THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Economic and Social Research Council

Demand–Supply Imbalance Risk and Long-Term Swap Spreads

Samuel G. Hanson
Aytek Malkhozov
Gyuiri Venter

SRC Discussion Paper No 118
April 2022

Systemic Risk Centre

Discussion Paper Series

ISSN 2054-538X

Abstract
We develop a model in which long-term swap spreads are determined by end users' demand for swaps, constrained dealers' supply of swaps, and the risk of future imbalances between demand and supply. Exploiting the sign restrictions implied by our model, we estimate these unobserved demand and supply factors using data on swap spreads and a proxy for dealers' swap arbitrage positions. We find that demand and supply play equally important roles in driving the observed variation in swap spreads. Yet, as predicted by the model, demand plays a more important role in shaping the expected returns on swap spread arbitrage, which embed a premium for bearing future demand-supply imbalance risk. Hedging activity from mortgage investors seems to play a key role in driving the demand for swaps. By contrast, the supply of swaps is closely linked to proxies for the tightness of dealers' constraints. Finally, our analysis helps explain the relationship between swap spreads and other no-arbitrage violations.

Keywords: limits to arbitrage, intermediary capital constraints, swap spreads, covered interest parity

JEL Classification: G12, E43

This paper is published as part of the Systemic Risk Centre's Discussion Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/R009724/1].

Samuel G. Hanson, Harvard Business School

Aytek Malkhozov, Queen Mary University London and Systemic Risk Centre, London School of Economics

Gyuri Venter, Warwick Business School

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

Demand–Supply Imbalance Risk and Long-Term Swap Spreads*

Samuel G. Hanson
Harvard Business School

Aytek Malkhozov
Queen Mary University of London

Gyuri Venter
Warwick Business School

April 4, 2022

Abstract

We develop a model in which long-term swap spreads are determined by end users' demand for swaps, constrained dealers' supply of swaps, and the risk of future imbalances between demand and supply. Exploiting the sign restrictions implied by our model, we estimate these unobserved demand and supply factors using data on swap spreads and a proxy for dealers' swap arbitrage positions. We find that demand and supply play equally important roles in driving the observed variation in swap spreads. Yet, as predicted by the model, demand plays a more important role in shaping the expected returns on swap spread arbitrage, which embed a premium for bearing future demand-supply imbalance risk. Hedging activity from mortgage investors seems to play a key role in driving the demand for swaps. By contrast, the supply of swaps is closely linked to proxies for the tightness of dealers' constraints. Finally, our analysis helps explain the relationship between swap spreads and other no-arbitrage violations.

Keywords: limits to arbitrage, intermediary capital constraints, swap spreads, covered interest parity

JEL Classification: G12, E43

*Emails: shanson@hbs.edu, a.malkhozov@qmul.ac.uk, and gyuri.venter@wbs.ac.uk.
We thank Ajay Subramanian, and seminar and conference participants at Bayes Business School, Queen Mary University of London, Warwick Business School, the German Finance Association Annual Meeting, the Econometric Society European Winter Meeting, and the Paris December Finance Meeting for helpful comments. We are also grateful to Chris Anderson, Sven Klingler, and Vlad Sushko for sharing their data.

1 Introduction

An interest rate swap is an agreement to exchange a series of floating interest payments based on the future realizations of a short-term reference rate—the "floating leg"—for a series of fixed interest payments—the "fixed leg". With more than \$370 trillion of outstanding notional value as of 2021, the market for interest rate swaps is the world's largest derivatives market.[1] Investors and borrowers use these swaps to manage their exposures to interest-rate risk.

Since their advent in the late 1980s until 2008, the fixed rates on swaps had always remained above the corresponding rates on like-maturity government bonds. In other words, swap spreads, the difference between swap rates and government bond rates, had been uniformly positive. However, beginning in the Global Financial Crisis, the fixed rates on long-dated swaps—for instance, the 30-year swaps—have fallen below government bond yields, resulting in negative long-dated swap spreads.[2] Negative swap spreads seemingly represent a pure arbitrage opportunity, a puzzle in a large and liquid market. As such, negative spreads have been alternatively attributed to large increases in the demand for long-dated swaps from end users or to rising balance sheet costs at the financial intermediaries that supply swaps.[3] Disentangling these forces in the swap market can help distinguish between periods in which intermediaries withdraw from arbitrage trades, and periods in which the large swings in investor demand outstrip intermediaries' capacity and cause market dislocations.[4] In addition, negative swap spreads provide a context for understanding the government borrowing costs through the lens of the differences in investor base for interest rate swaps and that of Treasury securities.[5]

In this paper, we develop a theoretical framework to study the effect of demand and supply forces on swap spreads. Using this framework, we separately identify investors' demand for swaps and financial intermediaries' supply of swaps, and decompose the variation in swap spreads into the respective contributions of these two forces. Moreover, we highlight a key determinant of swap spreads that has been overlooked in the literature: Financial intermediaries need to be compensated for the risk that spreads may temporarily widen because of future shocks to demand or supply. We show that compensation for this risk explains a significant fraction of the returns to swap arbitrage.

We build a model in which the demand of preferred habitat investors, such as mortgage investors and pension funds, for receiving the fixed rate (and paying the floating rate) in long-term interest rate swaps is not naturally offset by opposing demands from other investors and has to be accommodated by risk-averse and leverage-constrained financial intermediaries. Intermediaries' binding leverage constraints limit the arbitrage between the swap market and the

[1] See https://stats.bis.org/statx/srs/table/d5.1.

[2] See Boyarchenko et al. (2018) for an overview of the negative swap spreads witnessed since 2008.

[3] See Klingler and Sundaresan (2019) and Jermann (2020) for, respectively, a demand and a supply perspective on negative swap spreads.

[4] The latter situation corresponds, for instance, to the March 2020 stress in fixed-income markets; see Duffie (2020) and Vissing-Jorgensen (2021).

[5] This point was recently underscored by the Treasury Borrowing Advisory Committee; see TBAC (2021).

Treasuries market, where intermediaries hedge the interest rate risk of their swap positions. These binding leverage constraints open the door to nonzero swap spreads, which are a failure of the Law of One Price (LoOP). In equilibrium, swap spreads are driven by both shocks to end users' net demand for swaps and shocks to intermediaries' wealth, which can either relax or tighten their leverage constraints and thereby shift the supply of swaps.

Demand shocks and supply shocks naturally induce different co-movements between swap spreads and intermediaries' swap positions. Specifically, positive shocks to end-user demand for swaps increase the magnitude of both swap spreads and intermediaries' swap positions, whereas shocks to intermediaries' supply of swaps—shocks to intermediaries' wealth—decrease the magnitude of swap spreads but increase the magnitude of swap positions. Indeed, our model has a natural representation as a structural vector auto-regression (VAR) in which demand and supply shocks can be identified with sign restrictions on the responses of swap spreads and swap positions to these shocks. Estimating the structural VAR implied by our model, we find that demand and supply shocks both contribute significantly to the variation in swap spreads after 2009, each capturing a starkly distinct aspect of this variation. Moreover, shifting hedging demand from mortgage investors appears to be the main driver of net end-user demand for swaps.

In addition to the interplay between demand and supply, our model points to *arbitrage risk* as a key determinant of long-dated swaps spreads: Arbitraging long-dated swap spreads is not only balance sheet consuming, but also risky for financial intermediaries. Specifically, future shocks to either demand or supply may temporarily widen swap spreads, leading to capital losses for financial intermediaries. As a result, intermediaries will require compensation for this arbitrage risk which takes the form of *demand–supply imbalance risk*, increasing the magnitude of swap spreads at longer maturities. In other words, once LoOP violations arise because of limits to arbitrage, they are amplified by arbitrage risk. The presence of arbitrage risk implies that measures of intermediaries' positions in the swap arbitrage trade should predict the returns to this trade, even after controlling for measures of intermediaries' balance sheet costs. We verify this proposition in the data using predictive regressions.

While demand and supply both play important roles in driving the variation in swap spreads, we should expect end-user demand to be a stronger predictor of the returns to swap arbitrage than supply. Intuitively, positive demand shocks simultaneously increase the compensation intermediaries require for committing their scarce balance sheet to swap arbitrage and the compensation they require for bearing future swap spread risk. By contrast, a negative supply shock—a negative shock to intermediaries' wealth—reduces intermediaries' exposure to swap spread risk and the associated premium while increasing the compensation required for using the balance sheet. Because these two effects partially offset each other, supply shocks have a smaller impact on the returns to swap arbitrage than do demand shocks. We verify this predicted asymmetry in the data, providing further evidence in support of the arbitrage risk channel.

Interestingly, the model also points to the possible determinants of the balance sheet cost associated with intermediating swaps. To the extent that intermediaries allocate their scarce balance sheet between swap arbitrage and other risky investments, the market price of risk represents the opportunity cost of committing the balance sheet to swap arbitrage. Thus, in the model, there is a link between risk premia and swap spreads, which does not rely on intermediaries being the marginal investor in the market for the risky assets. Our findings are in line with this prediction.

Finally, we show that decomposing swap spreads into demand and supply components helps understand the relationship of negative swap spreads to other LoOP violations, such as the deviations from the covered interest parity. In particular, we find that swap supply is primarily correlated with short-maturity covered interest parity deviations whereas swap demand—with long-maturity covered interest parity deviations.

Our model builds on Vayanos and Vila (2021) and, as a result, is related to the growing literature on demand factors in the government bond market; see Greenwood and Vayanos (2014), Hanson (2014), Malkhozov et al. (2016), Gourinchas et al. (2020), and Greenwood et al. (2020), among others. Our model is different from the above papers along two dimensions. First, we consider long-maturity swap spreads that arise because of limits to arbitrage rather than long-maturity bond yields. Second, we allow for the variation in both demand and supply in the swap market, and use model predictions to disentangle these two forces empirically.

Our work is also related to De Long et al. (1990) who show that noise traders can create a risk in the price of an asset that deters rational arbitrageurs from aggressively betting against mispricings caused by these noise traders. In their model, an equilibrium with noise trader risk can exist alongside a more standard equilibrium in which arbitrageurs eliminate all mispricings and noise trader risk does not arise.[6] In contrast, we show that noise trader risk does not have to rely on the special type of equilibrium considered in De Long et al. (1990). In our model, the violations of the LoOP arise because intermediaries, who play the role of arbitrageurs, are subject to a leverage constraint. Once these violations are present, they are amplified by the risk of demand and supply shocks. Our results are also related to Spiegel (1998) as our model, which features demand shocks, has multiple equilibria that qualitatively differ from each other. While we discuss the existence of other equilibria of our model, we focus on the unique stable equilibrium to derive testable predictions.

Swap rates and Treasury yields have been extensively studied in previous literature. A stream of literature calibrates dynamic term structure models to understand the dynamics of swap spreads; see, for instance, Duffie and Singleton (1997), Lang et al. (1998), Collin-Dufresne and Solnik (2001), Liu et al. (2006), and Feldhütter and Lando (2008), among others. The ones closest to our paper are Klingler and Sundaresan (2019) and Jermann (2020) who focus on, respectively, the demand and the supply channels in the swap market in isolation.

Finally, our work is also related to Cohen et al. (2007), Chen et al. (2018), and Goldberg and

[6] Loewenstein and Willard (2006) argue that the equilibrium with noise trader risk in De Long et al. (1990) has several unappealing properties.

Nozawa (2021) who identify demand and supply shifts in shorting, index option, and corporate bond markets, respectively. In our paper, we focus on demand and supply forces in the swap market. Moreover, we show how shifts in demand and supply constitute a source of risk for the arbitrageurs in this market, who consequently require a premium reflected in long-maturity swap spreads.

The rest of the paper is organized as follows. Section 2 provides background on interest rate swap contracts and their market. Section 3 presents a theoretical framework to study the impact of demand and supply risk on swap spreads and derives testable predictions. Section 4 describes our data and Section 5 presents our main empirical findings. Section 6 concludes. An Online Appendix gathers additional results omitted from the main paper.

2 Background

Before delving into our theoretical and empirical analysis, we provide background to the market for interest rate swaps. After explaining the mechanics of these derivative contracts, we review the no-arbitrage pricing logic that makes negative swap spreads so surprising. We then discuss the swap market participants, dividing them into "end users" and specialized "intermediaries." Finally, we describe how the swap market evolved since its advent in the 1980s with a focus on the major changes this market has undergone since the Global Financial Crisis.

2.1 Interest rate swap mechanics

An interest rate swap is an agreement between two counterparties to exchange a series of interest payments over time based on some notional principal amount. In a plain-vanilla interest rate swap, the first counterparty pays the second a series of fixed and known payments based on the swap rate set at the contract's inception; the second counterparty pays the first a series of floating and initially unknown payments based on the future realizations of some short-term reference rate. The fixed swap rate is set so that the swap has zero value at inception. Thus, the fixed swap rate is equivalent to a par coupon yield derived from the underlying reference curve. As a result, the counterparty who is receiving (paying) the fixed swap rate acquires an financial exposure that is similar to that obtained by borrowing (investing) cash at the short-term interest rate and taking a long (short) position in long-term bonds. A par swap spread is the difference between the fixed swap rate and the par coupon yield of a Treasury bond with the same maturity.

Historically, the floating leg on most interest rate swaps was tied to the 3-month London Interbank Offer Rate (LIBOR), which is an indicative unsecured 3-month borrowing rate for major global banks. In recent years, swaps tied to overnight interbank unsecured borrowing rates and overnight secured borrowing have been replacing LIBOR-based swaps as the market standard, and LIBOR was discontinued at the end of 2021. Specifically, Overnight Index Swaps

(OIS) are tied to the effective federal funds rate—an overnight unsecured interbank borrowing rate—and SOFR swaps are tied to the Secured Overnight Financing Rate (SOFR)—a rate based on overnight repurchase agreements collateralized by Treasury securities.

Since its birth in the 1980s, the market for U.S. dollar interest rate swaps has grown into the largest derivatives market in the U.S. with over \$115 trillion of outstanding notional value and a gross market value exceeding \$2.3 trillion as of 2020.[7] The market for U.S. dollar interest rate swaps is highly liquid with an average daily trading volume of roughly \$500 billion in 2020.[8] A typical bid-ask spread for 10-year interest rate swaps is around 1 basis point whereas a typical bid-ask spread on a 10-year Treasury note is roughly 0.25 basis points.

2.2 No-arbitrage pricing bounds

To begin, note that, in the absence of frictions, a non-zero spread on a SOFR swap constitutes a failure of the LoOP—i.e., it implies the existence of a zero-cost portfolio that generates a riskless stream of positive cashflows—and hence an arbitrage. Suppose, for instance, that the SOFR swap spread is positive at some initiation time $t = 0$. Entering a receive-fixed SOFR swap and taking an offsetting short position in like-maturity Treasuries (and investing the short-sale proceeds at SOFR) generates a stream of riskless cashflows equal to this positive swap spread. In symbols, the cashflow on this zero-cost position that is a "long SOFR swap spread" at any time $t > 0$ is:

$$\begin{aligned} CF(\text{Long SOFR swap spread})_t &= \overbrace{(y_0^{\text{SOFR swap}} - \text{SOFR}_t)}^{CF(\text{Receive-fixed SOFR swap})_t} + \overbrace{(-y_0^{\text{Treasury}} + \text{SOFR}_t)}^{CF(\text{Short Treasury position})_t} \\ &= \underbrace{(y_0^{\text{SOFR swap}} - y_0^{\text{Treasury}})}_{\text{SOFR swap spread}_0 > 0} > 0. \end{aligned}$$

Alternately, if the SOFR swap spread is negative at $t = 0$, entering a pay-fixed SOFR swap and taking a long position in Treasuries (that is financed at SOFR) generates a stream of riskless cashflows equal to the negative one times the swap spread:

$$\begin{aligned} CF(\text{Short SOFR swap spread})_t &= \overbrace{(-y_0^{\text{SOFR swap}} + \text{SOFR}_t)}^{CF(\text{Pay-fixed SOFR swap})_t} + \overbrace{(y_0^{\text{Treasury}} - \text{SOFR}_t)}^{CF(\text{Long Treasury position})_t} \\ &= -\underbrace{(y_0^{\text{SOFR Swap}} - y_0^{\text{Treasury}})}_{\text{SOFR swap spread}_0 < 0} > 0. \end{aligned}$$

[7] See https://stats.bis.org/statx/srs/table/d5.1. The gross market value of the interest rate swap market is the absolute value of the market value of all outstanding receive fixed swap contracts. By way of comparison, at the end of 2020, there were roughly \$21 trillion of outstanding U.S. Treasury securities, \$16 trillion of corporate bonds, and \$9 trillion of MBS. Across all currencies, the BIS reports that there were over \$350 trillion of outstanding notional of interest rate swaps at the end of 2020 with a gross market value exceeding \$10 trillion.

[8] This is based on the total volume of swaps cleared through central counterparty clearing house (CCPs); see https://www.clarusft.com/2020-ccp-volumes-and-market-share-in-ird/. By way of comparison, according to SIFMA, the average daily trading volume in 2020 was roughly \$600 billion for U.S. Treasury securities, \$300 billion for U.S. MBS and \$40 billion for U.S. corporate bonds.

By contrast, assuming that the 3-month LIBOR will always exceed the secured financing rate for Treasuries—a highly plausible assumption since LIBOR contains compensation for credit risk—a positive spread on a LIBOR swap does not represent a failure of LoOP or of no arbitrage. For instance, receiving fixed on a LIBOR swap and shorting Treasuries amounts to receiving a fixed swap spread and paying the floating spread between LIBOR and the secured financing rate:

$$CF(\text{Long LIBOR swap spread})_t = \underbrace{(y_0^{\text{LIBOR swap}} - y_0^{\text{Treasury}})}_{\text{LIBOR swap spread}_0 > 0} - \underbrace{(\text{LIBOR}_t - \text{SOFR}_t)}_{\text{Short-rate differential}_t \geq 0}.$$

As a result, $CF(\text{Long LIBOR swap spread})_t$ will be negative if the short-rate differential is large enough. However, if $\text{LIBOR}_t - \text{SOFR}_t \geq 0$ in all possible states, a negative LIBOR swap spread is a violation of no arbitrage. In that case, a short position in the LIBOR swap spread—paying fixed on a LIBOR swap and taking an offsetting long position in Treasuries—is a zero-cost portfolio with strictly positive cashflows in all possible states given by:

$$CF(\text{Short LIBOR swap spread})_t = -\underbrace{(y_0^{\text{LIBOR swap}} - y_0^{\text{Treasury}})}_{\text{LIBOR swap spread}_0 < 0} + \underbrace{(\text{LIBOR}_t - \text{SOFR}_t)}_{\text{Short-rate differential}_t \geq 0} > 0.$$

2.3 Swap market participants

Who are the participants in the interest rate swap market? We find it useful to categorize market participants into "end users" and specialized " intermediaries." End users—for instance, financial institutions, corporations, and governments—generally use swaps to manage their pre-existing exposures to interest-rate risk. To clear the swap market, specialized intermediaries—who we associate with broker-dealers and fixed-income hedge funds—must accommodate the net end-user demand to either receive or pay the fixed swap rate. These specialized intermediaries hedge the interest rate risk associated with their swap positions by taking offsetting positions in the Treasuries market. As a result, these intermediaries are primarily concerned with the relative valuation of swaps and Treasuries—i.e., with the level of swap spreads and any mismatch between the short-term rate referenced by swaps and the short-term financing rate on their offsetting Treasuries positions. In addition to the cashflows and potential changes in the mark-to-market value of their swap spread positions, swap intermediaries must also weigh the fact that these swap spread positions will consume their scarce risk-bearing capital.

End users of interest rate swaps are agents who want long or short exposure to long-term bonds, but who for regulatory, accounting, or other frictional reasons prefer to obtain their desired bond exposure using interest rate swaps as opposed to Treasuries or other cash instruments. As explained in the U.S. Treasury Borrowing Advisory Committee's TBAC (2021) report on the swap market, there are several important groups of end users in the swap market.

Insurers and pensions use swaps to manage their pre-existing interest-rate risk exposures. Insurers and pensions typically receive the fixed swap rate on net—i.e., they use swaps to add duration—to offset the fact that the duration of their liabilities generally exceeds the duration of their preferred mix of on-balance sheet assets; see, e.g., Klingler and Sundaresan (2019) and TBAC (2021). In the case of pension funds, Klingler and Sundaresan (2019) argue that this desire to receive fixed increases when pensions become more underfunded.[9] Furthermore, since the convexity of their liabilities generally exceeds that of their assets, when long-term interest rates fall, insurers and pensions often need to enter into additional receive-fixed swaps to manage their exposure to interest rates; see Domanski et al. (2017).[10]

Commercial banks also typically receive fixed on net; see, for instance, Begenau et al. (2020) and TBAC (2021). Although banks borrow short-term and lend long-term, banks are generally hurt by declining interest rates because their loans reprice far more quickly than their deposit liabilities, which reprice very slowly if at all; see Driscoll and Judson (2013) and Drechsler et al. (2021), among others. To offset these pre-existing exposures, banks generally receive the fixed swap rate on net to add duration.

Non-financial corporations typically receive the fixed swap rate in order to convert fixed-rate bonds that they issued into synthetic floating-rate funding; see TBAC (2021). Chernenko and Faulkender (2011), in particular, argue that these firms are more likely to swap existing fixed rate debt to floating—i.e., to receive fixed—when the yield curve is steep.

Relative-value mortgage investors typically pay the fixed swap rate on net; see TBAC (2021). These investors—most prominently Fannie Mae and Freddie Mac, but also relative-value fixed-income asset managers—attempt to exploit the fact that passthrough mortgage-backed securities (MBS) sometimes trade cheap relative to a dynamic replicating portfolio of pay-fixed swaps.[11] These investors generally prefer to hedge their MBS holdings with swaps instead of Treasuries due to a combination of regulatory and accounting reasons and the fact that swaps have historically been a more effective hedge for MBS. When long-term rates fall, expected mortgage prepayments rise, causing MBS duration to decline. This prompts mortgage investors to enter receive-fixed swaps to reduce the net size of their pay-fixed swap hedge; see Perli and Sack (2003), Feldhütter and Lando (2008), Hanson (2014), and Malkhozov et al. (2016).

Mortgage servicers are institutions that earn a stream of fees to process mortgage payments; they collect monthly mortgage-related payments from homeowners and pass them along to MBS investors, local tax authorities, and property insurers. Mortgage servicers have an economic exposure similar to the holder of an interest-only (IO) MBS strip, which typically has a negative duration: their value declines when interest rates fall because the resulting rise in

[9] To see the idea, recall that a pension's amount of underfunding is the difference between the value of its liabilities and its assets $UF \equiv L - A$. As a result, a pension's dollar duration gap $DGAP \equiv D_L \cdot L - D_A \cdot A$ can be written as $DGAP = (D_L - D_A) \cdot L + D_A \cdot UF$, which is increasing in underfunding UF, all else equal.

[10] TBAC (2021) argues that the need to hedge variable annuity liabilities plays an especially important role in driving insurers' desire to pay the fixed swap rate.

[11] Because conventional fixed-rate mortgage embed a zero penalty prepayment option, MBS are akin to callable bonds—i.e., a combination of a long position of a regular non-callable bond and a short position in a call option on that same bond. As such, one can hedge a position in MBS with a dynamic replicating portfolio of pay-fixed swaps. However, in the case of MBS, the exercise behavior of the holders of this call option—individual homeowners—is quite complex and difficult to forecast; see Hanson (2014) and Malkhozov et al. (2016).

expected prepayments means that the IO holder expects to receive fewer interest payments. To offset the negative duration of their assets, servicers are typically net receivers of the fixed swap rate. Further, since IO strips are negatively convex just like passthrough MBS, servicers tend to increase their receive-fixed swap positions when long-term rates fall; see TBAC (2021).

Fixed-income money managers such as mutual funds use swap positions to adjust the duration of their portfolios and are typically net payers of the fixed swap rate; see TBAC (2021). These players typically hold cash bonds and prefer to use pay-fixed swap positions to attain their portfolio duration targets.

To summarize, most major groups of end users—including banks, insurers, pensions, corporations, and mortgage servicers—are typically net receivers of the fixed swap rate. The main net payers of the fixed swap rate are relative-value mortgage investors—for instance, Fannie Mae and Freddie Mac—and fixed-income money managers. Furthermore, since for many end users the duration of their interest rate exposure declines when rates fall, end-user demand to receive the fixed swap rate often rises when long-term rates decline; see TBAC (2021).

2.4 The evolution of the swap market

From the inception of the swap market in the 1980s until late 2008 at the height of the Global Financial Crisis (GFC), swap yields had always exceeded like-maturity Treasury yields—i.e., swap spreads had always been positive. The most straightforward explanation for positive swap spreads is that LIBOR always exceeded Treasury repo rates because (i) LIBOR is an unsecured 3-month bank borrowing rate that includes compensation for credit risk (Collin-Dufresne and Solnik, 2001) and (ii) Treasury yields and Treasury repo rates are depressed by a money-like convenience premium specific to these extremely safe and liquid assets (Feldhütter and Lando, 2008). However, there is also evidence that demand-and-supply forces in the swap market played a role in supporting the level of swap spreads before 2008.

Pre-GFC, hedged mortgage investors played a dominant role in the swap market and there was generally a net end-user demand to pay the fixed swap rate; see TBAC (2021). To accommodate this net demand to pay fixed, specialized intermediaries—broker-dealers and fixed-income hedge funds—were generally "long swap spreads," meaning that they received the fixed swap rate and took offsetting short positions in Treasuries; see, e.g., Duarte et al. (2006). Indeed, the 1998 bond market crisis involving Long Term Capital Management (LTCM) revealed that major fixed-income hedge funds (LTCM and D.E. Shaw) and large broker-dealers (Bank of America, Barclays, Goldman Sachs, Morgan Stanley, and Salomon Smith Barney) had substantial long positions in this swap spread trade; see, e.g., Lowenstein (2000).

This pattern is illustrated in Figure 1, which plots 30-year swap spreads alongside primary dealers' net position in U.S. Treasuries from 2001 to 2020. As shown in Figure 1, primary dealers were indeed net short in Treasuries pre-GFC when swap spread were positive. Our interpretation is that this net short Treasury position was driven, in significant part, by the dealers' long position in the swap spread trade. Indeed, we will use dealers' net position in Treasuries as an (admittedly noisy) proxy for the sign and magnitude of dealers' positions in the swap spread trade.

At the height of the GFC in late 2008, long-dated LIBOR swap spreads turned negative and have remained negative since, leading to an apparent violation of no arbitrage. Spreads' move into negative territory in the wake of the GFC occurred against the backdrop of three major changes in the swap market. First, there was a significant shift in the end-user demand for swaps, with net end-user demand swinging from a desire to pay fixed to a desire to receive fixed; see TBAC (2021). Second, financial intermediaries became far more concerned with husbanding their scarce loss-bearing capital. As a result, the shadow value of intermediary capital has been increasingly impounded into market prices, leading to a noteworthy rise in deviations from the LoOP across a number of intermediated markets; see, e.g., Gârleanu and Pedersen (2011), Du et al. (2018), and Siriwardane et al. (2021). Finally, there was a substantial increase in outstanding Treasury debt during the GFC and, historically, increases in Treasury supply have tended to reduce the money-like convenience premium in Treasury yields and hence swap spreads; see, e.g., Cortes (2003), and Krishnamurthy and Vissing-Jorgensen (2012). We next discuss these three forces in greater detail.

First, as shown in Figure 1, primary dealers' net position in Treasuries switched from net short to net long in early 2009—just around the time when 30-year swap spread turned negative. We interpret this expanding long Treasury position as being driven, in significant part, by the growing short position dealers were taking in the swap spread trade in response to a swing in net end users' desire from paying fixed to receiving fixed.

What drove the shift in net end-user demand? TBAC (2021) points to a decline in the demand to pay the fixed swap rate from hedged mortgage investors. Following their placement into conservatorship in late 2008, Fannie Mae and Freddie Mac began gradually shrinking the size of their on-balance-sheet or "retained" mortgage portfolios; see, e.g., Frame et al. (2015). The combined size of these retained portfolios shrank from roughly $1.5 trillion in 2008 (about 14% of outstanding residential mortgages) to $350 billion by 2020 (3% of outstanding mortgages), leading to a sizeable reduction in this traditional source of end-user demand to pay fixed. At the same time, the Federal Reserve's large-scale purchases of MBS—the Fed built its MBS holdings from zero in 2008:Q2 to $1.9 trillion in 2014, or roughly 19% of outstanding mortgages—had the effect of further removing MBS from the hands of investors who were inclined to hedge their value using pay-fixed swaps. A complementary explanation comes from Klingler and Sundaresan (2019) who argue that an increase in underfunding following the GFC led pensions to manage their duration by receiving the fixed rate on long-term swaps rather than by buying long-term bonds. As a result, it appears that the balance of end-user demand swung from a net desire to pay the fixed swap rate to a net desire to receive fixed.

Second, during the GFC, many financial intermediaries suffered large losses and became concerned with husbanding their increasingly scarce loss-bearing capital. And then, motivated by a desire to safeguard financial stability in the aftermath of the GFC, regulators have subjected large dealer banks to more stringent capital regulations; see, e.g., Hanson et al. (2010). Because it is expensive for dealer banks to finance themselves with equity instead of debt and because, as argued by Boyarchenko et al. (2020), dealers effectively pass along their capital requirements to the hedge funds who rely on them for leverage, these heightened capital requirements have further increased the cost of intermediation. Indeed, the shadow value of intermediary capital has increasingly been impounded into market prices since the GFC, leading to a rise in persistent deviations from the LoOP in a range of intermediated markets; see Gârleanu and Pedersen (2011).[12] Most importantly for us, U.S. regulators introduced the supplementary leverage ratio (SLR) in early 2014 after nearly six years of public discussion. In a nutshell, the SLR requires large dealer banks to have Tier 1 capital equal to 5% of their "total leverage exposure," defined as the sum of on-balance-sheet assets plus an adjustment for off-balance-sheet exposures. Crucially, unlike the more traditional forms of risk-based capital regulation, the SLR depends only on the notional scale of dealers' exposures and not on their assessed risk. Overall, the SLR has become the binding capital constraint for most large dealer banks; see, e.g., Duffie (2017) and Greenwood et al. (2017).[13]

Finally, there was a substantial increase in outstanding Treasury debt following the onset of the GFC. Specifically, the ratio of marketable Treasury debt to GDP rose rapidly from 31% at the end of 2007 to 59% at the end of 2010.[14] Historically, increases in Treasury supply tended to reduce swap spreads, arguably because they reduce the money-like safety or liquidity premium commanded by Treasuries when the latter are scarce; see Cortes (2003) and Krishnamurthy and Vissing-Jorgensen (2012). As a result, it is conceivable that the substantial post-GFC expansion in Treasury supply largely sated this special demand for Treasuries, effectively eliminating the convenience premium on Treasuries and leading to a decline in swap spreads.[15]

In the next section, we propose a model which helps us disentangle the role that end-user demand for swaps and constrained intermediaries' supply of swaps played in driving negative swap spreads since the GFC.

[12] See also Du et al. (2018) for a detailed study of one widely-discussed post-GFC breakdown of the LoOP, namely the deviations from covered interest parity in the market for foreign exchange futures; Boyarchenko et al. (2020) for estimates of the required capital needed to exploit various deviations from the LoOP; and Siriwardane et al. (2021) for a empirical investigation of the commonalities in various deviations from the LoOP.

[13] In response to the March 2020 Treasuries market disruption, the Federal Reserve Board temporarily removed on-balance sheet positions in U.S. Treasuries (as well as central bank reserves) from dealers' total leverage exposure under the SLR. While this temporary exemption expired in March 2021, the Federal Reserve Board is considering proposed modifications to the SLR.

[14] Debt-to-GDP gradually rose further from 59% in 2010 to 78% in 2019. It then leapt to 100% at the end of 2020 due to the massive fiscal response to the COVID-19 pandemic and recession.

[15] Consistent with this view, the spread between the 3-month overnight index swap rate and the 3-month Treasury repo rate declined significantly in late 2008 and early 2009. This spread averaged +8 basis point from 2001 to 2008 and -7 basis points from 2009 to 2020.

3 Model

In our model, end users demand long-maturity interest rate swaps that are supplied by risk-averse and leverage-constrained intermediaries, who specialize in the swap market. These specialized intermediaries—who we associate with broker-dealers and fixed income hedge funds—hedge the interest rate risk associated with their swap positions in the Treasury market and, thus, are only concerned with the relative valuation of swaps and Treasuries—i.e., with the level of swap spreads. As a result, the level of swap spreads reflects the interplay between end-user demand for swaps and intermediaries' willingness to supply swaps.

3.1 Setting

Time is discrete and infinite, and is indexed by t. To begin, we consider a single perpetual interest rate swap and a single perpetual Treasury bond with their coupons declining at the same rate $0 < \delta < 1$. Later we extend the model to the entire term structure of swap spreads.

The swap arbitrage trade. Let y_t^S denote the fixed rate on this perpetual swap at time t and i_t^S the short-term rate referenced by the swap—i.e., the London Interbank Offer Rate in the case of LIBOR swaps, the effective federal funds rate in the case of OIS swaps, or the Secured Overnight Financing Rate in the case of SOFR swaps. The excess return on a receive-fixed interest rate swap from t to $t+1$ is given by

$$r_{t+1}^S \equiv \left(y_t^S - i_t^S\right) - \frac{\delta}{1-\delta}\left(y_{t+1}^S - y_t^S\right), \tag{1}$$

where $1/\left(1-\delta\right)$ represents the duration of the swap. The excess return on this receive-fixed swap consists of a carry term $y_t^S - i_t^S$ and a capital gain term $-\left(\delta/\left(1-\delta\right)\right)\left(y_{t+1}^S - y_t^S\right)$ that arises from any changes in the swap fixed rate from t to $t+1$. Similarly, the excess return on a position in perpetual Treasury bonds with yield y_t^T that is financed at the secured short-term financing rate applicable to Treasuries i_t^T—i.e., the rate on repurchase agreements backed by Treasury bonds (SOFR)—is

$$r_{t+1}^T \equiv \left(y_t^T - i_t^T\right) - \frac{\delta}{1-\delta}\left(y_{t+1}^T - y_t^T\right). \tag{2}$$

The corresponding perpetual swap spread is then the difference between the fixed rate on perpetual interest rate swaps and the yield on perpetual Treasury bonds: $s_t \equiv y_t^S - y_t^T$. The excess return from t to $t+1$ on a *swap arbitrage trade* that receives the fixed swap rate and hedges the associated interest rate by going short Treasury bonds becomes

$$r_{t+1}^{Spread} \equiv r_{t+1}^S - r_{t+1}^T = \left(s_t - m_t\right) - \frac{\delta}{1-\delta}\left(s_{t+1} - s_t\right), \tag{3}$$

where $m_t \equiv i_t^S - i_t^T$ is the *short rate differential*: the spread between the short-term rate referenced by the swap i_t^S and the short-term Treasury financing rate i_t^T.

We posit $m_t = \overline{m} + z_t^m$, where $\mathrm{E}[m_t] = \overline{m}$ is the unconditional mean level of the short rate differential and z_t^m is a mean-zero state variable that captures its fluctuations. A non-zero m_t might either derive from (i) the fact that the short-term interest rate referenced by the swap i_t^S contains a time-varying compensation for credit risk—as with 3-month LIBOR—or (ii) that short-term Treasury rates embed a special money-like convenience premium relative to other money-market rates. In both cases, we would expect fluctuations in m_t over time, and we generically expect to have $m_t \equiv i_t^S - i_t^T \geq 0$ almost surely, as it has been the case historically.[16]

Our primary focus is on the level of swap spreads, s_t, and the equilibrium expected returns on the swap spread arbitrage trade, $\mathrm{E}_t[r_{t+1}^{Spread}]$. Thus, for simplicity, we think of i_t^S, i_t^T, and y_t^T as given exogenously and pinned down by forces outside of our model. Furthermore, while we allow $m_t \equiv i_t^S - i_t^T$ to differ from zero and to fluctuate stochastically over time, we will routinely emphasize the case where $m_t = 0$ for all t almost surely—e.g., as would approximately be the case for swaps tied to the SOFR. In this case, long-term swaps and Treasuries have identical payoffs. Therefore, in the absence of frictions, we should have zero swap spreads by the LoOP. However, in our model we have non-zero swap spreads even if the short rate differential is always zero: The LoOP will fail because of binding intermediary leverage constraints, opening the door to purely demand-and-supply driven fluctuations in swap spreads.

Intermediaries in the swap market. At time t, risk-averse and leverage-constrained intermediaries who specialize in the swap market allocate their scarce capital w_t between the swap spread arbitrage trade and an outside risky investment opportunity. The excess return on this outside investment opportunity is r_{t+1}^o, and we assume that its first two moments are exogenously given by $\mathrm{E}_t\left[r_{t+1}^o\right] = \bar{r}_o > 0$ and $\mathrm{Var}_t\left[r_{t+1}^o\right] = \sigma_o^2 > 0$.[17] For simplicity, we assume an overlapping generations structure where date-t intermediaries are exogenously born with capital equal to $w_t = \overline{w} + z_t^w$, where $\overline{w} \equiv \mathrm{E}[m_t] > 0$ and z_t^w is an exogenous mean-zero state-variable that captures shifts in intermediary net worth.[18] Date-t intermediaries have mean-variance preferences over their one-period ahead wealth $w_{t,t+1}$ and have an absolute risk-aversion of α.

More formally, letting x_t denote intermediaries' position in the receive-fixed swap arbitrage

[16]For LIBOR swap spreads, we think of the short rate differential as $m_t = i_t^S - i_t^T = m_t^{\text{Conv.}} + m_t^{\text{Cred.}}$, where the first term ($m_t^{\text{Conv.}}$) reflects the money-like safety and liquidity premium on Treasuries and the second ($m_t^{\text{Cred.}}$) reflects the credit risk component of LIBOR. In the case of OIS swaps, we simply have $m_t = i_t^S - i_t^T = m_t^{\text{Conv.}}$.

[17]In the Online Appendix, we present a version of the model in which the outside opportunity is another violation of the LoOP rather than a risky investment; this is isomorphic to the baseline version of the model.

[18]Thus, we are not modelling the mechanism through which initial losses on the swap spread arbitrage trade could lead to amplification because these losses reduce intermediaries' net worth and tighten future leverage constraints as in Brunnermeier and Pedersen (2009). Specifically, we could model amplification of this sort if we assumed that intermediary wealth evolved endogenously according to $w_{t+1} = w_t + r_{t+1}^{Spread} x_t + r_{t+1}^o o_t$. This extension would lead to non-linear dynamics and would add significant complexity: s_t would depend on w_t as in our model with exogenous wealth shocks, but w_t would in turn also depend endogenously on s_t through the law of motion for intermediary wealth.

trade and o_t their position in the outside investment opportunity, date-t intermediaries solve:

$$\max_{x_t,\, o_t} \mathrm{E}_t \left[w_{t,\,t+1} \right] - \frac{\alpha}{2} \mathrm{Var}_t \left[w_{t,\,t+1} \right], \tag{4}$$

subject to the budget constraint

$$w_{t,\,t+1} = w_t + x_t r_{t+1}^{Spread} + o_t r_{t+1}^{o} \tag{5}$$

and the leverage constraint

$$\kappa_x \left| x_t \right| + \kappa_o \left| o_t \right| \leq w_t. \tag{6}$$

Here $\kappa_x, \kappa_o \in [0, 1]$ are the capital requirements associated with the swap arbitrage trade and the outside investment opportunity, respectively. For instance, in order to undertake a swap arbitrage trade of notional size $|x_t|$ intermediaries must commit $\kappa_x |x_t|$ of their scarce capital.

End-user demand for swaps. End users of interest rate swaps—such as banks, insurers, pensions, corporations, and mortgage servicers—are agents who demand exposure to long-term bonds, but who for regulatory, accounting, or other frictional reasons prefer to obtain their desired bond exposure using interest rate swaps as opposed to Treasuries. Importantly, financial intermediaries only need to accommodate the net demand from end users to the receive fixed swap rate—i.e., the end-user demand to receive fixed that is not offset by other end-user demand to pay fixed. Since end users can also substitute between interest rate swaps and Treasuries, we allow net end-user demand to receive the fixed rate to potentially be increasing in the swap spread. Specifically, we assume that net end-user demand takes the form

$$d_t = \bar{d} + z_t^d + \gamma s_t, \tag{7}$$

where $\gamma \geq 0$, and z_t^d is a mean-zero state variable capturing shifts in end-user swap demand.

Market clearing. Since interest rate swaps are in zero net supply, market clearing requires $d_t + x_t = 0$. In particular, if the net demand from end users to receive the fixed swap rate is positive, $d_t > 0$, then in equilibrium intermediaries must take on a short position in the swap arbitrage trade (paying the fixed swap rate and going long Treasuries) equal to $x_t = -d_t < 0$ to accommodate this demand. By contrast, if there is net end-user demand to pay the fixed swap rate, $d_t < 0$, then intermediaries must take on a long position in the swap arbitrage trade (receiving the fixed swap rate and going short Treasuries).

3.2 Equilibrium swap spreads

We begin by providing a general characterization of swap spreads that is applicable in environments where intermediary leverage constraints may only bind periodically and where end user net demand may change signs over time. We then provide a more precise characterization of swap spreads that is valid in settings when (i) the leverage constraint is always binding and (ii) the sign of end-user demand is constant over time. This case is particularly helpful in understanding the post-GFC period that is the main focus of our empirical analysis.

3.2.1 General characterization

Letting $\psi_t \geq 0$ denote the Lagrange multiplier associated with the leverage constraint in (6) and assuming $\mathrm{Cov}_t[\mathbf{z}_{t+1}, r^o_{t+1}] = \mathbf{0}$ where $\mathbf{z}_t = [z^m_t, z^d_t, z^w_t]'$, intermediaries' first-order condition for x_t is

$$\mathrm{E}_t[r^{Spread}_{t+1}] = \kappa_x sgn\,(x_t) \cdot \psi_t + \alpha V_t \cdot x_t, \tag{8}$$

where $V_t \equiv \mathrm{Var}_t[r^{Spread}_{t+1}] = \left(\frac{\delta}{1-\delta}\right)^2 \mathrm{Var}_t[s_{t+1}]$ is the conditional variance of r^{Spread}_{t+1}.[19] Similarly, the first order condition for o_t is

$$\bar{r}_o = \kappa_o sgn\,(o_t) \cdot \psi_t + \alpha\sigma^2_o \cdot o_t. \tag{9}$$

Since $\bar{r}_o > 0$ and $\psi_t \geq 0$, it must be the case that $o_t > 0$, so date-t Langrange multiplier is

$$\psi_t = \frac{\alpha\sigma^2_o}{\kappa_o}\left(\frac{\bar{r}_o}{\alpha\sigma^2_o} - o_t\right) \geq 0.$$

Thus, the shadow value of intermediary capital is proportional to the difference between the unconstrained investment in the outside opportunity, $\bar{r}_o/\left(\alpha\sigma^2_o\right)$, and intermediaries' current investment, o_t. Combining (6) and (9), we have

$$\psi_t = \psi\left(w_t, |x_t|\right) = \max\left\{0, \frac{\bar{r}_o - \alpha\sigma^2_o\frac{1}{\kappa_o}w_t + \alpha\sigma^2_o\frac{\kappa_x}{\kappa_o}|x_t|}{\kappa_o}\right\}. \tag{10}$$

Naturally, the shadow value of capital is greater when intermediaries' capital w_t is lower and when the scale of their position in the swap arbitrage trade $|x_t|$ is larger.

Combining (3), (8) and (10), and imposing market clearing, we find that the equilibrium expected return to swap arbitrate, $\mathrm{E}_t[r^{Spread}_{t+1}]$, satisfies:

$$\overbrace{(s_t - m_t) - \frac{\delta}{1-\delta}(\mathrm{E}_t\left[s_{t+1}\right] - s_t)}^{\text{Expected return to swap arbitrage: } \mathrm{E}_t[r^{Spread}_{t+1}]} = \overbrace{(-\kappa_x)sgn\,(d_t) \cdot \psi\left(w_t, |d_t|\right)}^{\text{Compensation for using scarce capital: } (-\kappa_x)sgn\,(d_t)\cdot\psi_t} + \overbrace{(-\alpha)V_t \cdot d_t}^{\text{Compensation for risk}}. \tag{11}$$

Since in equilibrium intermediaries must take positions that are equal in size and opposite

[19]We assume that outside investment opportunity returns are orthogonal to state variables describing the swap market for simplicity, without affecting our model predictions. In the general case, the risk associated with swap arbitrage trade has an additional term proportional to the conditional covariance between the outside investment opportunity and swap arbitrage trade returns $\mathrm{Cov}_t[r^{Spread}_{t+1}, r^o_{t+1}]$. However, this term is small if intermediaries are specialized, i.e, if they allocate their balance sheet primarily to swap arbitrage.

in sign to those of end users, the equilibrium expected return on a long swap arbitrage position, $\mathrm{E}_t[r_{t+1}^{Spread}]$, has the opposite sign of the net end-user demand to receive the fixed swap rate, d_t. For instance, if the net demand to receive fixed is negative—as was arguably the case prior to the GFC—then we must have $\mathrm{E}_t[r_{t+1}^{Spread}] > 0$ to induce intermediaries to take the required long position in the swap arbitrage trade. By contrast, if the net demand to receive fixed is positive—as appears to be the case since the GFC—then we must have $\mathrm{E}_t[r_{t+1}^{Spread}] < 0$ to induce intermediaries to take the required short position in the swap arbitrage trade.

Equation (11) also highlights the two key forces that shape the equilibrium expected returns on the swap arbitrage trade: compensation for using scarce intermediary capital and compensation for risk. When $\kappa_x > 0$, constrained intermediaries will require compensation for committing their scarce capital to the swap arbitrage trade even if this trade is completely riskless. And, being risk-averse, specialized intermediaries will require additional compensation for bearing the risk that they may suffer losses on their swap arbitrage trades due to unexpected changes in swap spreads.

Iterating (11) forward, we find an expression for the equilibrium swap spread level:

$$\begin{aligned} s_t \;=\; & \overbrace{(1-\delta)\sum\nolimits_{k=0}^{\infty}\delta^k \mathrm{E}_t\left[m_{t+k}\right]}^{\text{Expected short-rate differentials}} + \overbrace{(1-\delta)\sum\nolimits_{k=0}^{\infty}\delta^k \mathrm{E}_t\left[\left(-\kappa_x\right) sgn\left(d_{t+k}\right)\psi\left(w_{t+k},\left|d_{t+k}\right|\right)\right]}^{\text{Expected compensation for scarce capital}} \\ & + \underbrace{(1-\delta)\sum\nolimits_{k=0}^{\infty}\delta^k \mathrm{E}_t\left[\left(-\alpha\right) V_{t+k} d_{t+k}\right]}_{\text{Expected compensation for risk}}. \end{aligned} \tag{12}$$

The first term on the right-hand side of (12) is the fundamental component of swap spreads. Recalling that $m_t \equiv i_t^S - i_t^T$, this is simply the expected future differential between the short-term rate referenced by the swap, i_t^S, and the secured Treasury financing rate, i_t^T, averaged over the lifetime of the swap. Under the assumption that $m_t > 0$ almost surely, as would be the case for LIBOR-based swaps, this fundamental term pushes towards having positive swap spreads. However, this term is negligible for SOFR and other OIS swaps where $m_t \approx 0$.

The second term is the expected future compensation for consuming scarce intermediary capital over the life of the swap. Since $\psi_t = \psi\left(w_t, |d_t|\right) \geq 0$, this term has the opposite sign of the net end-user demand, d_t. For instance, if the net demand to receive the fixed swap rate is typically positive, this second term pushes towards having negative swap spreads.

The third and final term is the expected future compensation for bearing the risk associated with swap spread volatility over the life of the swap. Just like the second term, this final term has the opposite sign of the net end-user demand to receive the fixed swap rate.

This general characterization highlights the fact that our main theoretical conclusions do not rely on the assumption that intermediaries' leverage constraints are always binding. Even if leverage constraints are not binding at time t and $\psi_t = 0$, the mere potential for them to bind in the future makes swap spread arbitrage risky for intermediaries and they will only accommodate end-user demand for long-term swaps if they are compensated for this risk. Furthermore, even away from the constraint, fluctuations in intermediary capital will shape the likelihood that the constraint will bind in the future and hence the level of swap spreads. Thus, shifts in demand d_t and intermediary capital w_t will have qualitatively similar effects on swap spreads in a model with periodically binding constraints and where demand can change sign.

3.2.2 An affine equilibrium

To derive an affine equilibrium that we can readily take to the post-GFC data, we assume the model parameters are such that we almost surely have

$$0 < d_t < \frac{1}{\kappa_x} w_t < \frac{\kappa_o}{\kappa_x} \frac{\bar{r}_o}{\alpha \sigma_o^2} + d_t. \tag{13}$$

The first inequality ($0 < d_t$) means that end-user net demand to receive the fixed swap rate is always positive. This sign restriction is consistent with the prime dealers' net positioning in long-term Treasuries and the negative swap spreads that have been observed since 2008. The second inequality ($\kappa_x d_t < w_t$) says that intermediaries always have sufficient capital to accommodate end-user demand in the swap market. Thus, when their leverage constraints bind, intermediaries are forced to downsize their investments in the outside investment opportunity. And, the returns to swap spread arbitrage must then adjust to compensate intermediaries for consuming their scarce capital.

The third inequality ($w_t < \kappa_o \cdot \bar{r}_o / \left(\alpha \sigma_o^2\right) + \kappa_x \cdot d_t$) ensures that the intermediaries' leverage constraint (6) is always binding, as the unconstrained optimal positions are not available. Thus, $\psi_t > 0$ always. This condition rules out non-linearities that arise if leverage constraints only bind periodically.[20] Under these parametric restrictions, the equilibrium condition (11) becomes

$$(s_t - m_t) - \frac{\delta}{1-\delta} \left(\mathrm{E}_t \left[s_{t+1} \right] - s_t \right) = (-\kappa_x) \cdot \overbrace{\frac{\alpha \sigma_o^2}{\kappa_o^2} \left(\kappa_o \frac{\bar{r}_o}{\alpha \sigma_o^2} + \kappa_x d_t - w_t \right)}^{\psi_t} + (-\alpha) V \cdot d_t. \tag{14}$$

Further, we assume an autoregressive process for the three state variables governing the evolution of the short rate differential (z_t^m), end-user demand (z_t^d), and intermediary wealth (z_t^w). Specifically, we assume that the vector of state variables $\mathbf{z}_t = [z_t^m, z_t^d, z_t^w]'$ follows

$$\mathbf{z}_{t+1} = \varrho \mathbf{z}_t + \varepsilon_{t+1}, \tag{15}$$

where $\varrho = diag\left(\rho_m, \rho_d, \rho_w\right)$ is a diagonal matrix of AR(1) coefficients $\rho_m, \rho_d, \rho_w, \in [0,1)$ and $\varepsilon_t = [\varepsilon_t^m, \varepsilon_t^d, \varepsilon_t^w]'$ is the vector of structural shocks. Let $\mathrm{Var}\left[\varepsilon_{t+1}^i\right] = \sigma_i^2$, $i = m, w, d$ denote the variances of the structural shocks. For simplicity, we assume that the three structural shocks are orthogonal to each other, i.e., $\mathrm{Var}_t[\varepsilon_{t+1}] = diag\left(\sigma_m^2, \sigma_d^2, \sigma_w^2\right)$.

[20] We can construct a symmetric equilibrium where $0 < -d_t < \frac{w_t}{\kappa_x} < \frac{\kappa_o}{\kappa_x} \frac{\bar{r}_o}{\alpha \sigma_o^2} - d_t$ almost surely.

We conjecture that equilibrium swap spreads s_t are an affine function of the state vector $\mathbf{z}_t$:

$$s_t = A_0 + A_m z_t^m + A_d z_t^d + A_w z_t^w. \tag{16}$$

A rational expectations equilibrium of our model is a fixed point of an operator which gives the price-impact coefficients A_m, A_d, and A_w that clear the swap market when intermediaries believe the risk of swap spread arbitrage trade is determined by some initial set of price-impact coefficients. Combining the conjectured affine form (16), the end-user demand curve (7), and the equilibrium condition (14), we have:

Theorem 1 *In the affine equilibrium,*

$$A_0 = \frac{\overline{m} - \alpha \left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 \left[\frac{\kappa_o}{\kappa_x}\frac{\bar{r}_o}{\alpha\sigma_o^2} + \bar{d} - \frac{1}{\kappa_x}\overline{w}\right] - \alpha V \bar{d}}{1 + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right]}, \tag{17a}$$

$$A_m = \frac{1}{\frac{1-\rho_m\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right]}, \tag{17b}$$

$$A_d = -\frac{\alpha\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right]}{\frac{1-\rho_d\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right]}, \tag{17c}$$

$$A_w = \frac{\frac{1}{\kappa_x}\alpha\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2}{\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right]}, \tag{17d}$$

and

$$V = \left(\frac{\delta}{1-\delta}\right)^2 \left(A_m^2\sigma_m^2 + A_d^2\sigma_d^2 + A_w^2\sigma_w^2\right). \tag{18}$$

Equations (17b)-(17d) and (18) define a system of higher-order polynomial equations in A_m, A_d, and A_w. A solution to this system exists as long as risk aversion α is below a threshold $\overline{\alpha} > 0$. In general, there exist multiple solutions corresponding to multiple affine equilibria.

To provide intuition for the nature of model equilibria, we consider a range of special cases.

If there exists a non-trivial interest rate differential ($m_t \neq 0$ for all t due to $\overline{m} \neq 0$ and/or $\sigma_m^2 > 0$), and in absence of frictions ($\kappa_x = 0$) and of demand and supply shocks ($\sigma_d^2 = \sigma_w^2 = 0$), the model has a unique affine equilibrium with non-zero swap spreads. In particular, if interest rate differential is positive almost surely as in the case of LIBOR swaps, equilibrium swap spreads are also positive. Note that interest rate differentials on their own do not induce LoOP violations or equilibrium multiplicity, yet have a mechanical effect on the swap spread level. Therefore, to better illustrate the role of the other forces in the model, in the following examples we shut down the interest rate differential channel by setting $m_t = 0$ for all t. In other words, we consider SOFR swaps. We also begin by assuming that end-user demand is completely inelastic, $\gamma = 0$.

With limits to arbitrage in the form of binding leverage constraints ($\kappa_x > 0$) but no demand risk ($\sigma_d^2 = 0$), the model has a unique affine equilibrium, in which swap spreads are determined by the level and fluctuations of intermediaries' capital.

When end-user demand is stochastic ($\sigma_d^2 > 0$) but there are no frictions ($\kappa_x = 0$), the model can have two affine equilibria—a "low-volatility" equilibrium in which the LoOP holds and swap spreads are always zero, and a "high-volatility" equilibrium in which swap spread risk deters intermediaries from eliminating non-zero spreads.[21] With demand risk, the underlying fixed-point problem is more complicated because the risk of the swap arbitrage depends on how spreads react to future demand shocks. For example, if intermediaries believe that future demand shocks will have a large impact on swap spreads, they will perceive the swap arbitrage trade as being highly risky. As a result, intermediaries will only absorb current demand shocks if they are compensated by large spread changes and high expected future returns, making their initial belief self-fulfilling. Yet, only the low-volatility equilibrium in which swap spreads are always zero is stable in the sense that this equilibrium (i) is robust to small perturbations in investors' beliefs about equilibrium price impact and (ii) does not diverge in the limit when $\sigma_d^2 \to 0$.[22] Note that this special case is similar to De Long et al. (1990) and Spiegel (1998).

With both demand risk ($\sigma_d^2 > 0$) and frictions ($\kappa_x > 0$), the model can have two affine equilibria, only one of which is stable in the sense defined above. However, this stable equilibrium is no longer trivial and features non-zero swap spreads. Agents understand that binding leverage constraints will prevent intermediaries from enforcing LoOP in the future. As a result, shocks to both end-user demand and intermediary supply will impact future swap spreads, creating demand–supply imbalance risk the exposure to which risk-averse intermediaries must be compensated. Thus, unlike in De Long et al. (1990) and Spiegel (1998), arbitrage risk arises even in the stable equilibrium of our model.

Finally, our conclusions hold with a few modifications when end-user demand is elastic. With demand risk ($\sigma_d^2 > 0$), frictions ($\kappa_x > 0$), and elastic demand ($\gamma > 0$), the model can have three affine equilibria, only one of which is stable in the sense defined above.[23] This stable equilibrium features non-zero swap spreads. The three self-fulfilling equilibria, as opposed to the two in De Long et al. (1990) and Spiegel (1998), originate from the fact that end-user demand itself depends on the level of the swap spreads as in Vayanos and Vila (2021); when $\gamma \to 0$, these collapse to only two equilibria.

[21] In this case the system (17c)-(18) reduces to a quadratic equation in A_d which has real foots if investors' risk aversion α is sufficiently small.

[22] By contrast, similar to De Long et al. (1990), the high-volatility equilibrium is a knife-edge outcome which is not robust to small perturbations in investors' beliefs. In this equilibrium, the LoOP fails because all agents believe that LoOP will fail in the future. However, if a small number of investors believe that LoOP will hold in the future, then those investors will enforce LoOP and it will indeed hold.

[23] In this case, the fixed-point problem is equivalent to a cubic equation.

In conclusion, our model features multiplicity because of the demand shocks. In the rest of the analysis, we will focus exclusively on the unique stable equilibrium of the model. In this equilibrium, spreads are less volatile and demand shocks have a smaller price impact compared to the non-stable equilibria. Nevertheless, swap spreads in the stable equilibrium are non-zero, reflecting the combined effect of short rate differential, binding leverage constraints, and demand–supply imbalance risk.

3.3 Predictions

3.3.1 Average swap spread level

Our model naturally allows for non-zero long-maturity swap spreads and outlines the forces that determine the average swap spread level. We have:

Proposition 1 *Swap spreads are on average negative, $A_0 < 0$, if the average short rate differential satisfies $\overline{m} < \overline{\overline{m}}$ for some $\overline{\overline{m}} > 0$ constant given in the Appendix.*

From (17a), the average level of swap spreads, A_0, is determined by a combination of three forces. First, the spread is increasing in the short rate differential. This differential is positive and potentially large for LIBOR swaps, but negligible for OIS and SOFR swaps. The other two forces push the swap spreads in the negative territory as long as end users demand to receive the fixed swap rate, $d_t > 0$. In this case, intermediaries require a negative swap spread as compensation for (i) committing their scarce capital to a short position in the swap arbitrage trade and (ii) for bearing the risk associated with this position. Thus, our model predicts negative average OIS and SOFR swap spreads. Relative to those, the interest rate differential pushes the average level of the LIBOR swap spread towards positive territory, but as long as $\overline{m}$ is small, the balance sheet and risk considerations dominate the interest rate differential effect and LIBOR spreads at long maturities are negative.

3.3.2 Identification of demand and supply shocks

Our model provides restrictions that can help us identify the structural demand (ε_t^d) and supply (ε_t^w) shocks using data on short rate differential (m_t), swap spread level (s_t), and intermediaries' swap spread arbitrage trade position (x_t). In particular, we obtain the following result:

Proposition 2 *The short rate differential m_t, the equilibrium swap spread s_t, and intermediary positions in the long-term swap spread arbitrage trade x_t can be written as*

$$\overbrace{\begin{bmatrix} m_t \\ s_t \\ x_t \end{bmatrix}}^{\mathbf{y}_t} = \overbrace{\begin{bmatrix} \overline{m} \\ A_0 \\ -\left(\bar{d} + \gamma A_0\right) \end{bmatrix}}^{\mathbf{a}} + \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ A_m & A_d & A_w \\ -\gamma A_m & -\left(1 + \gamma A_d\right) & -\gamma A_w \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} z_t^m \\ z_t^d \\ z_t^w \end{bmatrix}}^{\mathbf{z}_t}. \tag{19}$$

Combined with (15), we have the structural VAR representation of the model's equilibrium:

$$\mathbf{y}_{t+1} = \left(\mathbf{I} - \mathbf{A}\varrho\mathbf{A}^{-1}\right)\mathbf{a} + \mathbf{A}\varrho\mathbf{A}^{-1}\mathbf{y}_t + \mathbf{A}\varepsilon_{t+1}, \tag{20}$$

where $\mathbf{I}$ *denotes the* 3×3 *identity matrix. Moreover, assuming that* $\gamma > 0$ *and* $\kappa_x > 0$*, we have* $A_m > 0$*,* $A_d < 0$*,* $A_w > 0$*,* $-\gamma A_m < 0$*,* $-(1 + \gamma A_d) < 0$*, and* $-\gamma A_w < 0$*.*

Proposition 2 implies that the matrix $\mathbf{A}$ of structural VAR coefficients can be identified using a combination of sign and zero restrictions.[24] In particular, positive demand shocks lower the equilibrium swap spread (i.e., make it more negative) and intermediaries' swap positions (i.e., increase the amount supplied), whereas positive shocks to intermediaries' wealth increase the swap spread (i.e., make it less negative) while decreasing swap positions (i.e., increase the amount supplied).

To better understand the sign restrictions, note that higher end-user demand naturally increases the size of intermediaries' short position in the swap arbitrage trade ($-(1 + \gamma A_d) < 0$) and the compensation they require for taking this position ($A_d < 0$).[25] On the other hand, higher intermediaries' capital relaxes their leverage constraints. If end-user demand is elastic ($\gamma > 0$), lower shadow cost of intermediary capital increases the size of intermediaries' short positions ($-\gamma A_w < 0$). The effect on the swap spread depends on the combination of lower compensation for using scarce capital and higher compensation for the risk associated with a larger short position. On net, we show that higher intermediaries' capital makes swap spreads less negative ($A_w > 0$).

3.3.3 Swap arbitrage returns

Our model outlines the drivers of expected returns to long-maturity swap arbitrage. Equation (14) directly implies the following result:

Proposition 3 *The expected one-period return to swap arbitrage,* $E_t[r_{t+1}^{Spread}] \equiv (s_t - m_t) - \frac{\delta}{1-\delta}(E_t[s_{t+1}] - s_t)$*, is increasing in intermediaries' position,* x_t *(i.e., decreasing in end-user demand* d_t*), even after controlling for the shadow cost of intermediary capital.*

Proposition 3 states that, in equilibrium, returns to swap arbitrage compensate investors for swap spread risk, in addition to the cost of committing their scarce capital to this trade. The required compensation for swap spread risk is increasing in the amount of risk that intermediaries hold, V. As a result, expected returns have to be more negative when intermediaries take a larger short position in the swap arbitrage trade, even after controlling for compensation these intermediaries require for using their scarce capital over the return period.

[24] When taking the model to the data, we will identify the structural shocks up to scale. Specifically, we will assume that the structural shocks are orthogonal and each have unit variance, and thus we will identify $\mathbf{A}\mathbf{\Sigma}^{1/2}$.

[25] These effects are mitigated when demand is elastic and sensitive to more negative swap spreads, $\gamma > 0$. To see this, consider the expression for the coefficient A_d in (17c) and $-(1 + \gamma A_d)$.

The model also characterizes the sign and the relative magnitude of the effects that shifts in short rate differential (z_t^m), demand (z_t^d), and supply (z_t^w) have on $\mathrm{E}_t[r_{t+1}^{Spread}]$. Using (19), we obtain the following results:

Proposition 4 *The expected one-period return to swap arbitrage can be expressed as*

$$E_t[r_{t+1}^{Spread}] = B_0 + B_m z_t^m + B_d z_t^d + B_w z_t^w, \tag{21}$$

where the loadings on state variables satisfy $B_0 = A_0 - \overline{m} < 0$, $B_m = \frac{1-\delta\rho_m}{1-\delta} A_m - 1 \leq 0$, $B_d = \frac{1-\rho_d\delta}{1-\delta} A_d < 0$, *and* $B_w = \frac{1-\rho_w\delta}{1-\delta} A_w > 0$. *Moreover, we have* $|B_d| > |\kappa_x B_w|$ *as long as either* $\rho_d \leq \rho_w$ *or* $0 \leq \gamma < \bar{\gamma}$ *for some* $\bar{\gamma} > 0$ *constant given in the Appendix.*

Besides describing the directional impact of the three types of shocks on returns, Proposition 4 also highlights that, for a wide range of parameter values, shifts in demand alter $\mathrm{E}_t[r_{t+1}^{Spread}]$ by more than shifts in supply of comparable size. Specifically, we compare the respective effects of a demand and a supply shift that move the two sides of (6) by the same amount, which corresponds to a comparison between $|B_d|$ and $|\kappa_x B_w|$.

Intuitively, note that demand shifts move both compensation for using scarce intermediary capital and compensation for swap spread risk in the same direction. By contrast, supply shifts move these two components of expected returns to swap arbitrage in opposite directions. For instance, a negative shift in intermediary capital reduces the magnitude of their position, thereby decreasing compensation for holding swap spread risk. This partly offsets the increased shadow cost of the now scarcer capital.[26] As a result, we can show that demand shifts have a larger effect than supply shifts, provided that demand and supply shifts have similar persistence or, alternatively, provided that demand is not too sensitive to the swap spread level.[27] In the special case when $\gamma = 0$, we simply have $|B_d| - |\kappa_x B_w| = \alpha V > 0$. In this case, inelastic demand determines the equilibrium position of intermediaries in the swap arbitrage trade. As a result, demand shifts alter both the compensation that intermediaries require for using scarce capital and compensation for the amount of swap spread risk they hold, whereas supply shifts only alter compensation for using scarce capital while the amount of the swap spread risk held by intermediaries and the compensation they require for it do not change. This intuition carries through as long as the demand-sensitivity parameter γ is not too large.[28]

[26]Formally, combining (14) and (19), we have

$$\frac{\partial \mathrm{E}_t[r_{t+1}^{Spread}]}{\partial z_t^d} = \alpha\left[\left(\frac{\sigma_o\kappa_x}{\kappa_o}\right)^2 + V\right] \times \overbrace{\frac{\partial x_t}{\partial z_t^d}}^{-(1+\gamma A_d)<0} \quad \text{and} \quad \frac{\partial \mathrm{E}_t[r_{t+1}^{Spread}]}{\partial z_t^w} = \frac{\alpha\sigma_o^2\kappa_x}{\kappa_o^2} + \alpha\left[\left(\frac{\sigma_o\kappa_x}{\kappa_o}\right)^2 + V\right] \times \overbrace{\frac{\partial x_t}{\partial z_t^w}}^{-\gamma A_w<0}.$$

[27]The latter condition ensures that exogenous demand and supply shocks do not induce large endogenous demand adjustments in response to changes in the swap spread level.

[28]Proposition 4 also implies that when end-user demand to receive fixed does not depend on the swap spread, $\gamma = 0$, we have $B_m = 0$, and hence changes in the short-rate differential m_t do not alter $\mathrm{E}_t[r_{t+1}^{Spread}]$, even if they

3.3.4 Term structure

Our model focuses on modelling the long end of the swap spreads curve. In order to derive predictions for the entire term structure, in the Appendix we introduce a series of n-period zero-coupon swaps alongside the perpetual swap. We then derive the equilibrium n-period swap spreads, $s_t^{(n)}$, assuming that corresponding zero-coupon swaps are not traded in equilibrium—specifically, we consider intermediaries' first order conditions in the limit as end-user demand for the zero-coupon swaps goes to zero. In particular, we find that the one-period swap spread, $s_t^{(1)}$, which corresponds to the riskless one-period arbitrage, simply depends on the short rate differential and the shadow cost of intermediary capital: $s_t^{(1)} = m_t - \kappa_x \psi_t$.

Using the term structure extension of our model, we derive several properties of the swap spread term structure slope $s_t - s_t^{(1)}$. First, we have

Proposition 5 *The slope of the swap spread term structure is on average negative:*

$$\mathrm{E}[s_t - s_t^{(1)}] = -\alpha V \mathrm{E}\left[d_t\right] < 0.$$

Proposition 5 states that, when end-user demand to receive fixed rate is positive, long-maturity swap spreads are on average lower than short-maturity spreads. In particular, the model allows for on average positive short-maturity and on average negative long-maturity spreads. Negative term structure slope arises because long-maturity swaps arbitrage is subject to swap spread risk, for which intermediaries require compensation. By contrast, neither the short rate differential nor the shadow cost of capital committed to arbitrage contribute to the average term structure slope.

Because the term structure slope depends on the compensation for swap spread risk, it has predictive power for returns to swap spread arbitrage. We obtain the following result:

Proposition 6 *A higher slope of the swap term structure forecasts higher returns to swap arbitrage trade, i.e., running the regression*

$$r_{t+1}^{Spread} = \alpha + \beta_1 (s_t - s_t^{(1)}) + \epsilon_{t+1}$$

we find that $\beta_1 > 0$ as long as $0 \leq \gamma \leq \bar{\bar{\gamma}}$ for some $\bar{\bar{\gamma}} > 0$ constant given in the Appendix.

The Appendix provides a range of additional results on the term structure of swap spreads.

contribute to swap cash flow. This is because, when $\gamma = 0$, the short rate differential has no effect on end-user demand and thus on the equilibrium amount of swap spread risk intermediaries must hold. However, when $\gamma > 0$, movements in the short rate differential m_t induce changes in demand and change the equilibrium amount of swap spread risk held by intermediaries. In this case, for instance, an increase in m_t must be accompanied by a more negative $\mathrm{E}_t[r_{t+1}^{Spread}]$ to induce intermediaries to accommodate this demand by engaging in pay-fixed swap arbitrage trade.

3.3.5 Swap spreads and risk premia

In the model, equilibrium swap spreads depend on the expected return ($\bar{r}_o$) and the risk (σ_o^2) of the outside investment opportunity. We have the following result:

Proposition 7 *Holding σ_o^2 constant, $\partial A_0/\partial\left(\bar{r}_o/\sigma_o^2\right) < 0$: the swap spread s_t decreases while the intermediary position x_t increases in the outside investment opportunity risk-return tradeoff ratio $\bar{r}_o/\sigma_o^2$. In turn, holding $\bar{r}_o/\sigma_o^2$ constant, $\partial A_0/\partial\sigma_o^2 < 0$: the spread s_t decreases while the intermediary position x_t increases in the outside investment opportunity risk.*

The relationship between swap spreads and risk premia in Proposition 7 arises because intermediaries allocate their scarce capital between swap spread arbitrage and the outside risky investment opportunity. Thus, $\bar{r}_o/\sigma_o^2$, which is proportional to the optimal unconstrained position in the outside investment option, represents the opportunity cost of committing scarce capital to the swap arbitrage trade and determines the level of swap spreads. Note that the direction of the causality is from the risky asset market to the intermediaries, unlike in the models that rely on intermediaries being the marginal investor in the risky asset markets; see, for instance, He and Krishnamurthy (2013), Adrian et al. (2014), and He et al. (2017).

Note that fluctuations in the outside investment opportunity risk-return tradeoff ratio, $\bar{r}_o/\sigma_o^2$, would induce supply shifts similar to those stemming from fluctuations in intermediary capital w_t. Intuitively, fluctuations in both $\bar{r}_o/\sigma_o^2$ and w_t influence the shadow cost of intermediary capital ψ_t and hence the opportunity cost of supplying swaps to end users.[29]

4 Data

We use three main types of time series data in our analysis: data on swap spreads and other relevant market prices, data on primary dealers' Treasury positions, and data that we use to proxy for end-user demand for swaps. Our main dataset is weekly and runs from July 2001 to December 2020.

First, we obtain LIBOR, Fed Funds, and SOFR swap rates from Bloomberg and constant maturity Treasury rates from the Federal Reserve Board's H.15 Statistical Release. Naturally, we have data on LIBOR swap rates throughout our full sample whereas data on Fed Funds and SOFR swaps become available to us only in September 2012 and December 2018, respectively. We compute swap spreads as the difference between the swap rate and the constant maturity

[29]The Online Appendix presents a formal extension of the model in which $\bar{r}_{o,t} = \mathrm{E}_t[r^o_{t+1}]$ fluctuates exogenously over time. In this extension, we have

$$\psi_t = \psi\left(\bar{r}_{o,t}, w_t, |x_t|\right) = \max\left\{0, \alpha\sigma_o^2 \frac{\frac{\bar{r}_{o,t}}{\alpha\sigma_o^2} - \frac{1}{\kappa_o}w_t + \frac{\kappa_x}{\kappa_o}|x_t|}{\kappa_o}\right\}. \tag{22}$$

Treasury rate at the same maturity. In addition, we obtain 3-month LIBOR and the 3-month Treasury Eurodollar (TED) spread—the difference between 3-month LIBOR and the 3-month Treasury bill yield—from the Federal Reserve Bank of St. Louis' FRED database. The mid-market rate on 3-month general collateral repurchase agreements—i.e., 3-month loans collateralized by Treasuries—is taken from Bloomberg. Finally, we compute the τ-week holding period returns on a LIBOR swap arbitrage trade following Boyarchenko et al. (2020):

$$\mathrm{R}_{t,:t+\tau} = \sum_{h=0}^{\tau-1} \left[\frac{\left(\text{Swap spread}_{t+h} - (\text{LIBOR}_{t+h} - \text{Repo}_{t+h})\right)}{52} - \text{DV01}_{t+h} \times \Delta\text{Swap spread}_{t+h+1} \right],$$

where DV01 is the dollar value of a basis point.[30] The term in square brackets is the weekly holding period return from $t+h$ to $t+h-1$, which naturally is the sum of a "carry" component known at $t+h$ and a mark-to-market component depending on the change in spreads from $t+h$ to $t+h-1$. The τ-week holding period returns is just the sum of the next τ weekly returns.

We obtain a variety of other market prices. We obtain an estimate of the term premium component of 10-year Treasury yields from Adrian et al. (2013) which is available from New York Fed's website. Our measure of yield curve noise—the root mean-squared yield fitting-error obtained from using a Svensson (1994) curve to fit cross-section of Treasuries at a given point in time—is from Hu et al. (2013) and is accessed through Jun Pan's website. We retrieve the VIX SP500 option-implied equity volatility index and the VXTY 10-year Treasury futures option-implied volatility index are from the Chicago Board of Options Exchange. The 10-year U.S. sovereign credit default swap spread is retrieved from Bloomberg. We obtain data on deviations from covered interest rate parity (CIP) from Borio et al. (2018).

Next, we collect data on primary dealers' positions in Treasury securities. These data are from Form A of the FR2004 Primary Government Securities Dealers Reports, are available at a weekly frequency, and detail positions as of market-close on Wednesday of the relevant week. Form FR2004 underwent a significant revision in July 2001, which is why our main sample begins at that date. The net position is calculated as the difference between long and short outright Treasury positions at market value. We focus on nominal coupon-bearing Treasury securities—which are most likely to be tied to swap spread positions— and therefore exclude bills, TIPS, and FRNs. For additional analysis, we collect data from Form A of the FR2004 on primary dealers' positions in federal agency and government sponsored enterprise mortgage-backed securities, federal agency and government sponsored enterprise securities other than

[30] The DV01 for a T-year swap arbitrage trade is simply the T-year annuity factor divided by 10,000. In other words, assuming the fixed swap leg is paid semi-annually as is the case with LIBOR swaps, we have

$$\text{DV01}_t = \frac{\frac{1}{2}\sum_{i=1}^{2T} \exp(-y_t^{(i/2)} \frac{i}{2})}{10,000},$$

where $y_t^{(i/2)}$ is the $(i/2)$-year zero-coupon yield at time t. And, this annuity factor is identical to the modified duration of a par coupon bond.

mortgage-backed securities, and corporate debt securities. We also collected the amount of Treasury securities collateralizing financing agreements where primary dealers are the cash lender—their "securities in" through financing arrangements—from Form C of the FR2004. These financing arrangements include reverse repos, securities borrowing, and securities received as margin collateral. Finally, we obtain the value-weighted average of primary dealers' interest rate Value-at-Risk from Anderson and Liu (2021).

Moving to the determinants of the demand for receiving fixed in the swap market, the MBS dollar duration is calculated by multiplying the level of the Barclays U.S. MBS market value index by its modified duration, both available through Datastream. We use the pension underfunding measure of Klingler and Sundaresan (2019). The gross domestic public issuance of bonds by U.S. non-financial corporations is from the Federal Reserve Board.

5 Empirical evidence

5.1 Swap spreads and dealers' positions

Our model connects long-term swap spreads and intermediaries' positions in the swap arbitrage trade. In our empirical tests, we focus primarily on the 30-year LIBOR swap spreads. The 30-year swap spread corresponds most closely to the long-term swap spread in the model: intermediaries who engage in swap arbitrage at a 30-year maturity both consume scarce capital and they can suffer large short-term fluctuations in the mark-to-market value of their swap positions. Moreover, 30-year LIBOR swaps represent a major source of duration for end users in the swap market.

Unfortunately, data on intermediaries' positions in the swap arbitrage trade are not directly available. Since broker-dealers are critical intermediaries in the swap market who generally enter into swap agreements with end users and then take offsetting positions in Treasuries to hedge the resulting interest rate risk, we use primary dealers' net position in Treasuries to proxy for their position in swap arbitrage. To support this choice, we note that primary dealers' net Treasury positions do not appear to be tightly linked to their net exposure to interest rate risk: for instance, the correlation of primary dealers' net position changes and their interest rate book Value-at-Risk changes is only about 0.05 between 2001 and 2018 and 0.08 between 2009 and 2018.[31] This implies that the direct interest rate exposure from dealers' net position in Treasuries is offset, in large part, by positions in swaps and other fixed income instruments, suggesting that dealer's net Treasury positions largely hedge these positions in swaps and other fixed-income instruments.

Table 1 reports summary statistics for our data on 30-year swap spreads, primary dealers'

[31] We also note that the Value-at-Risk fell considerably after 2009 and stayed at a stable low level ever since. The dollar duration of primary dealers' interest rate book, a more direct measure of their interest rate risk exposure, is unfortunately not available to us.

net position in Treasuries, and several of the other time-series we use in our analysis. Our data here are weekly and runs from July 2001 to December 2020. We report summary statistics for the 2001-2008 and 2009-2020 subsamples. Figure 1 plots 30-years swap spreads alongside primary dealers' net Treasury position.

From Table 1 and Figure 1, we observe a simultaneous switch in the signs of the swap spreads and primary dealers' positions around 2009. The 30-year spread averaged 45 basis points between 2001 and 2008. Starting from 2009, the spread has been consistently negative in an apparent violation of the LoOP, averaging -25 basis points between 2009 and 2020.[32] In addition, the term structure of swap spreads steepened considerably in 2009. While the difference between the 30-year and the 10-year spreads was negligible prior to 2009, it averaged -30 basis points between 2009 and 2020, with the 30-year spread consistently more negative than the 10-year spread. As the 30-year swap spread turned negative, the primary dealers' position in Treasuries simultaneously switched from net short to net long.

The combination of negative long-maturity swap spread, downward sloping swap spread term structure, and dealers' long net position in Treasuries is in line with our model's regime in which dealers are constrained and supply fixed rate in the swap market; see Propositions 1 and 5. Thus, in what follows we test the model's more specific predictions pertaining to this regime using the series of swap spreads and dealers' net Treasury positions starting in 2009.

5.2 Demand and supply decomposition

We use a structural vector auto-regression (VAR) to disentangle the effects of end-user demand and intermediary supply on the level of 30-year swap spreads. We set-identify this structural VAR using the pure sign restrictions approach of Uhlig (2005).

We begin with the reduced-form representation of the structural VAR implied by our model:

$$\begin{bmatrix} \text{30y swap spread}_t \\ -\text{PD net position}_t \end{bmatrix} = \mathbf{c} + \sum_{l=1}^{L} \mathbf{C}_l \begin{bmatrix} \text{30y swap spread}_{t-i} \\ -\text{PD net position}_{t-i} \end{bmatrix} + \xi_t. \tag{23}$$

Equation (23) corresponds to our model, assuming there is no short rate differential ($m_t = 0$) and allowing for a general autoregressive lag structure (for parsimony, we have $L = 1$ in the model); see Proposition 2. We choose a lag length of $L = 2$ in equation (23) which minimizes the Akaike Information Criterion (AIC).[33] Naturally, the shocks ξ_t in this reduced-form VAR do not have a meaningful structural interpretation: they reflect a combination of the structural shocks to end-user demand and intermediary supply.

[32]Figure 1 also shows the 30-year OIS and the 30-year SOFR swap spreads. These two spreads are highly correlated with the 30-year LIBOR swap spread but are even more negative on average, making the apparent violation of the LoOP during the shorter period for which OIS and SOFR swap rates are available even more clear. Indeed, the short rate differential is smaller for the OIS swaps compared to the LIBOR swaps, while non-zero SOFR swap spreads by definition represent a violation of the LoOP.

[33]Nonetheless, our results are robust to alternate choices of L as well as to the inclusion of a time trend.

Following Proposition 2, we posit that this reduced-form representation results from the equilibrium relationship between 30-year swap spreads and intermediaries' position in the swap arbitrage trade measured by $x_t = -\text{PD net position}_t$ on the one hand, and the unobserved structural demand and supply factors on the other:

$$\begin{bmatrix} \text{30y swap spread}_t \\ -\text{PD net position}_t \end{bmatrix} = \mathbf{a} + \mathbf{A} \begin{bmatrix} \text{Demand}_t \\ \text{Supply}_t \end{bmatrix}, \tag{24}$$

combined with the dynamics of the structural factors

$$\begin{bmatrix} \text{Demand}_t \\ \text{Supply}_t \end{bmatrix} = \sum_{l=1}^{L} \mathbf{D}_l \begin{bmatrix} \text{Demand}_{t-i} \\ \text{Supply}_{t-i} \end{bmatrix} + \varepsilon_t, \tag{25}$$

where ε_t is the vector of orthogonal structural shocks each with unit variance—i.e., Var[ε_t]= $\mathbf{I}$.[34] Together, equations (24) and (25) imply that

$$\begin{bmatrix} \text{30y swap spread}_t \\ -\text{PD net position}_t \end{bmatrix} = \left(\mathbf{I} - \sum_{l=1}^{L} \mathbf{A}\mathbf{D}_l\mathbf{A}^{-1}\right)\mathbf{a} + \sum_{l=1}^{L} \mathbf{A}\mathbf{D}_l\mathbf{A}^{-1} \begin{bmatrix} \text{30y swap spread}_{t-i} \\ -\text{PD net position}_{t-i} \end{bmatrix} + \mathbf{A}\varepsilon_t,$$

which corresponds to equation (20) in the model.

We identify the structural demand and supply shocks in ε_t by imposing the sign restrictions from Proposition 2 on the matrix $\mathbf{A}$ which maps structural shocks ε_t into reduced-form shocks ξ_t. Specifically, we assume that a positive shock to end-user demand to receive fixed rate makes swap spreads more negative and increases the scale of intermediaries' short position in the swap arbitrage (pushing down $-\text{PD net position}_t$). By contrast, we assume that a positive shock to intermediary supply makes swap spreads less negative, while increasing intermediaries' short position in the swap arbitrage. In other words, we assume that the reduced-form shocks are related to the structural shocks by some matrix $\mathbf{A}$ whose four elements satisfy the following sign restrictions:[35]

$$\xi_t = \underbrace{\begin{bmatrix} - & + \\ - & - \end{bmatrix}}_{\mathbf{A}} \varepsilon_t. \tag{26}$$

We set identify $\mathbf{A}$ using the pure sign restrictions approach of Uhlig (2005). We briefly discuss the intuition for this identification approach. Since Var[ε_t]= $\mathbf{I}$—the structural shocks are

[34] The assumption that the demand and supply shocks are orthogonal is necessary for our identification approach and is useful for these shocks to have a straightforward structural interpretation. While imperfect, the assumption that demand and supply shocks are orthogonal is reasonable for the swap market since the institutions that receive the fixed rate—for instance, end users such as pension funds—and those that pay the fixed rate—for instance, intermediaries such as broker-dealers on the pay-fixed side—have very different investment objectives and investment horizons and face different institutional constraints.

[35] Note also that the sign restrictions apply only to the contemporaneous responses of spreads and positions to demand and supply shocks; we do not impose restrictions on the responses at horizons of one week or more. That said, the properties of the identified demand and supply point to two clearly distinct forces as discussed below.

assumed to be orthogonal and we are only attempting to identify their impact "up to scale"—the variance of reduced form shocks provides us with three restrictions on the four elements of $\mathbf{A}$, namely Var[ξ_t]= $\mathbf{A}\mathbf{A}'$. Of the set of matrices $\mathbf{A}$ that satisfy these three restrictions, the identified set then consists of those matrices that also satisfy the sign restrictions in equation (26). Following Fry and Pagan (2005, 2011) and Cieslak and Pang (2020), we select the value of $\mathbf{A}$ within this identified set for which instantaneous responses to structural shocks are the closest to the median response. We then invert equation (24) using this "closest to median" matrix, enabling us to estimate the latent demand and supply factors.[36]

In Appendix C, we consider a tri-variate VAR which includes the LIBOR-repo spread and corresponds to the version of our model with a non-zero short rate differential ($m_t \neq 0$). As explained in the Appendix, structural shocks in this VAR are identified using a combination of sign and zero restrictions implied by Proposition 2. We find that this short-rate differential explains a very small portion of the variation in 30-year swap spreads since 2009. Furthermore, including this short-rate differential does not have a significant effect on our estimates of the latent demand and supply factors. See Figure A2 in the Appendix. Indeed, the fact that short-rate differential plays a small role in explaining 30-year swap spreads is consistent with our model which predicts that transient fluctuations in m_t—for instance, due to a temporary rise in concerns about the creditworthiness of large banks—should play only a minor role in explaining movements in long-dated swap spreads. Consequently, we use the more parsimonious specification in (23) whose identification requires fewer restrictions as our baseline.

Figure 2 shows the historical decomposition of (de-meaned) swap spreads in equation (24) implied by our VAR. As shown in Figure 2, fluctuations in end-user demand and in intermediary supply both play an important role in explaining variation in swap spreads over time. For instance, our estimates suggest that an inward shift in intermediary supply starting from the second half of 2014 helped push spreads well into negative territory. This inward supply shift coincides with a series of regulatory changes that arguably increased the balance-sheet costs faced by intermediaries, including the implementation of the supplementary leverage ratio through 2014 and the introduction of the Volker rule in July 2015; see also Boyarchenko et al. (2020). Then, in 2016, our estimates suggest that rising end-user demand to receive the fixed rate pushed swaps even further below zero.

The forecast error variance decomposition, which quantifies the relative contribution of structural shocks to the residual variation in swap spreads at different horizons, confirms that both demand and supply shocks play an important role. Demand and supply shocks each explain roughly 50% of the residual variation in 30-year swap spread at a weekly horizon. Our finding that both demand and supply shocks play an important role in driving swap spreads

[36]The "closest to median" value of $\mathbf{A}$ is preferred to a simple median because it corresponds to a particular VAR model in the identified set. By contrast, the median $\mathbf{A}$ mixes different models and lacks a structural interpretation. We note that demand and supply factors implied by the "closest to median" value of $\mathbf{A}$ are effectively identical to the median factors and thus satisfy a basic check on our identification suggested by Fry and Pagan (2005, 2011). See Figure A1.

contrasts with Goldberg and Nozawa (2021), who find that, in the corporate bond market, intermediary supply shocks play a dominant role in driving variation in market liquidity. Interestingly, the contribution of supply shocks increases to 65% at an annual horizon. Supply plays a more important role at longer horizons because our estimates imply that the latent supply factor is more persistent than the demand factor.[37]

Columns (1), (3), and (5) in Table 2 shows that demand and supply factors capture two distinct aspects of the variation in swap spreads. The columns report the regressions of swap spreads, demand factor, and supply factor on a range of variables that capture market conditions, namely the Adrian et al. (2013) estimate of the term premium component of 10-year Treasury yields, the VXTY Treasury yield option-implied volatility index, the 10-year U.S. sovereign credit default swap spread, the Hu et al. (2013) yield curve noise measure capturing the scarcity of intermediary capital in the cash Treasury market, and the VIX option-implied equity volatility index. We find that higher Treasuries term premium is associated with higher (less negative) level of swap spreads, and this through a negative relationship between the term premium and the demand for swaps by end users. In turn, higher implied volatility of Treasury yields is associated with lower (more negative) level of swap spreads, and this through a negative relationship between implied volatility and the supply for swaps by intermediaries. In addition, higher U.S. sovereign credit default swap spreads are associated with higher demand and lower (more negative) level of swap spreads. This is consistent with lower demand for Treasuries relative to interest rate swaps when investors perceive the sovereign risk of Treasuries to be elevated. That said, the sovereign risk channel accounts for only a small fraction of the variation in swap spreads, with univariate R^2 of 0.01.

We note that elevated risk premia—as proxied either by the term premium or by the implied volatility— seem to be associated with lower-than-normal end-user demand to receive the fixed swap rate (which makes spreads less negative) and lower-than-normal intermediary supply (which makes spreads more negative). In other words, market conditions appear to have a starkly different effect on swap spreads depending on whether they influence the demand or the supply of swaps. This different effect is also confirmed by looking at pairwise correlations of the demand and supply factors with variables capturing market conditions reported in Appendix Table A1. For a meaningful comparison, the supply factor is multiplied by minus one so higher values of both factors correspond to more negative swap spreads. Higher values of risk premia and intermediary funding illiquidity measures tend to be associated with lower values of our estimated demand factor and higher values of our inverse supply factor. As a result, the comovement of swap spreads with market conditions may be ambiguous and depend on whether the associated inward demand shift or the inward supply shift dominates.

The observed negative correlation between the intermediary supply of swaps and the level

[37] As we will see shortly, the estimated demand factor appears to be closed linked to mortgage-related hedging flows and these mortgage hedging flows are themselves thought to be fairly transient in nature; see Hanson (2014) and Malkhozov et al. (2016).

of risk premia outside the swap market is in line with our model. In particular, according to Proposition 7, when the expected return on intermediaries' outside investment opportunity ($\bar{r}_o$) increases, the shadow value of intermediary capital (ψ_t) increases, leading to a decline in both the equilibrium level of swap spreads and in the size of intermediaries' short position in the swap arbitrage trade. Thus, if we allowed the expected return on this outside investment to be time-varying, positive shocks to outside risk premia would be equivalent to negative shocks to intermediary capital, which naturally decrease supply.[38] Note that this relationship between dealers' supply and risk premia outside of the swap market does not rely on the idea that swap intermediaries are also marginal investors in the market for these outside risky assets. This contrasts to more "integrated-market" theories of intermediary-based asset pricing—including He and Krishnamurthy (2013), Adrian et al. (2014), and He et al. (2017)—where the exact same set of intermediaries are assumed to be marginal in very broad array of financial asset classes. Instead, the outside risk premium in our model simply impacts the opportunity cost of committing scarce balance sheet to the swap arbitrage.

Our model is silent about the drivers of end-user demand to receive the fixed swap rate. In fact, an advantage of the identification strategy based on sign restrictions is that we do not need to specify proxies for demand drivers or, for that matter, supply drivers ex ante. That said, it is instructive to investigate which of the candidates proposed in the literature have a stronger association with our demand factor.

First, Hanson (2014), Malkhozov et al. (2016) , and TBAC (2021) argue that the desire of hedged mortgage investors and mortgage servicers to receive the fixed swap rate rises when interest rates fall. The idea is that declines in interest rates increase expected prepayments, leading to a decline in the duration of outstanding mortgage-backed securities (and mortgage services rights). And, mortgage investors then want to receive the fixed swap rate to add back duration to their portfolios. Thus, we would expect that end-user demand to receive fixed will be negatively related to mortgage duration.

Second, insurers and pension funds receive fixed to manage the gap between the duration of their liabilities and their preferred mix of on-balance sheet assets. And, Klingler and Sundaresan (2019) argue that the pension's incentives to receive fixed increases when pension funds become more underfunded.

Finally, the bond issuance of corporates, whom TBAC (2021) also identifies as receivers of fixed, could influence their demand for swaps.

Columns (2) and (4) in Table 2 reports the regressions of 3-month changes in swap spreads and our estimated demand factor on changes in MBS dollar duration, pension underfunding measure, and corporate bond issuance. The results in Table 2 suggest that MBS investors are an important driver of the swap spreads and, more specifically, of end-user demand to receive fixed, echoing earlier findings in Feldhütter and Lando (2008) and Hanson (2014). Illustrating this finding, Figure 3 shows the time series of MBS dollar duration and the demand factor, and

[38] Appendix presents a model extension with a time-varying outside investment opportunity expected return.

the patterns are almost a mirror images of each other.

Several additional comments are in order. First, we expect the explanatory power of investor variables to be stronger for the demand factor compared to the swap spread itself. This is because the latter reflects the interplay of both demand and supply. We confirm this for our demand factor. Second, the preponderance of the MBS variable in the regressions does not imply the exclusive role of MBS investors in the swap market. Indeed, Domanski et al. (2017) argue that, similar to MBS investors, insurers and pension funds also buy duration when interest rates fall in order to offset the increasing mismatch between their assets and liabilities. Thus, MBS dollar duration may capture the demand of insurers and pension funds in addition to MBS investors' demand. Finally, the demand by MBS investors can explain the negative correlation between the demand factor and the term premium reported in Table 2 and illustrated on Figure 3. Indeed, Hanson (2014) and Malkhozov et al. (2016) show that MBS dollar duration predicts Treasury returns. In turn, we find that MBS duration is negatively related to the demand for receiving fixed in the swap market.

5.3 Arbitrage risk

We now use our model's predictions to gauge the importance of arbitrage risk for swap spreads.

Proposition 3 states that expected holding period returns to swap arbitrage is comprised of two terms: the cost of committing the balance sheet to swap arbitrage over the holding period and the premium for arbitrage risk proportional to dealers' aggregate swap position. In other words, quantities should predict returns to swap arbitrage even after controlling for the short-term arbitrage spreads. We test this prediction by regressing returns to swap arbitrage $\mathrm{R}_{t,t+\tau}$ on the primary dealers' net Treasuries position and the repo-OIS spread, a measure of short-term balance sheet cost:

$$\mathrm{R}_{t,\,t+\tau} = \beta_0 + \beta_1 \mathrm{PD\ net}_t + \beta_2\ \text{Repo-OIS}_t + \beta_c^{\mathsf{T}} \mathrm{Controls}_t + \epsilon_{t,\,t+\tau}.$$

As reported in Table 3, positions strongly predict returns at both τ =3-month and τ =12-month horizons even after controlling for the short-term balance sheet cost. The negative sign of the estimated coefficient β_1 implies that larger short positions in the swap arbitrage trade correspond to higher expected returns on this short position, in line with our model prediction.[39]

Institutions other than primary dealers — for instance, hedge funds — can also engage in the

[39]Note that here we consider returns on swap arbitrage trades implemented with par-coupon swaps and Treasuries. We have also examined the return on swap arbitrage trades involving zero-coupon swaps and Treasuries, which involves fitting zero-coupon curves for these two sets of instruments. We generally find noticeably stronger forecasting results for returns on spread trades involving long-dated zero-coupon instruments than spread trades involving long-dated par instruments. This is intuitive since (1) movements in more distant forward rates primarily reflect variation in expected future returns as opposed to expected future interest rate differentials and (2) par-coupon rates place less weight on more distant forward rates whereas zero-coupon rates equally weight all forward rates over an instrument's life. Nonetheless, we focus on trades involving par instruments in the main text since these trades are far more common in practice and so as to sidestep issues related to yield curve fitting.

swap arbitrage. To account for this, we consider collateralized Treasuries financing extended by primary dealers to other financial institutions — also referred to as “securities in” — which would allow the latter to enter the swap arbitrage:

$$\mathrm{R}_{t,\,t+\tau} = \beta_0 + \beta_1 \text{PD net}_t + \beta_1' \text{PD sec. in.}_t + \beta_2 \text{Repo-OIS}_t + \beta_c^{\mathsf{T}} \text{Controls}_t + \epsilon_{t,\,t+\tau}$$

This additional measure of positions also predicts swap arbitrage returns, albeit less strongly than primary dealers’ own position. The R^2 of the predictive regression for 3-month returns which includes both primary dealers’ own position and their financing operations reaches approximately 15%.

Table 4 reports several additional results. First, we confirm that primary dealers’ position predicts swap arbitrage returns when we use the TED spread instead of the repo-OIS spread as a measure of short-term balance sheet cost. Second, we find that primary dealers’ net position in Treasuries with maturities of 11 years and higher — closest to the maturity of the 30-year swaps — on their own predict swap arbitrage returns. Third, we note that primary dealers’ gross position in Treasuries has considerably less predictive power compared to the net position. This lends further support for the arbitrage risk channel: whereas the gross position measures the commitment of primary dealers’ balance sheet, the net position captures their exposure to the arbitrage risk. Finally, we find that primary dealers’ positions in agency MBS, agency securities other than MBS, and corporate debt securities have little incremental predictive power for swap arbitrage returns.

Next, Proposition 4 states that, in presence of arbitrage risk, demand should be a stronger predictor of the swap spread trade returns than supply. This is because negative supply shocks increase intermediaries’ balance sheet cost but, at the same time, reduces their exposure to arbitrage risk; in contrast, demand shocks increase both the balance sheet cost and the arbitrage risk exposure. To test this prediction, we regress swap arbitrage returns on the demand and supply factors identified in Section 5.2:

$$\mathrm{R}_{t,\,t+\tau} = \beta_0 + \beta_1 \text{Demand}_t + \beta_2 \text{ Supply}_t + \beta_c^{\mathsf{T}} \text{Controls}_t + \epsilon_{t,\,t+\tau}$$

As reported in Table 3, the demand factor strongly predict returns at both 3-month and 12-month horizons. The negative sign of the estimated coefficient β_1 implies that higher demand for receiving fixed corresponds to higher expected returns on the short position in the swap arbitrage trade, in line with our model prediction. In contrast, the supply factor is not statistically significant. The R^2 of the predictive regression for 3-month returns reaches approximately 15%, similar to that of the regression with primary dealer positions.

When we extend our sample up to end-2020, primary dealers’ net position is no longer significant in the return predictive regressions; see Appendix Table A2. The period starting from mid-2018 is characterized by the Tax Cuts and Jobs Act-related Treasuries issuance and the Fed balance sheet unwinding, and later in the sample by the exceptional fiscal and monetary policy response to the onset of COVID-19 pandemic. Both Treasuries issuance and changes in the Fed Treasuries holdings have an effect on primary dealers' inventory of Treasuries; see, for instance, Fleming and Rosenberg (2008). This effect is unrelated to the mechanism studied in our paper, which can explain the poor performance of primary dealers' position in the extended sample. Indeed, as also reported in Appendix Table A2, orthogonalizing primary dealers' net position with respect to issuance and Fed holdings, we recover significance in the extended with the same signs of the coefficients as in our baseline regressions.

Finally, Propositions 5 and 6 rely only on the information in the swap spreads and not the information in dealers' positions. Testing these propositions allows us to further support the importance of arbitrage risk, including in the later part of the sample. In line with Proposition 5, we observe that the term structure of swap spreads is on average downward-sloping through our sample starting from 2009. In particular, the 30-year spread remained consistently more negative than the 10-year spread. Recall that in the model the on average downward-sloping term structure reflects the premium required by intermediaries for holding swap spread risk. Next, we test Proposition 6 using a version of the Fama-Bliss regression. Specifically, we regress returns to swap arbitrage trade on the slope of the swap spread term structure defined as the difference between the 30-year and the 10-year swap spreads:

$$\mathrm{R}_{t,\,t+\tau} = \beta_0 + \beta_1\left(\text{30y swap spread}_t - \text{10y swap spread}_t\right) + \beta_2 \text{Repo-OIS}_t + \beta_c^{\intercal}\text{Controls}_t + \epsilon_{t,\,t+\tau}.$$

As reported in Table 3, the slope is a significant predictor of swap arbitrage returns over the entire period between 2009 and 2020.[40] The positive sign of the estimated coefficient β_1 implies that a more negative slope corresponds to higher expected returns on the short position in the swap arbitrage trade.

Additionally, we find that primary dealers' positions in Treasuries and swap spread term structure slope predict returns to swap arbitrage prior to the GFC, albeit somewhat less strongly than in our baseline sample; see Appendix Table A3. Interestingly, in contrast to our baseline sample, we find that primary dealers' positions in agency MBS have predictive power for swap arbitrage returns in addition to their positions in Treasuries. This is consistent with a larger importance of MBS in investors' portfolios prior to the large scale purchases of MBS by the Federal Reserve in the aftermath of the GFC.

5.4 Term structure

Our analysis so far focused on the long end of the swap spread curve, and this for several reasons. First, long-maturity swap arbitrage is subject to more swap spread risk and, as a result, long-maturity swap spreads are more informative about the arbitrage risk channel. Second,

[40] We find very similar results when using alternative measures of the slope.

additional factors, distinct from those most relevant for long-maturity swaps, may influence the end-user demand for shorter-maturity swaps. Finally, while the SLR applies to all positions regardless of their risk, other capital requirements—for instance, leverage ratio and margin requirements for swaps—may be less stringent for shorter-maturity instruments, with the 5-year maturity being a typical cut-off; see Boyarchenko et al. (2020).

That said, it is nonetheless instructive to see whether the properties of the entire term structure of swap spreads are in line with our model predictions. First, Figure 4 shows the average term structures of LIBOR, OIS, and SOFR spreads. Starting from 2009 or from the date for which data become available, the term structures are on average downward-sloping. Further, average OIS and SOFR swap spreads are lower compared to the LIBOR swap spreads of the same maturity. In particular, at the short end, the 2-year LIBOR swap spreads are on average positive, whereas the 2-year OIS and SOFR swap spreads are on average negative. These patterns fit well with our model. Indeed, intermediaries require a negative swap spread as compensation for committing their scarce capital to a short position in the swap arbitrage and for bearing the risk associated with this position. Thus, our model predicts unambiguously negative average OIS and SOFR swap spreads. Relative to those, the difference between LIBOR and repo rates pushes the average level of the LIBOR swap spread towards positive territory. In addition, arbitrage risk and associated risk premium is larger for longer-maturity swaps, resulting in downward-sloping average term structures.[41]

Next, we verify that primary dealers' positions, demand and supply factors, and swap spread term structure slope predict returns to swap arbitrage if we consider 10-year swap spreads instead of 30-year swap spreads. As reported in Appendix Table A4, results for 10-year maturity are in line with those for the 30-year maturity. At the same time, consistent with 30-year swap arbitrage being subject to more swap spread risk compared to the 10-year swap arbitrage, we find that predictive variables are more strongly significant for returns to 30-year swap arbitrage compared to 10-year swap arbitrage.

Finally, we consider maturity-specific demand factors. Our model does not provide a precise way to disentangle multiple demand factors. As a result, the analysis that follows is more exploratory in nature. Specifically, we assume that end-user demand to receive fixed rate has two separate components—demand for short-maturity swaps and demand for long-maturity swaps—and, importantly, that intermediaries who accommodate the end-user demand for swaps hedge the resulting interest rate risk using Treasuries primarily in the maturity range close to that of the swap they supply. As explained in the Appendix C, we use these additional assumptions to identify short-maturity demand, long-maturity demand, and supply shocks in a structural VAR that includes 5-year swap spread, 30-year swap spread, primary dealers' position in Treasuries with maturities of up to 6 years, and primary dealers' position in Treasuries

[41] In the Appendix we generalize Proposition 5 to the "local" slope $s_t^{(n)} - s_t^{(n-1)}$ and provide a range of additional technical results for the term structure of swap spreads.

with maturities of more than 6 years.[42] The identified short-maturity demand shock, which by construction has a larger effect on primary dealers' position in Treasuries with shorter maturities, explains 39% of the variation in the 5-year swap spread and only 6% of the variation in the 30-year swap spread at the one-year horizon. By contrast, the long-maturity demand shock, which by construction has a larger effect on primary dealers' position in Treasuries with longer maturities, explains only 3% of the variation in the 5-year swap spread and 70% of the variation in the 30-year swap spread at the one-year horizon. To shed light on the properties of short-maturity and long-maturity demand for swaps, Appendix Table A5 reports the regressions of 3-month changes in our estimated short-maturity and long-maturity demand factors on changes in MBS dollar duration, Klingler and Sundaresan (2019) pension underfunding measure, and corporate bond issuance. Our results suggest that MBS investors drive the demand for both short-maturity and long-maturity swaps. In addition, pension underfunding plays a role for long-maturity swap demand.

5.5 Relationship to other arbitrages

We now test whether negative swap spreads are related to LoOP violations in other markets and whether it is end-user demand or intermediary supply that account for any such relationship. Specifically, we consider the deviations from covered interest parity (CIP) documented in Du et al. (2018). Similar to negative swap spreads, large deviations from CIP appeared in the aftermath of the GFC. Also similar to swap spreads, the CIP deviations have a term structure dimension, with different magnitude of deviations observed at different maturities.

In Table 5, we regress the CIP deviations at, respectively, 3-month and 5-year maturities τ on swap demand and swap supply factors identified in Section 5.2:[43]

$$\text{CIP}_t^{\tau} = \beta_0 + \beta_1 \text{Demand}_t + \beta_2 \text{ Supply}_t \ + \beta_c^{\top} \text{Controls}_t + \epsilon_t$$

First, we note that the 30-year swap spread itself is correlated with both 3-month and 5-year CIP deviations. However, there is a stark contrast between the drivers of this relationship at different maturities. At the short end, only the swap supply factor is statistically significant, whereas the opposite is generally true at the long end. Interestingly, demand and supply remain strongly significant in respective regressions even after we control for the CIP at the other maturity (5-year in the 3-month CIP regression and vice versa). This finding points to a multi-factor CIP term structure and suggests that swap demand and supply contain non-trivial information about this term structure.

The relationship between 3-month CIP deviation and the supply factor is perhaps not surprising as the latter captures the tightness of intermediaries' balance sheet constraints. In turn, the relationship between long-maturity CIP deviations and the demand factor could reveal a

[42] The choice of maturity buckets is determined by data availability.

[43] In choosing 3-month and 5-year CIP maturities we follow Du et al. (2018).

possible correlation between the demand for interest rate swaps and the demand for FX derivatives. Alternatively, large exposure to swap spread risk resulting from the higher demand for swaps could make dealers less willing to take on correlated long-maturity CIP deviation risk, leading to larger CIP deviations at the long end. Whereas the latter explanation is in line with our model and specifically with the arbitrage risk channel, a formal analysis of the demand for FX derivatives is outside the scope of this paper.

6 Conclusion

We find that both end-user demand and intermediary supply play an important yet starkly distinct role in explaining the variation in long-maturity swap spreads. In considering demand and supply forces jointly, our work speaks to the different nature of observed episodes of financial market dislocations, as argued by policymakers and academics, such as during the GFC and at the onset of COVID-19 pandemic in early 2020.

In addition to the interplay between the demand and the supply forces, we show that arbitrage risk is a key determinant of longer-maturity LoOP violations. Indeed, our model highlights that once LoOP violations arise because of limits to arbitrage, they are amplified by the presence of arbitrage risk. We also provide empirical support for all of these channels.

Our framework does not only speak to negative swap spreads, but can also be applied to other limited arbitrage settings—for instance, to deviations from the covered interest parity (CIP). We leave this for future research.

References

Adrian, T., R. K. Crump, and E. Moench (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics 110*(1), 110–138.

Adrian, T., E. Etula, and T. Muir (2014). Financial Intermediaries and the Cross-Section of Asset Returns. *The Journal of Finance 69*(6), 2557–2596.

Anderson, C. and W. Liu (2021). Inferring intermediary risk exposure from trade. Working paper, Federal Reserve Board.

Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica 86*(2), 685–720.

Begenau, J., M. Piazzesi, and M. Schneider (2020). Banks' risk exposures. Working paper, Stanford University.

Borio, C., R. N. McCauley, P. McGuire, and V. Sushko (2018). The failure of covered interest parity: Fx hedging demand and costly balance sheets. BIS Working Papers 590, Bank for International Settlements.

Boyarchenko, N., T. M. Eisenbach, P. Gupta, O. Shachar, and P. Van Tassel (2020). Bank-intermediated arbitrage. Staff Report 858, Federal Reserve Bank of New York.

Boyarchenko, N., P. Gupta, N. Steele, and J. Yen (2018). Negative swap spreads. *Federal Reserve Bank of New York Economic Policy Review 24*(2), 1–14.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies 22*(6), 2201–2238.

Chen, H., S. Joslin, and S. X. Ni (2018). Demand for crash insurance, intermediary constraints, and risk premia in financial markets. *Review of Financial Studies 32*(1), 228–265.

Chernenko, S. and M. Faulkender (2011). The two sides of derivatives usage: Hedging and speculating with interest rate swaps. *Journal of Financial and Quantitative Analysis 46*(6), 1727 – 1754.

Cieslak, A. and H. Pang (2020). Common shocks in stocks and bonds. *Journal of Financial Economics forthcoming*.

Cohen, L., K. B. Diether, and C. J. Malloy (2007). Supply and demand shifts in the shorting market. *Journal of Finance 62*(5), 2061–2096.

Collin-Dufresne, P. and B. Solnik (2001). On the term structure of default premia in the swap and libor markets. *Journal of Finance 56*, 1095–1115.

Cortes, F. (2003). Understanding and modelling swap spreads. Quarterly bulletin 2003 q4, Bank of England.

De Long, B. J., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Noise trader risk in financial markets. *Journal of Political Economy 98*(4), 703–738.

Domanski, D., H. S. Shin, and V. Sushko (2017). The hunt for duration: Not waving but drowning? *IMF Economic Review 65*(1), 113–153.

Drechsler, I., A. Savov, and P. Schnabl (2021). Banking on deposits: Maturity transformation without interest rate risk. *Journal of Finance 76*(3), 1091–1143.

Driscoll, J. C. and R. A. Judson (2013). Sticky deposit rates. Finance and economics discussion series, Board of Governors of the Federal Reserve System.

Du, W., A. Tepper, and A. Verdelhan (2018). Deviations from Covered Interest Rate Parity. *Journal of Finance 73*(3), 915–957.

Duarte, J., F. A. Longstaff, and F. Yu (2006, 07). Risk and return in fixed-income arbitrage: Nickels in front of a steamroller? *Review of Financial Studies 20*(3), 769–811.

Duffie, D. (2017). Post-crisis bank regulations and financial market liquidity. Technical report, Banca d'Italia.

Duffie, D. (2020). Still the world's safe haven? Working paper, Hutchins Center.

Duffie, D. and K. J. Singleton (1997). An econometric model of the term structure of interest-rate swap yields. *Journal of Finance 52*, 1287–1321.

Feldhütter, P. and D. Lando (2008). Decomposing swap spreads. *Journal of Financial Economics 88*, 375–405.

Fleming, M. J. and J. V. Rosenberg (2008). How do treasury dealers manage their positions? Technical report.

Frame, W. S., A. Fuster, J. Tracy, and J. Vickery (2015). The rescue of fannie mae and freddie mac. *Journal of Economic Perspectives 29*(2), 25–52.

Fry, R. and A. Pagan (2005). Some issues in using vars for macroeconometric research. Technical report.

Fry, R. and A. Pagan (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature 49*(4), 938–960.

Gârleanu, N. and L. H. Pedersen (2011). Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies 24*, 1980–2022.

Goldberg, J. and Y. Nozawa (2021). Liquidity supply in the corporate bond market. *Journal of Finance 76*(2), 755–796.

Gourinchas, P.-O., W. Ray, and D. Vayanos (2020). A preferred-habitat model of term premia and currency risk. Working paper, London School of Economics.

Greenwood, R., S. G. Hanson, J. C. Stein, and A. Sunderam (2017). Strengthening and streamlining bank capital regulation. *Brookings Papers on Economic Activity 48*(2), 479–565.

Greenwood, R., S. G. Hanson, J. C. Stein, and A. Sunderam (2020). A quantity-driven theory of term premiums and exchange rates. Working paper, Harvard Business School.

Greenwood, R. and D. Vayanos (2014). Bond supply and excess bond returns. *Review of Financial Studies 27*(3), 663–713.

Hanson, S. G. (2014). Mortgage convexity. *Journal of Financial Economics 113*(2), 270–299.

Hanson, S. G., A. K. Kashyap, and J. C. Stein (2010). A macroprudential approach to financial regulation. *Journal of Economic Perspectives 25*(1), 3–28.

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

He, Z. and A. Krishnamurthy (2013). Intermediary Asset Pricing. *American Economic Review 103*(2), 732–770.

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

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

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

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

Lang, L. H., R. H. Litzenberger, and A. L. Liu (1998). Determinants of interest rate swap spreads. *Journal of Banking and Finance 22*, 1507–1532.

Liu, J., F. A. Longstaff, and R. E. Mandell (2006). The market price of risk in interest rate swaps: The roles of default and liquidity risks. *Journal of Business 79*, 2337–2359.

Loewenstein, M. and G. A. Willard (2006). The limits of investor behavior. *Journal of Finance 61*(1), 231–258.

Lowenstein, R. (2000). *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. New York: Random House.

Malkhozov, A., P. Mueller, A. Vedolin, and G. Venter (2016). Mortgage risk and the yield curve. *Review of Financial Studies 29*(5), 1220–1253.

Perli, R. and B. P. Sack (2003). Does mortgage hedging amplify movements in long-term interest rates? *Journal of Fixed Income 13*(3), 7 – 17.

Siriwardane, E., A. Sunderam, and J. Wallen (2021). Segmented arbitrage. Working paper, Harvard Business School.

Spiegel, M. (1998). Stock price volatility in a multiple security overlapping generations model. *Review of Financial Studies 11*, 419–447.

TBAC (2021, February). Treasury borrowing advisory committee report to treasury. Technical report.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics 52*, 381–419.

Vayanos, D. and J.-L. Vila (2021). A preferred-habitat model of the term structure of interest rates. *Econometrica 89*, 77–112.

Vissing-Jorgensen, A. (2021). The treasury market in spring 2020 and the response of the federal reserve. *Journal of Monetary Economics 124*, 19–47.

A Proofs and derivations

Proof of Theorem 1. We write the swap spread as

$$s_t = A_0 + \mathbf{a}_s'\mathbf{z}_t$$

where $\mathbf{a}_s = [A_m, A_d, A_w]'$. The generalization of our affine equilibrium to a generic VAR(1) data generating process ($\mathbf{z}_{t+1} = \varrho\mathbf{z}_t + \varepsilon_{t+1}$ with ϱ and $\mathrm{Var}_t[\varepsilon_{t+1}] = \mathbf{\Sigma}$ potentially non-diagonal), yield the fixed-point condition

$$\mathbf{a}_s = \mathbf{F}(\mathbf{a}_s) \equiv [(1-\delta)^{-1}(\mathbf{I} - \delta\varrho') + \gamma\alpha\left(V(\mathbf{a}_s) + \left(\frac{\kappa_x}{\kappa_o}\right)^2\right)\sigma_o^2\mathbf{I}]^{-1}[\mathbf{e}_m - \alpha\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2\left(\mathbf{e}_d - \frac{1}{\kappa_x}\mathbf{e}_w\right) - \alpha V(\mathbf{a}_s)\mathbf{e}_d], \tag{A-1}$$

where $\mathbf{I}$ denotes the 3×3 identity matrix, $\mathbf{e}_m = (1,0,0)^T$, $\mathbf{e}_d = (0,1,0)^T$, $\mathbf{e}_w = (0,0,1)^T$,

$$V(\mathbf{a}_s) = \left(\frac{\delta}{1-\delta}\right)^2 \mathbf{a}_s'\mathbf{\Sigma}\mathbf{a}_s.$$

Letting $\mathbf{a}_s^* = \mathbf{F}(\mathbf{a}_s^*)$ denote a solution to this fixed-point problem, we then have

$$A_0^* = \frac{\overline{m} - \alpha\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2\left[\frac{\kappa_o}{\kappa_x}\frac{\bar{r}_o}{\alpha\sigma_o^2} + \bar{d} - \frac{\overline{w}}{\kappa_x}\right] - \alpha V(\mathbf{a}_s^*)\bar{d}}{1 + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V(\mathbf{a}_s^*)\right]}. \tag{A-2}$$

In the particular diagonal case, (A-1) and (A-2) simplify to the equations (17b)-(18).

To understand the relevant cases for equilibrium existence and multiplicity, consider first $\gamma = 0$. Equations (17b)-(18) then simplify to

$$A_m = \frac{1-\delta}{1-\rho_m\delta} > 0, \tag{A-3a}$$

$$A_d = -\frac{1-\delta}{1-\rho_d\delta}\alpha\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right] \leq 0, \text{ and} \tag{A-3b}$$

$$A_w = \frac{1-\delta}{1-\rho_w\delta}\frac{1}{\kappa_x}\alpha\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 \geq 0, \tag{A-3c}$$

where $V = \left(\frac{\delta}{1-\delta}\right)^2\left(A_m^2\sigma_m^2 + A_d^2\sigma_d^2 + A_w^2\sigma_w^2\right)$. Thus, A_m and A_w are known constants and we only need to solve for A_d, which in turn is the solution to the following quadratic equation:

$$0 = \left(\frac{\delta}{1-\delta}\sigma_d\right)^2 \cdot A_d^2 + \frac{1}{\alpha}\frac{1-\rho_d\delta}{1-\delta}\cdot A_d + \left[\left(\frac{\kappa_x}{\kappa_o}\sigma_o\right)^2 + \left(\frac{\delta}{1-\delta}\right)^2\left(A_m^2\sigma_m^2 + A_w^2\sigma_w^2\right)\right].$$

Assuming we are not in the deterministic demand case $\sigma_d^2 = 0$, which would naturally impose $A_d = 0$, we obtain two solutions, given by

$$A_d = \frac{-\frac{1}{\alpha}\frac{1-\rho_d\delta}{1-\delta} \pm \sqrt{\left(\frac{1}{\alpha}\frac{1-\rho_d\delta}{1-\delta}\right)^2 - 4\left(\frac{\delta}{1-\delta}\sigma_d\right)^2\left[\left(\frac{\kappa_x}{\kappa_o}\sigma_o\right)^2 + \left(\frac{\delta}{1-\delta}\right)^2\left(A_m^2\sigma_m^2 + A_w^2\sigma_w^2\right)\right]}}{2\left(\frac{\delta}{1-\delta}\sigma_d\right)^2},$$

will only exists as long as α is sufficiently small relative to σ_d^2, namely as long as

$$1 - \rho_d\delta \geq 2\delta\alpha\sigma_d\sqrt{\left(\frac{\kappa_x}{\kappa_o}\sigma_o\right)^2 + \left(\frac{\delta}{1-\delta}\right)^2\left[\left(\frac{1-\delta}{1-\rho_m\delta}\right)^2\sigma_m^2 + \left(\frac{1-\delta}{1-\rho_w\delta}\frac{1}{\kappa_x}\left(\frac{\kappa_x}{\kappa_o}\right)^2\alpha\sigma_o^2\right)^2\sigma_w^2\right]}.$$

Note that the two roots, as long as they exist, both satisfy $A_d \leq 0$. Importantly, however, they behave differently when $\sigma_d^2 \to 0$. The one with the negative sign satisfies $\lim_{\sigma_d^2 \to 0} A_{d,-} = -\infty$, while the one with the positive sign, i.e. the one with smaller absolute value (being closer to zero) has a finite limit

$$\lim_{\sigma_d^2 \to 0} A_{d,+} = -\frac{1-\delta}{1-\rho_d \delta} \alpha \left[\left(\frac{\kappa_x}{\kappa_o} \sigma_o \right)^2 + \left(\frac{\delta}{1-\delta} \right)^2 \left(A_m^2 \sigma_m^2 + A_w^2 \sigma_w^2 \right) \right]$$

The latter solution is also the stable solution of the model in the sense that it is robust to a small perturbation in investors' beliefs regarding equilibrium price impact.

It is also useful to solve for the fixed point in terms of V—in some cases this is more tractable mathematically. Still sticking to the $\gamma = 0$ case we have A_m and A_w constants, so substituting (17b)-(17d) into (18), the system implies

$$\begin{aligned} 0 &= \left(\alpha \sigma_d \frac{\delta}{1-\rho_d \delta} \right)^2 \cdot V^2 + \left[2 \left(\alpha \sigma_d \frac{\delta}{1-\rho_d \delta} \right)^2 \left(\kappa_x \frac{\sigma_o}{\kappa_o} \right)^2 - 1 \right] \cdot V \\ &\quad + \left[\left(\alpha \sigma_d \frac{\delta}{1-\rho_d \delta} \right)^2 \left(\kappa_x \frac{\sigma_o}{\kappa_o} \right)^4 + \left(\frac{\delta}{1-\delta} \right)^2 \left(A_m^2 \sigma_m^2 + A_w^2 \sigma_w^2 \right) \right] \\ &\equiv A_V V^2 + B_V V + C_V, \end{aligned}$$

which again has a unique root as long as $\sigma_d = A_V = 0$; and when $A_V > 0$, the two solutions are given by

$$V_{+,-} = \frac{-B_V \pm \sqrt{B_V^2 - 4 A_V C_V}}{2 A_V}.$$

Since $A_V, C_V \geq 0$, $\sqrt{B_V^2 - 4A_V C_V} \leq |B_V|$, implying that the only valid solution in terms of variance is V_+, which also turns out to be the stable solution.

Using the above results, consider the following special cases:

1. If $\kappa_x = 0$ and $m_t = \sigma_m = 0$, then we have $A_m = A_d = A_w = 0$ in the unique stable equilibrium. However, there is an unstable equilibrium with $A_{d,-} < 0$. This is equivalent to the De Long et al. (1990) knife-edge equilibrium in which LoOP fails because investors think it will fail in the future. However, this equilibrium is not robust to small perturbations in intermediary beliefs.
2. If $\kappa_x = 0$ and $\sigma_m > 0$, then we have $A_m > 0$, $A_d < 0$, and $A_w = 0$ in the unique stable equilibrium. This is just a textbook Greenwood and Vayanos (2014)-style equilibrium with a risky fundamental (short-rate differential) and random net supply.
3. If $\kappa_x > 0$ and $m_t = \sigma_m = 0$, then we have $A_m = 0$, $A_d < 0$, and $A_w > 0$ in the unique stable equilibrium. This is just a textbook Greenwood and Vayanos (2014)-style equilibrium with a risky frictional cost (shadow value of capital) that functions like a risky fundamental and random supply.
4. Finally, recall from (A-3a)-(A-3c) above that if $\kappa_x > 0$ and $\sigma_m > 0$, then we have $A_m > 0$, $A_d < 0$, and $A_w > 0$ in the unique stable equilibrium.

Next we consider the $\gamma > 0$ case. When $\sigma_d^2 = 0$, there is a unique equilibrium. Specifically, we can think about this system as a fixed point in terms of $V \in [0, \infty)$ of the form

$$V = \left(\frac{\delta}{1-\delta} \right)^2 \left(\sigma_m^2 \left(A_m [V] \right)^2 + \sigma_w^2 \left(A_w [V] \right)^2 \right), \tag{A-4}$$

where $A_m [V]$ and $A_w [V]$ are given by (17b) and (17d).The RHS of (A-4) is non-negative and decreasing in V, is strictly positive when $V = 0$, while it approaches 0as $V \to \infty$. The LHS is non-negative and increasing in V, equals zero when $V = 0$, and approaches ∞ as $V \to \infty$. Therefore, a solution always exists and is unique, and features $A_m > 0$ and $A_w > 0$.

When $\gamma > 0$ and $\sigma_d^2 > 0$, equilibria may or may not exist, and if one exists, it will generally not be unique. For example, assuming away the constraint on agents, i.e. setting $\kappa_x = A_w = 0$, the system of equations collapses to a single higher order polinomial—5th order in the general case, but cubic already in the special case when, e.g., $\rho_m = \rho_d$. In this setting, we can already show that (i) the trivial $V = 0$ is not one of the solutions, and (ii) at least two of the solutions are valid since they feature $V > 0$. In fact, this complexity, which exists in any model of this sort so long as there are (i) demand shocks ($\sigma_d^2 > 0$) and (ii) price-sensitive demand ($\gamma > 0$), is the standard multiplicity issue highlighted by Spiegel (1998) and others. In the bond market context, this complexity arises, for instance, in Vayanos and Vila (2021) with random supply. In the Online Appendix we discuss equilibrium stability and multiplicity further. ■

Proof of Proposition 1. From (17a), it is imminent that $A_0 < 0$ as long as $\overline{m} < \overline{\overline{m}}$, where this latter is defined as $\overline{\overline{m}} = \alpha \left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 \left[\frac{\kappa_o}{\kappa_x}\frac{\bar{r}_o}{\alpha\sigma_o^2} + \bar{d} - \frac{1}{\kappa_x}\overline{w}\right] + \alpha V \bar{d}$. ■

Proof of Proposition 4. The results on the B_i, $i = \{m, d, w\}$, coefficients follow directly from the the definition of swap returns and the equilibrium coefficients (17b)-(17a). Next, we write

$$\begin{aligned} |B_d| - |\kappa_x B_w| &= \left|\frac{1-\rho_d\delta}{1-\delta}A_d\right| - \left|\kappa_x \frac{1-\rho_w\delta}{1-\delta}A_w\right| \\ &= \frac{\frac{1-\rho_d\delta}{1-\delta}\alpha\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]}{\frac{1-\rho_d\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]} - \frac{\frac{1-\rho_w\delta}{1-\delta}\alpha\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2}{\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]} \\ &= \frac{\left\{\frac{1-\rho_d\delta}{1-\delta}\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right] - \frac{1-\rho_w\delta}{1-\delta}\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2\right\}\alpha^2\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right] + \frac{1-\rho_w\delta}{1-\delta}\frac{1-\rho_d\delta}{1-\delta}\alpha V}{\left(\frac{1-\rho_d\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]\right)\left(\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]\right)}, \end{aligned}$$

so $|B_d| > |\kappa_x B_w|$ is equivalent to

$$\left\{\frac{1-\rho_d\delta}{1-\delta}\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right] - \frac{1-\rho_w\delta}{1-\delta}\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2\right\}\alpha^2\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]\gamma + \frac{1-\rho_w\delta}{1-\delta}\frac{1-\rho_d\delta}{1-\delta}\alpha V > 0,$$

which in turn holds for all $\gamma \geq 0$ as long as the term in braces is positive—i.e. when

$$V > \frac{\delta}{1-\rho_d\delta}\left(\rho_d - \rho_w\right)\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2, \tag{A-5}$$

which is more likely to hold, e.g., when δ and $\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2$ are small, $\rho_d \leq \rho_w$, or σ_m^2, σ_d^2, and σ_w^2 are large—or when

$$0 < \gamma < \bar{\gamma} \equiv -\frac{1}{\left\{\frac{1-\rho_d\delta}{1-\delta}\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right] - \frac{1-\rho_w\delta}{1-\delta}\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2\right\}}\frac{\frac{1-\rho_w\delta}{1-\delta}\frac{1-\rho_d\delta}{1-\delta}\alpha V}{\alpha^2\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2 + V\right]} > 0,$$

as long as (A-5) does not hold, i.e. when δ and $\left(\frac{\kappa_x}{\kappa_o}\right)^2\sigma_o^2$ are large, $\rho_d \gg \rho_w$, or σ_m^2, σ_d^2, and σ_w^2 are small. ■

Proof of Proposition 7. From (17a), $\partial A_0/\partial\left(\bar{r}_o/\sigma_o^2\right) < 0$ since an increase in $\bar{r}_o/\sigma_o^2$ strictly lowers the nominator of A_0 but does not affect its denominator. On the other hand, an increase in σ_o^2 lowers the nominator and has a non-negative impact on the denominator of A_0, i.e., $\partial A_0/\partial\sigma_o^2 < 0$. ■

B Term structure model

Since it features a single long-term perpetual swap, our baseline model does not deliver specific implications for the full term structure of swap spreads. In order to study this term structure in a simple fashion, we introduce a series of n-period zero-coupon swaps alongside the perpetual swap introduced above. We then take the limit as net end-user demand for these zero-coupon swaps goes to zero. Using this approach, we can derive the spreads on these non-traded n-period zero-coupon swaps from intermediaries' first order conditions.

A position in the n-period swap spread trade receives the fixed rate on an n-period zero-coupon swap and hedges the associated interest rate risk by going short n-period zero-coupon Treasury bonds. As a result, the return on this n-period receive-fixed spread trade is

$$r_{t+1}^{Spread(n)} = ns_t^{(n)} - (n-1)\, s_{t+1}^{(n-1)} - m_t,$$

where $s_t^{(n)} \equiv y_t^{S(n)} - y_t^{T(n)}$—i.e., the n-period zero-coupon swap spread is defined as the between the n-period swap yield, $y_t^{S(n)}$, and the n-period Treasury yield, $y_t^{T(n)}$.[44]

General analysis. Intermediaries maximize

$$\max_{o_t, x_t, \{x_t^{(n)}\}_{n=1}^N} \mathrm{E}_t\left[w_{t,t+1}\right] - \frac{\alpha}{2}\mathrm{Var}_t\left[w_{t,t+1}\right], \tag{B-6}$$

subject to the budget constraint

$$w_{t,t+1} = w_t + x_t r_{t+1}^{Spread} + \sum\nolimits_{n=1}^{N} x_t^{(n)} r_{t+1}^{Spread(n)} + o_t r_{t+1}^{o} \tag{B-7}$$

and the leverage constraint

$$\kappa_x \left|x_t\right| + \kappa_x \sum\nolimits_{n=1}^{N} |x_t^{(n)}| + \kappa_o \left|o_t\right| \leq w_t. \tag{B-8}$$

In other words, we now allow intermediaries to take positions in the perpetual long-term swap as well as in n-period zero-coupon swaps for $n = 1, ..., N$.[45]

The first-order condition for the position in n-period swaps ($x_t^{(n)}$) is

$$\mathrm{E}_t[r_{t+1}^{Spread(n)}] = \kappa_x sgn(x_t^{(n)}) \cdot \psi_t + \alpha \mathrm{Cov}_t[r_{t+1}^{Spread(n)}, w_{t,t+1}]. \tag{B-9}$$

The first-order conditions for the position in the perpetual swap (x_t) and the outside investment opportunity (o_t) are

$$\begin{aligned} \mathrm{E}_t[r_{t+1}^{Spread}] &= \kappa_x sgn\left(x_t\right) \cdot \psi_t + \alpha \mathrm{Cov}_t[r_{t+1}^{Spread}, w_{t,t+1}] \text{ and} && \text{(B-10)} \\ \bar{r}_o &= \kappa_o \cdot sgn\left(o_t\right) \psi_t + \alpha \sigma_o^2 \cdot o_t. && \text{(B-11)} \end{aligned}$$

Since $o_t > 0$, we have

$$\psi_t = \max\left\{0, \frac{\bar{r}_o - \alpha\sigma_o^2 \frac{1}{\kappa_o} w_t + \alpha\sigma_o^2 \frac{\kappa_x}{\kappa_o}\left(|x_t| + \sum_{n=1}^{N} |x_t^{(n)}|\right)}{\kappa_o}\right\}.$$

To compute equilibrium swap spreads, we impose market clearing, setting $x_t = -d_t$ and $x_t^{(n)} = -d_t^{(n)}$ for all n. To close the model a parsimonious way, we assume the same demand process for the generic long-term swap

[44] The return on an n-period zero coupon swap from t to $t+1$ is $r_{t+1}^{S(n)} \equiv ny_t^{S(n)} - (n-1)\, y_{t+1}^{S(n-1)} - i_t^S$ where $y_t^{S(n)}$ is the n-period zero-coupon swap yield at time t. Similarly, the excess return on an n-period zero-coupon Treasury is $r_{t+1}^{T(n)} \equiv ny_t^{T(n)} - (n-1)\, y_{t+1}^{T(n-1)} - i_t^T$ where $y_t^{T(n)}$ is the n-period zero-coupon Treasury yield at time t.Thus, the return on the n-period swap spread arbitrage trade from t to $t+1$ is equal to $r_{t+1}^{Spread(n)} \equiv r_{t+1}^{S(n)} - r_{t+1}^{T(n)} = ns_t^{(n)} - (n-1)\, s_{t+1}^{(n-1)} - m_t$.

[45] A simplifying assumption here is that all swap spread arbitrage positions have the same margin requirement κ_x irrespective of maturity n. Allowing these margin requirements to vary by maturity would introduce additional complexity without qualitatively changing our results.

d_t introduced above in (7) and then take the limit as $d_t^{(n)}$ approaches zero from one side. Taking this one-sided limit as positions in the n-maturity swaps grow small is necessary to preserve the terms relating to balance sheet costs.[46]

Taking this limit , we obtain

$$\overbrace{ns_t^{(n)} - (n-1)\, E_t[s_{t+1}^{(n-1)}] - m_t}^{\mathrm{E}_t[r_{t+1}^{Spread(n)}]} = -\alpha \mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}]d_t - sgn(d_t^{(n)})\kappa_x \psi_t. \tag{B-12}$$

Iterating this equation forward, we obtain the following expression the n-period swap spread

$$s_t^{(n)} = \overbrace{n^{-1}\sum\nolimits_{k=0}^{n-1} \mathrm{E}_t[m_{t+k}]}^{\text{Expected short-rate differentials}} + \overbrace{n^{-1}\sum\nolimits_{k=0}^{n-1} \mathrm{E}_t[(-\kappa_x)\, sgn(d_t^{(n-k)})\psi\,(w_{t+k}, |d_{t+k}|)]}^{\text{Expected compensation for using scarce capital}} \underbrace{+n^{-1}\sum\nolimits_{k=0}^{n-1} (-\alpha)\, \mathrm{E}_t[\mathcal{C}_{t+k}^{(n-k)} d_{t+k}]}_{\text{Expected compensation for risk}}, \tag{B-13}$$

where $\mathcal{C}_t^{(n)} \equiv \mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}] = (n-1)\,\frac{\delta}{1-\delta}\mathrm{Cov}_t[s_{t+1}^{(n-1)}, s_{t+1}]$ is the risk of n-period swaps captured by their return co-movement with the return on the long-maturity swap. This expression naturally generalizes equation (12) above.

Consider the spreads on 1-period swaps, $s_t^{(1)}$. Since the return on a 1-period swap spread trade is riskless, we have $\mathcal{C}_t^{(1)} = 0$ implying that

$$s_t^{(1)} = m_t + (-\kappa_x)\, sgn(d_t^{(1)})\psi\,(w_t, |d_t|)\,. \tag{B-14}$$

Thus, the short-dated spread is the sum of the current short-rate differential (m_t) and a term that is proportional to the current shadow cost of intermediary capital ($\psi_t \geq 0$). If the net demand to receive fixed is always positive for all maturities n—i.e., $sgn(d_t^{(n)}) = sgn(d_t) = 1$ for all t and n, then $s_t^{(1)} = m_t + (-\kappa_x)\,\psi\,(w_t, d_t)$ and $s_t^{(n)} = n^{-1}\sum_{k=0}^{n-1}\mathrm{E}_t[s_{t+k}^{(1)}] + n^{-1}\sum_{k=0}^{n-1}(-\alpha)\,\mathrm{E}_t[\mathcal{C}_{t+k}^{(n-k)}d_{t+k}]$. In this case, the short-term swap spread $s_t^{(1)}$ plays a role that is analogous to that played by short-term interest rates in traditional term structure models.[47]

Assume that $m_t \equiv 0$ for all t as is the case for a SOFR swap, so that any frictionless model would predict that $s_t^{(n)} = 0$ for all n and t by the LoOP. Naturally, this obtains in our model in the limit where $\kappa_x = 0$. However, when $\kappa_x > 0$, swap spreads will no longer be zero due to the potential for binding intermediary capital constraints. Indeed, the current short-term spread ($s_t^{(1)}$) reveals the current shadow value of intermediary capital (ψ_t) up to a constant of proportionality. In particular, if $m_t \equiv 0$, short-dated spreads will be zero at t if the constraint is slack at t ($\psi_t = 0$).

Now consider the spreads on swaps with $n \geq 2$ periods and continue to assume that $m_t \equiv 0$ for all t. Crucially, $s_t^{(n)}$ can still be non-zero for $n \geq 2$ and will be impacted by supply and demand even when $\psi_t = 0$ today—i.e., even when the current shadow value of capital is zero. In order to have $s_t^{(n)} \neq 0$ for $n > 1$, we only need to have $\mathrm{E}_t[\psi_{t+k}] > 0$ for some $k \leq n-1$—i.e., the constraint must be expected to bind sometime over the life of the swap.

Furthermore, even if the constraint is slack at time t, the expected returns from t to $t+1$ for spread arbitrage on long-dated swaps will be non-zero if there can be news at $t+1$ about the severity of future constraints. Specifically, even if $\psi_t = 0$, we can have $\mathrm{E}_t[r_{t+1}^{Spread(n)}] = -\alpha\mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}]d_t \neq 0$, so long as $\mathrm{Var}_t[\mathrm{E}_{t+1}[\psi_{t+k}]] > 0$ for some $k \leq n-1$. That is, even if the constraint is slack at t, intermediaries need to be compensated for the risk of swap spread movements at $t+1$ due to news about the tightness of future constraints.

[46] As explained in the Online Appendix, there are other ways we could close the model. Specifically, we could set $d_t \equiv 0$ and specify a demand process for the set of swap maturities $\{d_t^{(n)}\}_{n=1}^N$. This approach would allow us to consider separate shocks to the demand for swaps with different maturities. We have opted to close the model in a more parsimonious way here since our focus is mainly on how the general demand for swaps of long maturities impacts the term structure of swap spreads as opposed to providing a general theory of the swap spread term structure.

[47] In the limiting case where end-user demand to receive fixed is completely inelastic ($\gamma = 0$), $s_t^{(1)}$ does not depend on the amount of swap spread risk—i.e., $s_t^{(1)}$ does not depend on σ_m^2, σ_d^2, or σ_w^2. However, when $\gamma > 0$, $s_t^{(1)}$ will depend on swap-spread risk since the equilibrium demand for the generic long-term swap d_t is influenced by risk.

As a result, intermediaries still need to be compensated for between t and $t+1$ for bearing demand-and-supply risk even when the constraint is not currently binding at t.

An affine equilibrium. To obtain an affine model of the term structure of swap spreads, we conjecture that

$$s_t^{(n)} = A_0^{(n)} + A_m^{(n)} z_t^m + A_d^{(n)} z_t^d + A_w^{(n)} z_t^w.$$

We then impose the restriction (13) and take the upper limit as $d_t^{(n)}$ approaches zero from above for all n—i.e., we assume that intermediaries have a vanishingly small short position in the pay-fixed swap spread arbitrage for all maturities n. We obtain the following result:

Theorem 2 *The equilibrium n-period swap spread is given by*

$$s_t^{(n)} = A_0^{(n)} + A_m^{(n)} z_t^m + A_d^{(n)} z_t^d + A_w^{(n)} z_t^w, \tag{B-15}$$

with the exact functional forms of $A_0^{(n)}$, $A_d^{(m)}$, $A_d^{(n)}$, and $A_w^{(n)}$ provided in the Appendix. When $\gamma = 0$, $A_0^{(n)}$ is negative and decreasing in maturity n, $A_m^{(n)}$ and $A_w^{(n)}$ are positive and decreasing in n, and $A_d^{(n)}$ is negative for all n and either increasing or U-shaped.

Swap spreads are the average returns on swap spread arbitrage throughout the lifetime of a swap, and hence are determined by the joint effect of expectations and risk premia until their maturity, as shown by (B-13). In particular, swap spreads are determined by expected future short-rate differentials, expected compensation for using scarce capital throughout the life of swaps, and expected compensation for risk.

A key implication of our model is that the swap spread curve is downward-sloping on average when end-user demand to receive fixed is positive ($d_t > 0$). This result arises because, at least when end-user demand is completely inelastic ($\gamma = 0$), $\mathcal{C}^{(n)} \equiv \mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}]$ is an increasing function of maturity n—i.e., the returns on longer-dated swaps covary more strongly with the returns on intermediaries' portfolios—implying that longer-dated swaps are riskier for intermediaries. The greater average magnitude of longer-dated spreads reflects the greater risk compensation that intermediaries require on these longer-dated swaps. Regarding the level of spreads, when $\mathrm{E}[m_t] = \overline{m} \geq 0$ is sufficiently small, such as for OIS swaps, $A_0^{(n)} < 0$ for all n. For LIBOR swaps where $\overline{m}$ is sufficiently large, $A_0^{(n)} > 0$ for all n. In the intermediate case where $\overline{m}$ is moderately positive, $A_0^{(n)} > 0$ for small n and $A_0^{(n)} < 0$ for larger n. These cases highlight that a higher interest rate differential pushes all swap spreads towards positive values, however, the balance sheet cost and the risk compensation terms push swap spreads towards negative levels. The relative magnitudes of these forces determine the overall sign of $A_0^{(n)}$, with longer maturity spreads more likely to be negative.

The term $A_m^{(n)}$ reflects variation in the expected future short rate differentials (m_t) over the life of the swap and, thus, is positive and locally downward-sloping across maturities n. Similarly, $A_w^{(n)}$ reflects variation in the impact of intermediary wealth on expected future balance sheet costs ($-\kappa_x \psi_t$) over the life of the swap and, thus, is positive and locally downward-sloping across maturities n.

Finally, $A_d^{(n)}$ is negative for all maturities n and reflects both the (i) expected balance sheet costs and (ii) compensation for risk over the life of the swap. When the volatility of swap spreads is sufficiently low, (i) dominates for all maturities, and $A_d^{(n)}$ is an increasing function of maturity n. When the volatility of swap spreads is higher, (ii) can became strong for longer maturities. In this case, $A_d^{(n)}$ is downward-sloping across maturities n when ρ_d is sufficiently high and is a U-shaped function of maturity n when ρ_d is lower.

When $\gamma > 0$, end users submit price-sensitive demands and hence movements in z_t^m and z_t^w affect intermediaries' equilibrium exposure to swaps; this implies that $A_m^{(n)}$ and $A_w^{(n)}$ now also reflect compensation for risk. However, by continuity, all analytical results of the $\gamma = 0$ case continue to hold when end-user demand is somewhat inelastic, i.e., for $\gamma > 0$ sufficiently small. However, as we show in the Appendix, the local shape of the swap spread curve can become more complex when $\gamma > 0$ is larger, i.e., when end-user demand is highly elastic, which creates strong feedback effects between current and expected future swap spreads and can lead to theoretical swap spread curves with undesirable oscillatory properties.

Using the term structure extension of our model, we can further study both the global as well as the local shape of the spread curve. Our measure of the global shape of the spread curve, $Slope_t \equiv s_t - s_t^{(1)}$, is simply the difference in spreads between the long-term generic swap—which we associate with a swap with the average duration of end-user demand, which then also equalts intermediaries' average exposure—and a 1-period swap.[48] Further, we also study the local shape of the term structure, i.e., the determinants of $Slope_t^{(n)} \equiv s_t^{(n)} - s_t^{(n-1)}$ for some n. The next two Propositions summarize our results about the global and local shape of the term structure:

Proposition 8 *Using (16) and (B-15), the* global *slope of the term structure of swap spreads is given by*

$$Slope_t = \left(A_0 - A_0^{(1)}\right) + \left(A_m - A_m^{(1)}\right) z_t^m + \left(A_d - A_d^{(1)}\right) z_t^d + \left(A_w - A_w^{(1)}\right) z_t^w. \tag{B-16}$$

The coefficients in (B-16) satisfy $A_0 - A_0^{(1)}, A_m - A_m^{(1)}, A_w - A_w^{(1)} < 0$, *while* $A_d - A_d^{(1)} < 0$ *so long as*

$$\frac{\delta}{1-\delta}\left(1-\rho_d\right)\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 < V.$$

Proposition 8 compares the level and shock sensitivity of the long-term generic swap to those of the 1-period swap. On average, we find that term structure is globally downward-sloping in the sense that $\mathrm{E}[Slope_t] = A_0 - A_0^{(1)} = -\alpha V \mathrm{E}\left[d_t\right] < 0$—see also Proposition 8. The average slope is more negative when intermediaries' risk aversion α is higher, the swap spread risk $V > 0$ is higher, and when the average demand from end users to receive the fixed swap rate, $\mathrm{E}\left[d_t\right] = \bar{d} + \gamma A_0 > 0$, is higher. We also find that positive shocks to the interest rate differential and to intermediaries' wealth increases the slope (since the slope is negative, this corresponds to the curve becoming flatter); for example, higher intermediary wealth relaxes their balance sheet constraint, leading to lower required compensation for supplying swaps, and this effect is stronger for longer maturities. Finally, we show that shocks to investor demand can have ambiguous impact on the slope, depending on whether the expectation hypothesis term or the risk compensation term dominate long-term swaps spreads: the first one has a larger impact on short-term swap spreads, while the second becomes important for longer maturities. This implies that positive demand shocks are more likely to lower long-term spreads more than short-term spreads and thus make the slope more negative when σ_o^2 is small and when ρ_d, σ_m^2, σ_d^2, and σ_w^2 are large.

While the global slope of the term structure only shows that the average level of the long-term generic swap is more negative than the average level of the 1-period swap, looking at the local properties can highlight non-monotonic effects of the forces that shape swap spreads:

Proposition 9 *Using (16) and (B-15), the* local *slope of the term structure of swap spreads is given by*

$$Slope_t^{(n)} = \left(A_0^{(n)} - A_0^{(n-1)}\right) + \left(A_m^{(n)} - A_m^{(n-1)}\right) z_t^m + \left(A_d^{(n)} - A_d^{(n-1)}\right) z_t^d + \left(A_w^{(n)} - A_w^{(n-1)}\right) z_t^w. \tag{B-17}$$

When end-user demand is completely inelastic ($\gamma = 0$), we find that $A_0^{(n)} - A_0^{(n-1)}, A_m^{(n)} - A_m^{(n-1)}, A_w^{(n)} - A_w^{(n-1)} < 0$ *for all maturities n, while* $A_d^{(n)} - A_d^{(n-1)}$ *is ambiguous: either* $A_d^{(n)} - A_d^{(n-1)} < 0$ *for all maturities n, or* $A_d^{(n)} - A_d^{(n-1)} < 0$ *for short maturities and* $A_d^{(n)} - A_d^{(n-1)} > 0$ *for long maturities.*

Expected returns also have a term structure, as follows:

Proposition 10 *The expected return on the n-period swap spread trade is given by*

$$E_t[r_{t+1}^{Spread(n)}] = B_0^{(n)} + B_m^{(n)} z_t^m + B_m^{(n)} z_t^d + B_w^{(n)} z_t^w, \tag{B-18}$$

with the exact functional forms given below.

When end-user demand is completely inelastic ($\gamma = 0$), (i) $B_0^{(n)}$ is negative and decreasing in maturity n, (ii) $B_m^{(n)} = 0$ for all n, (iii) $B_d^{(n)}$ is negative and decreasing in n, and (iv) $B_w^{(n)}$ is positive and constant across maturities.

When end-user demand is somewhat inelastic (γ is positive, but not too large), (i) $B_0^{(n)} < 0$, (ii) $B_m^{(n)} < 0$, (iii) $B_d^{(n)} < 0$, and (iv) $B_w^{(n)} > 0$, Since $\mathcal{C}^{(n)}$ is increasing in n when γ is not too large, all four coefficients are decreasing in maturity n.

[48] Alternatively, we could define $Slope_t^* \equiv s_t^{(n)} - s_t^{(1)}$ as the difference between the n- and the 1-period swaps; while the exact coefficients and certain conditions change, our qualitative results presented in Proposition 5 remain the same.

When $\gamma = 0$, interest rate differential shocks and supply shocks have the same effect on the expected swap arbitrage returns, as measured by $B_m^{(n)}$ and $B_w^{(n)}$, respectively. In fact, $B_m^{(n)} = 0$ for all maturities because, just like in the case of the perpetual long-term swap spread, changes in the short-rate differential m_t impact swap cash flows, but do not affect the equilibrium amount of swap spread risk intermediaries must hold. In turn, the constancy of $B_w^{(n)}$ reflects the fact that as long as $\gamma = 0$, the $B_w^{(n)} z_t^w$ term in (B-18) purely reflects compensation for consuming scarce capital, and all swaps, irrespective of maturity, consume the amount of capital per unit notional.[49] By contrast, the $B_d^{(n)} z_t^d$ term reflects both compensation for consuming scarce capital and compensation for risk. The fact that $B_d^{(n)}$ is decreasing in n reflects the fact that longer-term swaps are riskier for intermediaries; this in turn implies that average expected returns, $B_0^{(n)}$, must also increase (in absolute terms) for longer maturities to compensate intermediaries—i.e., $\mathcal{C}^{(n)}$ is increasing in n.

When end-user demand is somewhat inelastic (γ is positive, but not too large), $B_0^{(n)}$, $B_m^{(n)}$, $B_d^{(n)}$, and $B_w^{(n)}$ are each decreasing in n. This decline reflects the facts that (i) $\mathrm{E}_t[r_{t+1}^{Spread(n)}] = -\alpha \mathcal{C}^{(n)} d_t - \kappa_x \psi(w_t, d_t)$, (ii) $\mathcal{C}^{(n)}$ is increasing in n, and (iii) $d_t = \bar{d} + z_t^d + \gamma s_t$ is increasing in z_t^m, z_t^d, and z_t^w when $\gamma > 0$.

Our framework also has important implications predicting swap returns. To this end, we calculation Fama-Bliss-style regression coefficients. In particular, we obtain the following results

Proposition 11 *A higher slope of the swap term structure forecasts higher returns on the n-period swap spread trade, i.e., running the regressions*

$$r_{t+1}^{Spread(n)} = \alpha + \beta_1 (s_t - s_t^{(1)}) + \xi_{t+1}^{(n)}$$

or

$$r_{t+1}^{Spread(n)} = \alpha + \beta_2 (s_t^{(n)} - s_t^{(1)}) + \xi_{t+1}^{(n)},$$

we find that under mild conditions the regression coefficients β_1 and β_2 are positive.

Finally, our framework also allows to study swap spread volatility. In particular, we obtain the following result:

Proposition 12 *The conditional and unconditional volatilities of swap spreads are given by*

$$\mathit{Var}_{t-1}[s_t^{(n)}] = (A_m^{(n)})^2 \sigma_m^2 + (A_d^{(n)})^2 \sigma_d^2 + (A_w^{(n)})^2 \sigma_w^2$$

and

$$\mathit{Var}[s_t^{(n)}] = (A_m^{(n)})^2 \frac{\sigma_m^2}{1-\rho_m^2} + (A_d^{(n)})^2 \frac{\sigma_d^2}{1-\rho_d^2} + (A_w^{(n)})^2 \frac{\sigma_w^2}{1-\rho_w^2}.$$

When end-user demand is completely inelastic ($\gamma = 0$) or moderately inelastic (γ is not too large), and swap spread risk is meaningful (σ_m^2, σ_d^2, and σ_w^2 are large small relative to σ_o^2), (i) shocks to short-rate differentials and intermediary wealth drive more of the volatility of short-dated swaps (n small) whereas (ii) shocks to end-user demand drive more of the volatility of long-dated swaps (n large). When (i) dominates, swap spread volatility is a decreasing function of maturity n; when (ii) is strong enough, swap spread volatility is a hump-shaped function of maturity n.

Formally, this follows from the fact that $|A_m^{(n)}|$ and $|A_m^{(w)}|$ are decreasing in n, whereas $|A_d^{(n)}|$ is either decreasing or a hump-shaped function of n. Intuitively, the changing drivers of swap spread volatility at different maturities arises from the fact that shocks to short-rate differentials and intermediary wealth primarily affect the term structure by changing the expected short-dated spread over the life of a swap. Since these shocks are mean reverting, this expectations-hypothesis-like channel plays an greater role in driving movements in short-dated swap spreads. By contrast, shocks to demand primarily affect the term structure by changing the expected compensation for risk over the life of a swap. This term-premium-like channel plays a greater role in driving movements in long-dated spreads.

[49] Formally, we have (i) $\mathrm{E}_t[r_{t+1}^{Spread(n)}] = -\alpha \mathcal{C}^{(n)} d_t - \kappa_x \psi(w_t, d_t)$, (ii) $\mathcal{C}^{(n)}$ is increasing in n, and (iii) $d_t = \bar{d} + z_t^d$ is independent of z_t^w when $\gamma = 0$. Thus, $\partial \mathrm{E}_t[r_{t+1}^{Spread(n)}]/\partial z_t^w = -\kappa_x \cdot \partial \psi_t / \partial z_t^w$ for all n.

Proofs and derivations

Proof of Theorem 2. To derive the solution to our affine term structure model, we start from (14) and (B-12) in the case when $d_t > 0$ and $d_t^{(n)} \searrow 0$, and thus $sgn(d_t^{(n)}) = sgn(d_t^{(n)}) = 1$. Combining these equations, we obtain

$$\overbrace{ns_t^{(n)} - (n-1)\,\mathrm{E}_t[s_{t+1}^{(n-1)}] - m_t}^{\mathrm{E}_t[r_{t+1}^{Spread(n)}]} + \alpha \overbrace{\mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}]}^{\mathcal{C}^{(n)}} d_t = \mathrm{E}_t[r_{t+1}^{Spread}] + \alpha V d_t = -\kappa_x \psi_t. \quad \text{(B-19)}$$

Using our affine conjecture for the n-period swap spread (B-15) to express $\mathrm{E}_t[r_{t+1}^{Spread(n)}]$, combining it with (7) and (21), (B-19) implies

$$\begin{aligned}
& (nA_0^{(n)} - (n-1)\,A_0^{(n-1)} - \overline{m} + \alpha\left(\overline{d} + \gamma A_0\right)\mathcal{C}^{(n)}) + (nA_m^{(n)} - \rho_m\,(n-1)\,A_m^{(n-1)} - 1 + \alpha\gamma A_m \mathcal{C}^{(n)}) \cdot z_t^m \\
& + (nA_d^{(n)} - \rho_d\,(n-1)\,A_d^{(n-1)} + \alpha\,(1+\gamma A_d)\,\mathcal{C}^{(n)}) \cdot z_t^d + (nA_w^{(n)} - \rho_w\,(n-1)\,A_w^{(n-1)} + \alpha\gamma A_w \mathcal{C}^{(n)}) \cdot z_t^w \\
= \; & \left(B_0 + \alpha V\left(\overline{d} + \gamma A_0\right)\right) + (B_m + \alpha\gamma V A_m) \cdot z_t^m \\
& + (B_d + \alpha V\,(1+\gamma A_d)) \cdot z_t^d + (B_w + \alpha\gamma V A_w) \cdot z_t^w.
\end{aligned}$$

Letting $D_j^{(n)} \equiv nA_j^{(n)}$ for $j \in \{0, m, d, w\}$ and noting that

$$\mathcal{C}^{(n)} \equiv \mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}] = \frac{\delta}{1-\delta}\left(D_m^{(n-1)} A_m \sigma_m^2 + D_d^{(n-1)} A_d \sigma_d^2 + D_w^{(n-1)} A_w \sigma_w^2\right),$$

we can write this system of recursive equations more compactly as

$$\begin{bmatrix} D_0^{(n)} \\ D_m^{(n)} \\ D_d^{(n)} \\ D_w^{(n)} \end{bmatrix} = \Phi_1 \begin{bmatrix} D_0^{(n-1)} \\ D_m^{(n-1)} \\ D_d^{(n-1)} \\ D_w^{(n-1)} \end{bmatrix} + \Phi_0,$$

with

$$\Phi_1 = \begin{bmatrix} 1 & -\alpha\frac{\delta}{1-\delta}\sigma_m^2\left(\bar{d}+\gamma A_0\right)A_m & -\alpha\frac{\delta}{1-\delta}\sigma_d^2\left(\bar{d}+\gamma A_0\right)A_d & -\alpha\frac{\delta}{1-\delta}\sigma_w^2\left(\bar{d}+\gamma A_0\right)A_w \\ 0 & \rho_m - \alpha\gamma\frac{\delta}{1-\delta}\sigma_m^2 A_m^2 & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_d^2 A_d A_m & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_w^2 A_w A_m \\ 0 & -\alpha\frac{\delta}{1-\delta}\sigma_m^2\,(1+\gamma A_d)\,A_m & \rho_d - \alpha\frac{\delta}{1-\delta}\sigma_d^2\,(1+\gamma A_d)\,A_d & -\alpha\frac{\delta}{1-\delta}\sigma_w^2\,(1+\gamma A_d)\,A_w \\ 0 & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_m^2 A_m A_w & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_d^2 A_d A_w & \rho_w - \alpha\gamma\frac{\delta}{1-\delta}\sigma_w^2 A_w^2 \end{bmatrix} \quad \text{(B-20)}$$

and

$$\Phi_0 = \begin{bmatrix} B_0 + \alpha\gamma V A_0 + \overline{m} + \alpha V \overline{d} \\ B_m + \alpha\gamma V A_m + 1 \\ B_d + \alpha\gamma V A_d + \alpha V \\ B_w + \alpha\gamma V A_w \end{bmatrix} = \begin{bmatrix} (1+\alpha\gamma V)\,A_0 + \alpha V \overline{d} \\ \left(\frac{1-\delta\rho_m}{1-\delta} + \alpha\gamma V\right) A_m \\ \left(\frac{1-\rho_d\delta}{1-\delta} + \alpha\gamma V\right) A_d + \alpha V \\ \left(\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma V\right) A_w \end{bmatrix}, \quad \text{(B-21)}$$

and the initial conditions $D_0^{(0)} = D_m^{(0)} = D_d^{(0)} = D_w^{(0)} = 0$. The solution of this system of recursive equations is then

$$\begin{bmatrix} D_0^{(n)} & D_m^{(n)} & D_d^{(n)} & D_w^{(n)} \end{bmatrix}' = (\mathbf{1} - \Phi_1)^{-1}\,(\mathbf{1} - \Phi_1^n)\,\Phi_0. \quad \text{(B-22)}$$

For example, the 1-period swap rate at time t is given by

$$s_t^{(1)} = m_t - \kappa_x \psi_t = A_0^{(1)} + A_m^{(1)} z_t^m + A_d^{(1)} z_t^d + A_w^{(1)} z_t^w,$$

where $[A_0^{(1)}, A_m^{(1)}, A_d^{(1)}, A_w^{(1)}]' = [D_0^{(1)}, D_m^{(1)}, D_d^{(1)}, D_w^{(1)}]' = \Phi_0$ with $A_m^{(1)} > 0$, $A_d^{(1)} < 0$, and $A_w^{(1)} > 0$, and $A_0^{(1)}$ depending on

Since $B_0 + \overline{m} = A_0$, it follows that $\mathrm{E}[s_t - s_t^{(1)}] = A_0 - A_0^{(1)} = -\alpha V \mathrm{E}[d_t] < 0$, implying that the term structure of swap spreads—summarized using the difference between the generic long-term spread s_t and the 1-period spread—is downward-sloping on average. The average slope is a function of intermediaries' risk aversion α, swap spread risk V, and the average demand from end-users to receive the fixed swap rate, $\mathrm{E}[d_t] = \bar{d} + \gamma A_0 > 0$.

To understand how the $D_j^{(n)}$ coefficients behave as a function of n, note that (B-22) implies

$$\begin{bmatrix} D_0^{(n)} \\ D_m^{(n)} \\ D_d^{(n)} \\ D_w^{(n)} \end{bmatrix} - \begin{bmatrix} D_0^{(n-1)} \\ D_m^{(n-1)} \\ D_d^{(n-1)} \\ D_w^{(n-1)} \end{bmatrix} = \Phi_1^{n-1}\Phi_0,$$

or, ignoring the constant term,

$$\begin{bmatrix} D_m^{(n)} \\ D_d^{(n)} \\ D_w^{(n)} \end{bmatrix} - \begin{bmatrix} D_m^{(n-1)} \\ D_d^{(n-1)} \\ D_w^{(n-1)} \end{bmatrix} = \Gamma_1^{n-1} \begin{bmatrix} D_m^{(1)} \\ D_d^{(1)} \\ D_w^{(1)} \end{bmatrix},$$

where Γ_1 is the 3×3 obtained from Φ_1 by deleting the first row and the first column:

$$\Gamma_1 = \begin{bmatrix} \rho_m - \alpha\gamma\frac{\delta}{1-\delta}\sigma_m^2 A_m^2 & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_d^2 A_d A_m & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_w^2 A_w A_m \\ -\alpha\frac{\delta}{1-\delta}\sigma_m^2\left(1+\gamma A_d\right)A_m & \rho_d - \alpha\frac{\delta}{1-\delta}\sigma_d^2\left(1+\gamma A_d\right)A_d & -\alpha\frac{\delta}{1-\delta}\sigma_w^2\left(1+\gamma A_d\right)A_w \\ -\alpha\gamma\frac{\delta}{1-\delta}\sigma_m^2 A_m A_w & -\alpha\gamma\frac{\delta}{1-\delta}\sigma_d^2 A_d A_w & \rho_w - \alpha\gamma\frac{\delta}{1-\delta}\sigma_w^2 A_w^2 \end{bmatrix},$$

and where $D_m^{(1)} > 0$, $D_d^{(1)} < 0$, and $D_w^{(1)} > 0$.

In the general case where $\gamma > 0$, the behavior of the $D_j^{(n)}$, $j \in \{m, d, w\}$, coefficients can be quite complex as we explain below and in the Online Appendix. However, their behavior in the case when end-user demand is inelastic ($\gamma = 0$) is straightforward and intuitive. When $\gamma = 0$, letting $\phi_d \equiv \rho_d - \alpha\frac{\delta}{1-\delta}\sigma_d^2 A_d > \rho_d > 0$, we have

$$\begin{aligned} \begin{bmatrix} D_m^{(n)} \\ D_d^{(n)} \\ D_w^{(n)} \end{bmatrix} - \begin{bmatrix} D_m^{(n-1)} \\ D_d^{(n-1)} \\ D_w^{(n-1)} \end{bmatrix} &= \begin{bmatrix} \rho_m & 0 & 0 \\ -\alpha\frac{\delta}{1-\delta}\sigma_m^2 A_m & \phi_d & -\alpha\frac{\delta}{1-\delta}\sigma_w^2 A_w \\ 0 & 0 & \rho_w \end{bmatrix}^{n-1} \begin{bmatrix} D_m^{(1)} \\ D_d^{(1)} \\ D_w^{(1)} \end{bmatrix} \\ &= \begin{bmatrix} \rho_m^{n-1} & 0 & 0 \\ -\alpha\frac{\delta}{1-\delta}\sigma_m^2 A_m \cdot (\sum_{k=0}^{n-2}\phi_d^{n-2-k}\rho_m^k) & \phi_d^{n-1} & -\alpha\frac{\delta}{1-\delta}\sigma_w^2 A_w \cdot (\sum_{k=0}^{n-2}\phi_d^{n-2-k}\rho_w^k) \\ 0 & 0 & \rho_w^{n-1} \end{bmatrix} \begin{bmatrix} D_m^{(1)} \\ D_d^{(1)} \\ D_w^{(1)} \end{bmatrix}. \end{aligned}$$

It follows that

$$\begin{aligned} D_m^{(n)} - D_m^{(n-1)} &= \rho_m^{n-1} \cdot D_m^{(1)} > 0, \\ D_w^{(n)} - D_w^{(n-1)} &= \rho_w^{n-1} \cdot D_w^{(1)} > 0, \end{aligned}$$

and

$$\begin{aligned} D_d^{(n)} - D_d^{(n-1)} = -\alpha\frac{\delta}{1-\delta}\sigma_m^2 A_m \cdot (\textstyle\sum_{k=0}^{n-2}\phi_d^{n-2-k}\rho_m^k) \cdot D_m^{(1)} \\ + \phi_d^{n-1} \cdot D_d^{(1)} \\ - \alpha\frac{\delta}{1-\delta}\sigma_w^2 A_w \cdot (\textstyle\sum_{k=0}^{n-2}\phi_d^{n-2-k}\rho_w^k) \cdot D_w^{(1)} < 0. \end{aligned}$$

Thus, when $\gamma = 0$, $D_m^{(n)}$, $D_d^{(n)}$, and $D_w^{(n)}$ are monotonic functions of n with the same signs as A_m, A_d, and A_w. From here it also follows that $\mathcal{C}^{(n)}$ is increasing in n.

Further, we can also express the functions $A_m^{(n)}$, $A_d^{(n)}$, and $A_w^{(n)}$ After some algebra, we obtain the following closed-form expressions:

$$A_m^{(n)} = \frac{1}{1-\rho_m}\frac{1-\delta\rho_m}{1-\delta}A_m\frac{1-\rho_m^n}{n} > 0, \tag{B-23}$$

$$A_w^{(n)} = \frac{1}{1-\rho_w}\frac{1-\rho_w\delta}{1-\delta}A_w\frac{1-\rho_w^n}{n} > 0, \tag{B-24}$$

and

$$A_d^{(n)} = \theta_o \cdot \frac{1}{n}\frac{1-\phi_d^n}{1-\phi_d} + \theta_m \cdot S^{(n)}(\phi_d, \rho_m) + \theta_w \cdot S^{(n)}(\phi_d, \rho_w), \tag{B-25}$$

where

$$\begin{aligned}
\theta_o &= -\alpha\left(\kappa_x \frac{\sigma_o}{\kappa_o}\right)^2 < 0, \\
\theta_m &= -\alpha\frac{\delta}{1-\delta}\sigma_m^2 A_m = -\alpha\sigma_m^2\frac{\delta}{1-\delta\rho_m} < 0, \text{ and} \\
\theta_w &= -\alpha\frac{\delta}{1-\delta}\frac{1-\rho_w\delta}{1-\delta}A_w^2\sigma_w^2 = -\alpha\sigma_w^2\frac{\delta}{1-\rho_w\delta}\left(\alpha\kappa_x\left(\frac{\sigma_o}{\kappa_o}\right)^2\right)^2 < 0
\end{aligned}$$

and

$$S^{(n)}(a,b) \equiv \frac{1}{1-b}\frac{1}{n}\left(\frac{1-a^n}{1-a} - \frac{b^n-a^n}{b-a}\right) > 0.$$

■

In (B-25), the first term reflects a combination of (i) expected future compensation for using scarce capital and (ii) expected future compensation for risk due to shocks to z_t^d (recall that $\phi_d - \rho_d = -\alpha\frac{\delta}{1-\delta}\sigma_d^2 A_d > 0$) and (ii) . This term is only present when $\kappa_x > 0$, but exists even when $\sigma_m^2, \sigma_d^2, \sigma_w^2 \to 0$. The second two terms in square brackets reflect expected future compensation for risk due to shocks to z_t^m and z_t^w (as amplified amplified by the risk of shocks to z_t^d) and are only present when $\sigma_m^2 > 0$ and $\sigma_w^2 > 0$, respectively.

Turning to the shape of these these functions, $n^{-1}\sum_{k=0}^{n-1}\phi_d^k = n^{-1}(1-\phi_d^n)/(1-\phi_d)$ is a decreasing function of n when $\phi_d < 1$, increasing when $\phi_d > 1$, and constant when $\phi_d = 1$. $S^{(n)}(\phi_d, \rho_z)$ is a hump-shaped function of n when $\phi_d < 1$ and an increasing function of n when $\phi_d \geq 1$. Since $\phi_d > \rho_d$, this means that the last two terms tend to be increasing in n when ρ_d is large and hump-shaped in n when ρ_d is low. For instance, in the limit where $\sigma_d \to 0$, $\phi_d \to \rho_d < 1$, and both of these terms will be hump-shaped. This hump shape stems from the fact that $A_d^{(n)} = n^{-1}\sum_{k=1}^{n}\rho_d^{n-k}B_d^{(k)}$—i.e., the impact of a shift in demand on n-period swap spreads in an average of the expected risk premia (the $\rho_d^{n-k}B_d^{(k)}$) over the life of the swap. When demand shocks are sufficiently persistent (ρ_d is large), the $A_d^{(n)}$ coefficients will rise with n since the impact of demand on risk premia $B_d^{(n)}$ rises with n. However, when demand shocks are more transitory the $A_d^{(n)}$ coefficients become hump-shaped since these higher risk premia are only expected to persist for a short fraction over the life of the swap.

Proof. By continuity of the model's solution in γ, these same conclusions must also hold when $\gamma > 0$ is sufficiently near 0. However, these conclusions need hold for γ large. Specifically, when γ is large, one can construct extreme parameterizations where $D_w^{(n)}$ and $\mathcal{C}^{(n)}$ are not monotonic. For instance, take the special case of the model where $\sigma_m^2 = \sigma_d^2 = 0$ and $\sigma_w^2 > 0$—i.e., the limit with wealth shocks only. In this case, A_m and A_d are irrelevant, and A_w is the unique solution to

$$A_w = \frac{\alpha\kappa_x\left(\frac{\sigma_o}{\kappa_o}\right)^2}{\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma\left[\left(\kappa_x\frac{\sigma_o}{\kappa_o}\right)^2 + \left(\frac{\delta}{1-\delta}\right)^2 A_w^2\sigma_w^2\right]}.$$

Uniqueness follows from the fact that the right-hand side is positive, strictly decreasing in A_w, and converges to 0 as $A_w \to \infty$. In this case, the dynamics of the vector $\left[D_m^{(n)}, D_d^{(n)}, D_w^{(n)}\right]^T$ becomes

$$\begin{bmatrix} D_m^{(n)} \\ D_d^{(n)} \\ D_w^{(n)} \end{bmatrix} - \begin{bmatrix} D_m^{(n-1)} \\ D_d^{(n-1)} \\ D_w^{(n-1)} \end{bmatrix} = \overbrace{\begin{bmatrix} \rho_m^{n-1} & 0 & X^{(n-1)} \\ 0 & \rho_d^{n-1} & Y^{(n-1)} \\ 0 & 0 & \phi_w^{n-1} \end{bmatrix}}^{\Gamma_1^{n-1}} \begin{bmatrix} D_m^{(1)} \\ D_d^{(1)} \\ D_w^{(1)} \end{bmatrix},$$

where $\phi_w \equiv \rho_w - \alpha\gamma\frac{\delta}{1-\delta}\sigma_w^2 A_w^2$ and $D_w^{(1)} > 0$ (and $X^{(n-1)}$ and $Y^{(n-1)}$ are irrelevant constants obtained by computing Φ_1^{n-1}). Thus, if ρ_w is small and γ is large enough, we can construct equilibria where $\phi_w = \rho_w - \alpha\gamma\frac{\delta}{1-\delta}\sigma_w^2 A_w^2 < 0$. This means that $D_w^{(n)}$ will oscillate between increasing and decreasing as a function of n. As a result, $\mathcal{C}^{(n)} \equiv \frac{\delta}{1-\delta}D_w^{(n-1)}A_w\sigma_w^2$ will also oscillate between increasing and decreasing. One can construct similar examples where $D_m^{(n)}$ and $\mathcal{C}^{(n)}$ are not monotonic by, e.g., taking the limit where $\sigma_w^2 = \sigma_d^2 = 0$ and $\sigma_m^2 > 0$. Then, if ρ_m is sufficiently small and γ is sufficiently large, we obtain oscillatory patterns.

In the Online appendix, we provide closed-form solutions for $D_j^{(n)}$, $j \in \{m, n, w\}$, in the $\gamma > 0$ case, which illustrates the exact role of the elements of Φ_1^{n-1}. ■

Proof of Proposition 5. Combining (17b)-(18) with (B-21), we obtain

$$\begin{bmatrix} A_0 - A_0^{(1)} \\ A_m - A_m^{(1)} \\ A_d - A_d^{(1)} \\ A_w - A_w^{(1)} \end{bmatrix} = \begin{bmatrix} -\alpha\left(\bar{d} + \gamma A_0\right) V < 0 \\ -\left[\frac{\delta}{1-\delta}(1-\rho_m) + \alpha\gamma V\right] A_m < 0 \\ -\left[\frac{\delta}{1-\delta}(1-\rho_d) + \alpha\gamma V\right] A_d - \alpha V \\ -\left[\frac{\delta}{1-\delta}(1-\rho_w) + \alpha\gamma V\right] A_w < 0 \end{bmatrix}$$

From here, $A_d - A_d^{(1)} < 0$ iff

$$\frac{\delta}{1-\delta}(1-\rho_d)\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 < V, \tag{B-26}$$

which is more likely when $\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2$ is small and when ρ_d, σ_m^2, σ_d^2, and σ_w^2 are large. Thus, in our general model, the term structure is downward-sloping on average. Shocks to z_t^m and z_t^w have a larger impact on short-dated spreads than on longer-term spreads. Since these variables raise the level of spreads, positive values of z_t^m and z_t^w are associated with a more downward-sloping spread curve. By contrast, negative values of z_t^m and z_t^w are associated with a less downward-sloping spread curve. Finally, Shocks to z_t^d have an ambiguous implact the slope of the swap spread curve, with having a smaller (less negative) impact on short-dated spreads than longer-term spreads when (B-26). ■

Proof of Proposition 9. For the local slope of the swap spread curve, we are interested in

$$s_t^{(n)} - s_t^{(n-1)} = \overbrace{\left[A_0^{(n)} - A_0^{(n-1)}\right]}^{<0} + \overbrace{\left[A_m^{(n)} - A_m^{(n-1)}\right]}^{<0} z_t^m + \overbrace{\left[A_d^{(n)} - A_d^{(n-1)}\right]}^{>0} z_t^d + \overbrace{\left[A_w^{(n)} - A_w^{(n-1)}\right]}^{<0} z_t^w.$$

Focusing on the $\gamma = 0$ case, from (B-23)-(B-25), after some algebra, we obtain

$$A_m^{(n)} - A_m^{(n-1)} = \overbrace{\frac{1}{(1-\rho_m)}\left[\frac{1-\rho_m^n}{n} - \frac{1-\rho_m^{n-1}}{n-1}\right]}^{<0} \overbrace{\left[\frac{1-\delta\rho_m}{1-\delta} + \alpha\gamma V\right] A_m}^{>0} < 0, \tag{B-27}$$

$$A_w^{(n)} - A_w^{(n-1)} = \overbrace{\frac{1}{(1-\rho_w)}\left[\frac{1-\rho_w^n}{n} - \frac{1-\rho_w^n}{n-1}\right]}^{<0} \overbrace{\left[\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma V\right] A_w}^{>0} < 0,$$

and

$$\begin{aligned} A_d^{(n)} - A_d^{(n-1)} = {} & \overbrace{\frac{1}{(1-\phi_d)}\left[\frac{1-\phi_d^n}{n} - \frac{1-\phi_d^{n-1}}{n-1}\right]}^{<0} \overbrace{\left(\left[\frac{1-\rho_d\delta}{1-\delta} + \alpha\gamma V\right] A_d + \alpha V\right)}^{<0} \\ & - \overbrace{\frac{1}{(1-\phi_d)}\left[\begin{array}{c} \frac{1}{n}\left(\frac{1-\rho_m^n}{(1-\rho_m)} - \frac{\phi_d^n - \rho_m^n}{\phi_d - \rho_m}\right) \\ -\frac{1}{n-1}\left(\frac{1-\rho_m^{n-1}}{(1-\rho_m)} - \frac{\phi_d^{n-1} - \rho_m^{n-1}}{\phi_d - \rho_m}\right) \end{array}\right]}^{\text{ambiguous}} \cdot \overbrace{\alpha\frac{\delta}{1-\delta} A_m^2 \sigma_m^2 \left[\frac{1-\delta\rho_m}{1-\delta} + \alpha\gamma V\right]}^{>0} \\ & - \overbrace{\frac{1}{(1-\phi_d)}\left[\begin{array}{c} \frac{1}{n}\left(\frac{1-\rho_w^n}{(1-\rho_w)} - \frac{\phi_d^n - \rho_w^n}{\phi_d - \rho_w}\right) \\ -\frac{1}{n-1}\left(\frac{1-\rho_w^n}{(1-\rho_w)} - \frac{\phi_d^{n-1} - \rho_w^{n-1}}{\phi_d - \rho_w}\right) \end{array}\right]}^{\text{ambiguous}} \cdot \overbrace{\alpha\frac{\delta}{1-\delta} A_w^2 \sigma_w^2 \left[\frac{1-\rho_w\delta}{1-\delta} + \alpha\gamma V\right]}^{>0}, \end{aligned} \tag{B-28}$$

where the signs of the multiplicative constants are straightforward from (17b)-(17d), and the signs of the terms that depend on n follow from the following Lemma:

Lemma 1 *Suppose* $0 < a, b < 1$ *and* $n \in \mathbb{Z}, n \geq 1$. *Then*

$$\frac{1}{n}\frac{1-a^n}{1-a}$$

is positive and decreasing in n, *and*

$$f(n;a,b) \equiv \frac{1}{n}\left(\frac{1-a^n}{1-a} - \frac{b^n-a^n}{b-a}\right)$$

is positive, and either decreasing of hump-shaped in n.

Proof. Simple algebra shows that

$$\frac{1}{n}\frac{1-a^n}{1-a} = \frac{1}{n}\left[1+a+...+a^{n-1}\right], \tag{B-29}$$

where $1 > a^k > a^{k+1} > 0$ for all $k \geq 1$. Therefore, the RHS of (B-29) is trivially positive. Moreover,

$$\frac{1}{n+1}\left[1+a+...+a^{n-1}+a^n\right] = \frac{1}{n+1}\left[\frac{1+a+...+a^{n-1}}{n}n + a^n\right],$$

a higher n in (B-29) corresponds to taking the average of a series that includes smaller additional elements. Hence, $\frac{1}{n}\left[1+a+...+a^{n-1}\right] > \frac{1}{n+1}\left[1+a+...+a^n\right]$.

Similarly, we write

$$\frac{1}{n}\left(\frac{1-a^n}{1-a} - \frac{b^n-a^n}{b-a}\right) = \frac{1}{n}\left(1+a+...+a^{n-1} - \left[b^{n-1}+ab^{n-2}+...+a^{n-2}b+a^{n-1}\right]\right) \tag{B-30}$$

Since $0 < a, b < 1$, $0 < a^k b^l < a^k < 1$ for all $k, l \geq 1$. Therefore, the RHS of (B-30) is trivially positive. Moreover,

$$\begin{aligned}
\frac{df(n;a,b)}{dn} &= \frac{\left(\frac{-\ln a \cdot a^n}{1-a} - \frac{\ln b \cdot b^n - \ln a \cdot a^n}{b-a}\right)n - \left(\frac{1-a^n}{1-a} - \frac{b^n-a^n}{b-a}\right)}{n^2} \\
&= \frac{\left(\frac{1}{n}\ln 1 \cdot 1^n - \frac{1}{n^2}\cdot 1^n\right) - \left(\frac{1}{n}\ln a \cdot a^n - \frac{1}{n^2}a^n\right)}{1-a} - \frac{\left(\frac{1}{n}\ln b \cdot b^n - \frac{1}{n^2}\cdot b^n\right) - \left(\frac{1}{n}\ln a \cdot a^n - \frac{1}{n^2}a^n\right)}{b-a}
\end{aligned}$$

Note that these terms are actually the slopes of chords of the function

$$g(n;x) \equiv \frac{1}{n}\ln x \cdot x^n - \frac{1}{n^2}\cdot x^n$$

between 1 and a, and b and a, respectively. But the function $g(n;x)$ has (limit) values

$$\lim_{x \searrow 0} g(n;x) = \frac{1}{n}\lim_{x \searrow 0}\ln x \cdot x^n = \frac{1}{n}\lim_{x \searrow 0}\frac{\ln x}{x^{-n}} = \frac{1}{n}\frac{\frac{1}{x}}{-nx^{-n-1}} = -\frac{1}{n^2}\lim_{x \searrow 0} x^n = 0$$

and $g(n;1) = -\frac{1}{n^2}$, and its slope is

$$\begin{aligned}
\frac{d}{dx}g(n;x) &\equiv \frac{1}{n}\frac{d}{dx}(\ln x \cdot x^n) - \frac{1}{n^2}\cdot\frac{d}{dx}x^n = \frac{1}{n}\left(\frac{1}{x}x^n + n\ln x \cdot x^{n-1}\right) - \frac{1}{n^2}\cdot nx^{n-1} = \frac{1}{n}(1+n\ln x)x^{n-1} - \frac{1}{n}x^{n-1} \\
&= \ln x \cdot x^{n-1} < 0 \\
\frac{d^2}{dx^2}g(n;x) &= \frac{d}{dx}\left(\ln x \cdot x^{n-1}\right) = \left[1+(n-1)\ln x\right]x^{n-2} > 0 \text{ iff } x > e^{-\frac{1}{n-1}}
\end{aligned}$$

if $n \geq 2$ whereas if $n = 1$ we have

$$\frac{d^2}{dx^2}g(n;x) = \frac{d}{dx}\left(\ln x \cdot x^{n-1}\right) = \frac{1}{x} > 0$$

From here, we conclude that as long as a and b are close to 1, $f(n;a,b)$ is decreasing because the g function is convex on $(\min\{a,b\},1)$, otherwise $f(n;a,b)$ is hump-shaped. ■

From here, the sign of $A_d^{(n)} - A_d^{(n-1)}$ depends on the value of ϕ_d; as long as the highligted $f(n;\rho_m,\phi_d) - f(n-1;\rho_m,\phi_d)$ and $f(n;\rho_w,\phi_d) - f(n-1;\rho_w,\phi_d)$ terms in (B-28) are negative, $A_d^{(n)} - A_d^{(n-1)} > 0$; otherwise $A_d^{(n)}$ can be U-shaped and thus locally increasing (for long maturities). ■

Proof of Proposition 10. The expected returns on a position in the n-period swap spread arbitrage trade are given by

$$E_t[r_{t+1}^{Spread(n)}] = ns_t^{(n)} - (n-1)\,\mathrm{E}_t[s_{t+1}^{(n-1)}] - m_t = B_0^{(n)} + B_m^{(n)} z_t^m + B_d^{(n)} z_t^d + B_w^{(n)} z_t^w,$$

where

$$\begin{aligned}
B_0^{(n)} &= D_0^{(n)} - D_0^{(n-1)} - \overline{m} = B_0 + \alpha\left(\bar{d} + \gamma A_0\right)\cdot(V - \mathcal{C}^{(n)}) \\
B_m^{(n)} &= D_m^{(n)} - \rho_m D_m^{(n-1)} - 1 = B_m + \alpha\gamma A_m \cdot (V - \mathcal{C}^{(n)}) \\
B_d^{(n)} &= D_d^{(n)} - \rho_d D_d^{(n-1)} = B_d + \alpha\left(1 + \gamma A_d\right)\cdot(V - \mathcal{C}^{(n)}) \\
B_w^{(n)} &= D_w^{(n)} - \rho_w D_w^{(n-1)} = B_w + \alpha\gamma A_w \cdot (V - \mathcal{C}^{(n)}).
\end{aligned}$$

Thus, $B_z^{(n)} - B_z$ for $z \in \{0, m, d, w\}$ is proportional to $V - \mathcal{C}^{(n)} = \mathrm{Cov}_t[r_{t+1}^{Spread} - r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}]$. We have $B_0^{(1)} < 0$, $B_m^{(1)} \leq 0$ (< 0 when $\gamma > 0$), $B_d^{(n)} < 0$, and $B_w^{(1)} > 0$.

When $\mathcal{C}^{(n)} \equiv \mathrm{Cov}_t[r_{t+1}^{Spread(n)}, r_{t+1}^{Spread}]$ is a monotonically increasing function of n, it follows that $B_0^{(n)}$, $B_m^{(n)}$, $B_d^{(n)}$, and $B_w^{(n)}$ are all weakly decreasing in maturity n and strictly so when $\gamma > 0$. This result obtains because end-user demand for swaps, $d_t = \left(\overline{d} + \gamma A_0\right) + \gamma A_m z_t^m + \left(1 + \gamma A_d\right) z_t^d + \gamma A_w z_t^w$, is weakly increasing in all three state variables (strictly so when $\gamma > 0$). Since intermediaries must accomodate larger swap demand when any of these variables is high and because the returns on longer-date swaps co-move more strongly with intermediary wealth (under the assumption that $\mathcal{C}^{(n)}$ is increasing in n), expected returns must decline more steeply with maturity n when either z_t^m, z_t^d, and z_t^w are large. ■

Proof of Proposition 6. When we estimate regressions of the form

$$r_{t+1}^{Spread} = \alpha + \beta_1(s_t - s_t^{(1)}) + \xi_{t+1}^{(n)}$$

or

$$r_{t+1}^{Spread(n)} = \alpha + \beta_2(s_t^{(n)} - s_t^{(1)}) + \xi_{t+1}^{(n)},$$

we find that under mild conditions the slope of the spread curve positively predicts future returns to spread arbitrage on long-term swaps. For instance, a larger negative slope ($s_t^{(n)} - s_t^{(1)}$ more negative) is associated with more negative values of $r_{t+1}^{Spread(n)}$. Since

$$\begin{aligned}
\mathrm{E}_t[r_{t+1}^{Spread(n)}] &= -\alpha\mathcal{C}^{(n)}\left(\overline{d} + z_t^d\right) - \kappa_x\psi\left(\overline{w} + z_t^w, \overline{d} + z_t^d\right) \\
&= \underbrace{\left[-\alpha\mathcal{C}^{(n)}\overline{d} - \kappa_x\psi\left(\overline{w}, \overline{d}\right)\right]}_{B_0^{(n)} < 0.} + \underbrace{\left[-\alpha\mathcal{C}^{(n)} - \kappa_x\psi_2\right]}_{B_d^{(n)} < 0} z_t^d + \underbrace{\left[-\kappa_x\psi_1\right]}_{B_w^{(n)} = B_w > 0} z_t^w,
\end{aligned}$$

the model-implied regression results are given by

$$\beta_1 = \frac{\overbrace{[A_d - A_d^{(1)}]}^{\text{Ambiguous}} \times \overbrace{[B_d]}^{<0} \times \frac{\sigma_d^2}{1-\rho_d^2} + \overbrace{[A_w - A_w^{(1)}]}^{<0} \times \overbrace{[B_w]}^{>0} \times \frac{\sigma_w^2}{1-\rho_w^2}}{[A_d - A_d^{(1)}]^2 \times \frac{\sigma_d^2}{1-\rho_d^2} + [A_w - A_w^{(1)}]^2 \times \frac{\sigma_w^2}{1-\rho_w^2}}.$$

and

$$\beta_2 = \frac{\overbrace{[A_d^{(n)} - A_d^{(1)}]}^{\text{Ambiguous}} \times \overbrace{[B_d^{(n)}]}^{<0} \times \frac{\sigma_d^2}{1-\rho_d^2} + \overbrace{[A_w^{(n)} - A_w^{(1)}]}^{<0} \times \overbrace{[B_w^{(n)}]}^{>0} \times \frac{\sigma_w^2}{1-\rho_w^2}}{[A_d^{(n)} - A_d^{(1)}]^2 \times \frac{\sigma_d^2}{1-\rho_d^2} + [A_w^{(n)} - A_w^{(1)}]^2 \times \frac{\sigma_w^2}{1-\rho_w^2}},$$

respectively.

Coefficient β_1 is, for example, positive when

$$(A_d - A_d^{(1)})B_d \frac{\sigma_d^2}{1-\rho_d^2} + (A_w - A_w^{(1)})B_w \frac{\sigma_w^2}{1-\rho_w^2} > 0$$

or

$$\begin{aligned} 0 \quad < \quad & -\left(\frac{\delta\,(1-\rho_d)}{1-\rho_d\delta}\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right] - V\right)\left[\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 + V\right]\frac{\sigma_d^2}{1-\rho_d^2} \\ & -\frac{\delta\,(1-\rho_w)}{1-\rho_w\delta}\left[\frac{1}{\kappa_x}\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2\right]^2 \frac{\sigma_w^2}{1-\rho_w^2}. \end{aligned} \tag{B-31}$$

In particular, notice that as long as

$$\frac{\delta}{1-\delta}\,(1-\rho_d)\left(\frac{\kappa_x}{\kappa_o}\right)^2 \sigma_o^2 > V,$$

which is more likely to hold when ρ_d and V are small, (B-31) cannot hold as the RHS is negative for sure.

For β_2, if an increase in z_t^d reduces $s_t^{(n)} - s_t^{(1)}$ (i.e., $A_d^{(n)} - A_d^{(1)} < 0$)—which is more likely when V is large, ρ_d is fairly persistent, and σ_o^2 is not too large, then β is the sum of a negative piece and a positive piece. Which term dominates depends on whether demand or supply shocks are more volatile. ■

C Additional VARs

To check that our identification of demand and supply shocks is not sensitive to the inclusion of the LIBOR-repo spread as an additional variable in the VAR, we consider the following tri-variate specification:

$$\begin{bmatrix} \text{LIBOR-repo}_t \\ \text{30y swap spread}_t \\ -\text{PD net position}_t \end{bmatrix} = \mathbf{c} + \sum_{i=1}^{L} \mathbf{C}_i \begin{bmatrix} \text{LIBOR-repo}_{t-i} \\ \text{30y swap spread}_{t-i} \\ -\text{PD net position}_{t-i} \end{bmatrix} + \xi_t.$$

Following Proposition 2, we identify the structural demand, supply, and LIBOR-repo shocks by imposing a combination of sign and zero restrictions. Specifically, in addition to the structural shock orthogonality and the sign restrictions that we imposed in our baseline VAR, we assume that LIBOR-repo spread does not respond on impact to demand and supply shocks, and that both spreads' on-impact responses to LIBOR-repo shock have the same sign. Thus, structural shocks ε_t are related to reduced-form VAR residuals ξ_t by the mapping

$$\xi_t = \begin{bmatrix} + & 0 & 0 \\ + & - & + \\ . & - & - \end{bmatrix} \varepsilon_t.$$

We estimate the structural VAR using the sign and zero restriction approach of Arias et al. (2018) with lag length set to $L = 4$. The historical decomposition implied by the estimated VAR and shown on Figure A2 suggests that LIBOR-repo shocks contribute very little to the swap spread variation. This is confirmed by the forecast error variance decomposition: LIBOR-repo shocks account for just 2.88% of the swap spread variance in the long run.

To study maturity-specific end-user demand to receive fixed rate, we consider the following VAR specification:

$$\begin{bmatrix} \text{5y swap spread}_t \\ \text{30y swap spread}_t \\ -\text{PD net position}_t \\ -\text{PD net position LT - ST}_t \end{bmatrix} = \mathbf{c} + \sum_{i=1}^{L} \mathbf{C}_i \begin{bmatrix} \text{5y swap spread}_{t-i} \\ \text{30y swap spread}_{t-i} \\ -\text{PD net position}_{t-i} \\ -\text{PD net position LT - ST}_{t-i} \end{bmatrix} + \xi_t.$$

We identify the structural short-maturity demand, long-maturity demand, and supply shocks by imposing a combination of sign restrictions. First, we assume that positive demand shocks (both short-maturity and long-maturity) make swap spreads (both 5-year and 30-year) more negative and increase the scale of intermediaries' overall short position in swap arbitrage, pushing down $-$PD net position$_t = -$ (PD net position>6y$_t$ + PD net position<6y$_t$). Moreoever, we assume that shocks to short-maturity demand have a larger effect on intermediaries' position in short-maturity Treasuries relative to long-maturity Treasuries and vice versa: a positive short-maturity demand shock pushes $-$PD net position LT - ST$_t = -$ (PD net position>6y$_t$ $-$ PD net position<6y$_t$) up, while a positive long-maturity demand shock pushes it down. We also assume that a positive shock to intermediary supply makes swap spreads (both 5-year and 30-year) less negative, while increasing intermediaries' overall short position in the swap arbitrage. The fourth shock in the VAR is not identified. Thus, structural shocks ε_t are related to reduced-form VAR residuals ξ_t by the mapping

$$\xi_t = \begin{bmatrix} - & - & + & . \\ - & - & + & . \\ - & - & - & . \\ + & - & . & . \end{bmatrix} \varepsilon_t.$$

We estimate the structural VAR with the pure sign restrictions approach of Uhlig (2005) with lag length $L = 2$.

D Tables

Table 1: **Summary statistics:** This table reports the means, the standard deviations, and the correlations with the 30-year swap spread of the 30-year swap spread (Swap spread), the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), the Primary Dealers' net position in coupon-bearing Treasury securities (PD net), the Primary Dealers' gross position in coupon-bearing Treasury securities (PD gross), and the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - repo). Data are weekly and run from July 2001 to December 2008 (01-08) and from January 2009 to December 2020 (09-20).

	Mean		St. dev.		Corr. with swap spread	
	01-08	09-20	01-08	09-20	01-08	09-20
Swap spread, bps	45	-25	16	15	1.00	1.00
Repo - OIS, bps	-8	7	12	8	-0.41	-0.47
PD net, bln	-95.63	59.32	36.12	64.46	-0.27	-0.31
PD gross, bln	350.02	392.41	64.73	60.09	0.15	0.05
LIBOR - repo, bps	35	18	44	18	-0.35	-0.05

Table 2: **Demand and supply drivers:** This table reports the slope and intercept coefficients from regressions of, respectively, the 30-year swap spread (regressions 1-2), the demand factor (regressions 2-4) and the supply factor (regression 5) on the Adrian et al. (2013) term premium (Term premium), the VXTY Treasury volatility index (Treasury volatility), the 10-year U.S. sovereign credit default swap spread (Sovereign risk), the Hu et al. (2013) yield curve fitting error (Yield curve noise), the VIX volatility index (Stock volatility), the aggregate dollar duration of mortgage-backed securities (Mortgage duration), the Klingler and Sundaresan (2019) pension fund underfunding factor (Pension underfunding), and the U.S. corporate bond issuance (Corporate issuance). In regressions (1), (3) and (5), data are weekly and run from January 2009 to June 2018. In regressions (2) and (4), data are quarterly and run from January 2009 to June 2018, and variables are in 3-month changes. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Swap spread		Demand		Supply
	(1)	(2)	(3)	(4)	(5)
Term premium	10.552***		−4.561***		0.553
	(2.032)		(0.606)		(0.506)
Treasury volatility	−3.441***		0.634*		−1.037***
	(1.309)		(0.375)		(0.351)
Sovereign risk	−0.205**		0.079***		−0.020
	(0.083)		(0.023)		(0.023)
Yield curve noise	−0.357		0.104		−0.069
	(0.852)		(0.252)		(0.180)
Stock volatility	0.175		−0.051		0.034
	(0.245)		(0.065)		(0.069)
Mortgage duration		0.624***		−0.257***	
		(0.153)		(0.081)	
Pension underfunding		30.075		15.230	
		(67.376)		(23.573)	
Corporate issuance		−8.427		12.540	
		(34.612)		(10.451)	
Constant	−7.993	0.361	−1.445	0.175	6.068***
	(5.248)	(1.469)	(1.458)	(0.401)	(1.391)
Observations	495	37	495	37	495
Adjusted R^2	0.232	0.020	0.461	0.113	0.162

Table 3: **Swap spread trade returns predictability:** This table reports the slope and intercept coefficients from regressions of, respectively, the 3-month holding period return on the 30-year swap spread trade (regressions 1-4), and the 12-month holding period return on the 30-year swap spread trade (regressions 5-8) on the Primary Dealers' net position in coupon-bearing Treasury securities (PD net), the Primary Dealers' outstanding balances of Treasury securities in through financing arrangements (PD securities in), the demand (Demand) and supply (Supply) factors, the difference between the 30-year swap spread and the 10-year swap spread (Slope), the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), and the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - repo). Data are weekly and run from January 2009 to June 2018. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	3-month returns				12-month returns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PD net	−1.597***	−1.618***			−5.308***	−5.414***		
	(0.384)	(0.383)			(1.004)	(0.825)		
PD securities in		−0.242*				−1.230***		
		(0.137)				(0.197)		
Demand			−12.920***				−52.809***	
			(2.876)				(5.252)	
Supply			−2.905				6.192	
			(3.197)				(7.665)	
Slope				6.502***				20.974***
				(1.346)				(2.516)
Repo - OIS	3.176	2.585		1.960	3.329	0.333		2.282
	(3.218)	(2.976)		(2.043)	(5.888)	(5.162)		(3.475)
LIBOR - repo	−1.577	−1.750	−1.451	0.458	−5.660***	−6.539***	−4.766***	−0.762
	(1.267)	(1.250)	(1.249)	(1.121)	(1.479)	(1.375)	(1.366)	(1.635)
Constant	46.232*	506.807*	8.583	162.787***	217.122***	2555.930***	47.988	603.727***
	(27.905)	(258.198)	(27.795)	(39.312)	(63.203)	(389.578)	(45.193)	(83.539)
Observations	495	495	495	615	495	495	495	576
Adjusted R^2	0.118	0.145	0.148	0.123	0.291	0.437	0.552	0.331

Table 4: **Swap spread trade returns predictability, additional results:** This table reports the slope and intercept coefficients from regressions of, respectively, the 3-month holding period return on the 30-year swap spread trade (regressions 1-4) and the 12-month holding period return on the 30-year swap spread trade (regressions 5-8) on Primary Dealers' net position in coupon-bearing Treasury securities (PD net), Primary Dealers' net position in coupon-bearing Treasury securities with maturity of 11 years and higher (PD net 11y+), Primary Dealers' gross position in coupon-bearing Treasury securities (PD gross), the Primary Dealers' net position in federal agency and government sponsored enterprise mortgage-backed securities (PD net, agency MBS), the Primary Dealers' net position in federal agency and government sponsored enterprise securities excluding mortgage-backed securities (PD net, agency ex. MBS), the Primary Dealers' net position in corporate debt securities (PD net, corporate), the difference between the 30-year swap spread and the 10-year swap spread (Slope), the spread between the 3-month LIBOR rate and the 3-month Treasury bill rate (TED), the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), and the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - repo). Data are weekly and run from January 2009 to June 2018 for regressions 1-3 and 5-7, and from January 2009 to December 2020 for regressions 4 and 8. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	3-month returns				12-month returns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PD net	−1.797***			−2.889***	−5.048***			−7.307***
	(0.392)			(0.571)	(1.101)			(1.221)
PD net 11y+		−0.633***				−1.202**		
		(0.159)				(0.527)		
PD gross			−0.530**				−1.999***	
			(0.238)				(0.392)	
PD securities in				−0.263*				−0.986***
				(0.141)				(0.199)
PD net, agency MBS				0.001				−0.002
				(0.001)				(0.002)
PD net, agency ex. MBS				0.002*				−0.001
				(0.001)				(0.002)
PD net, corporate				−0.004**				−0.003
				(0.002)				(0.003)
TED	4.175*				−0.733			
	(2.315)				(4.283)			
Repo - OIS		2.604	−1.164	−0.412		−2.855	−11.435*	−1.316
		(3.094)	(2.885)	(2.931)		(6.594)	(6.155)	(6.418)
LIBOR - repo	−5.395**	−1.366	−2.246*	−1.584	−5.036	−5.188***	−8.206***	−5.498***
	(2.390)	(1.290)	(1.269)	(1.122)	(4.780)	(1.693)	(1.805)	(1.264)
Constant	9.473	70.621**	232.566**	672.790***	236.110***	214.735***	935.760***	2494.352***
	(34.166)	(31.374)	(95.472)	(238.597)	(64.479)	(73.666)	(173.351)	(384.776)
Observations	495	495	495	495	495	495	495	495
Adjusted R^2	0.140	0.097	0.050	0.205	0.289	0.127	0.160	0.468

Table 5: **Covered interest parity violations:** This table reports the slope and intercept coefficients from regressions of, respectively, the deviation from the 3-month EURUSD covered interest parity (CIP 3m, regressions 1-4) and the deviation from the 5-year EURUSD covered interest parity (CIP 5y, regressions 5-8) on the 30-year swap spread, the demand (Demand) and supply (Supply) factors, the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - Repo), the deviation from the 3-month EURUSD covered interest parity (CIP 3m), and the deviation from the 5-year EURUSD covered interest parity (CIP 5y). Data are monthly and run from January 2009 to June 2018. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	CIP 3m				CIP 5y			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Swap spread	0.675***				0.610***			
	(0.172)				(0.078)			
Demand		−0.770	−0.863	1.563		−1.736***	−1.487***	−1.527***
		(0.524)	(0.563)	(0.931)		(0.259)	(0.234)	(0.196)
Supply		2.338***	2.257***	1.624***		0.531*	0.407	−0.103
		(0.673)	(0.657)	(0.468)		(0.311)	(0.292)	(0.244)
Repo - OIS			0.189				−0.381**	
			(0.370)				(0.168)	
LIBOR - Repo			−0.266				−0.265***	
			(0.227)				(0.066)	
CIP 3m								0.271***
								(0.041)
CIP 5y				1.344***				
				(0.451)				
Constant	−16.525***	−32.486***	−28.197***	7.931	−15.852***	−30.068***	−23.183***	−21.258***
	(4.232)	(3.468)	(4.020)	(11.618)	(2.521)	(1.364)	(1.563)	(1.609)
Observations	113	113	113	113	113	113	113	113
Adjusted R^2	0.192	0.215	0.244	0.496	0.481	0.528	0.654	0.697

E Figures

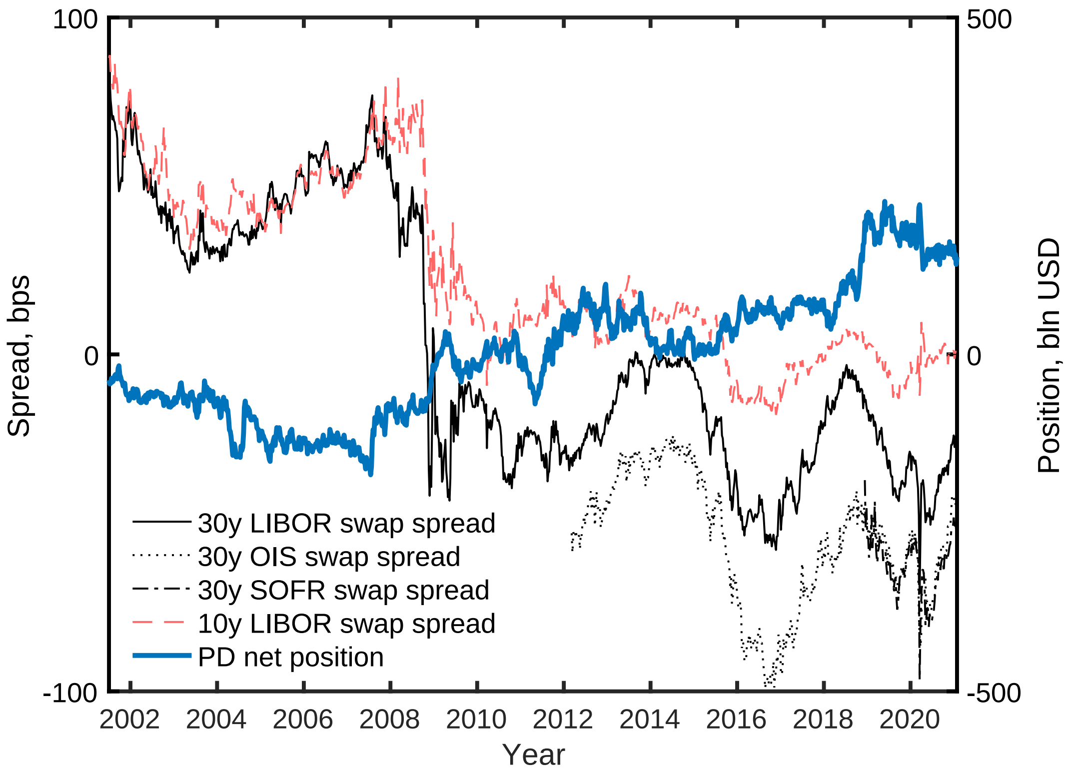

Figure 1: **Swap spread and Primary Dealers' position:** This figure shows the 30-year LIBOR swap spread (30y swap spread), the 10-year LIBOR swap spread (10y swap spread), the 30-year OIS swap spread (30y OIS swap spread), and the Primary Dealers' net position in coupon-bearing Treasury securities (PD net position). Data are weekly and run from January 2001 to December 2020.

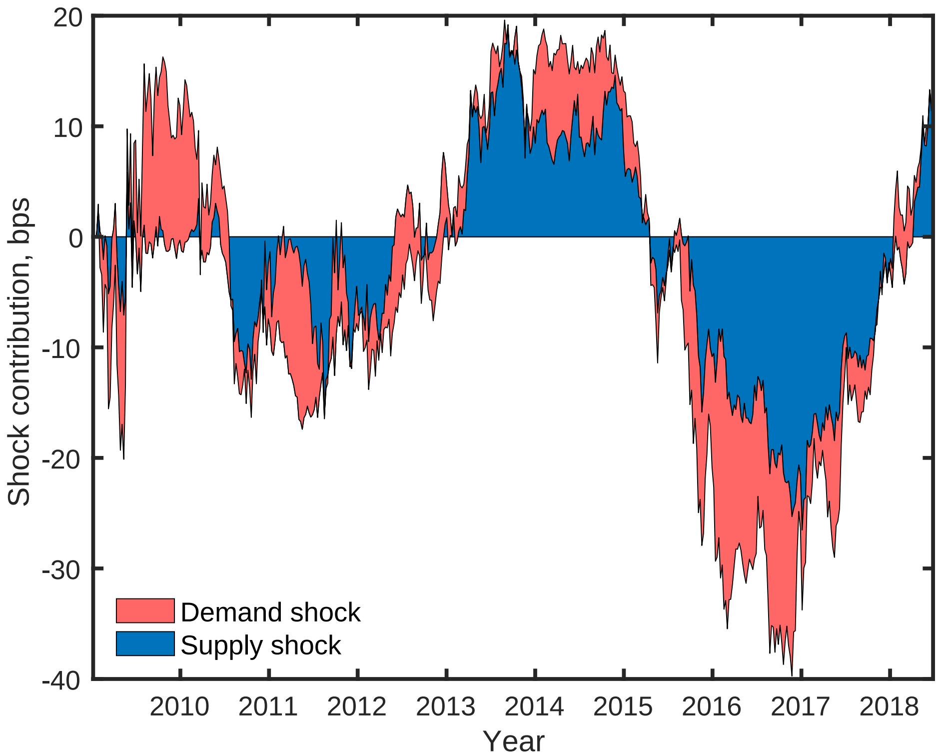

Figure 2: **Swap spread historical decomposition:** This figure shows the contribution of demand and supply shocks to the 30-year swap spread variation. Data are weekly and run from January 2009 to June 2018.

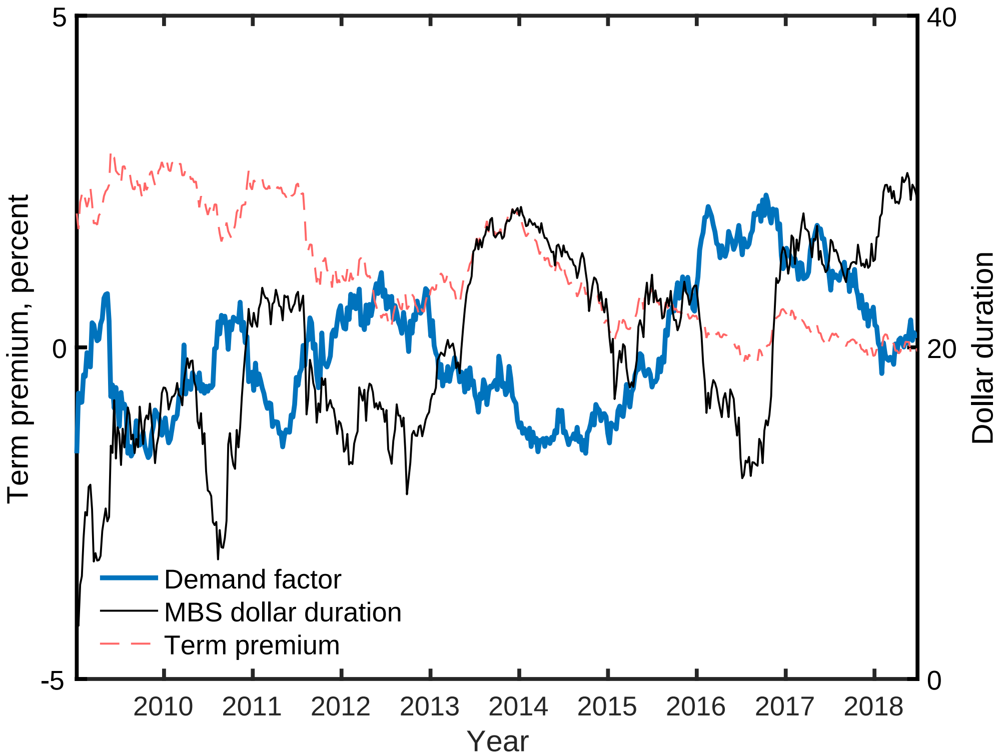

Figure 3: **Demand factor and aggregate MBS dollar duration:** This figure shows the series of the demand component of the 30-year swap spread (Demand factor), the aggregate dollar duration of U.S. mortgage-backed securities (MBS dollar duration), and the Adrian et al. (2013) term premium (Term premium). Data are weekly and run from January 2009 to June 2018.

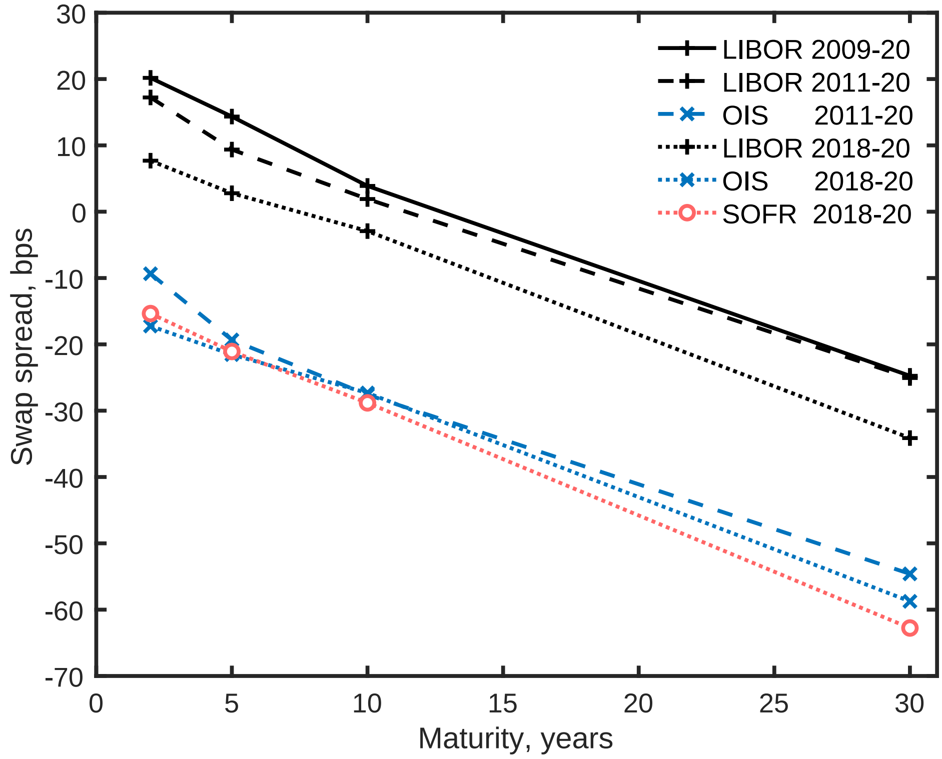

Figure 4: **Term structure of swap spreads:** This figure shows the average LIBOR, OIS, and SOFR swap spreads for 2-, 5-, 10- and 30-year maturities for periods starting in January 2009 (2009-20), September 2011 (2011-20) and December 2018 (2018-20), and ending in December 2020. Data are weekly.

F Additional tables and figures

Table A1: **Demand and supply properties:** This table reports the pairwise correlations of the 30-year swap spread (Swap spread), the demand factor (Demand) and minus one times the supply factor (-Supply) with, respectively, the Adrian et al. (2013) term premium (Term premium), the VXTY Treasury volatility index (Treasury volatility), the 10-year U.S. sovereign credit default swap spread (Sovereign risk), the Hu et al. (2013) yield curve fitting error (Yield curve noise), the VIX volatility index (Stock volatility). The supply factor is multiplied by minus one so higher values of both demand and supply factors correspond to more negative swap spreads. Data are weekly and run from January 2009 to June 2018.

	Swap spread	Demand	-Supply
Term premium	0.33	−0.62	0.19
Treasury volatility	−0.02	−0.27	0.39
Sovereign risk	−0.12	0.02	0.19
Noise	−0.05	−0.15	0.29
Stock volatility	−0.05	−0.16	0.30
Swap spread	1.00	−0.87	−0.76

Table A2: **Swap spread trade returns predictability, sample extended until end of 2020:** This table reports the slope and intercept coefficients from regressions of, respectively, the 3-month holding period return on the 30-year swap spread trade (regressions 1-4) and the 12-month holding period return on the 30-year swap spread trade (regressions 5-8) on Primary Dealers' net position in coupon-bearing Treasury securities (PD net), the Primary Dealers' outstanding balances of Treasury securities in through financing arrangements (PD securities in), Primary Dealers' net position in coupon-bearing Treasury securities orthogonalized with respect to a time trend, coupon-bearing Treasury securities held by the Federal Reserve and the 12-month average of coupon-bearing Treasury securities net issuance ($\overline{\text{PD net}}$), the Primary Dealers' outstanding balances of Treasury securities in through financing arrangements orthogonalized with respect to a time trend, coupon-bearing Treasury securities held by the Federal Reserve and the 12-month average of coupon-bearing Treasury securities net issuance ($\overline{\text{PD securities in}}$), the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), and the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - Repo). Data are weekly and run from January 2009 to December 2020. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	3-month returns				12-month returns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PD net	−0.116	−0.015			−0.789	−0.455		
	(0.382)	(0.370)			(0.800)	(0.785)		
PD securities in		−0.183				−1.083***		
		(0.122)				(0.208)		
$\overline{\text{PD net}}$			−0.842*	−0.724			−3.759***	−2.814***
			(0.467)	(0.487)			(0.884)	(0.727)
$\overline{\text{PD securities in}}$				−0.249**				−1.601***
				(0.122)				(0.202)
Repo - OIS	0.862	1.083	1.075	1.514	1.320	1.885	1.206	1.841
	(2.956)	(2.870)	(2.257)	(2.167)	(6.059)	(5.853)	(4.592)	(3.755)
LIBOR - repo	−1.073	−1.087	−0.778	−0.723	−6.231***	−7.327***	−4.473***	−5.178***
	(1.136)	(1.177)	(1.126)	(1.171)	(1.599)	(1.435)	(1.609)	(1.502)
Constant	12.941	353.357	−0.337	−4.342	144.273**	2182.976***	69.743	78.899
	(27.377)	(226.592)	(27.317)	(27.189)	(61.073)	(404.179)	(59.952)	(53.425)
Observations	615	615	614	614	576	576	576	576
Adjusted R^2	0.007	0.021	0.023	0.044	0.069	0.185	0.140	0.339

Table A3: **Swap spread trade returns predictability, pre-2008 sample:** This table reports the slope and intercept coefficients from regressions of, respectively, the 3-month holding period return on the 30-year swap spread trade (regressions 1-4) and the 12-month holding period return on the 30-year swap spread trade (regressions 5-8) on Primary Dealers' net position in coupon-bearing Treasury securities (PD net), the Primary Dealers' outstanding balances of Treasury securities in through financing arrangements (PD securities in), the Primary Dealers' net position in federal agency and government sponsored enterprise mortgage-backed securities (PD net, agency MBS), the Primary Dealers' net position in federal agency and government sponsored enterprise securities excluding mortgage-backed securities (PD net, agency ex. MBS), the Primary Dealers' net position in corporate debt securities (PD net, corporate), the difference between the 30-year swap spread and the 2-year swap spread (Slope), the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), and the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - repo). Data are weekly and run from December 2001 to December 2007. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	3-month returns				12-month returns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PD net		−1.426***	−1.420***			−1.465**	−1.474*	
		(0.402)	(0.386)			(0.690)	(0.769)	
PD securities in		−0.210***	−0.231***			−0.528***	−0.475***	
		(0.042)	(0.057)			(0.088)	(0.091)	
PD net, agency MBS			−1.709**				−2.804**	
			(0.837)				(1.091)	
PD net, agency ex. MBS			0.516				0.212	
			(1.305)				(1.606)	
PD net, corporate			0.214				−0.511	
			(0.672)				(0.863)	
Slope				5.067***				16.140***
				(1.492)				(2.464)
Repo - OIS	3.749**	0.889	1.105	5.978***	16.571***	−8.405**	−10.009**	19.959***
	(1.577)	(2.196)	(2.074)	(1.625)	(5.818)	(4.267)	(4.085)	(4.244)
LIBOR - repo	1.809***	2.821***	2.851***	3.397***	13.122	3.800	1.286	10.645*
	(0.673)	(0.527)	(0.768)	(0.753)	(8.398)	(5.675)	(5.553)	(6.012)
Constant	0.716	237.731***	248.375***	−51.292***	−130.465	752.079***	799.230***	−168.972*
	(18.926)	(61.364)	(85.317)	(19.642)	(119.411)	(155.517)	(156.551)	(89.232)
Observations	304	304	304	304	265	265	265	265
Adjusted R^2	0.042	0.237	0.260	0.168	0.108	0.488	0.521	0.455

Table A4: **Swap spread trade returns predictability, 10-year swap spread:** This table reports the slope and intercept coefficients from regressions of, respectively, the 3-month holding period return on the 10-year swap spread trade (regressions 1-4), and the 12-month holding period return on the 10-year swap spread trade (regressions 5-8) on the Primary Dealers' net position in coupon-bearing Treasury securities (PD net), the Primary Dealers' outstanding balances of Treasury securities in through financing arrangements (PD securities in), the Primary Dealers' net position in coupon-bearing Treasury securities with maturity of 11 years and higher (PD net 11y+), the demand (Demand) and supply (Supply) factors, the difference between the 10-year swap spread and the 2-year swap spread (Slope), the spread between the 3-month general collateral repo rate and the 3-month OIS rate (Repo - OIS), and the spread between the 3-month USD LIBOR rate and the 3-month general collateral repo rate (LIBOR - repo). Data are weekly and run from January 2009 to June 2018. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	3-month returns				12-month returns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PD net	-0.195				-1.098^{***}			
	(0.141)				(0.222)			
PD net 11y+		-1.653^{***}				-2.800^{***}		
		(0.492)				(0.999)		
PD securities in	-0.149^{***}	-0.164^{***}			-0.334^{***}	-0.353^{***}		
	(0.049)	(0.049)			(0.074)	(0.080)		
Demand			-2.492^{**}				-13.113^{***}	
			(1.044)				(1.442)	
Supply			1.923				12.167^{***}	
			(1.512)				(2.718)	
Slope				1.027^{**}				4.328^{***}
				(0.511)				(0.689)
Repo - OIS	-0.750	0.236		0.623	0.464	-0.717		2.013
	(0.967)	(0.874)		(1.034)	(1.614)	(1.634)		(1.485)
LIBOR - repo	-0.412	-0.375	-0.237	0.063	0.610	0.701	-0.307	2.410^{***}
	(0.330)	(0.316)	(0.319)	(0.365)	(0.518)	(0.542)	(0.487)	(0.559)
Constant	294.413^{***}	333.373^{***}	5.249	14.116	651.079^{***}	690.455^{***}	10.590	22.967
	(93.312)	(94.366)	(9.092)	(10.531)	(145.890)	(160.739)	(15.436)	(17.557)
Observations	495	495	495	495	495	495	495	495
Adjusted R^2	0.093	0.122	0.072	0.026	0.273	0.190	0.445	0.165

Table A5: **Short-term and long-term demand factors:** This table reports the slope and intercept coefficients from regressions of, respectively, the 5-year swap spread (5y spread), the 30-year swap spread (30y spread), the short-term demand factor (ST demand) and the long-term demand factor (LT demand) on the aggregate dollar duration of mortgage-backed securities (Mortgage duration), the Klingler and Sundaresan (2019) pension fund underfunding factor (Pension underfunding), and the U.S. corporate bond issuance (Corporate issuance). Data are quarterly and run from January 2009 to June 2018, and variables are in 3-month changes. Newey and West (1987, 1994) standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	5y spread	30y spread	ST demand	LT demand
Mortgage duration	-0.278	0.624^{***}	-0.131^{*}	-0.222^{**}
	(0.311)	(0.153)	(0.070)	(0.093)
Pension underfunding	214.389^{***}	30.075	-37.326	48.972^{**}
	(71.112)	(67.376)	(26.344)	18.413
Corporate issuance	24.597	-8.427	-0.217	15.848
	(38.446)	(34.612)	(8.675)	(11.866)
Constant	-0.216	0.361	0.197	0.053
	(1.065)	(1.469)	(0.365)	(0.447)
Observations	37	37	37	37
Adjusted R^2	0.137	0.020	0.010	0.177

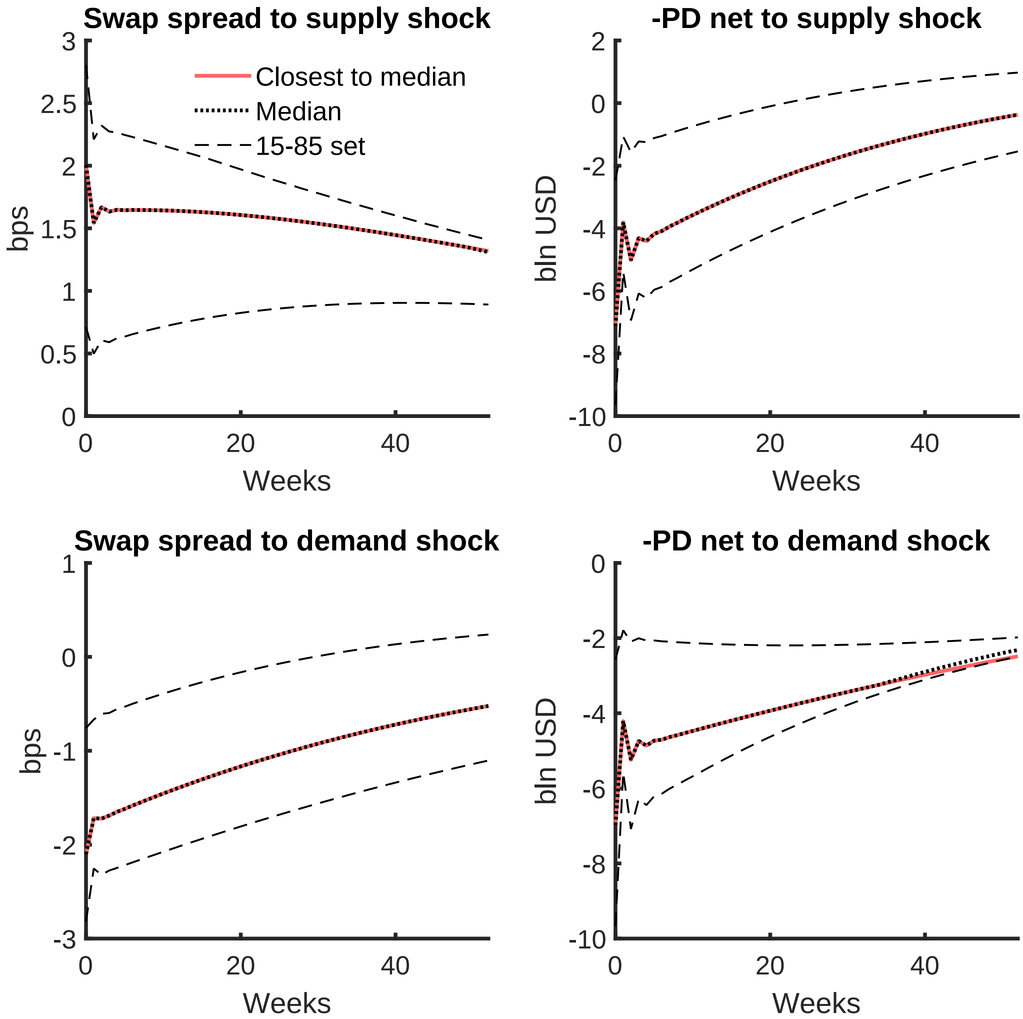

Figure A1: **Impulse response functions and factors:** The figure show the impulse response functions from the structural VAR. Median, "closest to median," and 15th-85th percentiles correspond to the set of responses that satisfy the identification restrictions. Data are weekly and run from January 2009 to June 2018.

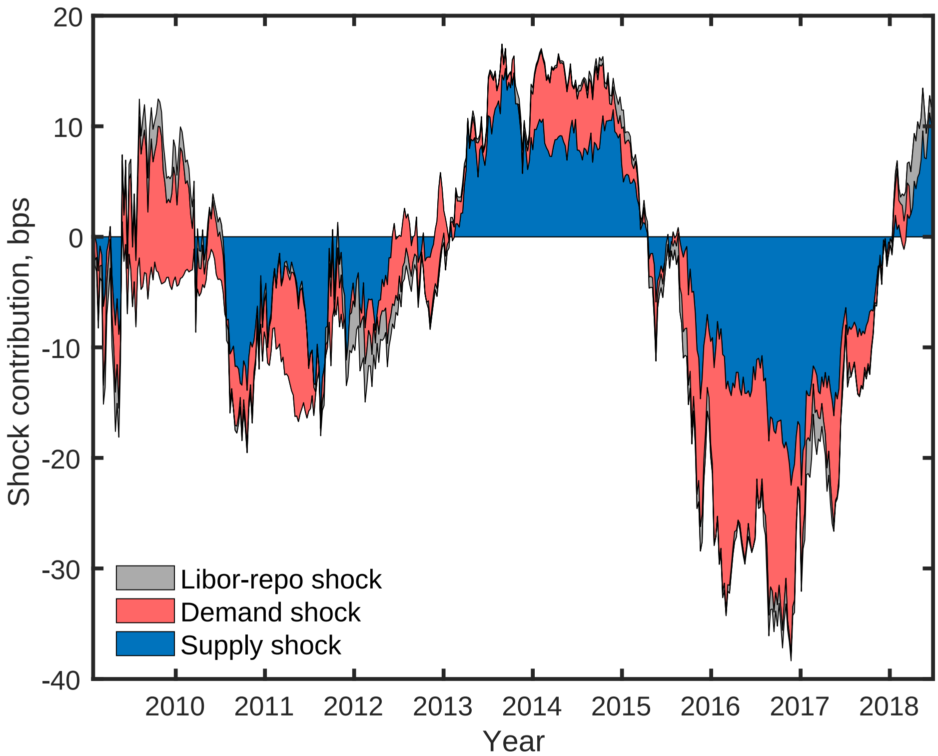

Figure A2: **Swap spread historical decomposition:** This figure shows the contribution of LIBOR-repo, demand, and supply shocks to the 30y swap spread. Shocks are identified using the structural VAR described in Appendix C. Data are weekly and run from January 2009 to June 2018.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Economic and Social Research Council

Systemic Risk Centre

The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk