, we show that under relaxed assumptions, if $y_{n,t} > y^l_{n,t}$, dealers will perceive an arbitrage opportunity, even if it is not guaranteed that future yields are within the bounds, i.e., if it is possible that $y_{n-1,t+1} \leq y^l_{n-1,t+1}$. The net short curve does not share this property. However, we are able to derive a lower value for yields which we call "partial equilibrium net short curve" such that the dealer will always be willing to go short, and show that the net short curve described in the main text is a close approximation of this partial equilibrium net short curve.

Pre-GFC and Post-GFC. Pre-GFC, synthetic lending rates were close to swap rates (i.e. CIP violations were roughly zero), and security lending rates (i_t^s) were roughly 25 basis points below one-month swap rates. As a result, during this period, $x_{2,t}$ was roughly the one-month swap rate less 25 basis points. It follows that net short curve yields $y_{n,t}^s$ were lower than matched-maturity swap rates by about the same amount, which is to say there was a significant positive swap spread. In contrast, the net long curve during this period was only about five basis below swap rates, reflecting the difference between the long financing rate (the tri-party repo rate) and swap rate.

Post-GFC, synthetic lending rates are well above swap rates, and the spread between the one-month swap rate and the long financing rate is small. In this period, $x_{1,t}$ is approximately equal to the synthetic lending rate. As a result, net long curve yields $y_{n,t}^l$ are approximately the same as synthetic swap rates (i.e. EUR swap rates converted to dollars), which are swap rates plus the CIP basis of the same maturity. This leads to a significant "negative swap spread" between swap rates and the net long curve. In contrast, the net short curve now features rates far below swap rates (large and positive swap spreads), reflecting both the spread between the short financing rate and one-month swap rates and the CIP violations.

Actual swap spreads went from being large and positive pre-GFC to negative post-GFC. This fact, combined with the discussion above, previews our result that Treasury yields went from being close to the net short curve pre-GFC to close to the net long curve post-GFC. We will discuss the consequences of this shift in Sections 3 and 4.

Hedging with Swaps. It is standard practice to hedge a Treasury position with an interest rate swap.[18] Prior to the GFC, hedging using LIBOR swaps was typical; following the GFC, OIS rose in prominence, and more recently swaps based on repo rates (SOFR swaps) have begun to trade. Our analysis will focus on OIS swaps, which are available for a reasonably long sample (unlike SOFR swaps) and insensitive to credit risk (unlike LIBOR swaps).[19]

Hedging Treasury bonds with interest rates swaps leaves the dealer exposed to mark-to-market risk associated with fluctuations in the Treasury-swap spread. In our log-linearized equations (10) and (14), we can see that the swap spreads for the net-long and net-short curve depend on the risk-neutral expectations of future balance sheet constraints and residual basis spreads between money

[18] There are a range of approaches: hedging can be static or dynamic, and based on matching maturity, duration, or cashflows.

[19] See more details in Internet Appendix Section A.

market rates.

Thus, the net long curve $y^l_{n,t}$ can be hedged with a synthetic dollar swap (by swapping EUR rates to dollars), under the assumption that the long financing rate vs. short-term swap spread is stable. The net short curve $y^s_{n,t}$ can be hedged via a combination of swaps and synthetic dollar swaps, under the assumption that the spread between short interest rates and short-term swap rates is stable.

If one assumes that a Treasury yield will always be at one end of these boundaries, then the Treasury bond itself can be hedged. But note that the hedge will differ depending on which of the two boundaries is assumed to apply, and in neither case is the hedge a single swap. The intuition for this result is that balance sheet costs matter, and hedging the Treasury bond with a swap hedges interest rate risk but does not hedge balance sheet risk.

Yields and Positions. The net short and net long curves we construct are estimates of yields at which the dealer should be willing to take a net position in a Treasury bond, after accounting for financing and balance sheet costs. This definition leads naturally the prediction that if the yield is at the net long yield, dealers should be net long, and if the yield is at the net short yield, dealers should be net short. These bounds are constructed from possibly unreasonable beliefs about the stochastic process driving bond prices. As a result, we should not be at all surprised if dealers are willing to go long at yields below the net long yield or go short at yields above the net short yield.

Moreover, the yield of a Treasury bond relative to these curves does not directly determine the scale of dealer positions. For example, if the yield is at the net long yield, dealers should be net long, but the quantity by which they will be net long will depend on their balance sheet capacity, risk tolerance, and other considerations that cannot be inferred directly from yield curves. Nevertheless, we should expect net dealer positions to increase when yields move close to the net long yield and to decrease (become net short) when yields move close to the net short yield. We can construct a heuristic mapping from yields to position via this intuition.

Dealers and Levered Clients. In developing the net long and net short curves, we have adopted the perspective of a dealer. In Internet Appendix Section E.1, we argue that these curves are also arbitrage bounds from the perspective of the dealer's levered clients (i.e. hedge funds). Dealers' balance sheet costs are in effect transmitted to these clients via bilateral lending markets, a point emphasized by Boyarchenko et al. (2018). As a result, balance sheets costs will influence a substantial segment of the Treasury market, even though dealers are on their own hold a relatively small quantity of Treasury bonds on net. Internet Appendix Section E.1 also contains some suggestive evidence supporting this perspective. In what follows, we will treat dealers and their levered clients as a consolidated entity.

2.4 Term Structure Estimation

To estimate the net long and net short curves, we need to construct the risk-neutral expectations of $x_{1,t+j}$ and $x_{2,t+j}$. This is where our assumption that interest rate swaps and cross-currency basis swaps are priced by a common SDF applies. The risk-neutral expectations we need are determined primarily by the swap curve and the term structure of CIP violations. We will proceed by constructing an SDF (in particular, a term structure model) that fits swap rates and CIP violations, and then use that SDF to construct the net long and net short curves.

At the heart of our calculations is a comparison between a Treasury hedged with a swap and a CIP violation of the same maturity. A rough version of this comparison can be done without a model: one simply compares the swap spread with the CIP violation. Our term structure model allows for a more careful version of this comparison. First, it allows us to consider Treasury bonds with maturities and coupon structures that do not exactly line up with the available points of the swap and CIP curves. Second, it allows us to explicitly account for the residual basis risk between financing rates. Third, it allows us to model the unwinding of the trade when the bond has six months remaining maturity (which helps fit the short-end of the yield curve). Fourth, it smooths the swap and CIP curves, reducing micro-structure-induced noise. Lastly, it accounts for covariances that are omitted from the naive spread calculation.

Our term structure model largely follows the standard approach in the no-arbitrage term structure model literature (Joslin, Singleton, and Zhu, 2011; Joslin, Priebsch, and Singleton, 2014). At first glance, this might seem strange, given that our model necessarily features arbitrage. In particular, our term structure model must match both swap rates and CIP violations, which is equivalent to matching dollar swap rates and synthetic dollar swap rates. Our term structure model therefore features two different short rates, and as a result two different yield curves, as in Augustin, Chernov, Schmid, and Song (2020).[20] We use a no-arbitrage approach (as opposed to other methods of

[20]We follow a similar approach to those authors when constructing our term structure model.

yield curve interpolation) on the grounds that such an approach is consistent with our assumption of no-arbitrage across derivatives.

We will use lower case symbols to denote scalars or vectors and uppercase symbols to denote matrices, and assume each time period is one month. We follow the convention that all rates and yields are expressed at yearly frequency, so we will scale them by 1/12 to obtain the monthly yield. Let z_t denote a vector of N risk factors (our empirical exercise will have $N = 5$) for our term structure model. We assume that, under the physical (actual) probability measure $\mathbb{P}$, z_t follows a Gaussian AR(1) process,

$$z_{t+1} = k^{\mathbb{P}}_{0,z} + K^{\mathbb{P}}_{1,z} \cdot z_t + (\Sigma_z)^{1/2} \varepsilon^{\mathbb{P}}_{z,t+1}, \varepsilon^{\mathbb{P}}_{z,t+1} \sim N(0, I_N), \tag{15}$$

where I_N is the $N \times N$ identify matrix, $k^{\mathbb{P}}_{0,z}$ is an $N \times 1$ vector of constants, $K^{\mathbb{P}}_{1,z}$ is an $N \times N$ matrix, and Σ_z is a symmetric positive semi-definite $N \times N$ matrix. The intermediary's log stochastic discount factor that prices derivatives is

$$m_{t+1} = -(\delta_0 + \delta_1^T \cdot z_t) - \frac{1}{2} \lambda_t^T \lambda_t + \lambda_t^T \varepsilon^{\mathbb{P}}_{z,t+1}, \tag{16}$$

where $\lambda_t = (\Sigma_z^{-1})(\lambda_0 + \Lambda_1 z_t)$ is the price of risk associated with each shock. We will assume that the profits of derivatives trades are discounted using the OIS curve, consistent with industry practice.[21] That is, $r_t = \delta_0 + \delta_1^T \cdot z_t$, where r_t is the log of the annualized fixed rate associated with a one-month overnight index swap.

This standard specification leads to a risk-neutral ($\mathbb{Q}$) measure dynamics for the state vector z_t,

$$z_{t+1} = k^{\mathbb{Q}}_{0,z} + K^{\mathbb{Q}}_{1,z} \cdot z_t + (\Sigma_z)^{1/2} \varepsilon^{\mathbb{Q}}_{z,t+1}, \varepsilon^{\mathbb{Q}}_{z,t+1} \sim N(0, I_N), \tag{17}$$

where the parameters $k^{\mathbb{Q}}_{0,z}$ and $K^{\mathbb{Q}}_{1,z}$ are functions of the physical measure parameters and the SDF parameters. It also leads zero coupon-swap rates that are affine in the state vector,

$$r_{n,t} = -\frac{12}{n} \ln(E^{\mathbb{Q}}_t[\exp(\sum_{j=0}^{n-1} -\frac{1}{12} r_{t+j})]) = a_n^{swap} + (b_n^{swap})^T z_t, \tag{18}$$

[21] The choice to use OIS rather than some other discount rate does not substantially impact our results, as we use the SDF only to price zero-NPV derivatives, whose value is not sensitive to shifts in the level of the discount rate.

where $r_{n,t}$ denotes log of the annual fixed rate associated with an n-month swap.

The one-month log synthetic dollar rate is likewise assumed to be affine,

$$r_t^{syn} = \hat{\delta}_0 + \hat{\delta}_1 z_t, \tag{19}$$

which generates the recursion

$$r_{n,t}^{syn} = -\frac{12}{n} \ln(E_t^{\mathbb{Q}}[\exp(\sum_{j=0}^{n-1} -\frac{1}{12} r_{t+j}^{syn})]) = a_n^{syn} + (b_n^{syn})^T z_t. \tag{20}$$

We assume that the rates $x_t = (x_{1,t}, x_{2,t}, \frac{1}{12} y_t^{bill})$ are affine functions of our state variable, where $x_{1,t}$ and $x_{2,t}$ are the discount rates associated with our net short and net long curves, and y_t^{bill} is the log annualized yield on a six-month Treasury bill. We assume that

$$x_t = \gamma_0 + \Gamma_1 z_t + (\Sigma_x)^{\frac{1}{2}} \varepsilon_{x,t}, \varepsilon_{x,t} \sim N(0, I_3). \tag{21}$$

These additional variables can be thought of as akin to the "macro factors" often included in standard term-structure models. The key assumption is that the measurement errors $\varepsilon_{x,t}$ do not enter the SDF (and hence have the same distribution under P and $\mathbb{Q}$). The coefficients γ_0 and Γ_1 can be identified from regressions of these factors on the state variables.[22]

Finally, we assume that the dealer unwinds their position when the remaining maturity reaches six months, at which point the Treasury bond is equivalent to a Treasury bill. We make this assumption to capture the empirical observation that, pre-GFC, bill yields fell below other short-term risk-free rates (see e.g. Nagel (2016)). Incorporating this effect is important when constructing bounds on a one-year bond (which will become equivalent to a bill relatively soon) but has a min-

[22] The tri-party repo and secured lending rates (which are essentially the long and short financing rates) are overnight rates. Given data limitations, we use overnight tri-party repo rates and overnight security lending rates to construct x_t. The 1-month CIP basis data are available, but to avoid the quarter-end effect (Du et al., 2018b), we instead use the 3-month CIP basis to obtain the synthetic rate in x_t. Our estimation reveals that there are unit-root elements in the z_t process. A more sophisticated approach is to restrict that the spreads $x_{1,t} - r_t - r_t^{cip}$, $x_{2,t} - r_t + r_t^{cip}$, and $x_{3,t} - r_t$ are stationary, i.e., zero loadings on the unit-root element. Our main approach is the direct regression of x_t on z_t, but we show in the internet appendix that results are similar if we impose stationarity on the spreads. See Internet Appendix Section C for more details.

imal effect on long-maturity bonds. The curves we construct are thus

$$e^{-ny^l_{n,t}} = E^{\mathbb{Q}}_t[\exp(-\sum_{j=0}^{n-7} x_{1,t+j})\exp(-\frac{6}{12}y^{bill}_{t+n-6})], \tag{22}$$

$$e^{-ny^s_{n,t}} = E^{\mathbb{Q}}_t[\exp(-\sum_{j=0}^{n-7} x_{2,t+j})\exp(-\frac{6}{12}y^{bill}_{t+n-6})]. \tag{23}$$

Under the assumptions of our term structure model, these curves are also affine in the state variables.

2.5 Term Structure Estimation Results and Predictions

We obtain data from various public sources, as documented in Internet Appendix Section A. We then estimate the term structure model to fit dollar swap rates and synthetic dollar swap rates, using a standard maximum likelihood approach. The goal of our estimation procedure is to accurately fit and interpolate these curves. Figures 6 and 7 illustrate the fit of our model. For details on the estimation procedure and on related issues such as coupon vs. zero-coupon bonds, see Internet Appendix Section C.6.

In Figure 8, we show the model-implied net long and net short curves, in comparison with the Treasury yield curve. In Figure 9, we subtract matched-maturity OIS rates from all the yield curves in Figure 8.

Several patterns are immediately apparent. Prior to the GFC, yields for one, three, five, and ten-year bonds were often close to the net short curve, consistent with the net position data. In contrast, twenty-year maturity bonds are often close to the net long curve. Our model therefore predicts that dealers would be short intermediate-maturity bonds and long longer-maturity bonds. We validate this prediction in more detailed position data below. We should also note that, because there are substantially more intermediate-maturity than long-maturity bonds outstanding, this pattern naturally generates an overall net short position.[23] After the GFC, all yields for bonds of one-year-maturity or greater are close to the net long yield, suggesting that dealers will be long coupon bonds. This is again consistent with the position data.

[23] Our model has no specific predictions about how dealers should allocate their long/short positions across various arbitrages. In making this argument, we are assuming that, all else equal, larger markets lead to larger positions.

Note that the six-month yield is constructed by regressing bill yields on the factors of our term structure model, and by assumption the net long and net short curves are identical at this maturity. We include it in these graphs to illustrate that, for the most part, our model captures the fluctuations in the bill-OIS spread. These fluctuations play an important role in the movements of the net long and net short curves at the one-year maturity point; they play a relatively minimal role for longer maturities. Intuitively, movements in the six-month bill-OIS spread are amortized over longer maturities and hence have only small effects on longer-maturity yields.

We next compare Treasury yields relative to the net short and net long curves to position data for specific bond maturity buckets. As discussed above, a bond's yield relative to the net short and net long curve bounds serves as a heuristic proxy for dealer positions. We define a "relative yield index" by

$$pos_{n,t} = 2 * \frac{y_{n,t} - y^s_{n,t}}{y^l_{n,t} - y^s_{n,t}} - 1. \tag{24}$$

This index takes on a value of one if the yield of the n-month maturity Treasury bond is equal to the net long yield, negative one if it is equal to the net short yield, and zero if it is equal to the average of the net long and net short yields.

We then plot this relative yield index against the net dealer position by maturity bucket. We obtain the net primary dealer coupon Treasury bond position in maturity buckets of <3 years, 3-6 years, 6-11 years, and >11 years from the FR2004 primary dealer statistics published by the Federal Reserve Bank of New York. We then normalize each of these by the total assets of primary dealers, and plot them with the relative yield index at the 2-year, 5-year, 10-year, and 20-year maturities. The bond position and relative yield index are plotted on different axes (because they are not in comparable units), with the zero points on each axis aligned.

For a variety of reasons, we do not expect these series to perfectly align. First, as discussed above, the mapping between how close a yield is to the net long or net short curves and the predicted net dealer position in that maturity is unclear, and may change over time. Second, the arbitrage bounds we construct are motivated by trading strategies that (at least potentially) hold the bond almost to maturity. Dealers also intermediate bonds between clients, and may be willing to buy a bond for the purpose of selling it quickly even if they view the bond as overpriced (close to the net short yield). This kind of intermediation activity acts as a kind of noise in the relationship between net dealer positions and the relative yield index, and is likely accentuated when looking at specific maturity buckets as opposed to overall net dealer positions. Despite these caveats, there is a non-trivial correspondence between the relative yield index and actual positions, as shown in Figure 10. We also construct a weighted average version of the relative yield index, where the weight for each maturity is the fraction of dealer Treasury bond holding at that maturity over total dealer Treasury bond holding. Then we plot this aggregate relative yield index together with total dealer Treasury holding scaled by dealer balance sheet size, as shown in Figure 11.

Summarizing our analysis thus far, Treasury yields have moved from being close to net short arbitrage bounds pre-GFC to being close to net long arbitrage bounds post-GFC, and net primary dealer positions have responded by switching from being net short to net long. Strikingly, the net short and net long curves are constructed by assuming that dealers will remain net short or long going forward. Our results therefore imply that, pre-GFC, dealers were expected by the market to maintain a net short position, and that following the GFC, this expectation flipped and the market now anticipates dealers maintaining a net long position going forward. The relationship between yields and positions we document is consistent with the view that balance-sheet-constrained dealers act as arbitrageurs between the Treasury and swap markets.[24] We next consider the implications of this perspective, with an emphasis on the causes and consequences of the regime shift we have documented.

3 A Model of the Treasury Market

Thus far, we have said little about how the Treasury regime is determined. In this section, we build a supply-and-demand model to endogenize dealers' net position, as a function of their balance sheet constraint, demand for Treasury bonds from non-dealers, and the overall supply of Treasury bonds. This model helps explain the change in the Treasury regime pre- and post-GFC, the striking correlation between the slope of the yield curve and the dealer position in Figure 2, and fragility of the Treasury market when dealers' balance sheet constraints are tight.

[24] We interpret our results as showing that Treasury yields are often at or near arbitrage bounds given swap prices. However, one could equally say that swap yields are at or near arbitrage bounds given Treasury yields, adopting the perspective of Hanson, Malkhozov, and Venter (2022).

3.1 Model Setup

The model has two dates (zero and one), and a single n-period Treasury bond. Date one exists only for the purpose of determining payoffs; all of our analysis will focus on date zero, and we will omit time subscripts for all date zero rates and yields. We will take as exogenous the date zero log interest rates y^{bill}, i^l, i^s, and r, as well as two different expectations concerned future bond prices. We define the dealer's date zero risk-neutral ($\mathbb{Q}$-measure) expectation of date one bond prices as

$$\begin{aligned} p_{\mathbb{Q}} \equiv \exp(-(n-1)y_{\mathbb{Q}}) &\equiv E^{\mathbb{Q}}[\exp(-(n-1)y_{n-1,1})] \\ &= E^{\mathbb{Q}}[\exp(-(n-1)r_{n-1,1} + (n-1)(r_{n-1,1} - y_{n-1,1}))] \end{aligned} \tag{25}$$

where $r_{n-1,1}$ and $y_{n-1,1}$ denote the $(n-1)$-period swap rates and Treasury bond yields at date one, and $y_{\mathbb{Q}}$ denotes the risk-neutral expectation-implied yield at date zero. We define the corresponding physical measure ($\mathbb{P}$) counterpart as $y_{\mathbb{P}}$, so the physical-measure expected future bond price is $\exp(-(n-1)y_{\mathbb{P}})$. Note that the dealer's SDF and associated risk-neutral measure $\mathbb{Q}$ price derivatives, but will generally not price positive-investment assets (such the Treasury bills).

We have written the definition of $y_{\mathbb{Q}}$ in this way to highlight the possible interpretations of comparative statics with respect to $y_{\mathbb{Q}}$. One interpretation, which we emphasize, is that a decrease in $y_{\mathbb{Q}}$ represents a change in the swap curve holding constant the risk-neutral expectation of future swap-Treasury spreads. An equally valid interpretation is as a change in the risk-neutral expectation of future swap-Treasury spreads holding the swap curve constant. It is important to distinguish between comparative statics that hold $y_{\mathbb{P}}$ constant and comparative statics that change both $y_{\mathbb{P}}$ and $y_{\mathbb{Q}}$. The first of these represents a change in risk premia, that latter a change in expected future rates.

These interest rates and expected bond prices will allow us to compute the dealer's net short and net long curves. The key endogenous variables are y, the yield of an n-period zero-coupon bond at date zero, and r^{syn}, the one-period synthetic unsecured risk-free rate at date zero. We focus on the following asset prices that are closely related to the empirical motivations in Figure 1.

- The Treasury term spread, $y - y^{bill}$.
- The synthetic–swap spread, $r^{syn} - r$, which maps to CIP deviations.

- The n-period swap spread $r_n - y$, where

$$\exp(-nr_n) \equiv E^{\mathbb{Q}}[\exp(-r-(n-1)r_{n-1,1})] \tag{26}$$

- The swap term spread, $r_n - r$, which contains both an expectation component and a risk premium component.

We will treat dealers and their levered clients as a single consolidated entity, based on the analysis of Internet Appendix Section E.1. In the Treasury market, dealers will interact with two kinds of investors. Hedged investors purchase the Treasury bond and swap it to their local currency. Unhedged investors choose between the Treasury bond and Treasury bills. Dealers also have other counterparties in the synthetic lending market; these other counterparties do not participate in the Treasury market. This structure implicitly assumes that the tri-party repo, bill, and interest rate swap markets are infinitely elastic, whereas the Treasury and synthetic borrowing markets are elastic but not infinitely so; we make these assumptions to simplify our exposition.

Our model of dealers is exactly that of section 2.1, with $p_{\mathbb{Q}}$ defined as in (25). Recall that q^{syn} is the quantity (in dollars) of synthetic loans supplied by the dealers at date zero, and q^{bond} be the quantity (in dollars) of bonds owned (positive) or short-sold (negative), and that the dealer balance sheet constraint is $q^{syn} + |q^{bond}| = \bar{q}$.

Let S^{bond} be the supply of bonds (in notional quantities) at date zero, and let D_U^{bond} and D_H^{bond} be the demand (in dollars) for bonds from unhedged and hedged investors, respectively. Market clearing in the bond market requires

$$q^{bond} + D_U^{bond} + D_H^{bond} = \exp(-ny)S^{bond}. \tag{27}$$

The bond price $\exp(-ny)$ enters this expression to convert the notional supply S^{bond} into dollars.

Unhedged investor demand is a continuously differentiable, strictly positive and increasing function of the expected log excess return of the bond over bills,[25]

$$D_U^{bond} = D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}). \tag{28}$$

[25] As documented in Haddad and Sraer (2020), typical bank portfolios behave like unhedged investors in our model in that their position in long-term bonds is increasing the expected excess return of long-term Treasury bonds.

Likewise, hedged investor demand is a continuously differentiable, strictly positive and increasing function of the expected log excess return in dollars[26] on the hedged strategy:

$$D_H^{bond} = D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}). \tag{29}$$

When a dealer helps a hedged investor exchange e.g. yen into dollars and hedge using forwards, the dealer will end up with a yen asset (cash) and a dollar liability. As a result, this activity increases the dealer's balance sheet, and is functionally equivalent, from the dealer's perspective, to lending synthetic dollars.

There is an important role for market segmentation in these equations. Both types of Treasury clients are assumed to care about $y_{\mathbb{P}}$ and not $y_{\mathbb{Q}}$, because they trade Treasury bonds but not swaps. Segmentation between constrained agents (dealers) and unconstrained agents (clients), whether endogenous (e.g., Alvarez and Jermann (2000); Chien, Cole, and Lustig (2011); Biais, Hombert, and Weill (2021)) or exogenous (e.g., Gertler and Kiyotaki (2010); He and Krishnamurthy (2013)), is necessary to generate Treasury-swap arbitrage.

Dealers can also lend synthetic dollars to other counterparties (hedged investors buying corporate bonds, for example). We assume that the demands of these other investors for synthetic dollars are $D^{syn}(r^{syn} - r)$, where D^{syn} is a continuously differentiable, non-negative and strictly decreasing function of the spread between synthetic dollars and risk-free rates. Market clearing in the synthetic dollar market requires that

$$q^{syn} = D_H^{bond} + D^{syn}(r^{syn} - r). \tag{30}$$

These market clearing conditions, together with the net long and net short inequalities in our simple model (equations (3) and (4)) and the associated implications for dealer positions, define our model.

To guarantee that an equilibrium exists in our model, we need interior solutions for (y, r^{syn}) to satisfy the two market clearing conditions in (27) and (30). Thus, we make the following technical assumptions.

Assumption 1. *We assume that the demand functions are well-behaved:*

[26]For simplicity, we use the hedged return in dollars, as opposed to in local currency; this allows us to ignore second-order terms associated with the covariance between interest rates and exchange rates.

- *Excess synthetic loan demand is possible:* $D^{syn}(0) > \bar{q}$.

- *Excess synthetic loan supply is possible: For all y,* $\lim_{r^{syn} \longrightarrow \infty} D^{syn}(r^{syn} - r) + D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) = 0$.

- *Excess bond supply is possible:* $\lim_{y \longrightarrow -\infty} D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) + D_H(ny - r - (n-1)y_{\mathbb{P}}) = 0$.

Note that excess bond demand is also possible without the need to impose additional assumptions. This is because if we take $y \to \infty$, the right hand side of (27) becomes zero while the left hand side is strictly positive, causing an excess bond demand. We will show that with Assumption 1, the equilibrium solution to the model exists and is unique.

Depending on the sign of q^{bond}, our static model features three possible kinds of equilibria, which we refer to as regimes: a long regime $q^{bond} > 0$, a short regime $q^{bond} < 0$, and an intermediate regime $q^{bond} = 0$. We discuss each of these regimes in turn.

Our focus, when analyzing these regimes, will be on the spread between synthetic dollars and the swap rate, $r^{syn} - r$, which is endogenously determined. In our quantitative analysis, this spread was a key input to the model, and we measured it with CIP violations. In this two-market market, the rate r^{syn} should be understood as the risk-free return the dealer requires for assets held on balance sheet, as opposed to specifically a euro rate swapped to dollars. Under this interpretation, the spread $r^{syn} - r$ is the kind of financial intermediation spread that plays a key role in macro-finance models (Brunnermeier and Pedersen, 2009; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014; Gertler and Kiyotaki, 2010). For this reason, we emphasize how this financial intermediation spread responds to shocks.

3.2 The Long Regime

We first consider a regime in which dealers are long ($q^{bond} > 0$), in which case $e^{-ny} = e^{-ny^l}$, where the net-long yield y^l is defined in (3) and $p_{\mathbb{Q}}$ can be expressed in terms of $y_{\mathbb{Q}}$ as in (25). Thus,

$$e^{-ny} = \frac{e^{-(n-1)y_{\mathbb{Q}}}}{e^{i^l} + e^{r^{syn}} - e^r}. \tag{31}$$

Equation (31) generates a dealer indifference condition between the two endogenous variables r^{syn} and y. This indifference condition suggests that r^{syn} strictly increases with y. Intuitively, the more attractive it is to buy the Treasury bond and hedge with swaps (higher y), the higher the return on synthetic lending (r^{syn}) must be to generate indifference between these two activities.

Combining the market clearing conditions in (27) and (30), the intermediary balance sheet constraint in (1), and using $q^{bond} > 0$, we have

$$\bar{q} - e^{-ny} S^{bond} + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) = D^{syn}(r^{syn} - r). \tag{32}$$

The left-hand side of (32) represents the residual balance sheet available for synthetic lending, which in equilibrium must equal the residual demand for synthetic lending. Note that the demand from hedged investors does not appear in this equation, because the dealer balance sheet is unaffected by changes in their demand, holding all else constant.[27] The left-hand side of (32) is strictly increasing in y, while the right-hand side is strictly decreasing in r^{syn}. Therefore, equation (32) generates a kind of market indifference condition, where r^{syn} strictly decreases with y. Intuitively, higher yields lead to more investor demand for bonds, which reduces the balance sheet dealers allocate to bonds, thereby increasing the balance sheet allocated to synthetic dollar lending and reducing the synthetic dollar lending spread.

A long-regime equilibrium (y, r^{syn}) is a point where these two indifference curves intersect and $q^{bond} > 0$, which requires

$$D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) < e^{-ny} S^{bond}. \tag{33}$$

Because the two indifference curves have opposite slopes, such an equilibrium is unique if it exists.

We next consider various comparative statics associated with a long-regime equilibrium: an increase in bond supply (S^{bond}), dealer balance sheet ($\bar{q}$), an increase in unhedged bond demand (a parallel increase in D_U), an increase in hedged bond demand (a parallel increase in D_H), an increase in the swap market term premium (an increase in $y_{\mathbb{Q}}$ holding $y_{\mathbb{P}}$ constant), and an increase in expected future rates (a parallel increase in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$). The following proposition summarizes our results.

[27] An increase in hedged investor demand reduces the quantity of Treasury bonds dealers must hold, but increases the amount of synthetic borrowing they must finance, and hence has no effect on their balance sheet usage, provided that dealers have a net long bond position.

Proposition 1. *In a long regime equilibrium, holding all else constant,*

1. *An increase in S^{bond} leads to an increase in y and an increase in r^{syn},*
2. *A decrease in $\bar{q}$ or a parallel decrease in D_U is equivalent to a expansion of the same size in the dollar supply of bonds.*
3. *A parallel increase in D_H does not change either y or r^{syn},*
4. *An increase in $y_{\mathbb{Q}}$ leads to an increase in y and a decrease in r^{syn},*
5. *An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ leads to an increase in y of less than $\frac{n-1}{n}dy$ and a decrease in r^{syn}.*
6. *A parallel increase in $D^{syn}(\cdot)$ increases both y and r^{syn}.*

Proof. See Internet Appendix Section D.1. $\square$

An increase in bond supply means that, holding yields constant, less dealer balance sheet is available for synthetic lending. As a result, the market indifference curve increases for each y, which leads to an increase in both synthetic lending spreads and bond yields in equilibrium. A decrease in either dealer balance sheet capacity or in unhedged bond demand is equivalent to an increase in the supply of bonds, and hence generates the same comparative statics. In contrast, an increase in hedged demand has no effect on the equilibrium, because this demand has no net effect on dealer balance sheet constraints; it merely transforms dealer bond holdings into dealer synthetic lending.

An increase in $y_{\mathbb{Q}}$ holding $y_{\mathbb{P}}$ constant is equivalent to an increase in forward swap rates holding expected future swap rates constant, and hence to an increase the swap term premium. Such a change makes owning the Treasury bond and hedging with swaps less attractive to dealers, and as a result the dealer indifference curve decreases for each y. This in turn leads to an increase in y and a decrease in r^{syn}.

An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ represents an increase in expected future rates holding the swap term premium constant. Holding fixed the dollar supply of bonds, such a change would have no effect on r^{syn} and would lead to an increase in y of $\frac{n-1}{n}dy$. However, because it is the notional and not dollar supply of bonds that is held constant, an increase in yields generates a contraction in the dollar supply of bonds. This contraction, in the long regime, has the effect of reducing synthetic lending spreads and dampening the increase in bond yields.

The two indifference curves derived from (31) and (32), and the first and fourth comparative statics in Proposition 1, are illustrated in Figure 12. The market indifference curves are truncated when yields are sufficiently high; this truncation highlights that the long regime equilibrium ceases to exist when client demand exceeds the bond supply.

3.3 The Short Regime

We next consider a regime in which dealers are short ($q^{bond} < 0$). In this regime, the bond yield must exactly equal to the net short yield, $e^{-ny} = e^{-ny^s}$, where the short yield y^s is defined in (4). Expressing $p_{\mathbb{Q}}$ with $y_{\mathbb{Q}}$ as in (25), we obtain

$$e^{-ny} = \frac{e^{-(n-1)y_{\mathbb{Q}}}}{e^{r^{sec}} + e^{r} - e^{r^{syn}}} \tag{34}$$

Equation (34) generates a dealer indifference condition, where r^{syn} strictly decreases with y. Intuitively, the less attractive it is to sell the Treasury bond and hedge with swaps (higher y), the lower the return on synthetic lending (r^{syn}) must be to generate indifference between these two activities. This relationship has the opposite sign compared to the long regime.

Combining the market clearing conditions in (27) and (30), balance sheet constraint (1), and using $q^{bond} < 0$, we obtain

$$\bar{q} + e^{-ny}S^{bond} - D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) - 2D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) = D^{syn}(r^{syn} - r). \tag{35}$$

The left-hand side of (35) again represents the residual balance sheet available for synthetic lending, which in equilibrium must equal the residual demand for synthetic lending. Note that demand from hedged investors has a double impact on the dealer's balance sheet. All else equal, an increase in this demand will result in dealers taking larger short positions and providing more synthetic financing to hedged investors, both of which use up dealer balance sheet.

The left-hand side of (35) is strictly decreasing in y: unlike the long regime, more demand from investors and less supply require dealers to take larger short positions, using up more balance sheet. Equation (35) therefore generates a kind of market indifference condition, where r^{syn} strictly increases in y.

A short-regime equilibrium (y, r^{syn}) is a point where these two indifference curves intersect and $q^{bond} < 0$, which requires

$$D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) > e^{-ny}S^{bond}. \tag{36}$$

Because the two indifference curves have opposite slopes, such an equilibrium is again unique if it exists.

Our next proposition considers the same set of comparative statics studied previously in the context of the short regime.

Proposition 2. *In a short regime equilibrium, holding all else constant,*

1. *An increase in S^{bond} leads to an increase in y and a decrease in r^{syn};*
2. *An increase in $\bar{q}$ or a parallel decrease in D_U is equivalent to an expansion of the same size in the dollar supply of bonds;*
3. *A parallel increase in D_H leads to a decrease in y and an increase in r^{syn};*
4. *An increase in $y_{\mathbb{Q}}$ leads to an increase in y and an increase in r^{syn};*
5. *An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ leads to an increase in y by less than $\frac{n-1}{n}dy$ and an increase in r^{syn}.*
6. *A parallel increase in $D^{syn}(\cdot)$ increases r^{syn} and decreases y.*

Proof. See Internet Appendix Section D.2. □

An increase in bond supply in the short regime increases yields (like the long regime) but decreases synthetic lending spreads (unlike the long regime). In the short regime, the larger the supply of bonds the smaller the dealer's required short position, and hence more balance sheet is available for synthetic lending.

Like the long regime, a decrease in unhedged bond demand is equivalent to an increase in bond supply. Unlike the long regime, an increase in dealer balance sheet capacity is equivalent to an increase in supply, because dealers are short bonds instead of being long bonds. Also unlike the long regime, an increase in hedged bond demand leads to a decrease in yields and an increase in synthetic lending spreads; it is equivalent to a contraction in the dollar supply of bonds of twice the magnitude of the demand increase.

An increase in the swap curve term premium makes shorting bonds and hedging with swaps more attractive. The dealer therefore requires a higher synthetic lending spread to be indifferent between synthetic lending and shorting Treasury bonds, which leads in equilibrium to higher yields and higher synthetic lending spreads (the opposite of the long regime).

Figure 13 illustrates the dealer indifference curve (34), the market indifference curve (35), and the first and fourth comparative statics discussed in Proposition 2.

3.4 The Intermediate Regime

The last regime we consider is one in which $q^{bond} = 0$.[28] In this regime, the yield must fall between the net short and net long yields,

$$y^s \leq y \leq y^l, \tag{37}$$

the bond market must clear without dealers taking a position,

$$D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) = e^{-ny}S^{bond}, \tag{38}$$

and the dealer balance sheet constraint is reduced to $q^{syn} = \bar{q}$. Equating q^{syn} with synthetic lending demand, we obtain

$$\bar{q} = D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) + D^{syn}(r^{syn} - r). \tag{39}$$

The following proposition summarizes the comparative statics in this case. We restrict attention to interior intermediate equilibria, for which $y^s < y < y^l$, and discuss the determinants of regime boundaries below.

Proposition 3. *If an interior intermediate regime exists, it is the only such equilibrium. In an interior intermediate regime equilibrium, holding all else constant,*

1. *An increase in S^{bond} leads to an increase in y and an increase in r^{syn};*

[28] In position data, dealers will never have an exactly zero net Treasury position, for reasons (for example, intermediation activities) that are outside the scope of our model. We view the intermediate regime as describing a situation in which dealers are targeting a roughly net flat position.

2. *An increase in $\bar{q}$ or a parallel increase in D_U leads to a decrease in both y and r^{syn};*

3. *A parallel increase in D_H leads to a decrease in y and an increase in r^{syn};*

4. *An increase in $y_{\mathbb{Q}}$ leaves both y and r^{syn} unchanged;*

5. *An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ leads to an increase in y by less than $\frac{n-1}{n}dy$ and a decrease in r^{syn}.*

6. *A parallel increase in $D^{syn}(\cdot)$ increases both r^{syn} and y.*

Proof. See Internet Appendix Section D.3. □

When dealers are not active in the Treasury market, increases in supply lead to higher yields as a consequence of clients demanding higher expected returns in exchange for holding larger positions. Some of these clients are hedged clients, whose increase position size requires additional synthetic dollar financing from dealers, reducing those dealer's ability to lend to other synthetic dollar clients, which results in increasing synthetic lending rates.

As usual, an increase in unhedged demand is equivalent to a decrease in supply. However, the comparative statics with respect to balance sheet and hedged demand are unlike either the long or short regime; both effects are driven by the role of the hedged demand in the Treasury market.

Unlike either the long or the short regime, in the intermediate regime the OIS term premium is disconnected from Treasury yields. Dealers are not actively arbitraging bonds and swaps, and neither type of client trades swaps; as a result, changes in the swap market term premium do not affect the Treasury market. In contrast, increases in expected yields lead to higher yields, exactly as in both the long and the short regime.

3.5 Equilibrium

Let us next consider the factors that determine which of three regimes occur in equilibrium. The following proposition summarizes how some of the comparative statics discussed thus far can change the equilibrium regime.

Proposition 4. *The equilibrium exists and is unique given the exogenous parameters. Holding all else constant,*

1. *There exists an $0 \leq S_S \leq S_B \leq \infty$ such that a short regime equilibrium exists for all $S^{bond} < S_S$, a long regime equilibrium exists for all $S^{bond} > S_B$, and an intermediate regime equilibrium exists for $S^{bond} \in [S_S, S_B]$.*

2. *There exists an $0 \leq y_{Q,S} \leq y_{Q,B} \leq \infty$ such that a short regime equilibrium exists for all $y_{\mathbb{Q}} < y_{Q,S}$, a long regime equilibrium exists for all $y_{\mathbb{Q}} > y_{Q,B}$, and an intermediate regime equilibrium exists for $y_{\mathbb{Q}} \in [y_{Q,S}, y_{Q,B}]$.*

Proof. See Internet Appendix Section D.4. □

Intuitively, when bonds are scarce, dealers will be short to meet client demands, while when bonds are abundant dealers will be long to fill the shortfall in client demand.

Less intuitively, when the yield curve is steep dealers will be short bonds. Considering the dealer's bond position in isolation, this looks like a money-losing strategy: expected returns are higher when the yield curve is steeper. However, if the swap curve has a higher term premium than the Treasury curve and the dealer is hedging with swaps, then selling bonds and hedging is in fact a profitable strategy. Similarly, buying bonds when the curve is flat looks like a money-losing strategy, but is in fact profitable if the swap curve is even flatter and the dealer hedges. In our model, client demand for Treasury bonds is driven by the expected returns on those bonds; spreads must therefore move in a way that induces dealers to take the opposite position.

Figure 14 illustrates the comparative statics of the model with respect to changes in the swap curve term premium. The swap-Treasury spread is an increasing function of the swap term premium, but the rate at which it increases depends on the regime. In contrast, the synthetic lending spread is U-shaped: it is high in both the long and short regimes, and low in the intermediate regime. Lastly, the dealer Treasury position as a fraction of its capacity $\bar{q}$ decreases with the OIS term premium, consistent with the motivating facts in Figure 2.

3.6 Model vs. Data

We now return to the motivating facts illustrated in Figures 1 and 2 and discuss them through the lens of our model. We divide our discussion between the pre-GFC and post-GFC periods.

Yield Curve Slope vs. Dealer Position The model is able to explain why dealers' net long position is larger when the term spread (slope of the yield curve) is lower. This pattern is documented in Figure 2, and is at first puzzling in light of the fact that the slope positively predicts future bond returns (Campbell and Shiller (1991)). That is, if dealers held more Treasury bonds when the expected returns on those bonds are higher (as in Jermann (2020)), we would expect a positive relationship between bond holdings and the term spread.

Our model resolves the puzzle by assuming that client demand is increasing in expected excess bond returns. Dealers accommodate client demand by acting as arbitrageurs between the bond and swap markets, which leads naturally to the result that their Treasury positions declines in expected excess bond returns, and hence the term spread. Meanwhile, dealers justify a larger position by charging a larger swap spread.

Pre-GFC. Prior the GFC, CIP violations were close to zero, swap spreads were positive, and dealers were net short (Figure 1). During this period, dealers were able to use large amounts of leverage for the purpose of engaging in arbitrage trades, and Treasury bonds were comparatively scarce (relative to the supply in the post-GFC period).

Our model reconciles these facts in the following way. First, because of the scarce supply of Treasury bonds, dealers ended up net short pre-GFC, absorbing the excess client demand for Treasury bonds (Proposition 4 part 1). Because dealers had ample balance sheet capacity, CIP violations remained small (Proposition 2 part 2). Because dealers were net short, Treasury yields ended up on the net short curve (34), which is below the OIS swap curve due to funding spreads ($i^s < r$) and low Treasury bill yields (which lead to a lower $y_{\mathbb{Q}}$). Consequently, swap spreads were positive.

To illustrate this point, in Figure 15, we replicate Figure 14, with a relatively high balance sheet capacity for dealers, relatively low bill yields, and a relatively small supply of Treasury bonds. Figure 15 shows that our model, with these parameters, predicts a small and roughly constant CIP spread, a positive swap spread, and a negative dealer Treasury position.

Thus, given a scarce bond supply, ample dealer leverage, and low Treasury bill yields (relative to other rates), our model can replicate the stylized facts illustrated in Figure 1 for the pre-GFC period.

Post-GFC. After the GFC, CIP violations were significant, swap spreads were negative, and there was a strong correlation between the two (Figure 1). Dealers were net short, and the size of their position was strongly negatively correlated with the slope of the yield curve (Figure 2). Dealers faced tight leverage constraints as part of the post-GFC regulatory regime (Duffie (2017)), and there was a large increase in the supply of Treasury bonds.

Our model reconciles these facts in the following way. First, because of the large supply of Treasury bonds, dealers ended up net long post-GFC, absorbing the demand shortfall (relative to supply) from clients (Proposition 4 part 1). Because dealers had tight balance sheet constraints, CIP violations became large (Proposition 1 part 2). Because dealers were net long, Treasury yields ended up on the net short curve (31), which is above the OIS swap curve mainly due to balance sheet costs (funding spreads ($i^l < r$) are small, and in the post-GFC regime bill yields are close to r). Consequently, swap spreads are negative. Moreover, because swaps spreads are now driven primarily by balance costs, they are now strongly correlated with other measures of balance sheet costs (CIP violations). Lastly, because Treasury yields are on the net long curve, dealers are willing to absorb changes in client demand for Treasury bonds. Because clients seek Treasury returns ((28) and (29)), and therefore demand more when the term spread is higher, dealer net positions are negatively correlated with the term spread (by market clearing).

We again replicate Figure 14 in Figure 16, but with a low balance sheet capacity for dealers, no spread between bill and funding rates, and a relatively large supply of Treasury bonds, as a way of capturing the post-GFC period. Figure 16 shows that our model, with these parameters, predicts significant CIP violations, negative swap spreads, co-movement between these two in response to shocks, and a positive dealer Treasury position.

Thus, given a large bond supply, limited dealer leverage, and Treasury bill yields comparable to other rates, our model can replicate the stylized facts illustrated in Figures 1 and 2 for the post-GFC period.

Pre- vs. Post-GFC. Our analysis shows that a combination of three factors (a large increase in Treasury supply, a tightening of dealer balance sheet constraints, and an increase in bill yields relative to other rates) are sufficient in our framework to explain the differences we document between the pre- and post-GFC periods.

Figure 3 illustrates the first two of these changes. The U.S. government borrowed a large amount during the financial crisis and continued running substantial deficits in the years that followed. As a result, the outstanding of marketable Treasury securities grew from \$4.7 trillion in 2008 to \$22.5 trillion as of June 2022. This resulted in a large increase in the supply of Treasury bonds, even after accounting for the Federal Reserve's bond purchases. At the same time, regulatory reforms lead to a reduction in the size of dealer balance sheets. We speculatively attribute the third change (a relative increase in bill yields) to the payment of interest on excess reserves.[29]

There are of course many other differences between the pre- and post-GFC periods. There are also relevant factors in the model (in particular, the shape of the demand curves of clients) that may have changed between these period and about which we have little information. For these reasons, we can show that the three changes we emphasize above are sufficient to explain the observed differences between the pre- and post-GFC periods, but cannot prove that all of the differences were caused by these forces and not some other changes between the two periods.

Counterfactuals. Our model allows us to consider two related counter-factual scenarios. We ask (i) what if balance sheet constraints were tight pre-GFC, and (ii) what if Treasury supply remained scarce post-GFC, in both cases holding all else equal.

Both of these counter-factual scenarios involve the short regime, in the first case because dealers were in fact short pre-GFC and in the second case as result of the hypothetical Treasury scarcity post-GFC. Both scenarios also involve tight balance sheet constraints, in the first case by assumption and in the second because balance sheet capacity was scarce post-GFC.

The combination of balance sheet scarcity and the short regime would lead our model to predict large and positive swap spreads. These swap spreads would be caused both by funding spreads ($i^s < r$) and balance sheet costs, and could easily be 100 basis points higher than the negative swap spreads observed post-GFC (under the assumption of a 20bps difference between i^s and i^l and a 40bps CIP violation). With regards to the post-GFC counterfactual, our model therefore implies that, treating swap rates as fixed, Treasury yields are substantially higher than they would have been in the absence of a large increase in supply.

[29] In the pre-GFC period, banks were significant holders of Treasury bills at rates well below tri-party repo rates. After the Fed began paying interest on reserves at rates higher than tri-party repo rates, banks substantially reduced their bill ownership, and government-only money market funds substantially increased their bill ownership. Such funds can invest in both bills and tri-party repo; as result, bill yields rose to roughly the level of tri-party repo rates.

3.7 Regimes and Treasury Market Fragility

During the financial crisis of 2008-2009, Treasury yields fell by more than matched maturity swap rates. During the COVID-induced financial turmoil of March 2020, the reverse was true: Treasury yields did not fall by as much as swap rates, and in fact briefly rose (Duffie, 2020; Haddad, Moreira, and Muir, 2021; He, Nagel, and Song, 2022). The different comparative statics across the long and short regimes in our model offer an explanation for this pattern.

In the short regime, an increase in balance sheet costs (as measured by the spread $r^{syn} - r$), all else equal, will lead to lower Treasury yields (Proposition 2). In contrast, in the long regime, an increase in balance sheet costs will lead, again all else equal, to higher Treasury yields (Proposition 1). Both crises were characterized by large increases in arbitrage spreads; the difference was that the market was in the short regime pre-GFC and in the long regime post-GFC.

He, Nagel, and Song (2022) attribute the differences between these two episodes to client demand for Treasury bonds (a dash-for-cash in COVID, a flight-to-safety in the GFC). Our story is compatible with theirs, in the sense that Treasury selling in COVID would increase balance sheet costs (the long regime) and Treasury buying in the GFC would also increase balance sheet costs (the short regime). However, our story does not rely customer demand for Treasury bonds as the causal factor behind the increase in balance sheet costs. In both the GFC and COVID episodes, even if clients had not bought or sold Treasury bonds on net, we expect that balance sheet costs would have risen, and as a result predict that Treasury yields would have moved relative to swap rates upwards in COVID but downwards in the GFC. Quantifying the role of Treasury demand, as opposed to other forces, in explaining the tightening of balance sheet constraints in these episodes is an interesting direction for future research.

4 Implications for Policy

In this section we consider a variety of Federal Reserve policies, and study how the effects of those policies depend on the regimes we have identified in the Treasury market. The specific policies we consider are interest rate policies (including forward guidance), swap lines with other central banks, supplementary leverage ratio (SLR) exemptions, and quantitative easing/tightening (QE/QT). Our analysis will emphasize the way in which the direction of the effects of these policies depends on the Treasury market regime; drawing quantitative conclusions would require estimating client demand curves. We will use the comparative statics described previously, and focus on the effects of the aforementioned policies on Treasury yields and on synthetic dollar rates. As discussed above, we view synthetic dollar rates as a proxy for financial intermediation spreads more generally.

Our framework treats the term structure of swap rates as exogenous. We first discuss of the effects of interest rate policies that change the level and slope of the swap curve. Then in our analysis of the subsequent three policies (swap lines, SLR exemptions, QE/QT), we focus our discussion on the direct quantity and balance sheet effects of these policies, without discussing the additional effects of these policies on the swap curve and expectations of future rates (i.e. keeping $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ unchanged in our static model).

Before discussing these policies, we should emphasize that all of these policies have effects on inflation, real economic activity, and financial stability that are outside the scope of our model. Policies that increase arbitrage spreads and other financial market distortions can be justified on these grounds. However, combinations of the policies we discuss might achieve these same objectives while avoiding financial market distortions, and it is the goal of our analysis to highlight these possibilities.

Table 1: Summary of Policy Implications

Policy Type	Long Regime		Short Regime	
	Tsy Yield	Lending Rate	Tsy Yield	Lending Rate
$\downarrow$ Term premium	$\uparrow$	$\uparrow$	$\uparrow$	$\downarrow$
Swap line	$\downarrow$	$\downarrow$	$\uparrow$	$\downarrow$
SLR Exemptions	$\downarrow$	$\downarrow$	$\uparrow$	$\downarrow$
QT (purchasing bills, selling bonds)	$\uparrow$	$\uparrow$	$\uparrow$	$\downarrow$

Notes: Tsy yield is the long-term Treasury yield y. Lending rate is the synthetic lending rate r^{syn}.

Table 1 summarizes our results, which we subsequently explain in more detail. Setting aside the specifics associated with each policy, our main message is that the effects of each of the policies depend on the Treasury market regime.

4.1 Interest Rate Policy

We first consider the effects of policies that affect swap rates and rate expectations, including both rate hikes and forward guidance. That is, we define interest rate policy as controlling the current level and future expectations of the federal funds rate (the floating rate for interest-rate swaps).[30] Our analysis will consider the relationship between shocks to current and future federal funds rates and shocks to Treasury yields.

The comparative statics of our model distinguish between changes in term premium (changes in $y_{\mathbb{Q}}$ holding $y_{\mathbb{P}}$ constant) and changes in expectations (equal changes in both values). Monetary policy likely changes both expectations and risk premia. Hanson and Stein (2015) and Hanson, Lucca, and Wright (2021) argue that rate hikes increase both expected future rates and term premia. There is also evidence that forward guidance affects risk premia in addition to rate expectations (e.g. Rogers, Scotti, and Wright (2018)).

To avoid taking a stand on the exact decomposition between interest rate policy and risk premia, we will describe interest rate policy in terms of the level and slope of the swap curve, under the premise that a steep slope implies a high risk premium and a flat or inverted slope implies a low risk premium.

A high level of expected rates (a parallel increase in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$) is equivalent, in our model, to a contraction in the dollar supply of bonds, because the supply of bonds is assumed to fixed in notional terms. This effect is potentially offset by forces outside our model– for example, the federal government might issue more debt to cover the additional interest costs associated with high rates. For this reason, we do not emphasize it as the main effect.

Instead, we focus on the slope of the term structure and the term premium. A low term premium in our model in the long regime leads to a contraction in the client demand for Treasury bonds, a further build-up in the dealer's long position, an increase in bond yields, more negative swap spreads, and an increase in synthetic lending spreads (see part four of Proposition 1 or Figure 16). In the short regime, a low term premium instead leads to a decline in synthetic lending spreads (see part four of Proposition 2).

[30] This sidesteps the issue of how administered rates (such as the interest on reserves rate and ONRRP rates) transmit to the federal funds market, which is beyond the scope of our model.

4.2 Central Bank Swap Lines

We next consider the policy of establishing swap lines between the Federal Reserve and other central banks. These swap lines allow foreign central banks to borrow dollars from the Federal Reserve, using their own currency as collateral. The foreign central banks then lend those dollars to their local banks, typically for the purpose of financing a position in dollar-denominated assets.

This procedure allows non-U.S. banks to borrow dollars, and is a substitute for borrowing synthetic dollars via a dealer.[31] We therefore incorporate swap lines into our static model as equivalent to a demand shift in the synthetic lending market (a parallel shift in D^{syn}). The swap lines establish by the Federal Reserve generally have rates determined by policy. If the rate is higher than the prevailing market rate, the facility will go unused, and the equivalent demand shift is zero. If the rate is appealing, it is equivalent to a rate ceiling in the synthetic loan market, and hence to an endogenously sized decrease in the demand for synthetic dollars.[32]

Any demand decrease in the synthetic lending market will lead to reduced synthetic lending spreads in both the long and short regimes (see part six of Propositions 1 and 2). In the long regime, these reduced lending spreads will also lead to reduced Treasury yields (and hence swap spreads), whereas in the short regime reduced lending spreads would lead to increased Treasury yields. Again, both of these effects operate through the relaxation of balance sheet constraints, and the regime determines the relationship between balance sheet tightness and Treasury yields. In both cases, there would be an increase in the demand for Treasury bond financing (either tri-party repo or security lending). Our model assumes these rates are fixed, but in a more complex model these rates might also adjust.

We should note that, in our model, a swap line with a rate equal to the swap rate and large capacity could drive the synthetic lending spread to zero. This would endogenously result in all of the dealer's balance sheet being allocated to the Treasury trade, and violate Assumption 1. Although we do not formally analyze this case, we should emphasize what does *not* happen: it is not the case that the balance sheet cost faced by dealers goes to zero. Instead, the CIP violation

[31] We are assuming that dealers are not themselves borrowing using the swap lines; providing subsidized dollar funding to dealers would have additional effects not described in this section.

[32] During normal times, because of stigma associated with tapping central bank liquidity facility and moral suasion from central banks discouraging banks to use swap line to fulfill their routine funding needs, the take-up in the swap line is extremely low even when the swap line rate borrowing rate is temporarily below the implied dollar rate from the FX swap market.

ceases to be a meaningful measure of balance sheet costs. Other financial intermediation spreads would decline (due to the relaxation of dealer balance sheet constraints), but would not go to zero.

Summarizing, in both regimes, swap lines can substitute for dealer balance sheet in the synthetic dollar market, thereby reducing synthetic lending spreads. However, swaps lines have opposite effects on Treasury yields in different regimes: they decrease Treasury yields in the long regime, but increase Treasury yields in the short regime.

4.3 Leverage Ratio Exemptions

We next consider changes to regulatory policy that involve exempting certain kinds of low-risk assets from the SLR calculation. We consider two possible exemptions: exempting Treasury bonds and repo loans against Treasury collateral (exempting Treasurys, for short), and exempting reserves. Similar policies were implemented during the most acute parts of the COVID-induced market disruptions in 2020.

Recall in our static model that we have consolidated dealers and their levered clients into a single entity, based on the analysis of Internet Appendix Section E.1. This consolidation is based on the fact that both repo loans that dealer provide to their levered clients to hold Treasury bonds and direct holdings of Treasury bonds increase the size of the dealer's balance sheet. For this reason, it is simpler to consider a policy that exempts both repo loans against Treasury bonds and Treasury bonds directly owned by dealers.[33]

Exempting Treasurys will both free up dealer balance sheet capacity for synthetic lending and remove the need to reduce CIP arbitrage activity when taking a net position in Treasury bonds. This will lead to Treasury yields that are a function only of financing rates (i^l in the long regime, i^s in the short regime), and therefore have the effect of reducing yields in the long regime and increasing yields in the short regime. In both regimes, the SLR exemption will allow dealers to allocate the regulated portion of their balance sheet entirely to synthetic lending ($q^{syn} = \bar{q}$) and lead to a reducing in financial intermediation spreads.

Exempting reserves (or any other assets) frees up the dealer balance sheet space for Treasury holding and synthetic lending, and thus is equivalent to expanding the balance sheet capacity $\bar{q}$ in our static model. In the long regime, this would result in a decline in bond yields and synthetic

[33]Exempting one but not the other would shift the net holdings of Treasury bonds from dealers to their levered clients or vice versa, in addition to relaxing balance sheet constraints.

lending spreads; in the short regime, bond yields would rise while synthetic lending spreads fall (see part two of Propositions 1 and 2).

Both SLR exemption policies will lead to a reduction in financial intermediation spreads, whose magnitude depends on the extent to which balance sheet constraints are relaxed. Exempting Treasurys will, in the long regime, reduce Treasury yields by removing balance sheet factors from their pricing entirely, whereas exempting reserves will not have this effect.

4.4 Quantitative Easing and Quantitative Tightening

We define QE (QT) as the Federal Reserve's purchases (or sales/redemptions) of Treasury bonds in the secondary market.[34] The Treasury bonds can ultimately come from (QE) or go to (QT) dealer inventory, bank portfolios (outside the broker-dealer subsidiary), or other non-bank (non-dealer) clients, and are traded in exchange for reserves. Here, we will separately consider the Treasury demand and reserve supply channels of QE/QT. We assume that no SLR exemption is applied to reserves or Treasury securities, and set aside the signaling effects of QE (which are covered in Section 4.1).

To isolate the effects of the Treasury demand channel, we consider a hypothetical version of QE in which the Fed purchases Treasury bonds in exchange for Treasury bills. This operation (which is somewhat akin to "Operation Twist") leaves the supply of reserves unchanged. We model this operation, which isolates the Treasury demand channel of QE, as a parallel outward shift in the demand curve for Treasury bonds (D_U) in our static model. Holding fixed money market yields and swap rates, in the long regime quantitative easing will reduce both yields and synthetic lending spreads (see part two of Proposition 1). In contrast, in the short regime, quantitative easing will reduce yields while increasing synthetic lending spreads (see part two of Proposition 2). Both of these effects operate through the balance sheet mechanism; the regime matters because it determines whether balance sheet constraints are tightened or loosened by QE. If the goal of quantitative easing is to lower financial intermediation spreads, then our results imply that QE is effective via the Treasury demand channel in the long regime but not in the short regime.

To isolate the reserve supply channel, we consider a hypothetical purchase of Treasury bills

[34] The Federal Reserve and other central banks have at times purchased mortgage, corporate, and other bonds as part of quantitative easing programs. The model described thus far considers only Treasury bonds, and for this reason we restrict attention to Treasury purchases.

in exchange for reserves, which leaves the supply of longer-maturity Treasury bonds unchanged. Note that the combination of these two operations is QE: the purchase of Treasury bonds in exchange for reserves. Note also that the effects of QT are exactly the opposite of those of QE. The effects of QE through the reserve supply channel are more complex, and depend in particular on whether the Fed's overnight reverse repo (ONRRP) facility is actively used.

Consider first the case without an active ONRRP facility. In this case, the reserves the Fed creates via QE must end up on bank balance sheets.[35] This will be true regardless of whether the ultimate seller of the bonds to the Fed is a bank, dealer, or non-bank. We can incorporate this effect into our model as a reduction in $\bar{q}$.

Suppose the reduction in balance sheet capacity is equal to the increase in bond demand (i.e. that all reserves end up on bank balance sheets). In the long regime, the reserve supply channel ($\bar{q}$) will exactly offset the Treasury demand channel ($D_U(\cdot)$); see (32). In contrast, in the short regime, both the reserve supply channel and the Treasury demand channel will lead to more tightly constrained dealer balance sheets (see (35)).

With an actively used ONRRP facility, the reserve supply channel is significantly muted. The ONRRP facility allow money market funds to make repo loans to the Federal Reserve. If the Fed exchanges reserves for bills, these funds will sell bills and receive deposits at their clearing banks, and then lend those deposits back to the Federal Reserve using the ONRRP facility. Thus, with an active ONRRP facility, the effects of QE can operate entirely through the Treasury demand channel.

In summary, in the long regime with an active ONRRP facility and binding balance sheet constraints (the situation as of June 2022), we expect QE/QT to have strong effects on Treasury yields and financial intermediation spreads.

4.5 Implications for Monetary Policy Tightening Cycles

In June 2022, the Federal Reserve began to normalize its large balance sheets from the extraordinary response to the COVID pandemic. In addition, the Fed has been increasing short-term interest rates substantially while engaging in quantitative tightening. Our framework has important implications for the dynamics of the Treasury market during such a tightening cycle.

[35] If the Fed's purchases under QE coincide with Treasury issuance, then some reserves might temporarily go into the Treasury's general account before eventually winding up on bank balance sheets.

We first note that tightening cycles are often associated with with flat or inverted Treasury yield curve, and low expected returns on long-term Treasury bonds. This dampens real money investors' demand for Treasury bonds, and it is particularly challenging for dealers and levered investors to accommodate a reduction in Fed holdings of Treasury bonds (QT) when client demand is weak.

Consider the experience of the 2017-2019 tightening cycle, in which the Fed normalized its balance sheet for the first time post-GFC and increased the short-term interest rates from the zero-lower-bound to 2.5 percent. During that tightening cycle, dealers' increased their Treasury holdings by about \$100 billion and hedge funds increased their holdings by about \$350 billion, together accounting for the entirety of the \$390 billion Fed balance sheet normalization from October 2017 to September 2019. The swap-Treasury spread and the Treasury cash-futures basis widened considerably over the period. Moreover, the increasingly crowded dealer balance sheet and significant build-up of the levered investor positions may have contributed to the repo market distress in September 2019 and Treasury market dislocation in March 2020.

Consistent with this experience, our model suggests that the combination of QT with an active ONRRP facility and a flattening curve, in the long regime, can lead to higher yields, more negative swap spreads, and higher financial intermediation spreads. SLR exemptions and the use of the swap lines established with foreign central banks have the potential to ameliorate these effects.

Finally, it is important to note that there are two main factors that distinguish the current tightening cycle from the 2017-2019 cycle. First, there is currently over \$2 trillion cash at the ONRRP facility (the ONRRP facility was not very active during 2017-2019). The cash at the ONRRP can, potentially, be deployed to finance Treasury supply without further crowding out dealers' balance sheet. Second, current interest rate volatility is significantly higher than during the previous tightening cycle, due to greater uncertainty in the future path of policy rates. Greater interest rate volatility can discourage the build-up of dealers' inventory and the levered investors' position due to additional value-at-risk-type constraints, which we have been abstracted from our framework.

5 Conclusion

We have documented a regime change in the U.S. Treasury bond market. Prior to the 2008-2009 financial crisis, dealers were net short-sellers of Treasury bonds, swap spreads were positive, and CIP violations were small. Following the GFC, dealers became net long Treasury bonds, swap spreads turned negative, and covered interest parity violations emerged. Our analysis ties these observations together by constructing arbitrage bounds, the net short and net long curves, and providing evidence of dealers-as-arbitrageurs in the Treasury market.

We then discuss the causes and consequences of this regime change. We view the large increase in Treasury supply and the tightening of leverage constraints on dealers as the primary drivers of this regime change. Using a stylized static model, we have argued that this regime shift has amplified the effects of quantitative easing and of the yield curve slope on borrowing spreads. In the post-GFC dealer-long regime our model predicts tighter dealer balance constraints in response to Fed quantitative tightening and a flat or inverted Treasury yield curve, and more elevated financial intermediation spreads. Our analysis suggests that other polices, including the use of swap lines and of exemptions to SLR calculations, can help offset these effects.

References

Fernando Alvarez and Urban J Jermann. Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica*, 68(4):775–797, 2000.

Patrick Augustin, Mikhail Chernov, Lukas Schmid, and Dongho Song. The term structure of cip violations. *NBER Working Paper No.27231*, 2020.

Patrick Augustin, Mikhail Chernov, Lukas Schmid, and Dongho Song. Benchmark interest rates when the government is risky. *Journal of Financial Economics*, 140(1):74–100, 2021.

Viktoria Baklanova, Cecilia Caglio, Marco Cipriani, and Adam Copeland. The use of collateral in bilateral repurchase and securities lending agreements. *Review of Economic Dynamics*, 33: 228–249, 2019.

Daniel Barth and R Jay Kahn. Hedge funds and the treasury cash-futures disconnect. *OFR WP*, pages 21–01, 2021.

Bruno Biais, Johan Hombert, and Pierre-Olivier Weill. Incentive constrained risk sharing, segmentation, and asset pricing. *American Economic Review*, 111(11):3575–3610, 2021.

Nina Boyarchenko, Thomas M Eisenbach, Pooja Gupta, Or Shachar, and Peter Van Tassel. Bank-intermediated arbitrage. *FRB of New York Staff Report*, (858), 2018.

Markus K Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *The review of financial studies*, 22(6):2201–2238, 2009.

Markus K Brunnermeier and Yuliy Sannikov. A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421, 2014.

John Y Campbell and Robert J Shiller. Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514, 1991.

YiLi Chien, Harold Cole, and Hanno Lustig. A multiplier approach to understanding the macro implications of household finance. *The Review of Economic Studies*, 78(1):199–234, 2011.

Wenxin Du, Alexander Tepper, and Adrien Verdelhan. Deviations from covered interest rate parity. *The Journal of Finance*, 73(3):915–957, 2018b.

Wenxin Du, Benjamin M Hébert, and Amy Wang Huber. Are intermediary constraints priced? *Review of Financial Studies*, 2022.

Darrell Duffie. *Post-crisis bank regulations and financial market liquidity*. Banca d'Italia, 2017.

Darrell Duffie. Still the world's safe haven? *Redesigning the US Treasury market after the COVID-19 crisis'. Hutchins Center on Fiscal and Monetary Policy at Brookings. Available online: https://www. brookings. edu/research/still-the-worlds-safe-haven (accessed on 28 November 2021)*, 2020.

Darrell Duffie, Nicolae Gârleanu, and Lasse Heje Pedersen. Over-the-counter markets. *Econometrica*, 73(6):1815–1847, 2005.

Giovanni Favara, Sebastian Infante, and Marcelo Rezende. Leverage regulations and treasury market participation: Evidence from credit line drawdowns. *Working Paper*, 2022.

Peter Feldhütter and David Lando. Decomposing swap spreads. *Journal of Financial Economics*, 88(2):375–405, 2008.

Matthias Fleckenstein and Francis A Longstaff. Treasury richness. *NBER Working Paper*, 2021.

Mark Gertler and Nobuhiro Kiyotaki. Financial intermediation and credit policy in business cycle analysis. In *Handbook of monetary economics*, volume 3, pages 547–599. Elsevier, 2010.

Jonathan Goldberg. Liquidity supply by broker-dealers and real activity. *Journal of Financial Economics*, 136(3):806–827, 2020.

Valentin Haddad and David Sraer. The banking view of bond risk premia. *The Journal of Finance*, 75(5):2465–2502, 2020.

Valentin Haddad, Alan Moreira, and Tyler Muir. When selling becomes viral: Disruptions in debt markets in the covid-19 crisis and the fed's response. *The Review of Financial Studies*, 34(11): 5309–5351, 2021.

Samuel G Hanson and Jeremy C Stein. Monetary policy and long-term real rates. *Journal of Financial Economics*, 115(3):429–448, 2015.

Samuel G Hanson, David O Lucca, and Jonathan H Wright. Rate-amplifying demand and the excess sensitivity of long-term rates. *The Quarterly Journal of Economics*, 136(3):1719–1781, 2021.

Samuel G Hanson, Aytek Malkhozov, and Gyuri Venter. Demand-supply imbalance risk and long-term swap spreads. Working paper, 2022.

Zhiguo He and Arvind Krishnamurthy. Intermediary asset pricing. *American Economic Review*, 103(2):732–70, 2013.

Zhiguo He, Stefan Nagel, and Zhaogang Song. Treasury inconvenience yields during the covid-19 crisis. *Journal of Financial Economics*, 143(1):57–79, 2022.

Benjamin M Hébert. Externalities as arbitrage. *NBER Working Paper No. 27953*, 2020.

Urban Jermann. Negative swap spreads and limited arbitrage. *The Review of Financial Studies*, 33 (1):212–238, 2020.

Scott Joslin, Kenneth J Singleton, and Haoxiang Zhu. A new perspective on gaussian dynamic term structure models. *The Review of Financial Studies*, 24(3):926–970, 2011.

Scott Joslin, Marcel Priebsch, and Kenneth J Singleton. Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, 69(3):1197–1233, 2014.

Sven Klingler and Suresh Sundaresan. An explanation of negative swap spreads: Demand for duration from underfunded pension plans. *The Journal of Finance*, 74(2):675–710, 2019.

Arvind Krishnamurthy and Annette Vissing-Jorgensen. The aggregate demand for treasury debt. *Journal of Political Economy*, 120(2):233–267, 2012.

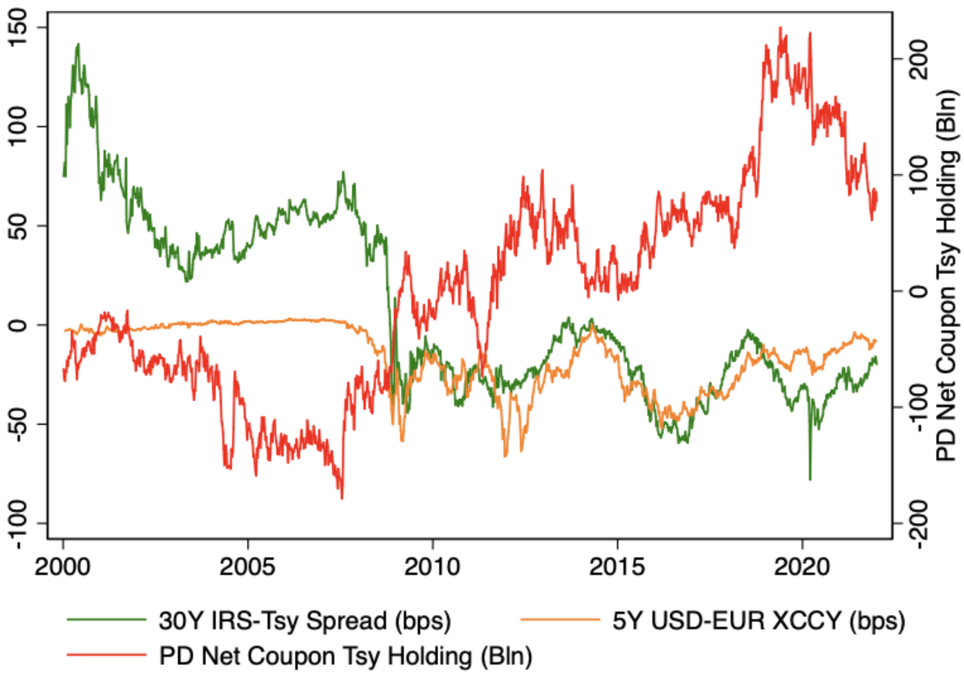
Mathias S Kruttli, Phillip Monin, Lubomir Petrasek, and Sumudu W Watugala. Hedge fund treasury trading and funding fragility: Evidence from the covid-19 crisis. 2021.

Stefan Nagel. The liquidity premium of near-money assets. *The Quarterly Journal of Economics*, 131(4):1927–1971, 2016.

John H Rogers, Chiara Scotti, and Jonathan H Wright. Unconventional monetary policy and international risk premia. *Journal of Money, Credit and Banking*, 50(8):1827–1850, 2018.

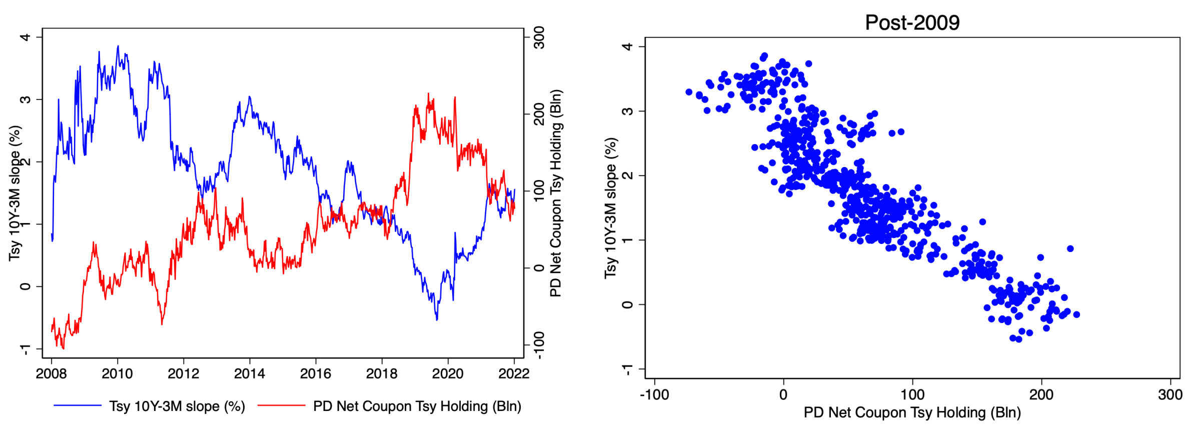
Emil Siriwardane, Aditya Sunderam, and Jonathan Wallen. Segmented arbitrage. *Available at SSRN 3960980*, 2021.

Figure 1: Primary Dealer Treasury Holing, Swap Spreads, and Cross-Currency Basis.



Notes: This figure plots the spread between the 30-year Libor-linked interest rate swap and the U.S. Treasury yield (in green), and the 5-year USD-EUR cross-currency basis (in orange), and net holdings of coupon Treasury bonds. The pricing data are from Bloomberg, and the primary dealer position data are from the publicly available primary dealer statistics published by the Federal Reserve Bank of New York. The quote on the cross-currency basis swap effectively measures the direct dollar interest rate minus the synthetic dollar interest by swapping EUR interest rate into dollars (Du, Tepper, and Verdelhan (2018b)).

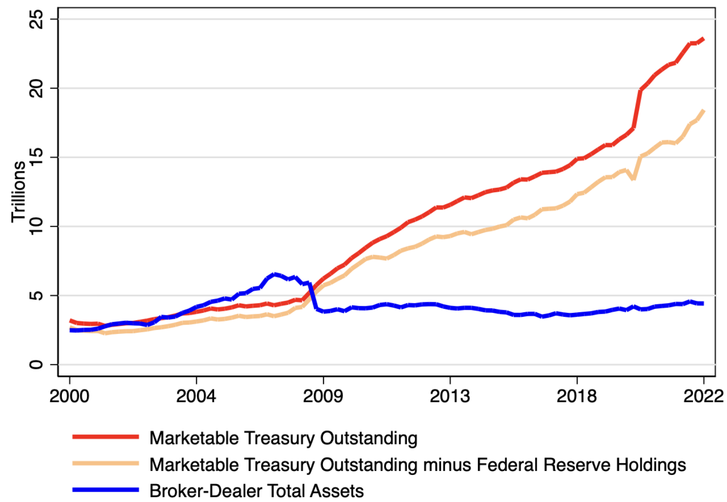
Figure 2: Term Spreads and Primary Dealer Treasury Holdings



(a) Tresaury Term Spread and Dealer Treasury Holding (b) Treasury Term Spread and Dealer Treasury Holding

Notes: Panel (a) plots the yield spread between the 10-year Treasury bond and the 3-month Treasury bill (in blue), and the primary dealers' net holdings of Treasury bonds. Panel (b) plots the relationship between the two variables post-2009 in a scatter plot. The pricing data are from Bloomberg, and the primary dealer position data are from the publicly available primary dealer statistics published by the Federal Reserve Bank of New York.

Figure 3: Treasury Supply and Broker-Dealer Total Assets.



Notes: This figure plots the total marketable Treasury securities outstanding (in red), the total marketable Treasury outstanding minus Federal Resereve holdings (in orange), and the financial assets of the U.S. broker-dealer sector (in blue) in trillions of dollars from Flow of Funds.

Figure 4: Balance Sheet Change for a Long-Treasury Trade in the Long Regime

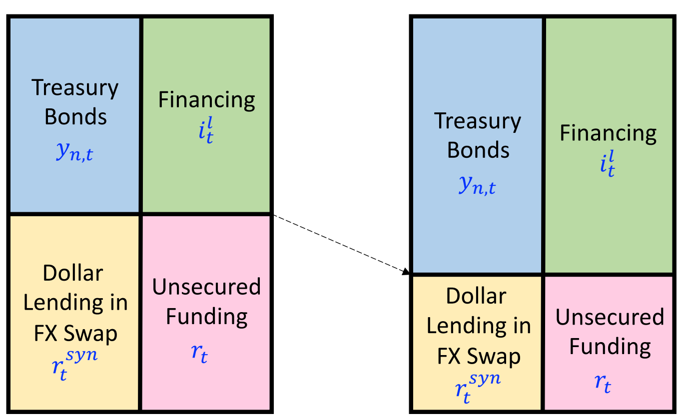


Figure 5: Balance Sheet Change for a Short-Treasury Trade in the Short Regime

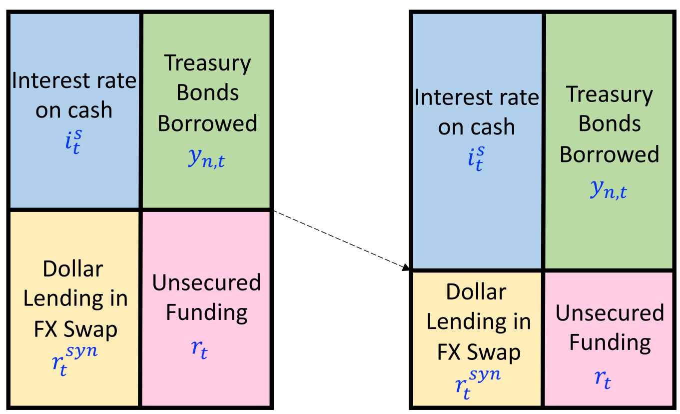
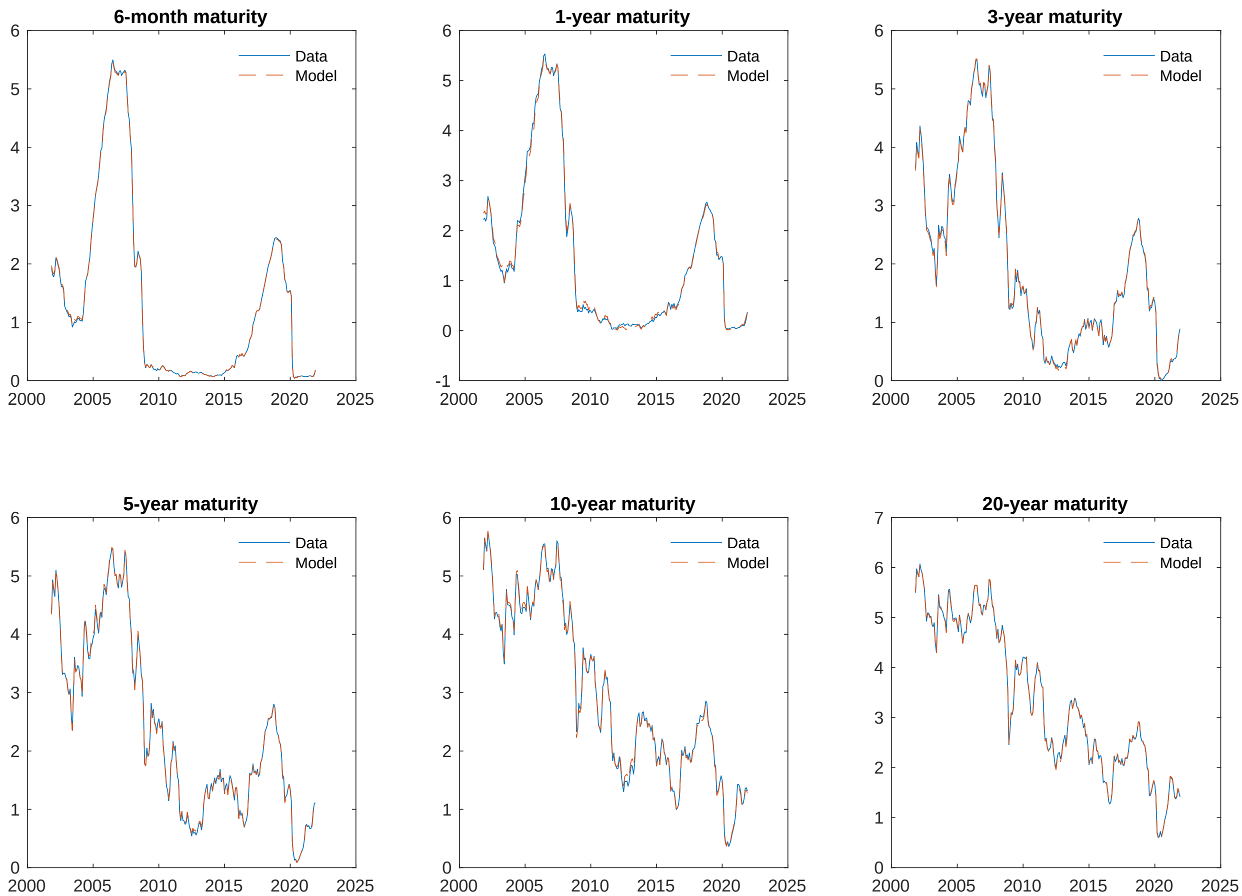
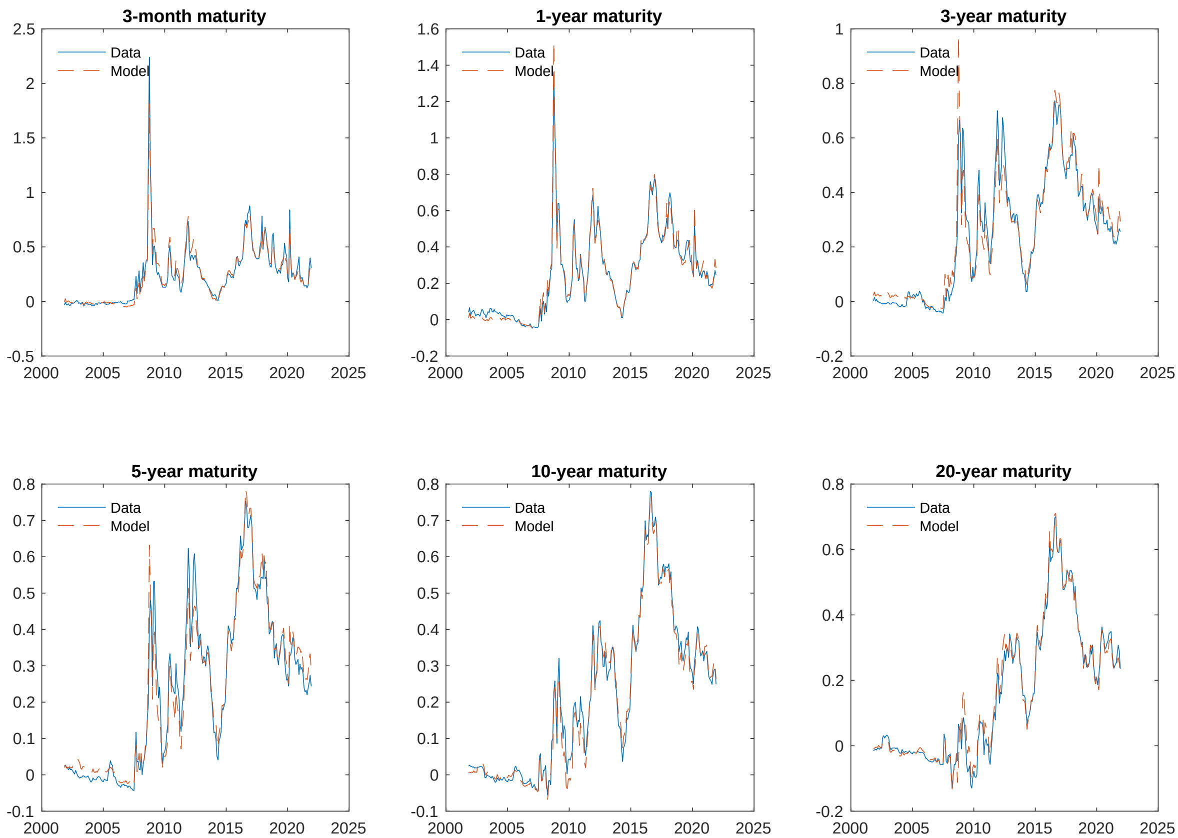


Figure 6: Fit of the TS Model to Dollar OIS Curves.



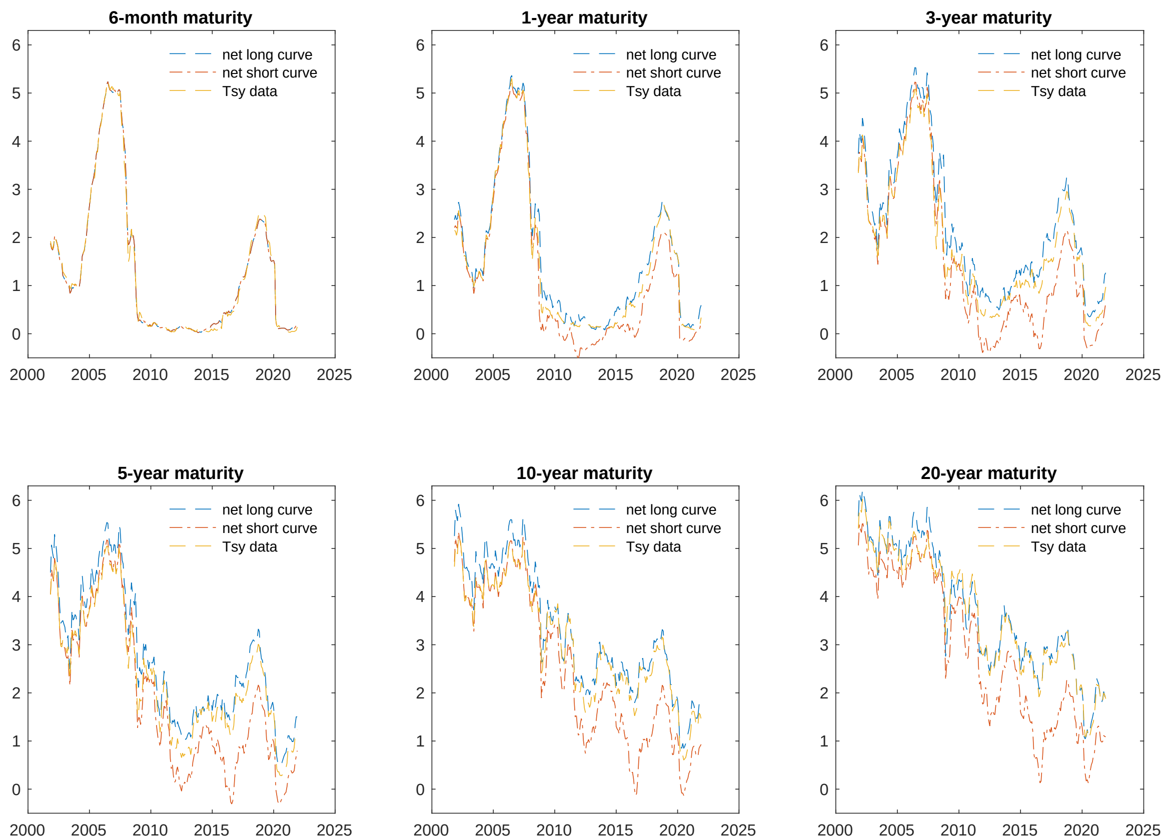
Notes: In this figure, we show the fitting of the dollar swap curve using our term-structure model. Data are from 2003 to 2021. More details on the term structure model can be found in Section 2.4.

Figure 7: Fit of the TS Model to USD-EUR CIP Deviations.



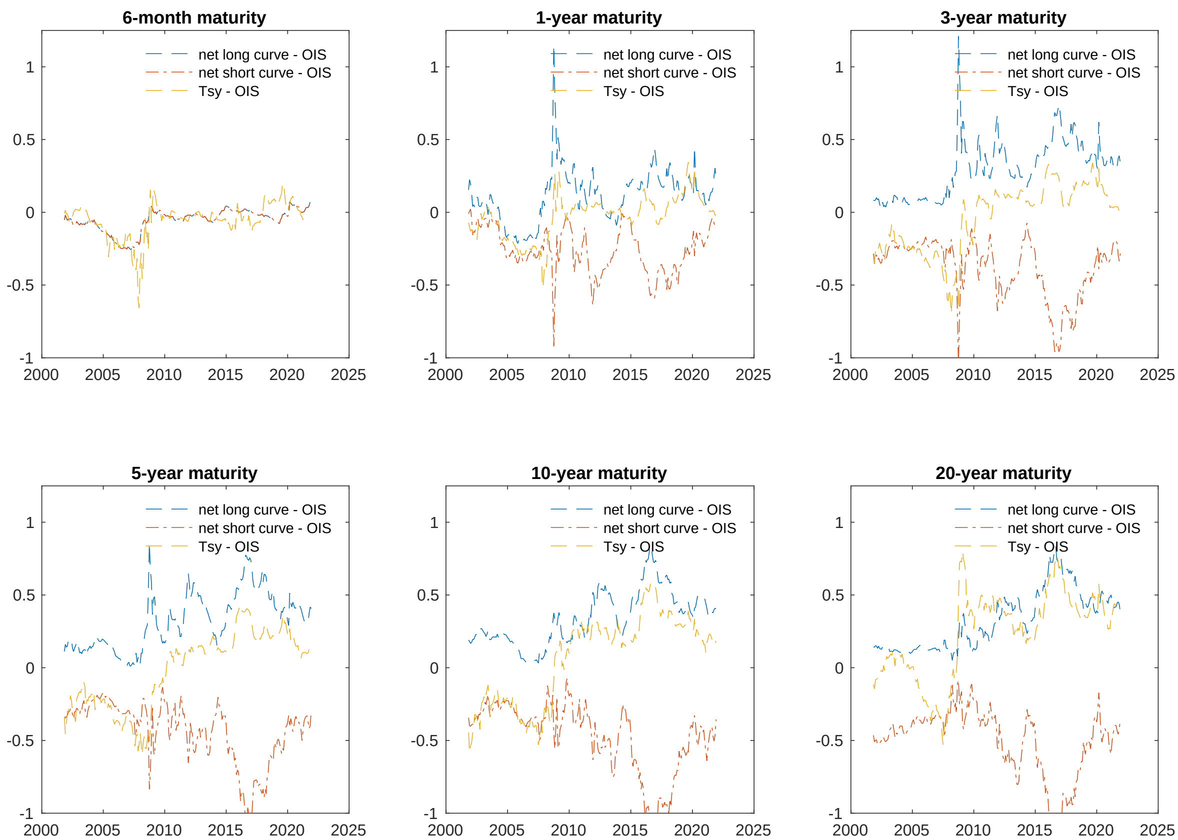
Notes: In this figure, we show the fitting of the OIS-Based EUR-USD CIP Deviations using our term structure model. The CIP deviations are shown as the spread between the synthetic dollar interest rate and the USD OIS rate (the negative of the quoted cross-currency basis). Data are from 2003 to 2021. More details on the term structure model can be found in Section 2.4.

Figure 8: Long and Short Curves.



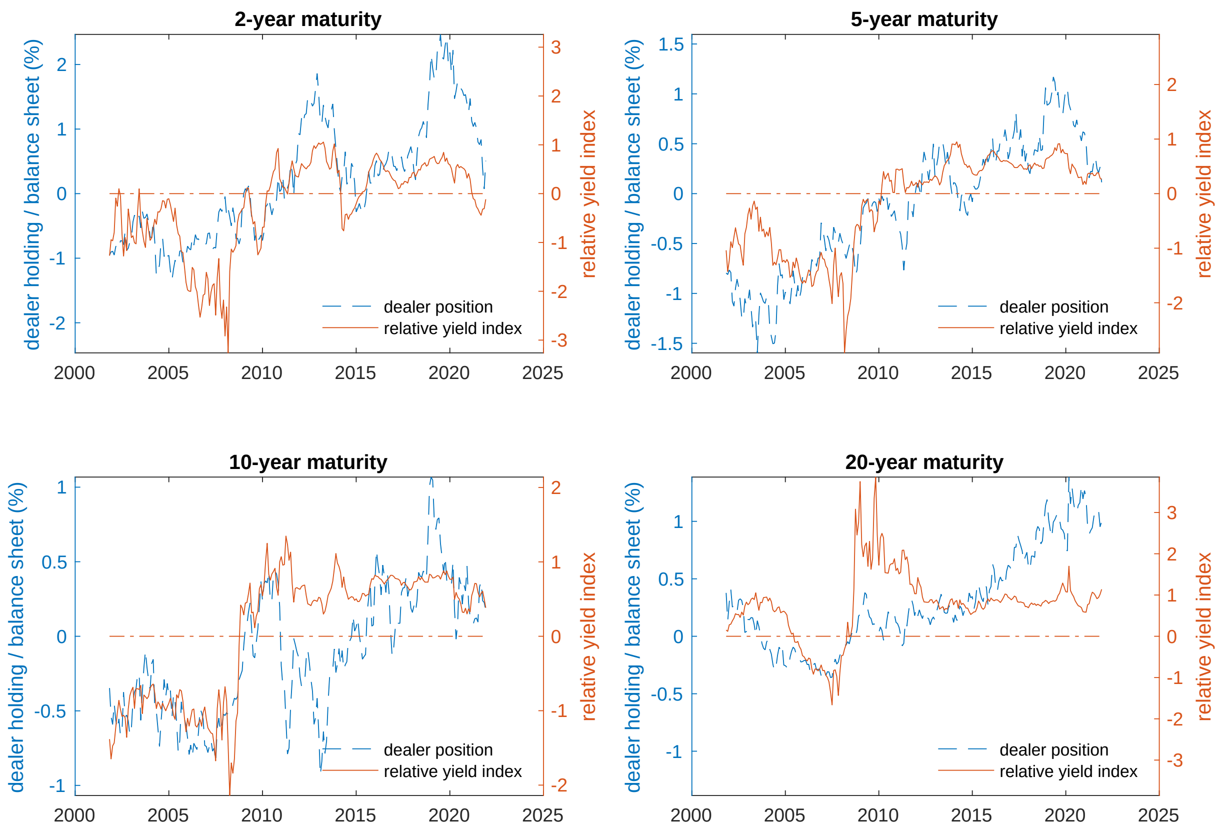
Notes: In this figure, we show the model-implied net-long and net-short curves for Treasury securities, together with the actual Treasury yields. Data are from 2003 to 2021. All yields are par yields. More details on the term structure model can be found in Section 2.4.

Figure 9: Long and Short Curves – OIS Spreads.



Notes: In this figure, we show the model-implied long and short Treasury curves minus the OIS rates for corresponding maturities, together with the actual Treasury–OIS spreads. Data are from 2003 to 2021. All yields are par yields. More details on the term structure model can be found in Section 2.4.

Figure 10: Relative Yield and Position by Maturity.

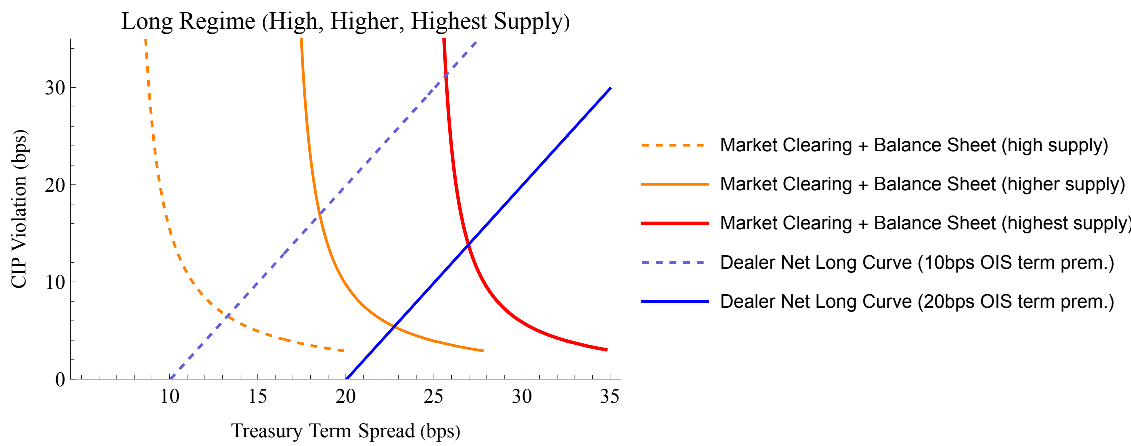


Notes: In this figure, we plot the relative yield index versus actual scaled primary dealer Treasury positions. The relative yield index is defined as 2*(long curve − Treasury curve)/(long curve − short curve) − 1. The scaled primary dealer Treasury position is calculated as the ratio of the primary dealer net position in the Treasury securities for the corresponding maturity bucket (published by Federal Reserve Bank of New York) to total financial assets of the broker-dealer sector (published by Flows of Funds). The position data corresponding to 2-year, 5-year, 10-year, and 20-year yields in the figure are defined based on the following maturity buckets, 1-to-3 years, 3-to-6 years, 6-to-11 years, over 11 years, respectively. Data are from 2003 to 2021.

Figure 11: Relative Yield and Position in Aggregate.

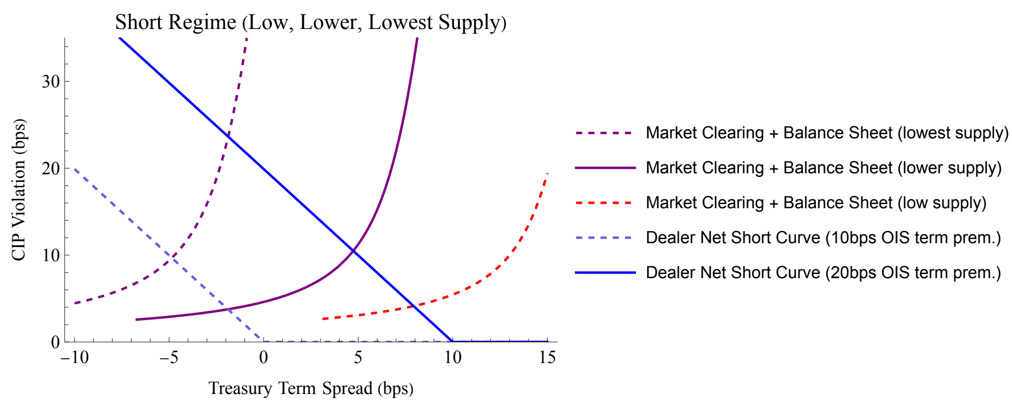
Notes: In this figure, we plot the weighted average of the relative yield index across maturities vs. primary dealer Treasury positions. For each maturity, the relative yield index is defined as 2*(long curve − Treasury curve)/(long curve − short curve) − 1. Each relative yield is then weighted by coupon Treasury outstanding for the corresponding buckets over total coupon Treasury securities, based on data from the Center for Research in Security Prices. The primary dealer Treasury position is the ratio of total primary dealer net position in all coupon Treasury securities (published by Federal Reserve Bank of New York) to total financial assets of the broker-dealer sector (published by Flows of Funds). Data are from 2003 to 2021.

Figure 12: Indifference Curves in the Long Regime.



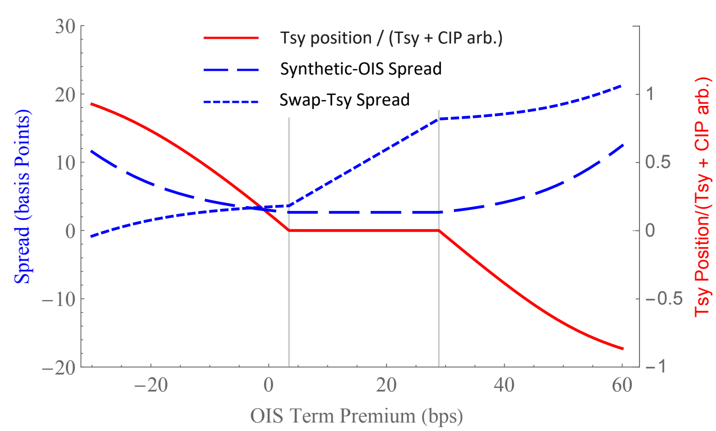
Notes: This figure illustrates the dealer indifference curve defined in (31) and the market indifference curve defined in (32), under three different levels of S^{bond} and two different levels of $y_{\mathbb{Q}}$. The functional forms and parameters used to generate the figure are described in Internet Appendix Section B.

Figure 13: Indifference Curves in the Short Regime.



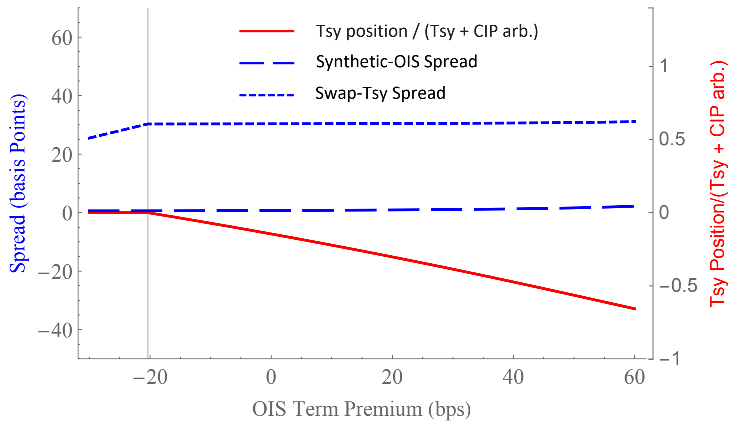
Notes: This figure illustrates the dealer indifference curve defined in (34) and the market indifference curve defined in (35), under three different levels of S^{bond} and two different levels of $y_{\mathbb{Q}}$. The functional forms and parameters used to generate the figure are described in Internet Appendix Section B.

Figure 14: All-Regimes and the swap Term Premium.



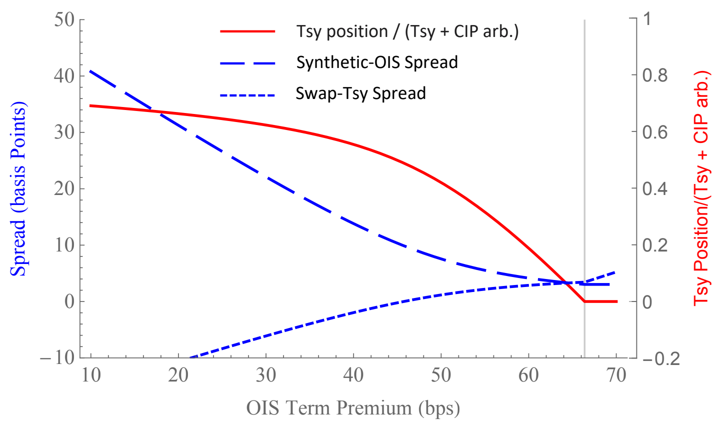
Notes: This figure plots the synthetic lending spread $r^{syn} - r$, swap-Treasury spread $r_n - y$, and the ratio of Treasury position over total Treasury and CIP arbitrage position $q^{bond}/\bar{q}$, as a function of the swap curve term premium $r_n - r$. We fix the expected future rates. The functional forms and parameters used to generate the figure are described in Internet Appendix Section B.

Figure 15: All-Regimes with Large Dealer Balance Sheet Capacity and Small Bond Supply.



Notes: This figure plots the synthetic lending spread $r^{syn} - r$, swap-Treasury spread $r_n - y$, and the ratio of Treasury position over total Treasury and CIP arbitrage position $q^{bond}/\bar{q}$, as a function of the swap curve term premium $r_n - r$. We fix the expected future rates and assume a relatively large dealer balance sheet capacity, a small bond supply, and a large spread between y^{bill} and r. The functional forms and parameters used to generate the figure are described in Internet Appendix Section B.

Figure 16: All-Regimes with Small Dealer Balance Sheet Capacity and Large Bond Supply.



Notes: This figure plots the synthetic lending spread $r^{syn} - r$, swap-Treasury spread $r_n - y$, and the ratio of Treasury position over total Treasury and CIP arbitrage position $q^{bond}/\bar{q}$, as a function of the swap curve term premium $r_n - r$, holding expected future rates fixed, with a relatively small dealer balance sheet capacity, large bond supply, and small spread between y^{bill} and r. The functional forms and parameters used to generate the figure are described in Internet Appendix Section B.

Internet Appendix
"Intermediary Balance Sheets and the Treasury Yield Curve"

A Data

A.1 Data Sources

We obtain the Treasury term structure from Bloomberg, for maturities 0.25, 0.5, 1, 3, 5, 10, 15, 20, and 30 years, all at daily frequency. The T-bills are from the ticker "GB", representing actively traded T-bill yields, and the non-bills are from the ticker "C082", representing the widely-used Bloomberg fair value Treasury yield curve.

We obtain OIS term structure denominated in USD from Bloomberg for maturities 0.25, 0.5, 1, 3, 5, 10, 15, 20, and 30 years, all at daily frequency. The ticker is "USSO" and data are from Nov 1996 to Dec 2021.

We construct the synthetic dollar lending rate from Euro (EUR). For this purpose, we obtain OIS term structure for EUR for maturities 1, 3, 5, 10, 15, 20, and 30 years. The EUR OIS data are from Aug 2009 to Dec 2021.

Then we obtain the above-one-year maturity LIBOR basis at a daily frequency for EUR-USD from Bloomberg. The EUR-USD LIBOR basis covers Nov 1999 to Dec 2021, and includes the following maturities (in years): 1, 3, 5, 10, 15, 20, 30. EUR 3-month LIBOR basis is from Bloomberg, and they are at daily frequency from Jan 2000 to Dec 2021.

To construct the OIS basis, we also collect EUR inter-bank interest-rate swap (IRS) term structure from Bloomberg at daily frequency. EUR IRS data are from Sep 1999 to Dec 2021. Then we construct the OIS basis for each maturity as

$$\begin{aligned}\text{EUR-USD OIS basis} = \text{EUR-USD LIBOR basis} &+ (\text{USD OIS - USD IRS}) \\ &- (\text{EUR OIS - EUR IRS}) \ ,\end{aligned}$$

where each term has the same maturity.

Due to data limitation, we use the “hybrid OIS basis”, defined as follows:

- Whenever OIS data are available, we construct the OIS basis from the LIBOR basis and LIBOR-OIS basis swap.
- When OIS data are not available (only happens before 2008), we use the LIBOR basis instead

This approach is essentially OIS basis throughout the whole sample period because OIS basis and LIBOR basis are almost the same before the global financial crisis. A comparison between the OIS basis and the LIBOR basis is shown in Figure A1.

Our data on the tri-party repo rate also comes from Bloomberg. We assume a two percent haircut (which is standard in the tri-party repo market), and define (at a daily frequency)

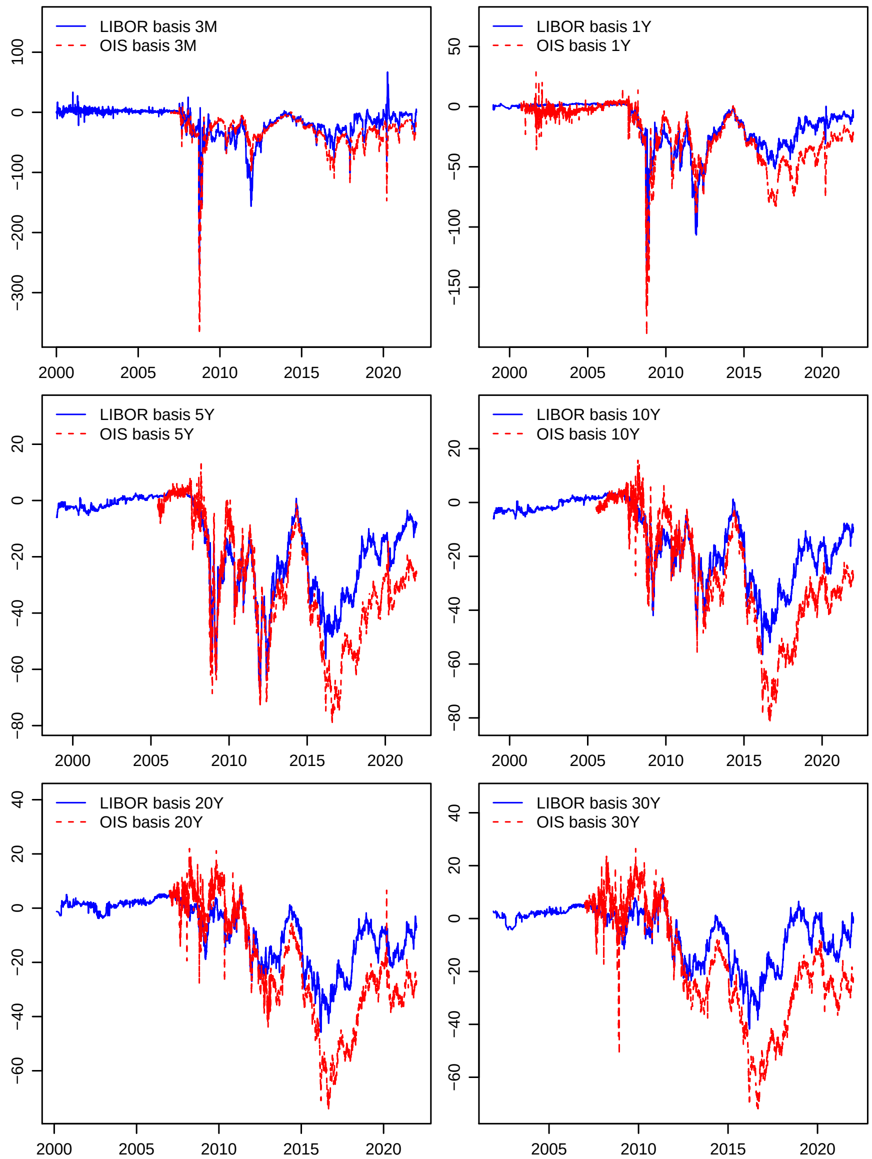
$$i_t^l = 0.98 * r_t^{tri} + 0.02 * ff_t \tag{A1}$$

where r_t^{tri} is the tri-party repo rate and ff_t is the effective federal funds rate (which is a proxy for unsecured borrowing costs and is also obtained via Bloomberg). Note that the federal funds rate is the floating rate associated with OIS swaps.

For the short-regime financing rate i_t^s, we use security lending rates, which we discussed in more detail in Internet Appendix Section A.2.

On the quantity side, dealer net holdings are Treasury securities are based on the primary dealer statistics published by the Federal Reserve Bank of New York.

Figure A1: Comparison of LIBOR EUR Basis and OIS EUR Basis



Notes: This figure illustrates the LIBOR EUR-USD basis and the OIS EUR-USD basis. The cross-currency basis is defined as the dollar rate minus the synthetic rate, which is exactly the opposite to the CIP violations we used in the model.

A.2 Treasury Securities Lending Rebate Rates

Dealers who wish to short-sell a Treasury bond must find someone to lend them the bond. There are, generally, two ways of borrowing a bond as a dealer. The first is when that bond is posted as collateral by a levered client (typically a hedge fund), and the second is by borrowing the bond from a security lender (an asset manager, insurance company, or similar institution).

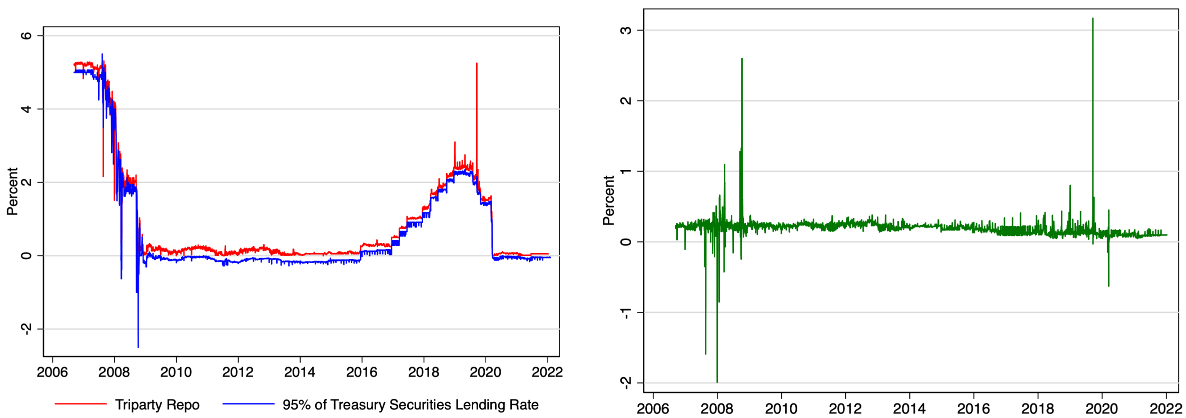
Dealers lend cash to clients against Treasury collateral in bilateral repo markets, at rates that are general higher than rates in the tri-party repo market (which is to say, the rate at which dealers can borrow against Treasury collateral). However, to earn these relatively high rates, dealers must be willing to accept whatever Treasury collateral their clients wish to borrow against. A dealer who wishes to short-sell a specific bond or bonds of a specific maturity might not be able to find a client who wants to take a levered long position in that same bond or maturity. In fact, if dealers and their levered clients generally share the same view on which bonds are relatively cheap or expensive, we should expect that the bonds dealers would like to short will not be the bonds their levered clients would like to long.

A dealer wanting to short a bond is therefore more likely to borrow that bond from a security lender. Security lenders require collateral from dealers. We will first assume cash collateral, although collateral swaps, discussed below, are also common. A dealer therefore lends cash when borrowing the bond. The security lender is borrowing this cash, and must invest it until the dealer returns the bond.

The security lender is willing to do this because it can earn a spread between the rate it pays to the dealer on the cash collateral (i^s in our model) and the rate it earns on its invested cash. Because the dealer can demand its cash back at any time, the security lender has a strong incentive to invest only in safe and liquid assets. For example, a security lender who reinvests the cash in a government money market fund, which in turn invests in the tri-party repo market, is likely to receive no more than the tri-party repo rate on its investments, and therefore will offer the dealer an even lower rate on its cash collateral. If the security lending market is competitive and the cash is re-invested in tri-party repo market, the spread between the tri-party repo rate and the rate on cash collateral is a measure of the costs to the security lender of running a security lending program.

We use data from Market Securities Finance to calculate the rebate rate on the cash collateral when the dealer is borrowing Treasury bonds from a security lender. Figure A2 shows that the 95 percentile of all Treasury securities lending rebate rate ("rebate rate", for short) is consistently below the tri-party repo rate. The spread between tri-party and rebate rate is about 20 basis points on average and quite stable throughout our sample. We emphasize the 95th percentile to highlight that this spread applies to all securities, and not merely those that are "special" (meaning that they are particularly hard to borrow).

Figure A2: Comparison Between Securities Lending Rebate and Triparty Repo Rates



(a) 95 pct of Rebate vs. Triparty (b) Spread between 95 pct of Rebate and Triparty

Notes: Panel (a) plots the yield spread between the 10-year Treasury bond and the 3-month Treasury bill (in blue), and the primary dealers' net holdings of Treasury bonds. Panel (b) plots the relationship between the two variables post-2009 in a scatter plot.

In practice, dealers might demand a haircut when lending cash to clients and be required to provide excess cash (a negative haircut) when borrowing from securities lenders. Baklanova et al. (2019) pools these two cases and demonstrate that the haircut is small in absolute value (a few percent at most). We therefore assume a zero haircut for want of better evidence, and would like to thank Sebastian Infante for making us aware of this issue.

Collateral swaps (in which the dealer posts a different bond as collateral) are also common. In a collateral swap, the dealers offers a bond of equal value as collateral in lieu of cash, and pays a cash fee. Our data from Market Securities Finance shows that the average fee in a collateral swap is roughly 20bps on an annualized basis (i.e. .002% of the bond value if borrowed for a whole year), consistent with our view of the tri-party repo rate as the security lender's outside investment option.

B Functional Forms and Parameters for Figures

This appendix section describes the functional form and parameter assumptions used to generate Figures 12, 13, 14, 15, and 16. These functional form and parametric assumptions are for illustrative purposes only and do not represent a calibration of the model.

We assume a constant elasticity functional form for the Treasury demand curves. Note that both of these demand curves are functions of the hedged bond log risk premium $\pi_{n,H}$ (the expected excess log return using r^{syn} as the risk-free rate). We assume that

$$D_H(\pi_{n,H}) = D_{H,0} \exp(\eta_H \pi_{n,H}) \tag{B-1}$$

where $D_{H,0} > 0$ represents the demand at zero risk premium. The parameter $\eta_H > 0$ is the semi-elasticity of bond demand to the log risk premium.

We similarly assume that

$$D_U(\pi_{n,U}) = D_{U,0} \exp(\eta_U \pi_{n,U}), \tag{B-2}$$

where $\pi_{n,U}$ is the log risk premium with respect to Treasury bills, with $D_{U,0} > 0$ and $\eta_U > 0$. Note that $\pi_{n,H}$ and $\pi_{n,U}$ are log risk premia; an η_U or η_H of 50 implies a roughly $1.35x$ change in demand given a 1% excess return.

For the synthetic demand curve, we assume that

$$D^{syn}(x) = D_0^{syn} x^{-\xi}, \tag{B-3}$$

where $x = r^{syn} - r$ is the spread in basis points and $\xi > 0$ is the elasticity of demand to the spread. This functional form imposes an Inada-type condition that ensures that demand is large as the spread becomes close to zero.

Note that these functional forms satisfy Assumption 1, irrespective of the parameters employed.

We use three sets of parameters to generate the figures used in the main text. The illustrative parameters are chosen to generate clear graphs, and in particular have the property that the regime can change given modest changes in term premium or bond supply. The pre-GFC parameter set perturbs this parameter set using a smaller Treasury supply, larger dealer balance sheet capacity, and larger repo-bill spread. The post-GFC parameter set uses instead a large bond supply, comparatively tight dealer balance sheet, and zero repo-bill spread.

The parameters are chosen under the assumption of an annual holding period and that the bond is a two-year bond ($n = 2$). The table below lists the sets of parameters we employ. Note that Figures 12 and 13 plot dealer indifference curves for different levels of $y_{\mathbb{Q}}$, holding all else constant. Likewise, Figures 14, 15, and 16 have the OIS term premium on the x-axis, which is equivalent to $y_{\mathbb{Q}}$ (holding $y_{\mathbb{P}}$ constant). For this reason, we do not list $y_{\mathbb{Q}}$ in the set of parameters below.

Table A1: Parameters for Figures

Parameter	Illustrative Value	Pre-GFC	Post-GFC
S^{bond}	10.5	9.5	14.5
$\bar{q}$	2	7	2
y^{bill} (bps)	95	65	95
$y_{\mathbb{P}}$ (bps)	95	65	95
r (bps)	100		
r^{long} (bps)	95		
r^{short} (bps)	75		
D_0^{syn}	4		
ξ	1		
$D_{U,0}$	9.5		
$D_{H,0}$	0.5		
$\eta_U = \eta_H$	50		

C Details of the Term Structure Model

The term structure model consists $\mathbb{P}$ and $\mathbb{Q}$ dynamics

$$z_{t+1} = k_{0,z}^{\mathbb{P}} + K_{1,z}^{\mathbb{P}} \cdot z_t + (\Sigma_z)^{1/2} \varepsilon_{z,t+1}^{\mathbb{P}}, \varepsilon_{z,t+1}^{\mathbb{P}} \sim N(0, I_N),$$

$$z_{t+1} = k_{0,z}^{\mathbb{Q}} + K_{1,z}^{\mathbb{Q}} \cdot z_t + (\Sigma_z)^{1/2} \varepsilon_{z,t+1}^{\mathbb{Q}}, \varepsilon_{z,t+1}^{\mathbb{Q}} \sim N(0, I_N)$$

The state variable vector z_t is 5-by-1, include the first three PCs of OIS term structure (r_t^{PC1}, r_t^{PC2}, and r_t^{PC3}) and the first two PCs of the cross-currency basis term structure ($r_t^{cip,PC1}$ and

$r_t^{cip,PC2}$),

$$z_t = \begin{bmatrix} r_t^{PC1} \\ r_t^{PC2} \\ r_t^{PC3} \\ r_t^{cip,PC1} \\ r_t^{cip,PC2} \end{bmatrix}.$$

The monthly OIS rate and the monthly synthetic rate are both affine functions of the state vector,

$$\frac{1}{12} r_t = \delta_0 + (\delta_1)^T z_t,$$
$$\frac{1}{12} r_t^{syn} = \hat{\delta}_0 + (\hat{\delta}_1)^T z_t,$$

For pricing Treasury securities, we also need the state vector $x_t = (x_{1,t}, x_{2,t}, x_{3,t})$, constructed from the data as

$$\begin{aligned} x_{1,t} &= \ln(e^{\frac{1}{12} i_t^l} - e^{\frac{1}{12} r_t} + e^{\frac{1}{12} r_t^{syn}}) \\ x_{2,t} &= \ln(e^{\frac{1}{12} i_t^s} + e^{\frac{1}{12} r_t} - e^{\frac{1}{12} r_t^{syn}}) \\ x_{3,t} &= \frac{1}{12} y_t^{bill} \end{aligned}$$

where i_t^l is the financing rate measured according to (A1), and i_t^s is the interest received on cash collateral.

To operationalize the term structure model and reduce dimensionality, we assume that the vector x_t is affine in the state vector z_t,

$$x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} = \gamma_0 + \Gamma_1 z_t + (\Sigma_x)^{\frac{1}{2}} \varepsilon_{x,t}, \varepsilon_{x,t} \sim N(0, I_3).$$

All state variables $x_{k,t}, k \in \{1,2,3\}$ represent yields at the monthly frequency. However, due to the lack of data, we use overnight tri-party rate and overnight security lending rate as proxies for the monthly counterparts. Furthermore, the one-month CIP basis is subject to a quarter-end effect, where the one-month CIP basis spikes at the end of each quarter due to capital regulation, as documented by Du et al. (2018b). To avoid such effect, we instead use the three-month CIP basis to construct the synthetic rate. The underlying assumption is that the rate difference due to maturity difference between one month and three months is negligible.

From our estimations, variance matrix Σ_x is close to zero (the maximum eigen value of Σ_x is about 7×10^{-5}, and much smaller than the maximum eigenvalue of Σ_z which is 4×10^{-3}). To simplify expositions, we set $\Sigma_z = 0$ and limit the actual state space to be five-dimensional. Thus, we will proceed with

$$x_t = \gamma_0 + \Gamma_1 z_t$$

In what follows, we first show the derivations of the OIS term structure and the basis term structure. Then we provide details on how the model generates dealer net long and net short curves. Next, we discuss the conversions between zero-coupon yields and par yields. Finally, we discuss how to estimate the model.

C.1 OIS Term Structure

The zero-coupon OIS term structure is the "risk-free rate" term structure in our model. Denote the swap rate as $r_{n,t}$. The swap exchanges floating payment pegged to the short-term OIS rate r_t to the fixed swap rate $r_{n,t}$. By construction, the floating leg and the fixed leg should have the same present value. Thus,

$$\exp(nr_{n,t})E_t^{\mathbb{Q}}[\exp(\sum_{k=1}^{n} -r_{t+k-1})] = 1$$

Conjecture

$$nr_{n,t} = A_n + B_n z_t$$

Then we have

$$\begin{aligned}
\exp(-n \cdot r_{n,t}) = \exp(-A_n - B_n z_t) &= E_t^{\mathbb{Q}}[\exp(-\sum_{k=1}^{n} r_{t+k-1})] \\
&= E_t^{\mathbb{Q}}[E_{t+1}^{\mathbb{Q}}[\exp(-\sum_{k=1}^{n-1} r_{(t+1)+k-1})]\exp(-r_t)] \\
&= E_t^{\mathbb{Q}}[\exp(-A_{n-1} - B_{n-1}z_{t+1} - \delta_0 - \delta_1 z_t)] \\
&= E_t^{\mathbb{Q}}[\exp(-A_{n-1} - B_{n-1}k_{0,z}^{\mathbb{Q}} - B_{n-1}K_{1,z}^{\mathbb{Q}}z_{t+1} + \frac{1}{2}B_{n-1}\Sigma_z(B_{n-1})^T - \delta_0 - \delta_1 z_t)]
\end{aligned}$$

which implies

$$
\begin{aligned}
A_n &= \delta_0 + A_{n-1} + B_{n-1}k^{\mathbb{Q}}_{0,z} - \frac{1}{2}B_{n-1}\Sigma_z(B_{n-1})^T \\
B_n &= \delta_1 + B_{n-1}K^{\mathbb{Q}}_{1,z}
\end{aligned}
$$

for all $n \geq 1$. The starting values are $A_0 = B_0 = 0$.

C.2 Synthetic-Rate Term Structure

We denote the synthetic rate as $r_{n,t} + r^{cip}_{n,t} \equiv r^{syn}_{n,t}$, i.e., composed of both OIS rate and the cross-currency basis. Conjecture that the cumulative synthetic rate is affine in the state vector,

$$
n(r^{cip}_{n,t} + r_{n,t}) = A^{syn}_n + B^{syn}_n z_t
$$

Then we have

$$
\begin{aligned}
&\exp(-n(r^{cip}_{n,t} + r_{n,t})) = \exp(-A^{syn}_n - B^{syn}_n z_t) \\
&= E^{\mathbb{Q}}_t[\exp(\sum_{k=1}^{n}(-r^{cip}_{t+k-1} - r_{t+k-1}))] \\
&= E^{\mathbb{Q}}_t[E^{\mathbb{Q}}_{t+1}[\exp(\sum_{k=1}^{n-1}(-r^{cip}_{(t+1)+k-1} - r_{(t+1)+k-1}))]\exp(-r^{cip}_t - r_t)] \\
&= E^{\mathbb{Q}}_t[\exp(-A^{syn}_{n-1} - B^{syn}_{n-1}z_{t+1} - (\delta_0 + \hat{\delta}_0) - (\delta_1 + \hat{\delta}_1)z_t)] \\
&= E^{\mathbb{Q}}_t[\exp(-A^{syn}_{n-1} - B^{syn}_{n-1}k^{\mathbb{Q}}_{0,z} - B^{syn}_{n-1}K^{\mathbb{Q}}_{1,z}z_{t+1} + \frac{1}{2}B^{syn}_{n-1}\Sigma_z(B^{syn}_{n-1})^T - (\delta_0 + \hat{\delta}_0) - (\delta_1 + \hat{\delta}_1)z_t)]
\end{aligned}
$$

The above equation is the present value of a CIP strategy that earns the CIP deviations, and the values is the same as the long-term CIP discounted at the long-term discount rate. Then we obtain the following iteration:

$$
\begin{aligned}
A^{syn}_n &= \delta_0 + \hat{\delta}_0 + A^{syn}_{n-1} + B^{syn}_{n-1}k^{\mathbb{Q}}_{0,z} - \frac{1}{2}B^{syn}_{n-1}\Sigma_z(B^{syn}_{n-1})^T \\
B^{syn}_n &= \delta_1 + \hat{\delta}_1 + B^{syn}_{n-1}K^{\mathbb{Q}}_{1,z}
\end{aligned}
$$

with the starting values $A^{syn}_0 = B^{syn}_0 = 0$.

C.3 Treasury Net Long Curve

Next, we derive the iteration steps for the Treasury net long curve.

$$e^{-\frac{n}{12}y^l_{n,t}}e^{x_{1,t}} = E^Q_t[e^{-\frac{n-1}{12}y^l_{n-1,t+1}}]$$

We use $\iota_1 = (1,0,0)$ to denote the indicator vector of the first element, so $x_{1,t} = \iota_1 x_t = \iota_1(\gamma_0 + \Gamma_1 z_t)$. Conjecture that the cumulative yield is affine in the state vector,

$$\frac{n}{12}y^l_{n,t} = A^l_n + B^l_n z_t$$

For all $n \geq 7$, the iteration is

$$\begin{aligned}
\exp(-\left(A^l_n + B^l_n z_t\right)) &= E^Q_t[e^{-\frac{n-1}{12}y^l_{n-1,t+1} - \iota_1(\gamma_0+\Gamma_1 z_t)}] \\
&= E^Q_t[\exp(-\left(A^l_{n-1} + B^l_{n-1}z_{t+1} + \iota_1(\gamma_0 + \Gamma_1 z_t)\right))] \\
&= E^Q_t[\exp(-\left(A^l_{n-1} + B^l_{n-1}(k^{\mathbb{Q}}_{0,z} + K^{\mathbb{Q}}_{1,z} \cdot z_t + (\Sigma_z)^{1/2}\varepsilon^{\mathbb{Q}}_{z,t+1}) + \iota_1(\gamma_0 + \Gamma_1 z_t)\right))] \\
&= E^Q_t[\exp(-\left(A^l_{n-1} + B^l_{n-1}k^{\mathbb{Q}}_{0,z} + \iota_1\gamma_0 - \frac{1}{2}B^l_{n-1}\Sigma_z(B^l_{n-1})' + (B^l_{n-1}K^{\mathbb{Q}}_{1,z} + \iota_1\Gamma_1)\cdot z_t\right))]
\end{aligned}$$

which implies the iteration equation

$$\begin{aligned}
A^l_n &= \iota_1\gamma_0 + A^l_{n-1} + B^l_{n-1}k^{\mathbb{Q}}_{0,z} - \frac{1}{2}B^l_{n-1}\Sigma_z(B^l_{n-1})^T \\
B^l_n &= \iota_1\Gamma_1 + B^l_{n-1}K^{\mathbb{Q}}_{1,z}
\end{aligned}$$

At $n = 6$, we have

$$\frac{6}{12}y^l_{6,t} = \frac{6}{12}y^{bill}_t = 6x_{3,t} = 6\iota_3(\gamma_0 + \Gamma_1 z_t)$$

with initial values

$$A^l_6 = 6\iota_3\gamma_0, \quad B^l_6 = 6\iota_3\Gamma_1$$

C.4 Treasury Net Short Curve

Next, we derive the iteration steps for the Treasury net short curve.

$$e^{-\frac{n}{12}y^s_{n,t}}e^{x_{2,t}} = E^Q_t\left[e^{-\frac{n-1}{12}y^s_{n-1,t+1}}\right]$$

Similar arguments as in the last section will lead to cumulative yield

$$\frac{n}{12}y^s_{n,t} = A^s_n + B^s_n z_t$$

where

$$\begin{aligned} A^s_n &= \iota_2\gamma_0 + A^s_{n-1} + B^s_{n-1}k^{\mathbb{Q}}_{0,z} - \frac{1}{2}B^s_{n-1}\Sigma_z(B^s_{n-1})^T \\ B^s_n &= \iota_2\Gamma_1 + B^s_{n-1}K^{\mathbb{Q}}_{1,z} \end{aligned}$$

At $n = 6$, we have

$$\frac{6}{12}y^s_{6,t} = \frac{6}{12}y^{bill}_t = 6x_{3,t} = 6\iota_3(\gamma_0 + \Gamma_1 z_t)$$

with initial values

$$A^s_6 = 6\iota_3\gamma_0, \quad B^s_6 = 6\iota_3\Gamma_1$$

C.5 Par Curve and Zero Curve Conversion

In our term structure model, all the yields are zero-coupon yields. In the data, on the other hand, yields are par yields. The ideal way to resolve the mismatch is asking the model to convert all zero-coupon yields into par yields. However, the model is solved thousands of times when we estimate it, and the extra conversion significantly slows the estimation process. Thus, we do the following:

- We convert the OIS term structure and the CIP basis term structure into zero-coupon yields for model estimation purpose.
- Once we finish estimating the model, then we generate the net long and net short zero-coupon curves, and convert them into par yields.

For the par-to-zero conversion, we follow the standard Svensson (1994) method that fits the whole yield curve with a parsimonious functional form and infer the zero yields.

For the zero-to-par conversion, we directly use the definition. We want to transform the annualized zero-coupon yields $r_{n,t}$ into annualized par yields $r_{n,t}^{par}$ with coupon payment every 6 months. Then for a coupon-bond of maturity n, the pricing relationship is

$$q_{n,t}^{par} = \frac{r_{n,t}^{par}}{2}\left(e^{-\frac{6}{12}r_{t,6}} + e^{-\frac{12}{12}r_{t,12}} + \cdots + e^{-\frac{n}{12}r_{n,t}}\right) + e^{-\frac{n}{12}r_{n,t}} \tag{C-1}$$

For a bond at the par, the price is $q_{n,t}^{par} = 1$, indicating the par yield as

$$r_{n,t}^{par} = 2 \times \frac{1 - e^{-\frac{n}{12}r_{n,t}}}{e^{-\frac{6}{12}r_{t,6}} + e^{-\frac{12}{12}r_{t,12}} + \cdots + e^{-\frac{n}{12}r_{n,t}}} \tag{C-2}$$

C.6 Model Estimation

We estimate the model to fit the OIS and basis term structure. Then we use regression-implied coefficients γ_0 and Γ_1 to obtain the model-implied net long and net short curves. Denote the observed OIS yield of maturity n at time t as

$$\hat{r}_{n,t} = r_{n,t} + \xi_{n,t}^{ois}, \quad \xi_t^{ois} \sim \mathcal{N}(0, \Sigma_{ois})$$

and the observed basis as

$$\hat{r}_{n,t}^{cip} = r_{n,t}^{cip} + \xi_{n,t}^{basis}, \quad \xi_t^{basis} \sim \mathcal{N}(0, \Sigma_{basis})$$

We denote the stacked OIS yields (across different maturities) as $\hat{r}_t$, and the stacked basis rates as $\hat{r}_t^{cip}$. For the estimation step, the set of parameters is $\Theta = \{k_{0,z}^{\mathbb{Q}}, K_{1,z}^{\mathbb{Q}}, k_{0,z}^{\mathbb{P}}, K_{1,z}^{\mathbb{P}}, \Sigma_z, \Sigma_{ois}, \Sigma_{basis}, \delta_0, \delta_1, \hat{\delta}_0, \hat{\delta}_1\}$. The objective of the estimation is to maximize the log likelihood that the observed yields are generated by the model,

$$\mathcal{L}(\{\hat{r}_t, \hat{r}_t^{cip}, z_t\}_{t\in\text{data}}; \Theta)$$

Denote the log likelihood an N-variable normal variable Z with mean μ and variance matrix Σ as $\mathscr{G}(Z,\mu,\Sigma)$. Then the objective function is

$$\mathscr{L}(\{\hat{r}_t,\hat{r}_t^{cip},z_t\}_{t\in\text{data}};\Theta)=$$
$$\sum_{t\in\text{data}}\left(\underbrace{\mathscr{G}(z_t-k_{0,z}^{\mathbb{P}}-K_{1,z}^{\mathbb{P}}\cdot z_{t-1},0,\Sigma_z)}_{\text{state variable physical dynamics}}+\underbrace{\mathscr{G}(\hat{r}_t-r_t,0,\Sigma_{ois})}_{\text{OIS fitting}}+\underbrace{\mathscr{G}(\hat{r}_t^{cip}-r_t^{cip},0,\Sigma_{basis})}_{\text{basis fitting}}\right)$$

The whole estimation problem is thus

$$\max_{k_{0,z}^{\mathbb{Q}},K_{1,z}^{\mathbb{Q}},k_{0,z}^{\mathbb{P}},K_{1,z}^{\mathbb{P}},\Sigma_z,\Sigma_{ois},\Sigma_{basis},\delta_0,\delta_1,\hat{\delta}_0,\hat{\delta}_1}\mathscr{L}(\{\hat{r}_t,\hat{r}_t^{cip},z_t\}_{t\in\text{data}};\Theta)\tag{C-3}$$

To reduce dimensionality, we assume that the covariance matrices for observation errors are in the form of $\Sigma_{ois}=\sigma_{ois}I$ and $\Sigma_{basis}=\sigma_{basis}I$.

Compared to the classical term structure estimation problem, the key challenge of this problem is that we need to estimate two inter-linked term structures simultaneously. However, the canonical form transformation in Joslin et al. (2011) only applies to one term structure. To resolve the challenge and at the same time taking advantage of the canonical form, we design the following two-step procedure that applies the canonical form to each individual term struccture as initialization (the initial values for this high-dimensional optimization problem are quite important):

1. Divide the state-space into two blocks, an OIS block, $z_t^{ois}=(z_{1,t},z_{2,t},z_{3,t})$, and a basis block $z_t^{basis}=(z_{4,t},z_{5,t})$. Similarly, we denote the associated sub-group risk-neutral dynamic parameters as $k_{0,z^{ois}}^{\mathbb{Q}},K_{1,z^{ois}}^{\mathbb{Q}}$ and $k_{0,z^{basis}}^{\mathbb{Q}},K_{1,z^{basis}}^{\mathbb{Q}}$. Denote the sub-group physical dynamic parameters as $k_{0,z^{ois}}^{\mathbb{P}},K_{1,z^{ois}}^{\mathbb{P}}$ and $k_{0,z^{basis}}^{\mathbb{P}},K_{1,z^{basis}}^{\mathbb{P}}$. Also divide the observations into the OIS group and basis group. Then apply the standard canonical form estimation procedure to two models separately,

$$\mathscr{L}(\{\hat{r}_t,z_t\}_{t\in\text{data}}^{ois};k_{0,z^{ois}}^{\mathbb{Q}},K_{1,z^{ois}}^{\mathbb{Q}},k_{0,z^{ois}}^{\mathbb{P}},K_{1,z^{ois}}^{\mathbb{P}},\Sigma_z^{ois},\Sigma_{ois},\delta_0^{ois},\delta_1^{ois})$$

$$\mathscr{L}(\{\hat{r}_t,z_t\}_{t\in\text{data}}^{basis};k_{0,z^{basis}}^{\mathbb{Q}},K_{1,z^{basis}}^{\mathbb{Q}},k_{0,z^{basis}}^{\mathbb{P}},K_{1,z^{basis}}^{\mathbb{P}},\Sigma_z^{basis},\Sigma_{basis},\delta_0^{basis},\delta_1^{basis})$$

where the short rate in the first estimation is $\delta_0^{ois}+\delta_1^{ois}*z_t^{ois}$, and the short rate in the second estimation is $\delta_0^{basis}+\delta_1^{basis}*z_t^{basis}$ is a two-dimensional vector that loads on z_t^{basis}. The covariance matrix Σ_z^{ois} is 3×3 and Σ_z^{basis} is 2×2. After estimating the above dynamics, we construct an initialization of the original problem as

$$k_{0,z}^{\mathbb{Q}} = \begin{pmatrix} k_{0,z^{ois}}^{\mathbb{Q}} \\ k_{0,z^{basis}}^{\mathbb{Q}} \end{pmatrix}, \quad K_{1,z}^{\mathbb{Q}} = \begin{pmatrix} K_{1,z^{ois}}^{\mathbb{Q}} & \\ & K_{1,z^{basis}}^{\mathbb{Q}} \end{pmatrix}, \quad \Sigma_z = \begin{pmatrix} \Sigma_z^{ois} & \\ & \Sigma_z^{basis} \end{pmatrix}$$

$$\delta_0 = \delta_0^{ois}, \quad \delta_1 = \begin{pmatrix} \delta_1^{ois} \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\delta}_0 = \delta_0^{ois} + \delta_0^{basis}, \quad \hat{\delta}_1 = \begin{pmatrix} \delta_1^{ois} \\ \delta_1^{basis} \end{pmatrix}$$

We initialize the physical dynamic parameters $(k_{0,z}^{\mathbb{P}}, K_{1,z}^{\mathbb{P}})$ simply from linear regressions,

$$z_t \sim k_{0,z}^{\mathbb{P}} + K_{1,z}^{\mathbb{P}} \cdot z_{t-1}$$

2. Then we feed these initial values to the whole estimation problem (C-3), and apply the optimization package in Matlab to optimize over the whole high-dimensional parameter space. We use the equivalent implementation of the CIP short rate (instead of the synthetic lending short rate), $r_t^{syn} - r_t$, and the corresponding loading $\hat{\delta}_0 - \delta_0 + (\hat{\delta}_1 - \delta_1)z_t$.

After we finish estimating the key parameter set Θ, we proceed to obtain γ_0 and Γ_1 via a simple linear regressions,

$$x_t \sim \gamma_0 + \Gamma_1 z_t$$

We find that the residual standard errors for this linear regression are one order of magnitude smaller than Σ_z. In other words, we are able to obtain very accurate approximation of x_t through the state vector z_t, so adding the extra estimation error to the above approximation in the model will not cause much difference, but it requires augmenting the state space. For this reason, we make the assumption that x_t is spanned by z_t in the main model.

Finally with estimated Θ and (γ_0, Γ_1), we are able to obtain the Treasury net long and net short curves. We convert these curves into par curves to be comparable with the Treasury yield data.

C.7 Stationarity Restrictions

Treasury yields can appear non-stationary (in sample), but the spread between an OIS rate and the matched-maturity Treasury yield cannot diverge due to arbitrage incentives in financial markets. Our main approach does not impose such a restriction for simplicity. In this subsection, we discuss how to impose stationarity on the process x_t and show that results are broadly similar.

First, our estimation reveals that z_t contains unit-root processes. In particular, the $\mathbb{Q}$-dynamics of z_t contains unit-root elements. Denote the eigenvalue decomposition of $K_{1,z}^{\mathbb{Q}}$ as

$$K_{1,z}^{\mathbb{Q}} = VDV^{-1}$$

where D is an diagonal matrix that contains all the eigenvalues of $K_{1,z}^{\mathbb{Q}}$, and V is the matrix of all the column eigenvectors for $K_{1,z}^{\mathbb{Q}}$. We find that two among the five eigenvalues have absolute values above 0.999, which is a strong sign of unit root.

To operationalize the stationarity restriction, we rotate the state vector z_t to $\tilde{z}_t = V^{-1}z_t$, and rewrite the $\mathbb{Q}$-dynamics in (17) as

$$\tilde{z}_{t+1} = V^{-1}k_{0,z}^{\mathbb{Q}} + D\tilde{z}_t + V^{-1}(\Sigma_z)^{1/2}\varepsilon_{z,t+1}^{\mathbb{Q}}, \varepsilon_{z,t+1}^{\mathbb{Q}} \sim N(0, I_N)$$

We denote the spread vector as

$$\hat{x}_t = x_t - \begin{pmatrix} r_t + r_t^{cip} \\ r_t - r_t^{cip} \\ r_t \end{pmatrix}$$

Then we project $\hat{x}_t$ on the stationary components of $\tilde{z}_t$, i.e., three of five with (absolute values of) eigenvalues below 0.999. The loadings on the non-stationary components are set as zeros. Then we denote the whole projection as

$$\hat{x}_t = \tilde{\gamma}_0 + \tilde{\Gamma}_1\tilde{z}_t$$

Next, we rotate back to z_t,

$$\hat{x}_t = \tilde{\gamma}_0 + \tilde{\Gamma}_1V^{-1}z_t$$

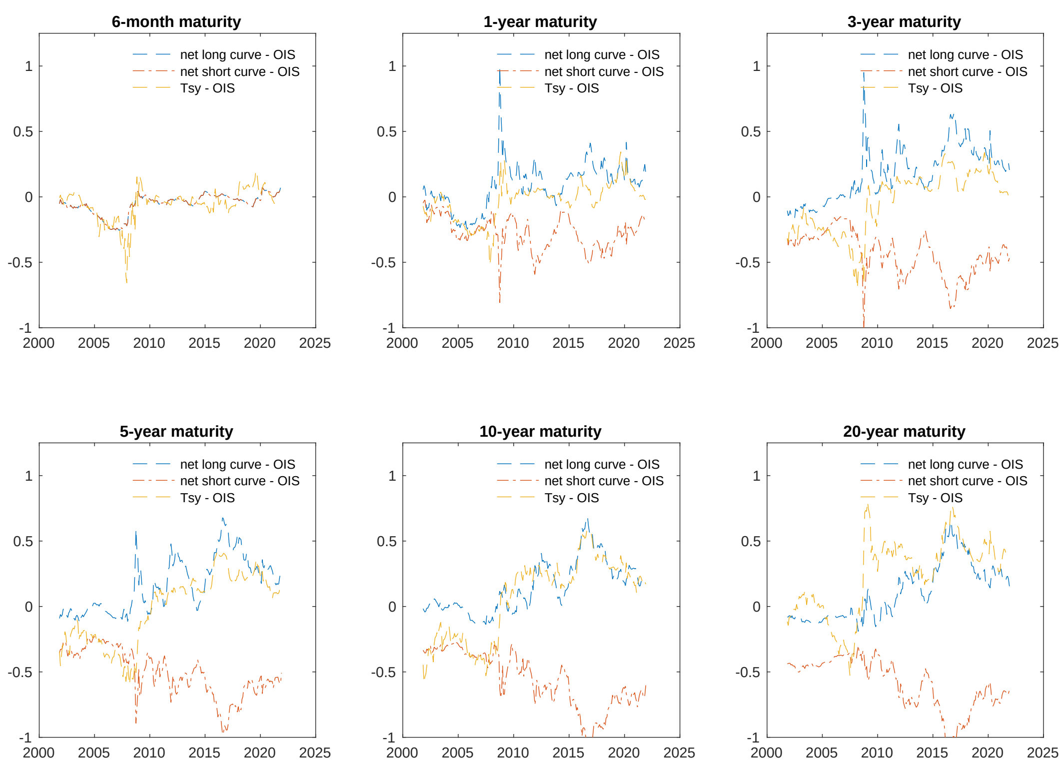
Thus, we obtain

$$x_t = \tilde{\gamma}_0 + \tilde{\Gamma}_1 V^{-1} z_t + \begin{pmatrix} r_t + r_t^{cip} \\ r_t - r_t^{cip} \\ r_t \end{pmatrix}$$

In the implementation, we find that there are complex-number eigenvalues, so the resulting $\tilde{\gamma}_0$ and $\tilde{\Gamma}_1 V^{-1}$ are also complex numbers. Nevertheless, the imaginary parts are quite small so we only keep the real parts.

With the projection of x_t on z_t, we are able to derive the Treasury net long and net short curves. We illustrate the results in Figure A3. We find that results are very close to the baseline results in Figure 9. Furthermore, all other results, such as the relative yield index matching the movements in dealer position, are quite similar. For conciseness, we omit other results in this appendix.

Figure A3: Long and Short Curves – OIS Spreads using the Alternative Projection Method.



Notes: In this figure, we show the model-implied long and short Treasury curves minus the OIS rates for corresponding maturities, together with the actual Treasury–OIS spreads. We use the alternative projection method as in Internet Appendix Section C.7. Data are from 2003 to 2021. All yields are par yields.

D Proofs for the Equilibrium Model

D.1 Proof of Proposition 1 (Long Regime)

Define the function

$$f_1^{long}(y, r^{syn}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}) = e^{-(ny-(n-1)\varepsilon)} - \frac{\exp(-(n-1)y_{\mathbb{Q}})}{e^{i^l} - e^r + e^{r^{syn}}}$$

and the function

$$\begin{aligned}
&f_2^{long}(y, r^{syn}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}) = \\
&\bar{q} - e^{-(ny-(n-1)(\varepsilon-\omega))} S^{bond} + D_U(ny - y^{bill} - (n-1)(y_{\mathbb{P}} + \varepsilon)) + \delta_U - (D^{syn}(r^{syn} - r) + \delta_{syn}).
\end{aligned}$$

By assumption, D_U and D^{syn} are continuously differentiable, and hence f_1 and f_2 are continuously differentiable.

Suppose there exists, given the exogenous values $y_{\mathbb{P}}, r, i^l, y^{bill}$ and some initial point ($S^{bond} > 0, \bar{q} > 0, y_{\mathbb{Q}}, \varepsilon = 0, \omega = 0, \delta_U = 0, \delta_H = 0, \delta_{syn} = 0$), a solution

$$\begin{bmatrix} f_1^{long}(y^*, r^{syn*}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, 0,0,0,0,0) \\ f_2^{long}(y^*, r^{syn*}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, 0,0,0,0,0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

such that

$$D_H(ny^* - r^{syn*} - (n-1)y_{\mathbb{P}}) + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) < e^{-ny^*} S^{bond}.$$

Such a point constitutes an equilibrium.

Observe that

$$\frac{\partial f_1^{long}(\cdot)}{\partial y} < 0, \frac{\partial f_1^{long}(\cdot)}{\partial r^{syn}} > 0, \frac{\partial f_2^{long}(\cdot)}{\partial y} > 0, \frac{\partial f_2^{long}(\cdot)}{\partial r^{syn}} > 0,$$

and consequently

$$\begin{bmatrix} \frac{\partial f_1^{long}(\cdot)}{\partial y} & \frac{\partial f_1^{long}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{long}(\cdot)}{\partial y} & \frac{\partial f_2^{long}(\cdot)}{\partial r^{syn}} \end{bmatrix}$$

is invertible (its determinant is strictly negative).

It follows that the equilibrium (y^*, r^{syn*}), if it exists, is unique. Suppose not and there exists another equilibrium (y, r^{syn}) in the long regime. If $r^{syn} > r^{syn*}$, then $\tilde{y} > y^*$ according to $f_1^{long}(\tilde{y}, r^{syn}) = 0$. By monotonicity of f_2^{long}, we have $f_2^{long}(y, r^{syn}) > f_2^{long}(y^*, r^{syn*}) = 0$, which contradicts to (y, r^{syn}) being an equilibrium. A symmetric argument rules out all $r^{syn} < r^{syn*}$. Thus, the equilibrium solution to y^* is unique. Strict monotonicity ensures the uniqueness of y^*.

By the implicit function theorem,

$$\begin{bmatrix} \frac{\partial y^*(\cdot)}{\partial x} \\ \frac{\partial r^{syn*}(\cdot)}{\partial x} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1^{long}(\cdot)}{\partial y} & \frac{\partial f_1(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2(\cdot)}{\partial y} & \frac{\partial f_2(\cdot)}{\partial r^{syn}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1^{long}(\cdot)}{\partial x} \\ \frac{\partial f_2^{long}(\cdot)}{\partial x} \end{bmatrix}$$

for any $x \in \{S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}\}$. Observe that the signs of the negative inverse matrix are

$$sgn\left(- \begin{bmatrix} \frac{\partial f_1^{long}(\cdot)}{\partial y} & \frac{\partial f_1^{long}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{long}(\cdot)}{\partial y} & \frac{\partial f_2^{long}(\cdot)}{\partial r^{syn}} \end{bmatrix}^{-1} \right) = sign\left(\begin{bmatrix} \frac{\partial f_2^{long}(\cdot)}{\partial r^{syn}} & -\frac{\partial f_1^{long}(\cdot)}{\partial r^{syn}} \\ -\frac{\partial f_2^{long}(\cdot)}{\partial y} & \frac{\partial f_1^{long}(\cdot)}{\partial y} \end{bmatrix} \right) = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}.$$

We solve for the comparative statics as follows:

1. An increase in S^{bond}: $\frac{\partial f_1^{long}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} < 0$, and therefore $\frac{\partial y^*(\cdot)}{\partial x} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

2. A decrease in $\bar{q}$ or a decrease in δ_U: $\frac{\partial f_1^{long}(\cdot)}{\partial x} = 0$ and

$$\frac{\partial f_2^{long}(\cdot)}{\partial x} = -\frac{\partial f_2^{long}(\cdot)}{e^{-ny} \cdot \partial(S^{bond})}.$$

   Thus, the decrease in $\bar{q}$ or δ_U is equivalent to the same same size expansion in the dollar supply of bonds.

3. An increase in δ_H has $\frac{\partial f_1^{long}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} = 0$, and therefore $\frac{\partial y^*(\cdot)}{\partial x} = 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} = 0$.

4. An increase in $y_{\mathbb{Q}}$ has $\frac{\partial f_1^{long}(\cdot)}{\partial x} > 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} = 0$ and thus $\frac{\partial y^*(\cdot)}{\partial x} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} < 0$.

5. An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ is equivalent to an increase ε by $\Delta\varepsilon = dy$ in both f_1 and f_2 and an increase in ω by $\Delta\omega = dy$. The increase in ε causes an $\frac{n-1}{n}\Delta\varepsilon$ increase in y and no change in r^{syn}. The increase in ω has $\frac{\partial f_1^{long}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} > 0$, and thus thus $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} < 0$. Taking the two effects together, clearly r^{syn*} will decrease. To determine the sign on y^*, we can evaluate the change of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ directly and obtain $\frac{\partial f_1^{long}(\cdot)}{\partial x} > 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} < 0$, which implies y^* will increase. In summary, we find that the increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ increases y^* by less than $\frac{n-1}{n}\Delta\varepsilon$ and decreases r^{syn*}. Furthermore, the absolute value of the effect of ω is smaller than that of ε, indicating that the total effect is still to increase bond yield.

   Taking the two effects together, we find that the increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ increase y^* by less than $\frac{n-1}{n}\Delta\varepsilon$ and decreases r^{syn*}.

6. An increase in δ_{syn} has $\frac{\partial f_1^{long}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} < 0$, and thus $\frac{\partial y^*(\cdot)}{\partial x} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

D.2 Proof of Proposition 2 (Short Regime)

Define the function

$$f_1^{short}(y, r^{syn}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}) = e^{-(ny-(n-1)\varepsilon)} - \frac{\exp(-(n-1)y_{\mathbb{Q}})}{e^{i^s} + e^r - e^{r^{syn}}}$$

and the function

$$\begin{aligned} f_2^{short}(y, r^{syn}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}) =& \bar{q} + e^{-(ny-(n-1)(\varepsilon-\omega))} S^{bond} \\ & - D_U(ny - y^{bill} - (n-1)(y_{\mathbb{P}} + \varepsilon)) - \delta_U \\ & - (D^{syn}(r^{syn} - r) + \delta_{syn}) \\ & - 2(D_H(ny - y^{syn} - (n-1)(y_{\mathbb{P}} + \varepsilon)) + \delta_H) \end{aligned}$$

By assumption, D_U, D_H, and D^{syn} are continuously differentiable, and hence f_1^{short} and f_2^{short} are continuously differentiable.

Suppose there exists, given the exogenous values $y_{\mathbb{P}}, r, i^s, y^{bill}$ and some initial point ($S^{bond} >$

$0, \bar{q} > 0, y_{\mathbb{Q}}, \varepsilon = 0, \omega = 0, \delta_U = 0, \delta_H = 0, \delta_{syn} = 0)$, a solution

$$\begin{bmatrix} f_1^{short}(y^*, r^{syn*}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, 0,0,0,0,0) \\ f_2^{short}(y^*, r^{syn*}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, 0,0,0,0,0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

such that

$$D_H(ny^* - r^{syn*} - (n-1)y_{\mathbb{P}}) + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) > e^{-ny^*} S^{bond}.$$

Such a point constitutes an equilibrium.

Observe that

$$\frac{\partial f_1^{short}(\cdot)}{\partial y} < 0, \frac{\partial f_1^{short}(\cdot)}{\partial r^{syn}} < 0, \frac{\partial f_2^{short}(\cdot)}{\partial y} < 0, \frac{\partial f_2^{short}(\cdot)}{\partial r^{syn}} > 0,$$

and consequently

$$\begin{bmatrix} \frac{\partial f_1^{short}(\cdot)}{\partial y} & \frac{\partial f_1^{short}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{short}(\cdot)}{\partial y} & \frac{\partial f_2^{short}(\cdot)}{\partial r^{syn}} \end{bmatrix}$$

is invertible (its determinant is strictly negative).

It follows that the equilibrium (y^*, r^{syn*}), if it exists, is unique. Suppose not and there exists another pair (r^{syn}, y) that satisfies the eqilibrium in the short regime. If $r^{syn} > r^{syn*}$, we must have $y < y^*$ due to $f_1^{short}(y, r^{syn}) = f_1^{short}(y^*, r^{syn*})$. (if no such $\tilde{y}$ exists, r^{syn} cannot be part of an equilibrium). It follows that $f_2^{short}(y, r^{syn}) > f_2^{short}(y^*, r^{syn*}) = 0$, and hence r^{syn} cannot be part of an equilibrium. A symmetric argument rules out all $r^{syn} < r^{syn*}$, and strict monotonicity ensures the uniqueness of y^*.

By the implicit function theorem,

$$\begin{bmatrix} \frac{\partial y^*(\cdot)}{\partial x} \\ \frac{\partial r^{syn*}(\cdot)}{\partial x} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1^{short}(\cdot)}{\partial y} & \frac{\partial f_1^{short}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{short}(\cdot)}{\partial y} & \frac{\partial f_2^{short}(\cdot)}{\partial r^{syn}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1^{short}(\cdot)}{\partial x} \\ \frac{\partial f_2^{short}(\cdot)}{\partial x} \end{bmatrix}$$

for any $x \in \{S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H\}$. Observe that the signs of the negative inverse matrix are

$$sgn\left(-\begin{bmatrix} \frac{\partial f_1^{short}(\cdot)}{\partial y} & \frac{\partial f_1^{short}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{short}(\cdot)}{\partial y} & \frac{\partial f_2^{short}(\cdot)}{\partial r^{syn}} \end{bmatrix}^{-1}\right) = sign\left(\begin{bmatrix} \frac{\partial f_2^{short}(\cdot)}{\partial r^{syn}} & -\frac{\partial f_1^{short}(\cdot)}{\partial r^{syn}} \\ -\frac{\partial f_2^{short}(\cdot)}{\partial y} & \frac{\partial f_1^{short}(\cdot)}{\partial y} \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

We solve for the comparative statics as follows:

1. An increase in S^{bond}: $\frac{\partial f_1^{short}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{short}(\cdot)}{\partial x} > 0$, and therefore $\frac{\partial y^*(\cdot)}{\partial x} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} < 0$. Thus, bond yield y^* increases, but the synthetic rate r^{syn*} decreases.

2. An increase in $\bar{q}$ or a decrease in δ_U (i.e., a parallel decrease in D_U): $\frac{\partial f_1^{short}(\cdot)}{\partial x} = 0$ and

$$\frac{\partial f_2^{short}(\cdot)}{\partial \bar{q}} = -\frac{\partial f_2^{short}(\cdot)}{\partial \delta_U} = \frac{\partial f_2^{short}(\cdot)}{e^{-ny} \cdot \partial(S^{bond})}.$$

   Thus, the increase in $\bar{q}$ or the same decrease in δ_U are equivalent to the same same size expansion in the dollar supply of bonds.

3. An increase in δ_H has $\frac{\partial f_1^{short}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{short}(\cdot)}{\partial x} < 0$, and therefore $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

4. An increase in $y_{\mathbb{Q}}$ has $\frac{\partial f_1^{short}(\cdot)}{\partial x} > 0, \frac{\partial f_2^{short}(\cdot)}{\partial x} = 0$ and thus $\frac{\partial y^*(\cdot)}{\partial x} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

5. An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$: this change is equivalent to an increase ε by dy in both f_1^{short} and f_2^{short} and an increase in ω by dy. The increase in ε causes an $\frac{n-1}{n}\Delta\varepsilon$ decrease in y and no change in r^{syn*}. The increase in ω has $\frac{\partial f_1^{short}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{short}(\cdot)}{\partial x} < 0$, and thus thus $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$. Taking the two effects together, clearly r^{syn*} will increase. To determine the sign on y^*, we can evaluate the change of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ directly and obtain $\frac{\partial f_1^{short}(\cdot)}{\partial x} > 0, \frac{\partial f_2^{short}(\cdot)}{\partial x} > 0$, which implies y^* will increase. In summary, we find that the increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ increases y^* by less than $\frac{n-1}{n}\Delta\varepsilon$ and increases r^{syn*}.

6. An increase in δ_{syn} has $\frac{\partial f_1^{long}(\cdot)}{\partial x} = 0, \frac{\partial f_2^{long}(\cdot)}{\partial x} < 0$, and thus $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

D.3 Proof of Proposition 3 (Intermediate Regime)

Define the functions

$$f_1^{int}(y, r^{syn}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}) = \bar{q} - e^{-(ny-(n-1)(\varepsilon-\omega))}S^{bond} - (D^{syn}(r^{syn} - r) + \delta_{syn}) \\ + D_U(ny - y^{bill} - (n-1)(y_{\mathbb{P}} + \varepsilon)) + \delta_U$$

and

$$f_2^{int}(y, r^{syn}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H, \delta_{syn}) = \bar{q} - (D_H(ny - r^{syn} - (n-1)(y_{\mathbb{P}} + \varepsilon)) + \delta_H) \\ - (D^{syn}(r^{syn} - r) + \delta_{syn}).$$

By assumption, D_U, D_H, and D^{syn} are continuously differentiable, and hence f_1^{int} and f_2^{int} are continuously differentiable.

Suppose there exists, given the exogenous values $y_{\mathbb{P}}, r, r^{long}, y^{bill}$ and some initial point ($S^{bond} > 0, \bar{q} > 0, y_{\mathbb{Q}}, \varepsilon = 0, \omega = 0, \delta_U = 0, \delta_H = 0, \delta_{syn} = 0$), a solution

$$\begin{bmatrix} f_1^{int}(y^*, r^{syn*}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, 0, 0, 0, 0, 0) \\ f_2^{int}(y^*, r^{syn*}; S^{bond}, \bar{q}, y_{\mathbb{Q}}, 0, 0, 0, 0, 0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

such that

$$y^s < y^* < y^l$$

Such a point constitutes an interior equilibrium.

Observe that

$$\frac{\partial f_1^{int}(\cdot)}{\partial y} = ne^{-ny}S^{bond} + nD_U' > 0$$

$$\frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} = -(D^{syn})' > 0$$

$$\frac{\partial f_2^{int}(\cdot)}{\partial y} = -nD_H' < 0$$

$$\frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} = D_H' - (D^{syn})' > 0$$

Then the determinant of the derivative matrix

$$\begin{bmatrix} \frac{\partial f_1^{int}(\cdot)}{\partial y} & \frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{int}(\cdot)}{\partial y} & \frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} \end{bmatrix}$$

is positive, which implies that the derivative matrix is invertible.

It follows that the equilibrium (y^*, r^{syn*}), if it exists, is unique. Suppose not and there exists another pair (r^{syn}, y) that satisfies the equilibrium in the intermediate regime. If $r^{syn} > r^{syn*}$, we must have $y > y^*$ due to $f_2^{int}(y, r^{syn}) = f_2^{int}(y^*, r^{syn*})$. It follows that $f_1^{int}(y, r^{syn}) > f_1^{int}(y^*, r^{syn*}) = 0$, and hence r^{syn} cannot be part of an equilibrium. A symmetric argument rules out all $r^{syn} < r^{syn*}$. Strict monotonicity also guarantees the uniqueness of y^*.

By the implicit function theorem,

$$\begin{bmatrix} \frac{\partial y^*(\cdot)}{\partial x} \\ \frac{\partial r^{syn*}(\cdot)}{\partial x} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1^{int}(\cdot)}{\partial y} & \frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{int}(\cdot)}{\partial y} & \frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1^{int}(\cdot)}{\partial x} \\ \frac{\partial f_2^{int}(\cdot)}{\partial x} \end{bmatrix}$$

for any $x \in \{S^{bond}, \bar{q}, y_{\mathbb{Q}}, \varepsilon, \omega, \delta_U, \delta_H\}$. Observe that the signs of the negative inverse matrix are

$$sgn\left(- \begin{bmatrix} \frac{\partial f_1^{int}(\cdot)}{\partial y} & \frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} \\ \frac{\partial f_2^{int}(\cdot)}{\partial y} & \frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} \end{bmatrix}^{-1} \right) = sign\left(- \begin{bmatrix} \frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} & -\frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} \\ -\frac{\partial f_2^{int}(\cdot)}{\partial y} & \frac{\partial f_1^{int}(\cdot)}{\partial y} \end{bmatrix} \right) = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}.$$

We solve for the comparative statics as follows:

1. An increase in S^{bond}: $\frac{\partial f_1^{int}(\cdot)}{\partial x} < 0, \frac{\partial f_2^{int}(\cdot)}{\partial x} = 0$, and therefore $\frac{\partial y^*(\cdot)}{\partial x} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$. Thus, both the bond yield y^* and the synthetic rate r^{syn*} increase.

2. An increase in $\bar{q}$: $\frac{\partial f_1^{int}(\cdot)}{\partial x} > 0$ and $\frac{\partial f_2^{int}(\cdot)}{\partial \bar{q}} > 0$. Thus, we have $\frac{\partial r^{syn*}(\cdot)}{\partial x} < 0$. To determine the sign of $\frac{\partial y^*(\cdot)}{\partial x}$, we note that

$$\frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} = D_H' - (D^{syn})' > \frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} = -(D^{syn})' > 0$$

$$\frac{\partial f_1^{int}(\cdot)}{\partial \bar{q}} = \frac{\partial f_2^{int}(\cdot)}{\partial \bar{q}} = 1$$

Thus,

$$\frac{\partial y^*(\cdot)}{\partial x} \propto \frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} - \frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} < 0$$

3. An increase in δ_U: $\frac{\partial f_1^{int}(\cdot)}{\partial x} > 0$ and $\frac{\partial f_2^{int}(\cdot)}{\partial \bar{q}} = 0$. Thus, we have $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} < 0$.

4. An increase in δ_H has $\frac{\partial f_1^{int}(\cdot)}{\partial x} = 0$, $\frac{\partial f_2^{int}(\cdot)}{\partial x} < 0$, and therefore $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

5. An increase in $y_{\mathbb{Q}}$ has $\frac{\partial f_1^{int}(\cdot)}{\partial x} = 0$, $\frac{\partial f_2^{int}(\cdot)}{\partial x} = 0$ and thus $\frac{\partial y^*(\cdot)}{\partial x} = 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} = 0$.

6. An increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$: this change is equivalent to an increase ε by dy in both f_1^{int} and f_2^{int} and an increase in ω by dy. The increase in ε causes an $\frac{n-1}{n}dy$ increase in y and no change in r^{syn*}. The increase in ω has $\frac{\partial f_1^{short}(\cdot)}{\partial x} > 0$, $\frac{\partial f_2^{short}(\cdot)}{\partial x} = 0$, and thus $\frac{\partial y^*(\cdot)}{\partial x} < 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$. To determine the total effect on y^*, we can evaluate the change of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ directly and obtain $\frac{\partial f_1^{int}(\cdot)}{\partial x} < 0$, $\frac{\partial f_2^{int}(\cdot)}{\partial x} > 0$, which implies that the total effect on y^* is positive. In summary, we find that the increase of dy in both $y_{\mathbb{Q}}$ and $y_{\mathbb{P}}$ increases y^* (by less than $\frac{n-1}{n}dy$) and increases r^{syn*}.

7. An increase in δ_{syn} has $\frac{\partial f_1^{long}(\cdot)}{\partial x} = \frac{\partial f_2^{long}(\cdot)}{\partial x} = -1$, and thus $\frac{\partial y^*(\cdot)}{\partial x} \propto \frac{\partial f_2^{int}(\cdot)}{\partial r^{syn}} - \frac{\partial f_1^{int}(\cdot)}{\partial r^{syn}} > 0$ and $\frac{\partial r^{syn*}(\cdot)}{\partial x} > 0$.

D.4 Proof of Proposition 4

Propositions 1, 2, and 3 establish that there is at most one equilibrium in each regime. To proceed, we first prove that across all possible regimes, the equilibrium is unique. Then we show the existence of an equilibrium. Finally, we will show how bond supply S^{bond} and the risk premium $y_{\mathbb{Q}}$ affects the equilibrium regime.

Define

$$f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = e^{-ny} - \frac{\exp(-(n-1)y_{\mathbb{Q}})}{e^{i^l} - e^r + e^{r^{syn}}}$$

$$f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = \bar{q} - e^{-ny}S^{bond} - D^{syn}(r^{syn} - r) + D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}).$$

$$f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = e^{-ny}S^{bond} - D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) - D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}).$$

$$f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = e^{-ny} - \frac{\exp(-(n-1)y_{\mathbb{Q}})}{e^{i^s} + e^r - e^{r^{syn}}}$$

$$
\begin{aligned}
f_5(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = \bar{q} + e^{-ny} S^{bond} - D^{syn}(r^{syn} - r) - D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) \\
- 2D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}).
\end{aligned}
$$

where f_1 is the residual of long-regime dealer indifference equation (31), f_2 is the residual of the long-regime market indifference curve (32), f_3 is the residual of the bond-market clearing condition in (27), f_4 is the residual of short-regime dealer indifference equation (34), and f_5 is the residual of the short-regime market indifference curve (35).

In equilibrium, bond market clearing (27) and synthetic lending market clearing (30) implies

$$f_3 = q^{bond}$$

$$D_H + D^{syn} = q^{syn}$$

By assumption, $r > i^l > i^s$, and in any equilibrium, $r^{syn} \geq r$. It follows that

$$2e^{r^{syn}} \geq 2e^r > e^{i^s} + 2e^r - e^{i^l} \geq e^{i^s} + e^r + (e^r - e^{i^l}),$$

and hence that

$$e^{r^{syn}} + (e^{i^l} - e^r) > e^{i^s} + e^r - e^{r^{syn}}.$$

It follows that

$$f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}).$$

In a long-regime equilibrium, $q^{bond} > 0$, so

$$\bar{q} = q^{bond} + q^{syn}$$

Therefore,

$$
\begin{aligned}
f_5(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) &= \bar{q} + f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) - D^{syn}(r^{syn} - r) - D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) \\
&= \bar{q} + q^{bond} - q^{syn} \\
&= 2q^{bond} \\
&> 0
\end{aligned}
$$

Furthermore, the equilibrium conditions in the long equilibrium indicates

$$f_1 = f_2 = 0, \quad f_3 = q^{bond} > 0$$

In a short-regime equilibrium, $q^{bond} < 0$, so

$$\bar{q} = -q^{bond} + q^{syn}$$

Therefore,

$$\begin{aligned}
f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) &= \bar{q} - f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) - D^{syn}(r^{syn} - r) - D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) \\
&= \bar{q} - q^{bond} - D^{syn}(r^{syn} - r) - D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}) \\
&= \bar{q} - q^{bond} - q^{syn} \\
&= -2q^{bond} \\
&> 0
\end{aligned}$$

Furthermore, the equilibrium conditions in the short equilibrium indicates

$$f_4 = f_5 = 0, \quad f_3 = q^{bond} < 0$$

In an intermediate-regime equilibrium, $f_3 = q^{bond} = 0$, so

$$\bar{q} = q^{syn}$$

and

$$f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = \bar{q} - q^{bond} - q^{syn} = 0$$

$$f_5(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = \bar{q} + q^{bond} - q^{syn} = 0$$

Furthermore, the intermediate-regime equilibrium requires that the yield is between the long and short thresholds, so

$$f_1 \geq 0 \geq f_4$$

Note that f_1, f_3, and f_5 are decreasing in y and increasing in r^{syn}, whereas f_2 is increasing in both y and r^{syn} and f_4 is decreasing in both in both y and r^{syn}.

We next show that the existence of either a long or a short equilibrium rules out the existence of another kind of equilibrium. Since all equilibria involve $q^{bond} > 0, q^{bond} < 0$, or $q^{bond} = 0$, it follows that an intermediate-regime equilibrium cannot coexist with other equilibria as well (i.e., the uniqueness of the intermediate-regime equilibrium holds once we prove the other two). Thus, the equilibrium if exists must be unique.

D.4.1 Uniqueness of a Long Regime Equilibrium

Suppose there is a $(y_{long}, r^{syn}_{long})$ that is a long equilibrium. Equilibrium conditions imply

$$f_4(y_{long}, r^{syn}_{long}; \cdot) < 0 = f_1(y_{long}, r^{syn}_{long}; \cdot) = f_2(y_{long}, r^{syn}_{long}; \cdot)$$

$$f_3(y_{long}, r^{syn}_{long}; \cdot) > 0, \quad f_5(y_{long}, r^{syn}_{long}; \cdot) > 0$$

The goal is to show that there cannot be another equilibrium in the short or the intermediate regime.

1. Now suppose there is another equilibrium (y, r^{syn}) that is in the intermediate regime, which implies

$$f_2(y, r^{syn}; \cdot) = f_3(y, r^{syn}; \cdot) = f_5(y, r^{syn}; \cdot) = 0$$

$$f_1(y, r^{syn}; \cdot) \geq 0 \geq f_4(y, r^{syn}; \cdot)$$

If $r^{syn} > r^{syn}_{long}$, we must have $y > y_{long}$ by $f_3(y, r^{syn}; \cdot) < f_3(y_{long}, r^{syn}_{long}; \cdot)$, but in this case,

$$f_2(y, r^{syn}; \cdot) > f_2(y_{long}, r^{syn}_{long}; \cdot) = 0,$$

which results in a contradiction.

If $r^{syn} < r^{syn}_{long}$, we have have $y < y_{long}$ by $f_1(y, r^{syn}; \cdot) > f_1(y_{long}, r^{syn}_{long}; \cdot)$, but in this case

$$f_2(y, r^{syn}; \cdot) < f_2(y_{long}, r^{syn}_{long}; \cdot) = 0,$$

whcih reuslts in a contradiction.

If $r^{syn} = r^{syn}_{long}$, it is not possible to simultaneously increase f_1 and decrease f_3 by changing y, and therefore no intermediate equilibrium exists.

Consequently, there is no alternative equilibrium in the intermediate regime.

2. Now suppose there is another equilibrium (y, r^{syn}) that is in the short regime,which implies

$$f_1(y, r^{syn}; \cdot) > 0 = f_4(y, r^{syn}; \cdot) = f_5(y, r^{syn}; \cdot)$$

$$f_2(y, r^{syn}; \cdot) > 0, \quad f_3(y, r^{syn}; \cdot) < 0$$

If $r^{syn} > r^{syn}_{long}$, we must have $y > y_{long}$ by $f_3(y, r^{syn}) < f_3(y_{long}, r^{syn}_{long}; \cdot)$, but in this case $f_4(y, r^{syn}; \cdot) < f_4(y_{long}, r^{syn}_{long}; \cdot) < 0$, which leads to a contradiction.

If $r^{syn} < r^{syn}_{long}$, we have have $y < y_{long}$ by $f_1(y, r^{syn}; \cdot) > f_1(y_{long}, r^{syn}_{long}; \cdot)$, but in this case $f_2(y, r^{syn}; \cdot) < f_2(y_{long}, r^{syn}_{long}; \cdot) = 0$, which again leads to a contradiction.

If $r^{syn} = r^{syn}_{long}$, it is not possible to simultaneously increase f_1 and decrease f_3 by changing y, and therefore no short equilibrium exists.

Consequently, there is no alternative equilibrium in the long regime.

D.4.2 Uniqueness of a Short Regime Equilibrium

Suppose there is a $(y_{short}, r^{syn}_{short})$ that is a short equilibrium. Equilibrium conditions imply

$$f_1(y_{short}, r^{syn}_{short}; \cdot) > 0 = f_4(y_{short}, r^{syn}_{short}; \cdot) = f_5(y_{short}, r^{syn}_{short}; \cdot)$$

$$f_2(y_{short}, r^{syn}_{short}; \cdot) > 0, \quad f_3(y_{short}, r^{syn}_{short}; \cdot) < 0$$

1. Now suppose there is another equilibrium (y, r^{syn}) in the intermediate regime, which implies

$$f_2(y, r^{syn}; \cdot) = f_3(y, r^{syn}; \cdot) = f_5(y, r^{syn}; \cdot) = 0$$

$$f_1(y, r^{syn}; \cdot) \geq 0 \geq f_4(y, r^{syn}; \cdot)$$

If $r^{syn} > r^{syn}_{short}$, we must have $y > y_{short}$ by $f_5(y, r^{syn}; \cdot) = f_5(y_{short}, r^{syn}_{short}; \cdot)$, but in this case $f_2(y, r^{syn}; \cdot) > f_2(y_{short}, r^{syn}_{short}; \cdot) > 0$, which leads to a contradiction.

If $r^{syn} < r^{syn}_{short}$, we have have $y < y_{short}$ by $f_5(y, r^{syn}; \cdot) = f_5(y_{short}, r^{syn}_{short}; \cdot)$, but in this case $f_4(y, r^{syn}; \cdot) > f_4(y_{short}, r^{syn}_{short}; \cdot) = 0$, which leads to a contradiction.

If $r^{syn} = r^{syn}_{short}$, then by $f_5(y, r^{syn}; \cdot) = 0$ we must have $y = y_{short}$. However, then this leads to $f_3(y, r^{syn}; \cdot) = f_3(y_{short}, r^{syn}_{short}; \cdot) < 0$, which is a contradiction.

Consequently, there is no alternative equilibrium in the intermediate regime.

2. Now suppose there is another equilibrium (y, r^{syn}) in the long regime, which implies

$$f_4(y, r^{syn}; \cdot) < 0 = f_1(y, r^{syn}; \cdot) = f_2(y, r^{syn}; \cdot)$$

$$f_3(y, r^{syn}; \cdot) > 0, \quad f_5(y, r^{syn}; \cdot) > 0$$

If $r^{syn} > r^{syn}_{short}$, we have have $y > y_{short}$ by $f_1(y, r^{syn}; \cdot) < f_1(y_{short}, r^{syn}_{short}; \cdot)$, but in this case $f_2(y, r^{syn}; \cdot) > f_2(y_{short}, r^{syn}_{short}; \cdot) > 0$, which is a contradiction.

If $r^{syn} < r^{syn}_{short}$, we must have $y < y_{short}$ by $f_3(y, r^{syn}) > f_3(y_{short}, r^{syn}_{short}; \cdot)$, but in this case $f_4(y, r^{syn}; \cdot) > f_4(y_{short}, r^{syn}_{short}; \cdot) = 0$, which is a contradiction.

If $r^{syn} = r^{syn}_{short}$, it is not possible to simultaneously increase f_3 and decrease f_1 by changing y, and therefore no long equilibrium exists.

Consequently, there is no alternative equilibrium in the short regime.

D.4.3 Equilibrium Existence

Next, we prove the existence of the equilibrium. The high-level idea is to construct the equilibrium as a convex mapping from a compact and convex set to itself, and then apply the Kakutani fixed-point theorem.

First, we show the compactness of the relevant space of (y, r^{syn}).

Compactness of the y dimension.

In any equilibrium, we must have

$$f_3(y, r; S^{bond}, y_{\mathbb{Q}}) \leq f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) \leq \bar{q}.$$

Because f_3 is decreasing in y, there is a y_{min} such that

$$f_3(y_{min}, r; S^{bond}, y_{\mathbb{Q}}) > \bar{q},$$

and any equilibrium must have $y \geq y_{min}$. We must also have, in any equilibrium,

$$f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) \geq -\bar{q},$$

which yields, by $D_H \geq 0$,

$$e^{-ny} S^{bond} - D_U(ny - y^{bill} - (n-1)y_{\mathbb{P}}) \geq -\bar{q}.$$

Defining y_{max} by

$$e^{-ny_{max}} S^{bond} - D_U(ny_{max} - y^{bill} - (n-1)y_{\mathbb{P}}) = -\bar{q},$$

it follows that $y \leq y_{max}$.

Compactness of the r^{syn} dimension.

Define r^{min} as

$$D^{syn}(r^{min} - r) = \bar{q},$$

By assumption $D^{syn}(0) > \bar{q}$ and D^{syn} is a strictly decreasing function, we have $r^{min} - r > 0$. For any $r^{syn} < r^{min}$,

$$D^{syn}(r^{syn} - r) > \bar{q},$$

which violates the synthetic market clearing condition in (30). Consequently, in any equilibrium, $r^{syn} \geq r^{min}$.

Next, we will find an upper bound r^{max} such that for any $r^{syn} > r^{max}$, one of the market clearing conditions are violated. First, we note that there exists a r_1^{max} such that for all $r^{syn} > r_1^{max}$, for any feasible y we consider, i.e. $y \in [y_{min}, y_{max}]$,

$$e^{-ny}(e^{i^l} - e^r) + e^{r^{syn}} > \exp(-(n-1)y_{\mathbb{Q}}) > e^{-ny}(e^{i^s} + e^r - e^{r^{syn}}).$$

which says that $y \in (y^s, y^l)$ and thus the equilibrium is in the intermediate regime and dealer chooses $q^{bond} = 0$, and supply $\bar{q}$ to the synthetic lending market. We will show that if r^{syn} is too large, the synthetic lending market demand will fall below this supply.

Define synthetic lending demand as

$$m(y, r^{syn}) = D^{syn}(r^{syn} - r) + D_H(ny - r^{syn} - (n-1)y_{\mathbb{P}}).$$

which decreases in r^{syn}. There exists a $r^{max} \geq r_1^{max}$ such that, for all $r^{syn} > r^{max}$ and $y \in [y_{min}, y_{max}]$,

$$m(y, r^{syn}) < \bar{q},$$

which breaks the synthetic lending market clearing condition.

Consequently, if $r^{syn} \geq r^{max}$, no equilibrium can exist.

Convex and Closed Correspondence.

So far we have found a compact and convex space $\mathscr{C} = [y_{min}, y_{max}] \times [r^{min}, r^{max}]$ where the equilibrium (y, r^{syn}) must belong. Next, we define the correspondence for the equilibrium and prove that it is convex and closed.

The mapping we construct will constitute four dimensions, including (y, r^{syn}), the dealer bond position q^{bond}, and dealer synthetic lending q^{syn}.

From the dealer optimization problem, the demand correspondence only depends on (y, r^{syn}) and is defined as follows

$$Q(y, r^{syn}) = \begin{cases} \{(\bar{q}, 0)\} & \text{if } f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < 0 \\ \{(q^{bond}, q^{syn}) \in \mathbb{R}_+^2 : q^{bond} + q^{syn} = \bar{q}\} & \text{if } f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0, \\ \{(q^{bond}, q^{syn}) \in \mathbb{R}_- \times \mathbb{R}_+ : -q^{bond} + q^{syn} = \bar{q}\} & \text{if } f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0, \\ \{(-\bar{q}, 0)\} & \text{if } f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) > 0, \\ \{(0, \bar{q})\} & otherwise. \end{cases}$$

The first case $f_1 < 0$ is the only-long region where $y > y^l$. The second case $f_1 = 0$ is the long region where $y = y^l$. The third case $f_4 = 0$ is the sell region where $y = y^s$. The fourth case $f_4 > 0$ is the sell-only region where $y < y^s$. The fifth case is the intermediate region where $y^s < y < y^l$.

Define the aggregate excess demand correspondence as

$$Z(y, r^{syn}) = \{(z_1, z_2\} \in \mathbb{R}^2 : (z_1 + f_3(y, r^{syn}; \cdot), m(y, r^{syn}) - z_2) \in Q(y, r^{syn})\}.$$

Here, z_1 represents the excess demand for bonds, and z_2 is the excess demand for synthetic loans. By definition, $f_3(\cdot)$ is the bond supply less non-intermediary demand, and hence $f_3(\cdot) + z_1$ must equal the intermediary demand q^{bond}. Likewise, $m(\cdot)$ is synthetic loan demand, and $m(\cdot) - z_2$ must equal the synthetic loan supply q^{syn}.

Note that this correspondence is non-empty, u.h.c. (by the u.h.c. property of q, which ultimately arises from the continuity of f_1, f_4, and the continuity of f_3 and m). Note that it is also convex-valued, a property it inherits from Q. Define the maximum and minimum possible excess demands by

$$z_{max}^{bond} = \max_{(y, r^{syn}) \in [y_{min}, y_{max}] \times [r^{min}, r^{max}]} \bar{q} - f_3(y, r^{syn}; \cdot),$$

$$z_{min}^{bond} = \min_{(y, r^{syn}) \in [y_{min}, y_{max}] \times [r^{min}, r^{max}]} -\bar{q} - f_3(y, r^{syn}; \cdot),$$

$$z_{max}^{syn} = \max_{(y, r^{syn}) \in [y_{min}, y_{max}] \times [r^{min}, r^{max}]} m(y, r^{syn}),$$

$$z_{\min}^{syn} = \min_{(y, r^{syn}) \in [y_{min}, y_{max}] \times [r^{min}, r^{max}]} m(y, r^{syn}) - \bar{q},$$

Now define a price player, who solves, given any vector $(z_1, z_2) \in [z_{min}^{bond}, z_{max}^{bond}]$,

$$\max_{(y, r^{syn}) \in [y_{min}, y_{max}] \times [r^{min}, r^{max}]} (y, r^{syn}) \cdot \begin{bmatrix} -z_1 \\ z_2 \end{bmatrix}.$$

Let $p^*(z)$ be the optimal policy correspondence, and note that it is non-empty, u.h.c., and convex-valued (which follows from the concavity of the objective).

Now define the correspondence

$$g(y, r^{syn}, z) = \begin{bmatrix} p^*(z) \\ Z(y, r^{syn}) \end{bmatrix}$$

which maps $[y_{min}, y_{max}] \times [r^{min}, r^{max}] \times [z_{min}^{bond}, z_{max}^{bond}] \times [z_{min}^{syn}, z_{max}^{syn}]$ to itself. Note that this set is compact, and by the u.h.c. properties of p^* and Z and the compactness of this set, g has a closed graph. Consequently, by Kakutani's fixed point theorem, a fixed point (y^*, r^{syn*}, z^*) exists.

By construction, at y_{min},

$$f_3(y_{min}, r; S^{bond}, y_{\mathbb{Q}}) > \bar{q},$$

and consequently all values $Z_1(y, r^{syn})$ are negative. The best response of the price player at this point would be y_{max}, and hence there cannot be a fixed point with $y^* = y_{min}$. Essentially the same logic rules out $y^* = y_{max}$. Similarly, if $r^{syn*} = r^{min}$, then all values of $Z_2(y^*, r^{syn*})$ are positive, and the price player's best response is r_{max}, and hence this cannot be a fixed point. Likewise, if $r^{syn*} = r^{max}$, then all values of $Z_2(y^*, r^{syn*})$ are negative, and the price player's best response is $r^{syn} = r^{min}$. It follows that the fixed point is interior, and hence that $z^* = Z(y^*, r^{syn*}) = (0,0)$. Note that a fixed point with $z^* = (0,0)$ cannot exist in which there is no supply of synthetic lending; consequently, the equilibrium is either a long regime equilibrium,

$$f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0,$$

a short regime equilibrium,

$$0 = f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}),$$

or an intermediate equilibrium,

$$f_4(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < 0 < f_1(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}).$$

D.4.4 Bond Supply and Equilibrium Regime

To prove that the existence of cutoffs S_S and S_B with $0 \leq S_S \leq S_B \leq \infty$, such that the short-regime, the intermediate regime, and the long-regime fall into the three regions, we simply prove that there is a ranking of the equilibrium along the supply of bonds S.

Consider $S^{bond} = S$. According to the previous proofs, an equilibrium (y, r^{syn}) exists and must be unique.

Long Equilibrium

First, we show that if S corresponds to a long-regime equilibrium, then for any $\bar{S} > S$, the equilibrium $(\bar{y}, \bar{r}^{syn})$ must also be a long equilibrium.

Suppose instead the equilibrium for $S^{bond} = \bar{S}$ is a short-regime equilibrium with $(\bar{y}, \bar{r}^{syn})$. Then we must have $y = y^l > y^s = \bar{y}$. Furthermore, $f_2(y, r^{syn}; S, y_{\mathbb{Q}}) = 0$ and $f_2(\bar{y}, \bar{r}^{syn}; S, y_{\mathbb{Q}}) > f_2(\bar{y}, \bar{r}^{syn}; \bar{S}, y_{\mathbb{Q}}) > 0$. By monotonicity of f_2, we must have $\bar{r}^{syn} > r^{syn}$. Therefore, by monotonicty of f_5, we have

$$f_5(\bar{y}, \bar{r}^{syn}; \bar{S}, y_{\mathbb{Q}}) > f_5(y, r^{syn}; \bar{S}, y_{\mathbb{Q}})$$

However, in the long regime, $f_5(\bar{y}, \bar{r}^{syn}; \bar{S}, y_{\mathbb{Q}}) = 0$, and in the short regime, $f_5(y, r^{syn}; \bar{S}, y_{\mathbb{Q}}) >$

$f_5(y, r^{syn}; S, y_{\mathbb{Q}}) > 0$, which leads to a contradiction. Thus, $\bar{S}$ cannot correspond to a short equilibrium.

Next, suppose that $\bar{S}$ corresponds to an intermediate-regime equilibrium. Then we have $y = y^l \geq \bar{y}$, $f_3(\bar{y}, \bar{r}^{syn}; \bar{S}, y_{\mathbb{Q}}) = 0$. For the equilibrium of S, we have $f_3(y, r^{syn}; \bar{S}, y_{\mathbb{Q}}) > f_3(y, r^{syn}; S, y_{\mathbb{Q}}) > 0$. By the monotonicity of f_3, $r^{syn} > \bar{r}^{syn}$. Thus, $f_2(\bar{y}, \bar{r}^{syn}; S, y_{\mathbb{Q}}) < f_2(y, r^{syn}; S, y_{\mathbb{Q}})$.

However, by the properties of long and intermediate regimes, we also have $f_2(y, r^{syn}; S, y_{\mathbb{Q}}) > 0$ and $f_2(\bar{y}, \bar{r}^{syn}; S, y_{\mathbb{Q}}) > f_2(\bar{y}, \bar{r}^{syn}; \bar{S}, y_{\mathbb{Q}}) = 0$, which leads to a contradiction.

In summary, if S is a long-regime equilibrium, for any $\bar{S} > S$, the equilibrium $(\bar{y}, \bar{r}^{syn})$ must also be a long-regime equilibrium.

Short Equilibrium

Second, we show that if S corresponds to a short-regime equilibrium, then for any $\underline{S} < S$, the equilibrium solution $(\underline{y}, \underline{r}^{syn})$ must also be a short-regime equilibrium.

Suppose that instead the equilibrium for $S^{bond} = \underline{S}$ is a long-regime equilibrium. Then we must have $y = y^s < y^l = \underline{y}$. Furthermore, $f_5(\underline{y}, \underline{r}^{syn}; S, y_{\mathbb{Q}}) > f_5(\underline{y}, \underline{r}^{syn}; \underline{S}, y_{\mathbb{Q}}) > 0$, and $f_5(y, r^{syn}; S, y_{\mathbb{Q}}) = 0$. By monotonicity of f_5, we get $\underline{r}^{syn} > r^{syn}$. Thus, by monotonicity of f_2, we obtain

$$f_2(y, r^{syn}; S, y_{\mathbb{Q}}) < f_2(\underline{y}, \underline{r}^{syn}; S, y_{\mathbb{Q}})$$

However, by the properties of long and short regimes, we must have $f_2(\underline{y}, \underline{r}^{syn}; S, y_{\mathbb{Q}}) < f_2(\underline{y}, \underline{r}^{syn}; \underline{S}, y_{\mathbb{Q}}) = 0$, and $f_2(y, r^{syn}; S, y_{\mathbb{Q}}) > 0$, which leads to a contradiction.

Suppose that the equilibrium for $S^{bond} = \underline{S}$ is an intermediate-regime equilibrium. Then we must have $y = y^s \leq \underline{y}$. Furthermore, $f_5(\underline{y}, \underline{r}^{syn}; S, y_{\mathbb{Q}}) > f_5(\underline{y}, \underline{r}^{syn}; \underline{S}, y_{\mathbb{Q}}) = 0$, and $f_5(y, r^{syn}; S, y_{\mathbb{Q}}) = 0$. By monotonicity of f_5, we get $\underline{r}^{syn} > r^{syn}$. Thus, by monotonicity of f_2, we obtain

$$f_2(y, r^{syn}; S, y_{\mathbb{Q}}) < f_2(\underline{y}, \underline{r}^{syn}; S, y_{\mathbb{Q}})$$

However, by the properties of short and intermediate regimes, we must have $f_2(y, r^{syn}; S, y_{\mathbb{Q}}) > 0$ and $f_2(\underline{y}, \underline{r}^{syn}; S, y_{\mathbb{Q}}) < f_2(\underline{y}, \underline{r}^{syn}; \underline{S}, y_{\mathbb{Q}}) = 0$, which leads to a contradiction.

In summary, if S is a short-regime equilibrium, for any $\underline{S} < S$, the equilibrium $(\underline{y}, \underline{r}^{syn})$ must also be a short-regime equilibrium.

Regime Ranking

From the above discussions, we know that there must be cutoffs S_S and S_B with $0 \le S_S \le S_B \le \infty$, such that a short-regime equilibrium exists in the left region, an intermediate-regime equilibrium exists in the middle region, and a long-regime equilibrium exists in the right region. However, we still have to prove whether the intervals are open or closed.

Intervals for Regimes

We now show that the interval of long-regime equilibrium should be (S_B, ∞) instead of $[S_B, \infty)$. Suppose that $S^{bond} = S$ is a long-regime equilibrium, with solutions (y, r^{syn}), and q^{bond}. By definition, $q^{bond} > 0$.

In the long regime, $f_3(y, r^{syn}; S, y_{\mathbb{Q}}) = q^{bond}$. We know that $q^{bond} > 0$ increases in the total supply of bond and the mapping is continuous. Therefore, there exists a smaller bond supply $S^{bond} = S - \varepsilon$ for $\varepsilon > 0$, such that the new equilibrium still has $q^{bond} > 0$. Consequently, the interval of S^{bond} for the long-regime equilibrium must be an open set.

Similarly, the interval for the short-regime equilibrium must also be an open set, $(-\infty, S_B)$.

D.4.5 Term Premium and Equilibrium Regime

Next, we study how the term premium $y_{\mathbb{Q}}$ affects the equilibrium. Consider $y_{\mathbb{Q}}$. According to previous proofs, an equilibrium solution (y, r^{syn}) exists and is unique.

Long Equilibrium

First, we show that if $y_{\mathbb{Q}}$ corresponds to a long-regime equilibrium, then for any $\bar{y}_{\mathbb{Q}} > y_{\mathbb{Q}}$, the equilibrium $(\bar{y}, \bar{r}^{syn})$ must also be a long equilibrium.

Suppose instead the equilibrium for $\bar{y}_{\mathbb{Q}}$ is a short-regime equilibrium with $(\bar{y}, \bar{r}^{syn})$. Then we must have $y = y^l > y^s = \bar{y}$. Furthermore, $f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0$ and $f_2(\bar{y}, \bar{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_2(\bar{y}, \bar{r}^{syn}; S^{bond}, \bar{y}_{\mathbb{Q}}) > 0$. By monotonicity of f_2, we must have $\bar{r}^{syn} > r^{syn}$. Therefore, by monotonicty of f_5, we have

$$f_5(\bar{y}, \bar{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) > f_5(y, r^{syn}; S^{bond}, y_{\mathbb{Q}})$$

However, in the short regime, $f_5(\bar{y}, \bar{r}^{syn}; S^{bond}, \bar{y}_{\mathbb{Q}}) = f_5(\bar{y}, \bar{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0$, and in the long regime, $f_5(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) > 0$, which leads to a contradiction. Thus, $\bar{y}_{\mathbb{Q}}$ cannot correspond to a short equilibrium.

Next, suppose that $\bar{y}_{\mathbb{Q}}$ corresponds to an intermediate-regime equilibrium. Then we have $y = y^l \ge \bar{y}$, $f_3(\bar{y}, \bar{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_3(\bar{y}, \bar{r}^{syn}; S^{bond}, \bar{y}_{\mathbb{Q}}) = 0$. For the long-regime equilibrium of $y_{\mathbb{Q}}$, we have $f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) > 0$. By the monotonicity of f_3, $r^{syn} > \bar{r}^{syn}$. Thus, $f_2(\bar{y}, \bar{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) < f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}})$.

However, by the properties of long and intermediate regimes, we also have $f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0$ and $f_2(\bar{y}, \bar{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_2(\bar{y}, \bar{r}^{syn}; S^{bond}, \bar{y}_{\mathbb{Q}}) = 0$, which leads to a contradiction.

In summary, if $y_{\mathbb{Q}}$ is a long-regime equilibrium, for any $\bar{y}_{\mathbb{Q}} > y_{\mathbb{Q}}$, the equilibrium $(\bar{y}, \bar{r}^{syn})$ must also be a long equilibrium.

Short Equilibrium

Second, we show that if $y_{\mathbb{Q}}$ corresponds to a short-regime equilibrium, then for any $\underline{y}_{\mathbb{Q}} < y_{\mathbb{Q}}$, the equilibrium $(\underline{y}, \underline{r}^{syn})$ must also be a short-regime equilibrium.

Suppose that instead the equilibrium for $\underline{y}_{\mathbb{Q}}$ is a long-regime equilibrium. Then we must have $y = y^s < y^l = \underline{y}$. Furthermore, $f_5(\underline{y}, \underline{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_5(\underline{y}, \underline{r}^{syn}; S^{bond}, \underline{y}_{\mathbb{Q}}) > 0$, and $f_5(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) = 0$. By monotonicity of f_5, we get $\underline{r}^{syn} > r^{syn}$. Thus, by monotonicity of f_2, we obtain

$$f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < f_2(\underline{y}, \underline{r}^{syn}; S^{bond}, y_{\mathbb{Q}})$$

However, by the properties of long and short regimes, we must have

$$f_2(\underline{y}, \underline{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_2(\underline{y}, \underline{r}^{syn}; S^{bond}, \underline{y}_{\mathbb{Q}}) = 0,$$

and $f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) > 0$, which leads to a contradiction.

Suppose that the equilibrium for $\underline{y}_{\mathbb{Q}}$ is an intermediate-regime equilibrium. Then we must have $y = y^s \leq \underline{y}$. Furthermore, $f_3(\underline{y}, \underline{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_3(\underline{y}, \underline{r}^{syn}; S^{bond}, \underline{y}_{\mathbb{Q}}) = 0$, and $f_3(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < 0$. By monotonicity of f_3, we get $\underline{r}^{syn} > r^{syn}$. Thus, by monotonicity of f_2, we obtain

$$f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) < f_2(\underline{y}, \underline{r}^{syn}; S^{bond}, y_{\mathbb{Q}})$$

However, by the properties of short and intermediate regimes, we must have $f_2(y, r^{syn}; S^{bond}, y_{\mathbb{Q}}) > 0$ and $f_2(\underline{y}, \underline{r}^{syn}; S^{bond}, y_{\mathbb{Q}}) = f_2(\underline{y}, \underline{r}^{syn}; S^{bond}, \underline{y}_{\mathbb{Q}}) = 0$, which leads to a contradiction.

In summary, if $y_{\mathbb{Q}}$ is a short-regime equilibrium, for any $\underline{y}_{\mathbb{Q}} < y_{\mathbb{Q}}$, the equilibrium $(\underline{y}, \underline{r}^{syn})$ must also be a short-regime equilibrium.

Regime Ranking

From the above discussions, we know that there must be cutoffs y_S and y_B with $0 \leq y_S \leq y_B \leq$

∞, such that a short-regime equilibrium, an intermediate-regime equilibrium, and a long-regime equilibrium exits in the left, middle and right regions. However, we still need to determine whether those intervals are open or closed sets.

Intervals for Regimes

We now show that the interval of long-regime equilibrium should be (y_B, ∞) instead of $[y_B, \infty)$. Suppose that $y_{\mathbb{Q}}$ is a long-regime equilibrium, with solutions (y, r^{syn}), and q^{bond}. By definition, $q^{bond} > 0$.

In the long regime, we know that $q^{bond} > 0$ is a continuous function of $y_{\mathbb{Q}}$. Therefore, there exists a smaller risk-neutral expectation $y_{\mathbb{Q}} - \varepsilon$, where the new equilibrium is still in the long regime with $q^{bond} > 0$. Consequently, the interval of $y_{\mathbb{Q}}$ for the long-regime equilibrium must be an open set.

Similarly, the interval of $y_{\mathbb{Q}}$ for the short-regime equilibrium must also be an open set.

E Additional Derivations

E.1 Dealers and Levered Clients

In Section 2, we developed net long and net short curves from the perspective of a securities dealer, yields at which the dealer would be willing to either net long or net short Treasury bonds. In this section, we extend our model to consider the perspective of a levered Treasury investor (e.g. a hedge fund) financed by a security dealer of the kind considered in that section. The main result is that levered clients will have the same net long and net short curves as the dealer that finances them. That is, the net long curve represents a yield at which the levered client would be willing to buy the Treasury bond, irrespective of its beliefs about the stochastic process driving Treasury yields, and a symmetric result holds for the net short curve. This result occurs in spite of the fact that the levered client is not itself directly affected by balance sheet constraints.

This result is important from a general equilibrium perspective. Dealers are never on net long or short a large quantity of Treasury bonds during our sample, relative to the overall Treasury supply. Dealers moved from a net short of roughly 100 hundred billion in 2005 to a net long of 200 hundred billion in 2020. The overall supply of Treasury securities rose from 4 trillion to 22 trillion over the same period.

However, dealers intermediate repo and reverse-repo for their levered clients in much greater quantities– on the order of trillions each day. In this section we will argue that the recipients of much of this financing will act like dealers, and subsequently provide some suggestive evidence on this point.

Consider the following trading strategy for the dealer: finance a client's entire purchase of a Treasury using bilateral repo, use the resulting collateral to raise financing, and reduce CIP activity so that the trade is balance sheet neutral. In a competitive market, the profits of such a strategy are zero:

$$\underbrace{\left(e^{\frac{1}{12}r_t^{bi}} - e^{\frac{1}{12}i_t^{l}}\right)}_{\text{Lending/Borrowing Spread}} - \underbrace{\left(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}r_t}\right)}_{\text{Forgone CIP profits}} = 0. \tag{E-1}$$

That is, the dealer must be indifferent between matched book repo lending and taking advantage of CIP arbitrage, as both activities use balance sheet.

Let's now consider the perspective of a levered client who can purchase a Treasury bond, financed by this intermediary, and can trade derivatives with the securities dealer. Because the levered client can trade derivatives with the dealer, the projection of its stochastic discount factor onto the space of derivative returns must agree with the same projection for the dealer's SDF. Equivalently, the risk-neutral measure Q is shared (within this space) by the levered clients and the dealer.

We will also assume that the levered client can engage in risk-free unsecured borrowing[1] from the unsecured dealer at the synthetic lending rate. The dealer is unwilling to lend at a rate lower than this, as otherwise it would be better off engaging in CIP arbitrage.

Under these assumptions, the levered client considers buying an n-month Treasury and then selling one month later:

$$\underbrace{e^{-ny_{n,t}} e^{r_t^{bi}}}_{\text{secured financing}} + \cdot \underbrace{e^{-ny_{n,t}} e^{r_t^{syn}}}_{\text{unsecured financing}} \geq E_t^{\mathbb{Q}}[e^{-(n-1)y_{n-1,t+1}}]. \tag{E-2}$$

Substituting in (E-1), this condition becomes identical to (7). It follows immediately that levered clients must be willing to go net long if the yield reaches the net long curve.

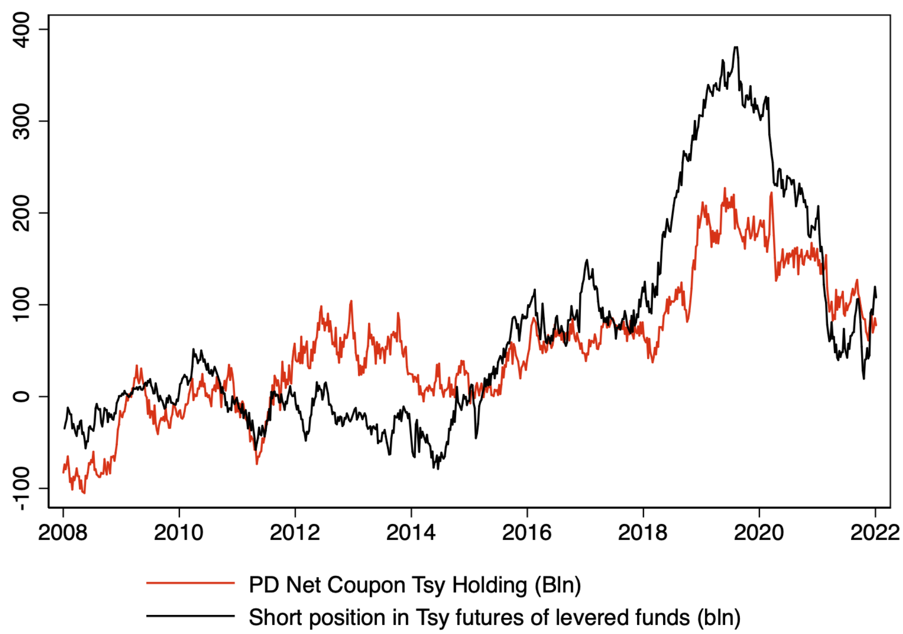
Essentially identical logic applies to the net short curve: the dealers indifference between

[1] It is probably better to think of this as secured borrowing using non-Treasury securities that the dealer cannot itself finance in a repo market.

matched book repo (in the net short case, intermediating between security lenders and short-sellers) and CIP arbitrage converts the levered client's indifference condition to the dealer's indifference condition.

We conclude that levered clients who are dependent on dealers for financing will act as if they face the same balance sheet costs that dealers face, even if they are not themselves directly regulated. As a result, balance sheets costs will influence a substantial segment of the Treasury market, even though dealers are on their own hold a relatively small quantity of Treasury bonds on net.

Figure A4: Primary Dealer Treasury Holdings and Implied Treasury Holding of Levered Investors



Notes: This figure plots the primary dealer's net position in coupon-bearing Treasury securities from Primary Dealer Statistics published by the Federal Reserve Bank of New York, and the short position in the Treasury futures market by levered funds from the Commitments of Traders Report published by the Commodity Futures Trading Commission.

In Figure A4, we provide evidence consistent with this perspective. While the Treasury positions of levered investors are not publicly available, we can infer the holdings of investors that engage in Treasury cash-future trades from Treasury futures positions. Figure A4 plots the primary dealer net coupon holdings and levered funds' short positions in Treasury futures contracts published by the CTFC. For relative value hedge funds that arbitrage Treasury cash-futures basis, a short position in Treasury futures corresponds to a long position in the cash Treasury bonds. We see that primary dealer positions and the levered funds' short Treasury futures position are strongly positively correlated, which is consistent with our result that dealers and levered investors take similar positions and can be considered as a consolidated intermediary.

E.2 Partial Equilibrium Arbitrage Bounds

In this appendix section, we construct the net short and net long curves described in the main text as arbitrage bounds under weaker assumptions than those employed in the main text. In particular, in the main text we assumed that zero-cost, zero-balance sheet trades are weakly unattractive under a common SDF (i.e., a version of the no-arbitrage assumption). That assumption leads to $y_{n-1,t+1} \leq y^b_{n-1,t+1}$ with probability one. Here, we instead assume that there could be profitable zero-cost, zero-balance sheet trading strategies under the intermediary's stochastic discount factor. Then we consider the question of whether this intermediary is willing to go net long or net short a Treasury bond, irrespective of the intermediary's preferences or beliefs about the stochastic process driving Treasury yields.

We will assume that this intermediary's SDF prices derivatives, and that the intermediary believes with probability one that $x_{1,t} \geq r_t \geq x_{2,t}$, where $x_{1,t}$ and $x_{2,t}$ are defined as in the main text. We discuss the role of this assumption below.

The Net Long Curve

Consider first the trade in which the intermediary buys a zero-coupon seven-month Treasury bond, and then sells it in one month, at which time the Treasury bond becomes a zero-coupon T-bill. The intermediary can finance this purchase with tri-party repo, up to the standard two percent haircut h, and finance the remainder with unsecured debt. This trade, in combination with a reduction in CIP activity, is a balance-sheet neutral, zero-financing trade. The intermediary is therefore willing to get net long if this strategy is weakly appealing under the SDF that prices derivatives. Let $\mathbb{Q}$ denote the risk-neutral measure associated with this SDF. We assume that r_t is the log risk-free rate associated with this SDF.[2]

Let $y^b_{7,t}$ denote a yield at which this trade is attractive to the dealer, and define $y^b_{6,t} = y^{bill}_t$. The dealer will be indifferent between employing and not employing this trading strategy if

$$\underbrace{e^{-\frac{7}{12}y^b_{7,t}}}_{\text{Purchase price}}\Big(\underbrace{e^{\frac{1}{12}i^l_t}}_{\text{Repo Financing}}+\underbrace{(e^{\frac{1}{12}r^{syn}_t}-e^{\frac{1}{12}r_t})}_{\text{Forgone CIP profits}}\Big)=\underbrace{E^{\mathbb{Q}}_t[e^{-\frac{6}{12}y^{bill}_{t+1}}]}_{\text{Sale price}}.$$

Cheap financing ($i^l_t < r_t$) makes the trade attractive and hence decreases the required yield, while the opportunity cost of using balance sheet ($r^{syn}_t > r_t$) has the opposite effect. We assume that $x_{1,t} \geq r_t$, which is consistent with the post-GFC data and implies that

$$e^{\frac{1}{12}r^{syn}_{t+j}} - e^{\frac{1}{12}r_{t+j}} \geq e^{\frac{1}{12}r_{t+j}} - e^{\frac{1}{12}i^l_{t+j}}.$$

This assumption states that the balance sheet cost exceeds the financing advantage. It can be justified on the grounds that, if it did not hold, it would be efficient for dealers to purchase Treasury bills from money market funds, financed by repo loans from those same money market funds. This would lead to large dealer balance sheets, causing the leverage constraint to tighten, and hence cannot be part of an equilibrium.

Let us now define a yield curve , $y^l_{n,t}$, such that the dealer will be certainly be willing to purchase an n-month Treasury bond, regardless of her preferences or beliefs, if its yield exceeds this value. This will be the net long curve. We will conjecture and verify that the curve defined recursively by

$$e^{-\frac{n}{12}y^l_{n,t}}\left(e^{\frac{1}{12}i^l_t} - e^{\frac{1}{12}r_t} + e^{\frac{1}{12}r^{syn}_t}\right) = E^{\mathbb{Q}}_t\left[e^{-\frac{n-1}{12}y^l_{n-1,t+1}}\right]$$

has this property. That is, the net long curve is defined by the discount rate $e^{x_{1,t}} = e^{\frac{1}{12}i^l_t} - e^{\frac{1}{12}r_t} + e^{\frac{1}{12}r^{syn}_t}$, as in the main text.

Fix some $n > 7$ and suppose $y^l_{m,t}$ is defined by this recursion for all $m \in \{6,\ldots,n-1\}$. Consider a trading strategy that purchases the bond, finances the trade with repo and unsecured borrowing,

[2] That is, we assume the one-month OIS swap rate is the intermediary's unsecured borrowing rate. This assumption is consistent with the empirical observation that the one-month OIS rate closely tracks other unsecured rates, for example the one-month highly rated financial commercial paper rate. It is also consistent with the industry practice of using the OIS curve to discount derivative cashflows. Lastly, it is consistent with the observation that the unsecured borrowing rate is the appropriate discount rate for off-balance-sheet cashflows, under our generalized no-arbitrage assumption.

offsets the balance sheet cost by reducing CIP activity, and unwinds at the first moment at which the bond yield becomes weakly lower than $y^l_{m,t}$. Let τ denote the months elapsed and let $y_{n-\tau,t+\tau} \leq y^l_{n-\tau,t+\tau}$ be the bond price at which the trade is unwound. According to the strategy, we have $y_{m,t} \geq y^l_{m,t}$ for all $m \in \{6,\dots,n-1\}$. Further, $\tau \leq n-6$ is guaranteed because by assumption, the intermediary always unwinds the trade once the bond has six-month remaining maturity.

The intermediary will be willing to engage in this strategy provided that

$$\underbrace{e^{-\frac{n}{12}y_{n,t}}}_{\text{purchase price}} + E^{\mathbb{Q}}_t[\sum_{j=0}^{\tau-1} \underbrace{e^{-\sum_{k=0}^{j} r_{t+k}}}_{\text{discount rate}} \underbrace{e^{-\frac{n-j}{12}y_{n-j,t+j}}}_{\text{interim bond price}} (\underbrace{(e^{\frac{1}{12}i^l_{t+j}} - e^{\frac{1}{12}r_{t+j}})}_{\text{repo financing benefits}} + \underbrace{(e^{\frac{1}{12}r^{syn}_{t+j}} - e^{\frac{1}{12}r_{t+j}})}_{\text{forgone CIP profits}})] \leq E^{\mathbb{Q}}_t[\underbrace{e^{-\sum_{k=0}^{\tau-1} r_{t+k}}}_{\text{discount rate}} \underbrace{e^{-\frac{n-\tau}{12}y_{n-\tau,t+\tau}}}_{\text{sale price}}]$$

Since derivatives are priced by the intermediary, hedging does not affect the economic profit in the above trade. We could add a hedging component to this equation, so that certain future fluctuations in the financing rate are fixed at the beginning of the trade. We omit this extra zero-cost component for simplicity.

However, this strategy cannot be fully hedged by interest rates swaps. First, the time τ at which the bond yield falls below $y^l_{m,t}$ is uncertain, as is the ultimate sale price. Second, the interim price of the bond before τ affects the size of the trade that needs to be financed, and consequently both the benefit of cheap financing via tri-party repo and the opportunity cost of the balance sheet. The effects of intermediate bond prices occur because the intermediary uses short term, as opposed to term, financing, and because the assets are marked to market. Thus, even if it were possible to perfectly hedge all of the relevant interest rates, the attractiveness of this trade would depend in part on the intermediary's beliefs about the stochastic process driving bond yields.

However, the worse case scenario for the sale price is that it is exactly equal to the unwinding threshold, $y_{n-\tau,t+\tau} = y^l_{n-\tau,t+\tau}$. Under the assumption that $x_{1,t} \geq r_t$, the worse case scenario for the intermediate bond yields is that they are as low as possible (i.e. $y_{n-j,t+j} = y^l_{n-j,t+j}$), which is to say that the trading strategy uses up the maximum possible balance sheet capacity. Consequently, the intermediary will definitely be willing to buy the bond if, for all possible stopping times τ ,

$$
\begin{aligned}
e^{-\frac{n}{12}y_{n,t}} \leq & -E_t^{\mathbb{Q}}[\sum_{j=0}^{\tau-1} e^{-\frac{n-j}{12}y^l_{n-j,t+j}} e^{-\sum_{k=0}^{j} r_{t+k}}\left(\left(e^{\frac{1}{12}i^l_{t+j}} - e^{\frac{1}{12}r_{t+j}}\right) + \left(e^{\frac{1}{12}r^{syn}_{t+j}} - e^{\frac{1}{12}r_{t+j}}\right)\right)] \\
& + E_t^{\mathbb{Q}}[e^{-\sum_{k=0}^{\tau-1} r_{t+k}} e^{-\frac{n-\tau}{12}y^l_{n-\tau,t+\tau}}]
\end{aligned}
$$

and this is in fact the tightest possible bound. Rewriting the definition of net long curve, we obtain

$$
\begin{aligned}
e^{-\sum_{k=0}^{j-1} r_{t+k}} e^{-\frac{n-j}{12}y^l_{n-j,t+j}} = & - e^{-\sum_{k=0}^{j} r_{t+k}} e^{-\frac{n}{12}y^l_{n-j,t+j}}\left(e^{\frac{1}{12}i^l_{t+j}} - e^{\frac{1}{12}r_{t+j}}\right) \\
& - e^{-\sum_{k=0}^{j} r_{t+k}} e^{-\frac{n}{12}y^l_{n-j,t+j}}\left(e^{\frac{1}{12}r^{syn}_{t+j}} - e^{\frac{1}{12}r_{t+j}}\right) \\
& + e^{-\sum_{k=0}^{j} r_{t+k}} E_{t+j}^{\mathbb{Q}}[e^{-\frac{n-j-1}{12}y^l_{n-j-1,t+j+1}}]
\end{aligned}
$$

for any j, and thus it also holds for any bounded stopping time τ. By the definition of the net long curve, this inequality is equivalent to

$$
e^{-\frac{n}{12}y_{n,t}} \leq e^{-\frac{n}{12}y^l_{n,t}}
$$

Thus, the intermediary will be willing to buy the bond, regardless of the nature of the intermediary's preferences and beliefs about the bond price process, if $y_{n,t} \geq y^l_{n,t}$.

We conclude that the intermediary's demand for a zero-coupon bond should be high if its yield exceeds the net long curve yield. This demand is limited only by the intermediary's leverage constraint: at some point, the intermediary will have switched entirely to doing the Treasury arbitrage as opposed to other arbitrages, at which point $r^{syn}_t - r_t$ is no longer a valid measure of the opportunity cost of balance sheet. We therefore predict that if a bond's yield exceeds the buy yield, the intermediary's demand should be substantial.

The Net Short Curve

We next develop parallel logic for the case of short-selling. In this case, we assume that the intermediary borrows the security from a securities lender in exchange for cash equal to the market value of the security, and receives a log interest rate $i^s_t < r_t$ on the cash lent.

The intermediary will be willing to short a seven-month bond at yield $y_{7,t}$ if

$$\underbrace{e^{-\frac{7}{12}y_{7,t}}}_{\text{Sale price}} \underbrace{e^{\frac{1}{12}i_t^s}}_{\text{Gross return on cash in sec. lending}} \geq \underbrace{E_t^{\mathbb{Q}}[e^{-\frac{6}{12}y_{t+1}^{bill}}]}_{\text{Repurchase price}} + e^{-\frac{7}{12}y_{7,t}} \underbrace{(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}r_t})}_{\text{Forgone CIP profits}}. \tag{E-3}$$

Note that the sign of the forgone CIP profits has changed, relative to the analogous equation for the net long curve, reflecting the fact that both buying and short-selling increase the size of the balance sheet. In equation (E-3), moving the right-hand-side OIS term to the left and dividing both sides by $\exp(\frac{1}{12}r_t)$, we obtain

$$e^{-\frac{7}{12}y_{7,t}} \geq e^{-\frac{7}{12}y_{7,t}} e^{-\frac{1}{12}r_t}\left(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}i_t^s}\right) + e^{-\frac{1}{12}r_t} E_t^{\mathbb{Q}}[e^{-\frac{6}{12}y_{t+1}^{bill}}] \tag{E-4}$$

Under the assumption that yields are weakly positive, $y_{7,t} \geq 0$, the intermediately is definitely willing to short if

$$e^{-\frac{7}{12}y_{7,t}} \geq e^{-\frac{7}{12}y_{7,t}^s} \equiv e^{-\frac{1}{12}r_t}\left(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}i_t^s}\right) + e^{-\frac{1}{12}r_t} E_t^{\mathbb{Q}}[e^{-\frac{6}{12}y_{t+1}^{bill}}]. \tag{E-5}$$

Following the same spirit, let us define $y_{n,t}^s$ recursively for $n \geq 8$ as

$$e^{-\frac{n}{12}y_{n,t}^s} = e^{-\frac{1}{12}r_t}\left(e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}i_t^s} + E_t^{\mathbb{Q}}[e^{-\frac{n-1}{12}y_{n-1,t+1}^s}]\right), \tag{E-6}$$

which can be interpreted as the pricing equation for a bond with a monthly coupon of $e^{\frac{1}{12}r_t^{syn}} - e^{\frac{1}{12}i_t^s}$, discounted using the OIS curve.

As above, fix some $n > 7$ and suppose $y_{m,t}^s$ is defined as above. Consider a trading strategy that short-sells the bond, borrows the bond from a securities lender, offsets the balance sheet cost by reducing CIP activity, and unwinds at the first moment at which the bond yield becomes weakly higher than $y_{m,t}^s$. Let τ denote this time and let $y_{n-\tau,t+\tau} \geq y_{n-\tau,t+\tau}^s$ be the bond price at which the trade is unwound. According to the strategy, we have $y_{m,t} \leq y_{m,t}^s$ for all $m \in \{6, 7, \cdots, n-1\}$. Further, $\tau \leq n-6$ is guaranteed by the assumption that dealers always unwinds the trade once the bond has six-month remaining maturity. The intermediary will be willing to engage in this strategy provided it is profitable,

$$e^{-\frac{n}{12}y_{n,t}} \geq E_t^{\mathbb{Q}}[\sum_{j=0}^{\tau-1} e^{-\frac{n-j}{12}y_{n-j,t+j}} e^{-\sum_{k=0}^{j} r_{t+k}} (e^{\frac{1}{12}r_{t+j}^{syn}} - e^{\frac{1}{12}i_{t+j}^{s}})] + E_t^{\mathbb{Q}}[e^{-\sum_{k=0}^{\tau-1} r_{t+k}} e^{-\frac{n-\tau}{12}y_{n-\tau,t+\tau}}]. \quad \text{(E-7)}$$

Note that, because $i_t^s < r_t^{syn}$, the worst-case scenario is the one that makes intermediate bond prices as high as possible. Unlike the net long curve, the fact that $y_{n-j,t+j} < y_{n-j,t+j}^s$ is of no help is generating a bound. In this case, we instead assume a lower bound on yields, $y_{m,t} \geq 0$, motivated the possibility of substitution to cash. In the worst-case scenario, the pricing condition becomes

$$e^{-\frac{n}{12}y_{n,t}} \geq E_t^{\mathbb{Q}}[\sum_{j=0}^{\tau-1} e^{-\sum_{k=0}^{j} r_{t+k}} (e^{\frac{1}{12}r_{t+j}^{syn}} - e^{\frac{1}{12}i_{t+j}^{s}})] + E_t^{\mathbb{Q}}[e^{-\sum_{k=0}^{\tau-1} r_{t+k}} e^{-\frac{n-\tau}{12}y_{n-\tau,t+\tau}^s}]. \quad \text{(E-8)}$$

For all stopping times τ (bounded above by $n-6$), this is equivalent to

$$e^{-\frac{n}{12}y_{n,t}} \geq e^{-\frac{n}{12}y_{n,t}^s}, \quad \text{(E-9)}$$

which is to say that the intermediary will be willing to short-sell if yields are below $y_{n,t}^s$, irrespective of intermediary's preferences or beliefs about future bond prices.[3]

Finally, we will illustrate that to a first-order approximation, the net-short curve in this appendix is the same as the net-short curve (23) in the main text. Ignoring the covariance terms, the net-short curve in this appendix is

$$1 + \frac{1}{12}r_t - \frac{n}{12}y_{n,t}^s \approx \frac{1}{12}r_t^{syn} - \frac{1}{12}i_t^s + E_t^{\mathbb{Q}}[1 - \frac{n-1}{12}y_{n-1,t+1}^s]$$

$$ny_{n,t}^s \approx i_t^s - (r_t^{syn} - r_t) - (n-1)E_t^{\mathbb{Q}}[y_{n-1,t+1}^s]$$

$$ny_{n,t}^s \approx E_t^{\mathbb{Q}}\left[\sum_{j=0}^{n-7}(i_t^s - (r_t^{syn} - r_t)) + \frac{6}{12}y_{t+n-6}^{bill}\right]$$

It is straightforward to show that equation (23) in the main text also leads to the same linear approximation.

[3]Subject to the caveat that the intermediary must believe in the zero lower bound. Our formulas can be readily generated to other (non-zero) lower bounds, at the expense of additional notation.