Quantitative Easing, the Repo Market, and the Term Structure of Interest Rates

RUGGERO JAPPELLI*

LORIANA PELIZZON†

MARTI G. SUBRAHMANYAM‡

ABSTRACT

We develop a quantity-driven general equilibrium model that integrates the term structure of interest rates with the repurchase agreements (repo) market to shed light on the combined effects of quantitative easing (QE) on the bond and money markets. We characterize in closed form the endogenous dynamic interaction between bond prices and repo rates, and show (i) that repo specialness dampens the impact of any given quantity of asset purchases due to QE on the slope of the term structure and (ii) that bond scarcity resulting from QE increases repo specialness, thus strengthening the local supply channel of QE.

JEL Codes: E43, E52, G12

Keywords: Term Structure of Interest Rates, Repo Specialness, Money Market, Quantitative Easing.

---

*Leibniz Institute for Financial Research SAFE, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 3, 60323, Frankfurt am Main, Germany, jappelli@safe-frankfurt.de.

†Leibniz Institute for Financial Research SAFE, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 3, 60323, Frankfurt am Main, Germany, Ca' Foscari University of Venice, Dorsoduro, 3246, 30123 Venezia, Italy, and CEPR, pelizzon@safe.uni-frankfurt.de.

‡New York University, Leonard N. Stern School of Business and NYU Shanghai, Kaufman Management Center, 44 West Fourth Street, 9-68, 10012, New York, msubrahm@stern.nyu.edu.

We thank Giovanni Dell'Ariccia, Wenxin Du, Darrell Duffie, Ester Faia, Matthias Fleckenstein, Robin Greenwood, Zhiguo He (discussant), Florian Heider, Yesol Huh, Sebastian Infante, Urban Jermann, Francis Longstaff, Elisa Luciano (discussant), Toomas Laarits, Errikos Melissinos, Andrea Modena, Cecilia Parlatore, Pietro Reggiani, Stephen Schaefer, Fabian Smetak (discussant), Claudio Tebaldi, Davide Tomio, Bruce Tuckman, Dimitri Vayanos, Ernst-Ludwig von Thadden, Olivier Wang, Geoffery Zheng, and seminar and conference participants at l'Association Française de Finance, the Bank of Italy, the Federal Reserve Board of Governors, Goethe University, the Leibniz Institute for Financial Research SAFE, the 21st International Conference CREDIT, the 9th International Conference on Sovereign Bond Markets, the International Risk Management Conference (IRMC), the LSE, NYU Stern, UMass Amherst, the Venice Finance Workshop, and the Wharton School for their insightful comments. Any remaining errors are ours. A previous version of the paper was titled "A Preferred-Habitat Model of Repo Specialness." Subrahmanyam gratefully acknowledges the Alexander von Humboldt Foundation and NYU Stern's Center for Global Economy and Business for generous financial support via the Anneliese Maier Award and the Faculty Grant Award, respectively, and the Ca' Foscari University of Venice for its hospitality and the collegial atmosphere in which the early stage of this research was conducted. The project was supported by the Leibniz Institute for Financial Research SAFE.

1 Introduction

The recent experience with quantitative easing (QE) programs around the world has starkly demonstrated the importance of demand factors in fixed-income markets and thus on the term structure of interest rates, highlighting two aspects that are worthy of analytical examination. First, the price of nearly riskless securities delivering known streams of payments rises persistently with large purchases by central banks (Bernanke, 2020). This calls into question the assumption of perfectly inelastic supply underlying both the standard bond valuation models in the financial economics literature and the Ricardian equivalence theories in the macroeconomic literature. Hence, a large demand shock affects the yield curve. Second, these asset purchases induce a scarcity of high-quality collateral and exert downward pressure on the rates at which the targeted securities trade in the repurchase agreements (repo) market, the main secured money market.[1] Thus, a large demand shock also affects the secured money market. Given the essential role of both money markets and the term structure of interest rates in the financial and economic system, these important stylized facts require an understanding of fixed-income markets in which durable assets such as bonds not only serve as investment vehicles but also as collateral for loans, in the spirit of Kiyotaki and Moore (1997).

Does the term structure of interest rates interact with secured money markets, where investors use bonds to collateralize loans? Intuitively, it should. For instance, in the traditional models of the term structure, bond prices reflect the current realizations of money market rates and a premium attached to the risk that these rates might change in the future. Generally, the money market is summarized by the stochastic behavior of a unique, exogenous interest rate. However, this partial equilibrium approach does not allow for shifts in the bond market itself to affect borrowing and lending rates, since the latter are presumed to be exogenous. Such a restriction is at odds with the growing recognition that demand and supply forces, particularly QE, affect *both* the prices in the bond market (D'Amico and King, 2013; Greenwood and Vayanos, 2010, 2014; Vayanos and Vila, 2021) and the repo rates associated with bonds in the secured money market in the US and EU, among others (D'Amico et al., 2018; Arrata et al., 2020; Corradin and Maddaloni, 2020; Pelizzon et al., 2022). These two robustly documented empirical facts have generally been considered in isolation, even though the bond scarcity generated by QE is the common driving force behind both of them. Our paper is the first attempt to offer a

[1] The repo, the main secured money market instrument, is a secured short-term loan that serves the dual role of providing collateral and obtaining cash. A repo contract achieves collateralized financing and consists of the spot sale of a cash bond combined with a forward agreement to repurchase the bond on a specified future trading day. The counterparty enters the reverse side of the trade (reverse repo) by buying the collateral on the spot market and stipulating a forward contract to sell the security.

comprehensive, quantity-driven model of the term structure of interest rates that integrates the two effects and endogenizes the secured money market with the bond market. In particular, we show that the impact that QE exerts on secured money markets dampens its effect on the compression of the term premium.

The importance of the repo market and its close connection to the bond market underscores the relevance of the quantity-driven framework in which we model both markets in general equilibrium. The repo market is the lifeblood of the financial system since it provides liquidity to holders of financial assets while providing an avenue to engage in short-term securities lending for those with cash. Repo contracts are the primary financial instrument for money market transactions, where institutional investors routinely obtain collateralized financing, and the size the repo market is simply enormous – much larger than the bond market itself. The average daily volume of outstanding repo transactions is about \$12 trillion, roughly 14% of the world's GDP, of which Treasury repo transactions constitute about \$8 trillion. By contrast, the daily volume in the US Treasury bond market averages around \$0.6 trillion.[2] Moreover, it is well known that the repo market is segmented (see, e.g., Buraschi and Menini, 2002) and elastic to demand (D'Amico et al., 2018), frictions that we leverage in our model. Any repo contract is a short-term loan collateralized by a bond. A particular government bond ("special collateral or SC repo") or any bond from a predefined basket ("general collateral or GC repo") can be used as collateral. GC repo agreements are often called "cash-driven" transactions, because their primary purpose is to achieve collateralized financing that provides liquidity. In these transactions, each bond in a certain basket can be delivered as collateral. On the other hand, repo transactions can be motivated by the demand for a particular bond; in that case, they are "security-driven." An issue of securities that is subject to excess demand compared with others with very similar cash flows is said to be "on special." Competition to buy or borrow a special issue, perhaps to cover short selling commitments, causes buyers in the repo market to accept a lower interest rate in exchange for cash in these SC repo transactions. By lowering the attainable financing rate, special bonds yield a "repo dividend" (Duffie, 1996) that varies with the tenor and type of the collateral (D'Amico and Pancost, 2022) and the demand for that *particular* bond.

Importantly, as Duffie (1996) shows, the price of a bond is connected by an arbitrage relation to its special repo rate, which describes its value as collateral. However, since Duffie's foundational contribution, most of the literature on money market rates has largely abstracted from

[2]Sources: Bank for International Settlements (BIS), https://www.bis.org/publ/cgfs59.htm; US Department of the Treasury, https://home.treasury.gov/system/files/136/IAWG-Treasury-Report.pdf.

term structure considerations and most research on term structure issues largely assumes that the money market rates are exogenous. Our model helps fill this gap between the bond and repo markets and shows that, dynamically, bond scarcity, repo specialness, and the term structure feature nontrivial and previously undocumented interactions with one another.

Our model delivers two key results. First, we show that repo specialness strongly influences the term premium along with the entire yield curve for both GC and SC bonds. High levels of repo specialness are evidence of the significant costs of carry trades and hedging strategies, both of which are limits to arbitrage that attenuate the response of the yield curve to demand forces such as QE. That is, repo specialness dampens the impact of any given quantity of asset purchases on the term premium of the term structure of interest rates. By inducing frictions in money markets, asset purchases become less effective in achieving their main purpose of reducing the term premium. We show that the effect of QE on the term premium becomes more impaired as the repo specialness QE generates in the SC segment of the secured money market becomes larger.

Second, bond scarcity increases repo specialness, strengthening the local supply channel of QE. As documented by D'Amico and King (2013), QE often brings about local supply effects, defined as relative-price anomalies of closely related assets induced by demand. Such effects are typically absent in equilibrium term structure models (TSMs), where bonds must be priced consistently with one another by arbitrage. The SC repo market structure offers a natural solution to this puzzle. When a bond is subject to exceptional demand pressure in the market, it becomes overpriced relative to instruments with equivalent cash flows. The lure of price deviations from economic fundamentals induces term structure arbitrageurs like hedge funds to borrow the overpriced bond and sell it short. Arbitrageurs must then deliver that specific security at the end of the contract. Their behavior gradually raises the demand for high-quality collateral in the repo market, exerting endogenous downward pressure on special repo rates and thus eliminating arbitrage. Bond scarcity generates strongly localized supply effects like kinks in the term structure. Therefore, the apparent anomaly of "overpriced bonds" disappears once the term structure is integrated with the repo market.[3]

To derive our results, we build on the Vayanos and Vila (2021) (VV) TSM of the bond market. Unlike VV, we focus on the preferences of investors for specific characteristics. For

[3]For instance, yield curve fitting errors of Treasury securities are widely used by academics, policymakers, and practitioners. In an influential paper, Hu et al. (2013) use the dispersion in Treasury yield curve fitting errors as a measure of pricing noise, which proxies for the shortage of arbitrage capital in the economy. One caveat for considering the Treasury market in isolation from the repo market is that bond mispricing might not be executable if the borrowing cost of a position in the repo market is large. Thanks to endogenizing specialness, our model is able to explain the yield curve fitting errors in a manner that is consistent with the absence of arbitrage.

example, in the US Treasury bond market, securities with the same cash flows can be *on-the-run* or *off-the-run*. Traders prefer the former and bid up their prices.[4] A more recent phenomenon involves QE, in which central banks purchased large quantities of several bonds, making many of them special (see Arrata et al., 2020; Ballensiefen et al., 2023). We designate bonds subject to excess demand as "special." In doing so, we introduce a new dimension to TSMs: bonds that share the same tenor might differ in their exposure to demand forces. The notion of demand forces inducing bond specialness puts us in a comfortable position to model QE. To ensure that equilibrium demand-driven price differences between instruments with equivalent cash flows are consistent with the classical notion of arbitrage, we must account for the *different* borrowing cost of the bond in the SC repo market, where investors borrow a specific bond and lend cash. Asset purchases exert direct price pressure on special bonds. On the opposite side of the bilateral purchases of preferred-habitat investors (e.g., the central bank), aggressive market participants that we refer to as arbitrageurs sell the special bonds short and reinvest the proceeds until maturity in the money market. As a group, arbitrageurs thus borrow long-term bonds and invest cash at the series of overnight short rates in the money market, replicating the securities for which asset purchases generate excess (QE) demand through their carry strategy. This endogenous response of the private sector induces three effects.

1. Arbitrageurs intensify their search for collateral on the repo market to borrow special bonds in the face of increasing scarcity. Since the supply of any special bond is finite, its repo specialness increases along with the bond price due to the arbitrage between general and special bonds presented by Duffie (1996). GC bonds, which can be exactly replicated through interest rate derivatives, are not directly affected and will only be affected indirectly through second-round risk adjustment effects. This mechanism leads to the presence of *local supply effects* when comparing bonds across different time-to-maturity buckets.

2. Asset purchases increase the exposure of arbitrageurs to their carry trade strategy, com-

[4]For instance, it is common for traders to roll over their positions into each successive *on-the-run* issue, perhaps because of their exceptional liquidity and because they are often the cheapest among the basket of deliverable bonds for the settlement of futures contracts (Merrick Jr et al., 2005). This pattern has been well documented empirically. Cornell and Shapiro (1989) were among the first to show the existence of mispricing between bonds with equivalent cash flows. Among others, Barclay et al. (2006) show, using clearing records, that both the trading volume and the market share of electronic intermediaries decline by about 90% when Treasury securities go *off-the-run*. The terms *on-the-run* and *off-the-run* do not carry as much significance in other markets – including most of Europe, Japan, and India – because sovereign bonds in those markets are often issued "on tap." Thus, in principle, all bonds can be reissued and their specialness cannot be ascribed to recency of issuance. What explains specialness in other markets? In general, securities go on special when they attract a significant degree of excess demand, which sometimes arises when a bond becomes the *cheapest to deliver* in the futures market, when a bond is used as a hedge, or when the issue is labelled Green or Islamic.

pressing the term premium, as pointed out by VV. This channel is often referred to as the *duration extraction channel* and affects both GC and SC bonds.

3. In general equilibrium, the above channels interact with each other. Bond scarcity in money markets induces a reduction in SC rates, which results in lower yields for the corresponding special bonds on the bond market (Duffie, 1996). A distinct but related effect of bond scarcity is reducing the willingness of arbitrageurs to carry their trades across the curve to meet the exceptional demand prompted by preferred-habitat investors. Gradually, special bonds become more costly to borrow in the SC repo market, and their scarcity brings about the concrete risk of short squeezes. Arbitrageurs continue to borrow long-term bonds in the repo market but reduce their exposures and thus the compression effect of asset purchases on the term premium for both GC and SC bonds. Intuitively, higher levels of specialness in the SC repo market partially unwind the QE effect on the term premium. Frictions in the money markets thus impair the transmission of asset purchases to the term premium, which is reduced by less than in the benchmark case, which is absent specialness.

In our closed-form solutions, the equilibrium price of bonds targeted by exceptional demand exceeds the price of otherwise equivalent bonds by the risk-adjusted present value of their stream of repo dividends. Repo specialness is stochastic, dynamic, and affected by excess demand in the bond market. As a result, the bond and repo markets feature non-trivial interactions with each other, over and above the known arbitrage connection studied by Duffie, warranting interest in a general equilibrium approach. Moreover, the expected return-risk ratio on the bond market is consistent with the absence of arbitrage only when the short riskless rate varies at the instrument level, requiring empirical studies of the bond market to consider a broader picture that includes the rates at which securities are financed (see Cherian et al. (2004) and Chen et al. (2022)).

We also extend the model to incorporate the concept of imperfect substitutability in the habitat preferences of investors. This enables us to effectively model the rebalancing induced by QE on the portfolios of buy-and-hold investors. Additionally, we assess the effects of credit risk and analyze how haircuts and borrowing constraints impact the balance sheets of heterogeneous arbitrageurs. We also allow for a different degree of specialness based on the Treasury auction cycle. By incorporating these supplementary elements that enhance our ability to capture market features, we establish the fundamental essence of our key findings that the term structure of interest rates and repo markets are strictly linked and influence each other and that this link plays an important role in the effectiveness of QE.

A calibration of our theory using realistic parameters quantitatively illustrates our main findings. For comparability with previous studies, we use the US Treasury bond data from Gürkaynak et al. (2007). We start from the simplest case, in which a bond is assumed to remain on special throughout its entire life cycle and its specialness features a certain time decay, an assumption that is later relaxed to describe important generalizations. From our setup, two distinct yield curves of general and special bonds are obtained by rolling over GC and SC repo contracts, consistent with the price premium commanded by near-money assets (Krishnamurthy and Vissing-Jorgensen, 2012; Nagel, 2016; Van Binsbergen et al., 2022). Subsequently, we demonstrate the diverse impacts of QE on the term premium through calibration. First, we examine the scenario where the repo market exhibits inelasticity to quantity, allowing us to capture the duration effect of QE on the term premium as modeled by VV. Second, we investigate the case where the supply of bonds on the repo market displays elasticity to quantity, in line with the documented empirical evidence. We establish that elasticity within the repo market impairs the effect of QE on the term premium. We also provide other calibrations to show the local effect due to repo specialness and the consequent effect on the term structure of repo specialness due to the Treasury auction cycle.

Our contribution nests traditional and more recent TSMs as distinctive cases arising from the specification of a particular pricing kernel. While we build on the preferred-habitat theory, the expectation and the liquidity premium hypotheses of the term structure are also consistent with our model. Overall, our framework proposes a paradigm shift from a focus on "conceptual" arbitrage, at the core of finance, to one on "executable" arbitrage, in the spirit of a recent strand in the literature (Gabaix et al., 2007; Du et al., 2018; Fleckenstein and Longstaff, 2020; Jermann, 2020; Pelizzon et al., 2022). Our theory is distinguished by the fact that price differences are not attributable to specialized or constrained marginal investors but rather stem from the holding cost of arbitrage – that is, the cost of repeatedly borrowing a position to sell it short – as documented by Fontaine and Garcia (2012).

The remainder of the paper is organized as follows. Section 2 surveys the related literature. To motivate our analysis, Section 3 discusses certain stylized facts, while Section 4 presents a simple theory of the term structure of interest rates integrating capital and money markets. Section 5 discusses selected extensions of our baseline model, and Section 6 shows its main theoretical predictions and a calibration with market data. Section 7 offers concluding remarks. All proofs are available in the Appendix.

2 Literature Review

The Term Structure of Interest Rates. There is a vast literature on modeling the term structure – the relation between time to maturity and bond yield – at a general level. However, we confine ourselves to a discussion of more recent research focusing on the impact of unconventional monetary policies, such as QE, on the term structure. Unconventional monetary policies have renewed efforts by researchers to explain the effects of demand pressure on fixed-income securities in general and sovereign bonds in particular (see, e.g., D'Amico and King, 2013 and Greenwood and Vayanos, 2014). Relatedly, Du et al. (2022) document that the term premium is endogenous to the portfolio holdings of intermediaries. The canonical framework for this recent literature is found in Vayanos and Vila (2021); it provides the analytical structure to harmonize recent empirical findings with the received preferred-habitat theory (pioneered by Culbertson, 1957 and Modigliani and Sutch, 1966), which accounts for the differences in investment horizons across investors.

Our point of departure from the recent literature on habitat preferences in fixed-income markets is concentrating on investors' preferences for special bonds within maturity buckets – rather than on maturities – which could arise from mutual fund investment mandates and liquidity considerations (Pasquariello and Vega, 2009). This is an important feature of financial markets captured by our model.[5] Moreover, we explicitly consider the two segments of the money market, the GC and the SC repo markets, and endogenize special repo rates by allowing arbitrageurs to finance their positions in the repo market. Furthermore, while arbitrage is regarded as a risky carry trade in the VV framework, we allow arbitrageurs to be immune with respect to interest rate risk in the classical sense; namely, by buying two bonds of the same tenor when demand forces induce relative price differences, a strategy commonly referred to as a *convergence trade*. The comparatively higher price of the sought-after special bond is reflected in its appropriate special repo rate by the endogenous search for the collateral necessary to sell the security short. This approach paves the way to the assessment of the effects of bond scarcity due to QE on the money market and to the quantitative evaluation of new policy tools such as securities lending facilities (SLFs). Risk adjustments, which are relevant to measure the term premium commanded by arbitrageurs when executing carry trades over long horizons, are endogenous to portfolio holdings. We show that the effect of QE on the term premium theoretically predicted by VV is attenuated by the repo specialness effect, which is itself due

[5]Naturally, bonds differ in many dimensions other than maturity. For example, Chen et al. (2022) use the constraints on Islamic financial institutions for their investments to comply with Shariah law to identify clientele effects on bond prices and repo rates, and D'Amico et al. (2022) focus on Green premia, the yield differences between maturity-matched conventional and Green bonds.

to QE, since the prevailing rates in markets are elastic to quantities. This elasticity in the repo market generates a lower compensation for arbitrageurs who enter into reverse repo contracts by lending cash and borrowing the bond in the SC repo market. This in turn reduces their willingness to carry their trades across the curve to meet the exceptional demand for particular securities prompted by QE.

Our analysis yields two important results: (i) a dynamic relationship between quantity variables and bond scarcity, the prices of bonds, and the entire cross-section of their money market rates; and (ii) a novel link between repo specialness and the term premium, which are connected through the holdings of arbitrageurs. Therefore, the repo specialness generated by QE itself attenuates the duration extraction channel of QE. We also consider the imperfect substitutability in the habitat preferences of investors, which enables us to model the rebalancing induced by QE on the portfolios of inelastic market participants such as buy-and-hold investors. We then assess the effects of credit risk and the impact of haircuts and borrowing constraints on the balance sheets of arbitrageurs that are heterogeneous in their attitude toward risk. These extensions preserve the qualitative nature of our main results. In addition, we analyze the dynamic effects induced by the Treasury auction cycle on the behavior of specialness over time.[6]

The Repo Market. Repo contracts are similar to collateralized loans. In a foundational paper, Duffie (1996) shows that bond prices and the rate on the loans they collateralize are connected by an arbitrage restriction and develops a model connecting the two in a static sense (empirically validated by Jordan and Jordan, 1997), where special repo rates – that is, those significantly below prevailing riskless rates – decrease as arbitrageurs intensify the search for collateral to sell a bond short on the secondary market. Unlike Duffie's paper, which is static in nature, we explore the repo specialness in a dynamic sense in both the time series and the cross-section of bonds, explaining it as the result of the interaction between demand forces and costly arbitrage.[7] To our knowledge, our paper is the first general equilibrium model formal-

[6]Recently, the elegant framework proposed by VV has been extended to the foreign exchange market in Greenwood et al. (2023) and Gourinchas et al. (2022), to the credit risk market in Costain et al. (2022), and to the interest rate swaps market in Hanson et al. (2022), by using arbitrage restrictions. However, none of these papers focuses on the effects of demand pressure on the repo market by abstracting from a distinguishing feature of the behavior of arbitrageurs in financial markets.

[7]Relatedly, Krishnamurthy (2002) documents the gradual convergence of systematic price differences between new and old bonds with the same 30-year tenor, showing that spreads in repo financing rates between these securities prevent arbitrage opportunities. Other contributions in this area include Fisher (2002), who describes the pattern of repo specialness over the auction cycle, and Buraschi and Menini (2002), who test whether current special repo rates discount the future collateral value of Treasury bonds. Cherian et al. (2004) document the joint cyclicality of special repo rates and bond specialness over the auction cycle and present a no-arbitrage model where *on-the-run* bonds are discounted at an exogenously modeled special repo rate. We derive such phenomena endogenously by building on recent advances in the literature on heterogeneity in asset demand across investors.

izing these ideas in a term structure framework where repo specialness arises endogenously due to (i) preferred-habitat investors' demand that generates bond scarcity and (ii) repo market elasticity to quantities.

He et al. (2022) propose a preferred-habitat model that explains the behavior of Treasury convenience yields in times of crisis, where dealers subject to regulatory constraints provide GC repo financing to leveraged investors. Our paper differs from theirs because we focus on endogenous SC rates and provide a unified framework to price-specific (e.g., Green, Islamic, *on-the-run*) and generic securities, giving rise to equilibrium price differences between bonds with identical cash flows. In models where the short rate is constrained in the cross-section of bonds, such price differentials would normally result in arbitrage opportunities. Instead, in our framework, the equilibrium creates the specialness of the specific bond and satisfies a generalized notion of the Sharpe ratio that allows the short financing rate to depend on the characteristics of the collateral.

3 Stylized Facts

Collectively, the consensus view in the literature on the effects of QE in fixed-income markets has highlighted stable empirical patterns that have proven robust across countries and over time.

Stylized Fact 1: QE significantly affects the term structure. To put this into perspective, Christensen and Rudebusch (2012) use data around policy announcements in the US Treasury market to estimate a reduction of the term premium on the order of 29 basis points (bps) (see also Gagnon et al., 2018). Moreover, QE generates local supply effects. D'Amico and King (2013) quantify this effect during the first large-scale asset purchase program in the United States at around 30 bps (for evidence in EU markets, see Altavilla et al., 2021; Koijen et al., 2021).

Stylized Fact 2: QE significantly affects repo specialness. In the context of US markets,

In the model of Vayanos and Weill (2008), search costs induce endogenous specialness: that is, between two assets with identical cash flows, the one where short sellers concentrate their trades is priced at a premium, reflecting a larger pool of buyers. Gârleanu et al. (2021) examine the stock and securities lending markets when beliefs are heterogeneous. We complement their stationary search-based contribution from a term structure perspective, which has the advantage of allowing for time series analyses. Copeland et al. (2014) and Mancini et al. (2016), who focus on the stability of the repo market, present extensive descriptions of the institutional aspects of the US and European markets for repurchase agreements. Consistent with our model, Graveline and McBrady (2011) and Maddaloni and Roh (2021) show that inelastic investors participate in the repo market substantially less than in the secondary market, increasing the scarcity of collateral.

D'Amico et al. (2018) document an effect of - 1.8 bps per billion dollars purchased by the Fed, with a stronger impact at the short end of the curve. Today, the Fed holds around $5 billion in US Treasury securities. In the European Union, Arrata et al. (2020) document that large-scale asset purchases affect repo specialness through the collateral scarcity channel and estimate that purchasing 1% of the outstanding bond increases its specialness by 0.78 bps. The Eurosystem held 39% of German government bond (Bundesanleihen or bunds) at the end of 2017.[8] The estimates in Corradin and Maddaloni (2020) are even larger.

Neither the theoretical nor the empirical literature provides any guidance as to the connections between these two quantitatively large stylized facts, especially when evidence of local supply effects makes it essential to understand the bond market and the repo market in combination. At the same time, repo specialness affects the behavior of arbitrageurs and thus must have an impact on the term premium itself. In Figure 1, we provide an illustration of the dynamic interaction between repo and bond markets, showing a novel feature in the data. In the chart, we show the term premium over the last decade, measured as the yield differential between *benchmark* 10-year and 2-year German Treasury bonds (as reported by Bloomberg), along with the volume-weighted GC and SC repo rates associated with the universe of German bunds. We are aware that there is a vast literature that documents that the term premium is driven by several factors that interact with one another in a complex way, and a formal econometric investigation of the relationship between the slope of the term structure and repo specialness is beyond of the scope of this paper.[9] Nonetheless, the figure shows two key patterns. First, the borrowing rates in the repo market may differ substantially across GC and SC bonds, even if their maturities are identical. This aspect of money markets has been ignored in most theories of the term structure, where there exists a unique inelastic short rate such as the GC (for instance, the secured overnight funding rate (SOFR) is a rate measuring the cost of overnight borrowing of cash collateralized by US Treasury securities), and bond prices result from its current and risk-adjusted expected future realizations of this rate. Second, and strikingly, the term premium co-moves much more strongly with SC rates than it does with the GC rate, a phenomenon that

[8]According to the BIS, the share of special trades in the German repo market increased from around 5% before the introduction of the Public Sector Purchase Programme in 2015 to more than 50% in 2016, peaking at the staggering level of 550 bps (`https://www.bis.org/publ/mktc11.pdf`, IV.13.)

[9]In a foundational paper, Vasicek (1977) derives a general model of the term structure based on the absence of arbitrage opportunities between bonds and the instantaneous short rate. Notable contributions in this area include Cox et al. (1985), who develop a general equilibrium model in which the interest rate follows a square root diffusion process, and Heath et al. (1992), who derive no-arbitrage bond prices by modeling the stochastic evolution of the forward rate curve. Duffie and Kan (1996) describe the necessary and sufficient conditions for an affine representation of multifactor models for the term structure. On the empirical side, Fama and Bliss (1987) show that the sign of bond risk premia depends on the slope of the spot yield curve, and Campbell and Shiller (1991) document that when term premia are high, long rates tend to fall and short rates tend to rise.

has not previously been appreciated in the literature. However, this is in line with the theoretical predictions of our model, where we demonstrate that if intermediaries matter for asset prices, special repo rates and the term premium of the GC term structure are correlated.[10]

4 The Model

4.1 Setup

In this section, we develop a model in discrete time $t \in [0, \ldots, \mathtt{T}]$ that features a market for default-free (riskless) zero coupon bonds (zeros). Bonds are indexed by their tenor $n \in [1, \ldots, \mathtt{N}]$ and by their status $i = \{\text{g}, \text{s}\}$; that is, as general as opposed to special bonds. General and special bonds of the same tenor have equivalent cash flows, but their prices can differ because of the demand effects detailed below. At time t, a zero with tenor n has a price $b_t^n(i)$ expressed in dollars per unit of notional principal. All stochastic processes are modeled under the equivalent martingale measure defined on the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, and all are adapted to the filtration $(\mathcal{F}_t)_{t \in \mathtt{T}}$.[11] The continuously compounded yield to maturity is

$$y_t^n(i) = -n^{-1} \log b_t^n(i). \tag{1}$$

We assume the short rate to satisfy a Vasicek process whose parameters incorporate mean reversion and where the innovations are distributed as standard normal variates.[12]

$$r_{t+1} = \varrho r_t + (1 - \varrho)\theta + \sigma_r \eta_{t+1}. \tag{2}$$

[10]The *Financial Times* suggests that long rates may have become hopelessly distorted – essentially, they are too low – and the curve therefore no longer sends a reliable signal about future economic conditions. The article, titled "Is the yield curve lying?" and dated 21 June 2023, then argues that it is common to blame the Fed's bond buying programs for the distortions and makes the point that low rates are not driven by rate fundamentals but instead by the demand for long-term, risk-free securities for use as collateral.

[11]The choice of this risk-neutral measure allows us to retain the general approach of Dai and Singleton (2003) and still obtain both the canonical Vasicek (1977) and the more recent VV affine TSMs by specifying different risk adjustments.

[12]The choice of a Gaussian model is standard and motivated by simplicity. An excellent treatment of non-Gaussian models appears in Berardi et al. (2021).

Bonds can be used as collateral to obtain overnight secured financing in the repo market.[13] As is standard in modeling repurchase agreements, we abstract from collateral rehypothecation and credit risk and assume that the repo market clears once a day (see, e.g., Duffie, 1996).[14] Therefore, the GC repo rate must coincide with the short rate r_t to prevent arbitrage opportunities. In our model, the short-rate process in Equation (2) can thus be interpreted as describing the GC repo rate dynamics (e.g., the SOFR in the US Treasury market). As noted above, the repo market is segmented. Arbitrageurs with overnight cash on their hands have two distinct riskless options to lend money against either SC or GC bonds at their respective market rates, namely:

1. Reverse any of a basket of generic bonds ($i = g$) in the GC market by entering an overnight agreement that earns the GC repo rate r_t.

2. Reverse the position in the SC market ($i = s$), which is elastic in supply, and earn the lower overnight SC repo rate r_t^n, to be determined in equilibrium.

While the GC secures higher interest rates, arbitrageurs might want to forgo loan returns to borrow the special bonds needed to meet any pending short selling commitments. Specialness premia $r_t - r_t^n$ do not result in any arbitrage opportunities, as we demonstrate below. However, the supply of special bonds is *elastic* to quantities, as are SC repo rates. In fact, the amount of outstanding special bonds is fixed, and the repo activity of buy-and-hold investors such as pension funds and insurance companies is limited, (as documented by Maddaloni and Roh, 2021). Thus, incremental quantities of special bonds grant financing at progressively lower repo rates. As an illustration, Appendix Figure OA.1 shows the volume-weighted monthly trailing average of the daily rates on repo transactions collateralized by Italian treasury bonds ranging from 2012 to 2018, using data from Mercato Telematico dei titoli di Stato (MTS) with millisecond precision. We distinguish between GC and SC transactions and plot the latter for benchmark time-to-maturity buckets; SC rates are generally lower than GC rates. Further, it is clear that SC rates can vary stochastically across tenors and over time. We endogenize the

[13]We focus on overnight repo transactions for ease of notation because the modeling of term repos would require an additional index. Empirically, the overnight tenor attracts by far the dominant proportion of volume. For instance, the Fed reports the share of overnight repos to be about 80% of the volume in the US triparty market. A recent description of this market can be found at https://www.federalreserve.gov/econres/notes/feds-notes/the-dynamics-of-the-us-overnight-triparty-repo-market-20210802.htm.

[14]In the baseline model, we consider unlimited overnight borrowing without default risk, but Section 5.2 discusses borrowing constraints and haircuts. The results hold under re-use of collateral as long as the passthrough of the rehypothecated collateral is less than one, as is well understood empirically and need not be discussed in detail here.

difference between the GC and the SC repo rates as a result of the demand effect on the bond market, which induces the search for collateral on the repo market.

4.1.1 Preferred-Habitat Investors

As a group, preferred-habitat investors such as bond market mutual funds have an elastic demand for the *special* bond of a certain tenor. These investors have habitat preferences, which we allow to be a function of tenor, toward bonds with specific characteristics.[15] Preferred-habitat investors are not active on the repo market, or at least they are less so than arbitrageurs.[16] We define as special those bonds that are targeted by preferred-habitat investors and index them through $i = s$; for clarification, think of *on-the-run* and *first-off-the-run* securities as obvious candidates for specialness.[17] Conversely, we refer to bonds of all maturities for which the excess demand is permanently zero as general and index them through their status $i = g$; one example is *far-off-the-run* bonds. The demand of preferred-habitat investors is expressed net of the size of the issue supplied by the government, which is normalized to zero, without loss of generality. Borrowing the structure from VV, we define the excess demand $Z_t^n(i)$ for bonds with tenor n by

$$Z_t^n(i) = \begin{cases} q_t^n - \alpha^n \log b_t^n(i) & i = s, \\ 0 & i = g, \end{cases} \tag{3}$$

with a price elasticity α^n and a stochastic intercept q_t^n that evolves as the Vasicek process

$$q_{t+1}^n = \varphi_n q_t^{n+1} + (1 - \varphi_n)\kappa_n + \sigma_{q,n} \upsilon_{t+1}^n. \tag{4}$$

Equation (3) is a definition of segmented markets according to which exceptional demand risk factors affect only special bonds. The process for demand risk in Equation (4) is autoregressive and mean-reverting.[18] The parameters φ_n, κ_n, and $\sigma_{q,n}$ have the usual interpretation of tenor-specific persistence, long-run mean, and standard deviation of a process that has normal

[15] We wish to emphasize that our focus on preferences for specific bond characteristics that are clearly observable in market data is not vulnerable to the criticism of the preferred-habitat view of the term structure based on the argument that interest rate derivatives allow for hedging maturity habitats.

[16] Bond market mutual funds often target bellwether indices composed of *on-the-run* bonds of selected maturities and have mandates preventing them from achieving leverage through the repo market because of the risk involved (Krishnamurthy, 2002). Fleckenstein and Longstaff (2021) document that Treasury convenience premia have discontinuities at specific annual maturities induced by clientele effects unrelated to fundamentals.

[17] The set of securities targeted by excess demand includes but is not limited to bonds that are targeted by the purchases of central banks to achieve local effects, *on-the-run*, cheapest to deliver, Green, and Islamic.

[18] The process shifts forward in time by replacing time t with $t + 1$ and tenor (time to maturity) n with $n - 1$.

innovations.[19] To express the model in full generality, we allow for demand shocks and GC rate innovations to be correlated with the tenor-specific coefficient ρ_n. Under normal market conditions, Equation (3) describes preferences for liquidity and those arising from coordination equilibria among investors. In the context of QE, this formulation captures the purchases by central banks of targeted bonds relative to non-targeted bonds. For simplicity, in the baseline scenario we allow for two types of bonds with the same tenor: general and special. Below, in Section 5.3, we show that similar results are obtained by generalizing the specification to allow for a broader set of bonds targeted by purchases.

4.1.2 Arbitrageurs

Arbitrageurs resort to short-term repo financing and engage in term structure trades to smooth out price differences that would otherwise arise in a segmented equilibrium.[20] For example, arbitrageurs such as hedge funds would sell a bond short that is overpriced as a result of substantial demand pressure. To this end, they would reverse their position in the n-th bond earning the repo rate and simultaneously sell outright the collateral exerting downward pressure on the bond price. The reverse repo contract would then be rolled over until the bond matures or the position is closed. The portfolio holdings of arbitrageurs are denoted through $X_t^n(i)$. In equilibrium, the market clearing condition is such that

$$Z_t^n(i) + X_t^n(i) = 0, \qquad \forall \quad t, n, i. \tag{5}$$

Due to market clearing, and since the demand for general bonds does not exceed their supply from the government, arbitrageurs are only active in special bonds in equilibrium. Thus, we drop the status i from $X_t^n(i)$ for simplicity. Of course, nothing prevents arbitrageurs from trading general bonds as well, so that in equilibrium these securities would be as profitable as special bonds from their perspective. Effectively, arbitrageurs issue synthetic n-maturity special bonds by accepting the rollover risk associated with short sales financed through SC repurchase agreements. Conversely, GC bonds are inherently financed at the overnight GC rate, since there is no excess demand for these securities. Intuitively, higher activity from preferred-habitat investors increases repo specialness by locking up the bond and symmetrically increasing the search for collateral to short the bond by arbitrageurs.[21] The next expression is

[19]Technically, demand risk does not depend separately on tenor and time, thus generalizing the VV formulation.

[20]We emphasize that VV arbitrageurs engage in risky carry trades across the term structure and thus differ from the traditional Vasicek interpretation of investors with interest rate-neutral exposures. We allow for both views.

[21]Our approach is consistent with Banerjee and Graveline (2013), who decompose the *on-the-run* premium of Treasury bonds into higher prices encountered by long investors and larger borrowing costs borne by short

the dynamics of arbitrageurs' wealth W_t, where Δ denotes the first difference operator.

$$\Delta W_{t+1} = W_t r_t + \sum_{n \in \mathtt{N}} X_t^n \left(\log \frac{b_{t+1}^{n-1}}{b_t^n} - r_t^n \right). \tag{6}$$

Equation (6) is *not* a standard law of motion of wealth, even though the restriction $r_t^n = r_t \; \forall n$ corresponds to the VV case in which the short rate is constant in the cross-section of bonds. Notably, our approach departs from the textbook portfolio allocation problem between a riskless money market account and a set of risky assets. Here, the holdings of leveraged arbitrageurs are financed on the repo market for collateralized lending.[22] The first term on the right side of the equation captures cash investments. Invested wealth W_t achieves the remuneration r_t offered by the GC rate, the highest among short rates. Similarly, cash shortages are inherently financed at the GC rate in the absence of SC bonds. The second term is the marked-to-market value of the portfolio of special bonds net of their financing costs, each represented by the respective SC repo rate r_t^n. Arbitrageurs establish a long position by buying the bond outright in the spot market and finance that purchase by using the bond as collateral to enter an overnight repo agreement. The next trading day, arbitrageurs must either close the outright position or roll over the short-term collateralized financing. A short position is obtained by reversing the position in the collateral market in exchange for cash and simultaneously selling the security in the spot market. This does not require any cash commitment. However, in the next period, arbitrageurs must either deliver the bond they have shorted or roll over the reverse repo contract. Unlike an opportunity cost interpretation, r_t^n thus denotes the cost of the collateralized loan (which repos the bond) to finance the position, in the spirit of Tuckman and Vila (1992).[23]

Why are repo rates more interesting than a simple exogenous process for the short rate? Market considerations aside, the hallmark of special repo rates is the exposure to demand forces (Duffie, 1996). From a theoretical asset pricing perspective, there is simply no room for demand pressure to impact the exogenously specified short rate process in Equation (2). In the model we propose, the demand forces that affect bond prices contribute to the endogenous determination of special repo rates r_t^n. Special repo rates are important from a quantitative viewpoint. For example, using data from the New York Fed, Copeland et al. (2014) estimate SC repo transactions to be about 60% of the daily volume in the US market, with the remaining

sellers.

[22]Without any loss of generality, we assume that arbitrageurs use the repo market to finance their bond portfolios since it is optimal to do so.

[23]For details on how institutional investors finance Treasury trades, see Fisher (2002). A similar insight on their budget constraint can be found in He et al. (2022), where the GC rate results from regulatory frictions. We complement their approach by focusing on SC rates that vary across bonds, induced by exceptional demand.

40% constituted by GC transactions. The SC daily volume share of the EU repo market is even larger; for instance, Arrata et al. (2020) report an average value of 87%. Thus, the TSMs that exogenously specify the process for the short rate are suitable for describing the GC repo market but leave the larger SC segment of the market unmodeled.

4.1.3 General Bonds, Special Bonds

Two issues of the same tenor may differ in terms of collateral value: for instance, bonds with the same time to maturity might be SC as *on-the-run* securities or GC as *far-off-the-run* ones. While both are exposed to the same duration risk, only the former is targeted by preferred-habitat investors and thus affected by demand pressure. To highlight this distinction in our model, we define as special those bonds that are exposed to two risk factors and as general those bonds exposed to one risk factor. Formally, let us conjecture that the price process is exponentially affine in the short rate and, conditionally on the bond status, in demand shocks.

$$-\log b_t^n(i) = \begin{cases} A_n r_t + B_n q_t^n + C_n^s & i = s, \\ A_n r_t + C_n^g & i = g. \end{cases} \tag{7}$$

Specific to our framework, bonds with identical cash flows can trade at different prices because of demand pressure. This feature adds a layer of realism to the TSMs and arises because the exposure of GC bonds to demand risk is restricted to zero (by construction), so that the price of these bonds reflects only the risk of changes in the short rate r_t.[24] Equation (7) reflects a view of segmented markets, as the compensation for (GC) interest rate risk r_t is common to GC and SC bonds, while the exceptional demand risk q_t^n only exerts pressure on the price of bonds targeted by preferred-habitat investors. Indeed, we regularly observe that bonds on special are overpriced with respect to general bonds with identical cash flows in Treasury markets. In Section 4.3, we also derive the implications of targeted demand pressure q_t^n on the rate r_t^n requested to lend special bonds in the repo market.

4.2 Equilibrium in the Bond Market

Definition 1. *The equilibrium is a set of bond prices* $\{b_t^n(i)\}_{t,n,i}$ *such that the market clears and arbitrageurs behave optimally, given the demand of preferred-habitat investors.*

The next few steps leading to a closed-form solution of the arbitrageurs' maximization program essentially follow the structure in VV, generalizing that model to an arbitrary equivalent

[24] The extension to a multifactor model for both SC and GC rates is conceptually straightforward.

martingale measure and multiple instantaneous rates r_t^n in the cross-section of bonds. Our contributions become clear thereafter. Replace Equations (4) and (2) into Equation (7) to derive the one-period log-price variation of both special and general bonds:[25]

$$\log \frac{b_{t+1}^{n-1}}{b_t^n} = m_t^n - \sigma^n U_{t+1}^n, \tag{8}$$

$$m_t^n = r_t \Delta A_n + q_t^n \Delta B_n + \Delta C_n^i - A_{n-1}(1-\varrho)(\theta - r_t) - B_{n-1}(1-\varphi_n)(\kappa_n - q_t^n),$$
$$\sigma^n U_{t+1}^n = [A_{n-1}\sigma_r \quad B_{n-1}\sigma_{q,n-1}][\eta_{t+1} \quad \upsilon_{t+1}^{n-1}]'.$$

As usual, m_t^n is interpretable as the deterministic change in the present log value of the bond. Moreover, $\sigma^n U_{t+1}^n$ is the stochastic bond return, which depends on two sources of randomness, innovations in the short rate η_{t+1} and in the demand risk factor υ_{t+1}^{n-1}, which are generally correlated. In equilibrium, we verify that Equation (8) holds for both special and general bonds. However, through market clearing, arbitrageurs' net exposures at the close of the business day are only short positions in special bonds.[26] Substituting Equation (8) into the arbitrageurs' wealth dynamics in Equation (6), we obtain

$$\Delta W_{t+1} = W_t r_t + \sum_{n \in \mathtt{N}} X_t^n \Big(m_t^n - \sigma^n U_{t+1}^n - r_t^n \Big). \tag{9}$$

In each period, arbitrageurs maximize the expected value of the next period's wealth change, where the first moment is taken with respect to the risk-neutral measure $\mathbb{Q}$,

$$\max_{\{X_t^n\}_{n\in\mathtt{N}}} \mathbb{E}_t^{\mathbb{Q}} \Big[\Delta W_{t+1} \Big]. \tag{10}$$

The formulation of the problem under the equivalent martingale measure has the advantage of implicitly including the compensation for risk while leaving unrestricted the preferences of arbitrageurs that uniquely pin down such a market price of risk.[27] Replacing Equation (9) into

[25]Log returns are appropriate because prices are exponentially affine; e.g., one-period returns are $\log \frac{1}{b_t^1(g)} = A_1 r_t + C_1$, whence $A_1 = 1, C_1 = 0$.

[26]Empirically, D'Amico et al. (2018) use the repo volume spread, calculated as the volume of reverse repo versus repo contracts, to measure excess demand for bonds and proxy for the number of short positions. Their estimates show that the repo volume spread is 10 times larger for *on-the-run* than *off-the-run* Treasury bonds.

[27]As demonstrated in Section 4.4, the specification in VV is, in discrete time, a particular case when arbitrageurs have mean-variance preferences with a risk-aversion coefficient a. Specifically, let $\mathbb{V}_t$ denote the variance conditional on $\mathcal{F}_t$, and rewrite the optimization program as $\max_{\{X_t^n\}_{n\in\mathtt{N}}} \mathbb{E}_t^{\mathbb{P}} \Big[\Delta W_{t+1} \Big] - \frac{a}{2} \mathbb{V}_t^{\mathbb{P}} \Big[\Delta W_{t+1} \Big]$.

Equation (10), we obtain

$$\max_{\{X_t^n\}_{n\in\mathtt{N}}} W_t r_t + \sum_{n\in\mathtt{N}} X_t^n \Big(\mu_t^n - r_t^n\Big),$$

where, analogously to a drift in continuous time, the expectation of the change in the log price of the bond has been adjusted by Jensen's correction term, so that

$$\mu_t^n = m_t^n - 0.5A_{n-1}^2\sigma_r^2 - 0.5B_{n-1}^2\sigma_{q,n-1}^2 - A_{n-1}B_{n-1}\rho_{n-1}. \tag{11}$$

The first-order condition with respect to the position in the n-th tenor bond on special is

$$\mu_t^n = r_t^n. \tag{12}$$

Equation (12) is an equilibrium term structure equation specified under the equivalent martingale measure $\mathbb{Q}$, where the drift of the bond price μ_t^n is equivalent to the rate at which arbitrageurs can exchange cash for the special bond r_t^n. This result is key to determining the linkage between the term structure of bond prices and the equilibrium rates in the repo market. Unlike the classical formulation in which the borrowing rate is the short rate, a long (short) position in the special bond must be financed (remunerated) at its own SC repo rate. Intuitively, this result suggests that in equilibrium the deterministic change in the risk-adjusted price of the bond must equal the repo rate against which the market allows arbitrageurs to finance their positions.

4.2.1 Change of Measure

The uniqueness of the equivalent martingale measure is guaranteed by the optimizing behavior of arbitrageurs, whose preferences are left unrestricted in the specification above. Specifying the appropriate market price of risk $\underset{1\times 2}{\lambda(\cdot)}$ enables several interesting cases to emerge. In Section 4.4, we detail the parameter choices that lead from our setup to the well-known models of Vasicek (1977), Brennan and Schwartz (1979), and Vayanos and Vila (2021). The drift term in the equilibrium condition can be expressed under the physical measure $\mathbb{P}$ as $\hat{\mu}_t^n = \mu_t^n + \sigma^n\lambda(\cdot)$ by applying a Girsanov transformation to the affine change in the log price of bonds.[28] Under this parametrization, Equation (12) closely resembles the familiar TSM arbitrage equation, with one difference that is our first important contribution: the riskless rate r_t is replaced by the

[28] The Girsanov theorem is well defined in discrete time; see Föllmer and Schied (2008). Heuristically, the reader is easily convinced by analogy with the change of measure in a binomial tree example.

cross-section of overnight special repo rates, r_t^n:

$$\underset{\text{Vasicek – Brennan and Schwartz}}{\hat{\mu}_t^n - r_t = \sigma^n \lambda(\cdot)} \qquad\qquad \underset{\text{First Order Condition}}{\hat{\mu}_t^n - r_t^n = \sigma^n \lambda(\cdot)} \tag{13}$$

Equation (13) compares the textbook equilibrium concept with ours. Since the foundational paper by Vasicek (1977) and the two-factor model of Brennan and Schwartz (1979), the characterization of TSMs by the absence of arbitrage is routinely based on the restriction $r_t^n = r_t$ $\forall n$. In practice, however, financing costs differ across bonds since they can be used for collateralized borrowing at a variety of special rates. Hence, we relax this assumption and propose a generalized equilibrium condition that allows the short rate to vary with the collateral value the bond grants to its holder. Canonical TSMs are based on the standard arbitrage restriction: Since a portfolio consisting of the appropriate combination of bond exposures achieves perfect immunization against interest rate risk, such a portfolio should realize the same return as an investment remunerated at the spot rate. Therefore, one should observe a constant ratio between mean return and standard deviation across all traded instruments.

Building on the idea of a constant excess return to risk (Sharpe) ratio, we note that in practice borrowing is often collateralized. Hence, it is necessary to employ our equilibrium concept that different bonds give rise to different costs of financing for market participants to fund their positions. Thus, we must adjust the Sharpe ratio, since the risk-free rate is not constant in the cross-section of bonds. That is natural once we recognize that special bonds are simply bonds with an additional stream of repo dividends.[29] We propose a paradigm shift from a focus on arbitrage to one on *executable* arbitrage. The TSM of VV reflects a portfolio allocation decision à la Merton between a riskless spot rate and risky bonds. In our interpretation, however, the equilibrium results from the choices of leveraged investors that use their positions as collateral to borrow cash. For market participants, differences in the collateral value between bonds are crucial determinants of portfolio choices. Our paper captures the simplicity of this idea in the theoretical term structure literature. The stochastic discount factor is unique, but the payoffs of the securities must be redefined on account of their holding costs, which our model determines endogenously as a result of market demand segmentation. An econometric test for the relative performance of the two TSMs is described in Section 4.5. Here, we focus on the close connection between the bond and repo markets across the term structure of interest rates and provide a general solution of the model that endogenizes repo specialness l_t^n, which is

[29] The equilibrium concept naturally extends to equity markets by replacing the special repo rate with the securities lending rebate rate.

defined as the difference between GC and SC rates conditional on time to maturity:

$$l_t^n = r_t - r_t^n. \tag{14}$$

4.2.2 Affine Representation

We cast our affine TSM using the terminology of Dai and Singleton (2003) by noting that the equivalent martingale measure $\mathbb{Q}$ is alternatively defined by the conditional Laplace transform

$$b_t^n(i) = \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\sum_{j=0}^{n} r_{t+j}^{n-j}\right)\right] = \exp\left(-A_n r_t - B_n q_t^n - C_n^i\right), \tag{15}$$

provided a parametrization is admissible (Duffie and Kan, 1996).[30] In the context of affine Markovian models, this representation is particularly useful. We note that the coefficients A_n, B_n, and C_n project the current value of the risk factors on the risk-adjusted rational expectations forecast of their future conditional realizations to impound their information into market quotes. The notional principal at maturity is priced using the appropriate bond-specific discount factor (Buraschi and Menini, 2002), with factors that are more persistent exerting a stronger impact on long-term yields.

4.3 Equilibrium in the Repo Market

Thus far, we have derived the equilibrium by using the absence of arbitrage in the time series of bond prices and interest rates. An important difference arises when we turn to their cross-section. While term structure carry trade portfolios require the risky rollover of short-term financing, the cross-sectional static arbitrage between GC and SC bonds is riskless, since both their prices and repo rates are known.[31] Hence, we can exploit this arbitrage restriction in order to obtain an explicit relation between the specialness in the bond and in the repo markets. Indeed, from the market clearing condition we know that demand pressure in the cash market has its mirror image in arbitrageurs' search for collateral in the repo market. Exploiting this idea, the next results generalize the static framework in Duffie (1996) to characterize endogenously and dynamically special repo rates in our affine TSM. To this end, let us specify as an auxiliary variable the difference between the pricing constants of bonds of different status and the same tenor in Equation (7) by defining $D_n = C_n^s - C_n^g$.

[30] Grasselli and Tebaldi (2008) establish conditions for closed-form bond prices in admissible TSMs.
[31] We abstract from search costs in over-the-counter markets (see Duffie et al., 2005; Jankowitsch et al., 2011).

Lemma 1. *In equilibrium,*

$$\exp\Big(B_n q_t^n + D_n\Big) = \mathbb{E}_t^{\mathbb{Q}}\Big[\exp\Big(-\sum_{j=0}^{n} r_{t+j}\Big)\Big]\mathbb{E}_t^{\mathbb{Q}}\Big[\exp\Big(-\sum_{j=0}^{n} r_{t+j}^{n-j}\Big)\Big]^{-1}.$$

Proof. Lemma 1 results from the ratio of the price of the general bond $b_t^n(g)$ to the price of the special bond $b_t^n(s)$. We refer the reader to Appendix A for details. *Q.E.D.*

Both general and special bonds promise the payment of equivalent cash flows at maturity. Therefore, their relative price (on the left side of the expression above) in equilibrium must be equal to the ratio of the holding cost of replicating the two bonds through a series of overnight repo contracts, in expected risk-adjusted terms (on the right side of the equation). Intuitively, absent this equivalence, arbitrageurs would earn a free lunch by selling short (purchasing outright) the bond overpriced (underpriced) relative to the other bond and to its own repo rate. Since both bond prices and their repo rates respond to quantities, the decrease (increase) in the price and in the special repo rate would then contribute to restoring equality. An example will clarify matters.

Example 1. *In Lemma 1, we make no assumptions about the correlation structure between the stochastic processes considered. If, however, the stochastic processes for r_t and l_t^n are assumed to be independent, Lemma 1 reduces to $e^{(B_n q_t^n + D_n)} = \mathbb{E}_t^{\mathbb{Q}}\big[e^{-\sum_{j=0}^{n} l_{t+j}^{n-j}}\big]$.*

In Example 1, demand pressure induces different valuations between bonds with equivalent cash flows. Such price differences equal the risk-adjusted present discounted value (PDV) of repo specialness from the pricing date until the bond matures. More generally, Lemma 1 shows that when the contemporaneous correlation between GC rates and repo specialness is unrestricted, the price of special bonds exceeds the price of general bonds of the same tenor by the PDV of the stream of GC repo rates divided by the PDV of the series of SC repo rates, with both computed under the equivalent martingale measure. Intuitively, the prices of special bonds reflect exposure to SC repo rates and their comovement with GC repo rates. Among others, Buraschi and Menini (2002) and Cherian et al. (2004) suggest that repo specialness must be included in the pricing of bonds on special. However, these papers are silent on what determines special repo rates.

Lemma 1 endogenizes repo specialness into an equilibrium TSM. In our model, the behavior of arbitrageurs connects demand pressure to bond prices and special repo rates, inducing repo specialness on those bonds that are targeted by preferred-habitat investors. Since clientele effects influence bond pricing, it is natural to establish the mapping between segmentation in the repo market and exposure to different factors (in particular, demand risk q_t^n) in the cash market that is the main content of this result. General and special bonds that are differentially targeted by demand pressure in the cash market result in a separation between the GC and SC rates used by market participants to discount the claim on the notional principal at maturity. An important consequence of the above discussion is that demand pressure impacts repo specialness. Setting $n = 1$ in Lemma 1 results in

$$q_t^1 B_1 + D_1 = -l_t^1, \tag{16}$$

which implies a constant relation between excess demand and repo specialness, since B_1 does not depend on time. To spell out the linkage between demand pressure in the bond market and specialness in the repo market, let us define $\mathcal{E}^i$ as the sensitivity of repo specialness to arbitrageurs' demand for bonds that are about to reach maturity:

$$\mathcal{E}^i = \begin{cases} \frac{\partial l_t^1}{\partial q_t^1} & i = s, \\ 0 & i = g. \end{cases} \tag{17}$$

Equation (17) characterizes the elasticity of collateral supply in the market for repurchase agreements. SC repo rates are sensitive to quantity, as they decrease (their specialness increases) with demand pressure in the bond market and the resulting short selling behavior of the arbitrageurs who consider the issue overpriced.[32] Conversely, GC repo transaction rates are insensitive to demand pressure, because any instrument within a basket of bonds can be delivered on the buyback day. Indeed, the GC rate follows the exogenous process in Equation (2) and is inelastic to quantities. The main friction in our model is thus segmentation in the repo market. The GC market, where each bond is substitutable by others included in the basket of deliverables, features a perfectly inelastic price elasticity to quantity. The SC market, where contracts command the delivery of specifically designated bonds, is instead characterized by positive loan price elasticity of supply, since the outstanding amount of the bond is fixed and the repo activity of buy-and-hold investors such as pension funds and insurance companies is limited by regulatory constraints (Duffie, 1996; Maddaloni and Roh, 2021). Central banks can also be thought of as preferred-habitat investors targeting and holding specific bonds on the cash market until maturity and increasing their specialness in the repo market. For example, the European Central Bank (ECB) offers a bond purchased during QE operations for lending in its

[32] This mechanism is in line with Duffie (1996), who argues that "the extent of specialness, for a given supply of the instrument, is increasing in the demand for short positions and in the degree to which the owners of the instrument are inhibited from supplying it as collateral," well before the recent advances in the financial literature that have shown how to price excess demand factors in the bond market.

cash-collateralized SLF at lower than prevailing market rates, generating mispricing between instruments with equivalent cash flows (Pelizzon et al., 2022).

Unlike in VV, the pricing of demand pressure does not result from the risk aversion of arbitrageurs. Rather, exceptional demand pressure affects asset prices by inducing short sellers to intensify their search for collateral on the repo market and increasing the specialness of the security. Thus, excess demand is priced on the secondary market even under $\mathbb{Q}$, reflecting structural frictions in the repo market that cause the supply of SC to slope upward (see Duffie, 1996, Figure 3). This point is illustrated in Figure 2, which shows that the supply of SC bonds is linear in its repo rate, with slope $\mathcal{E}^s$.[33] However, the demand is inelastic because arbitrageurs have committed to deliver the specific bond. With rightward shifts in the demand curve for SC bonds in the repo market, equilibrium specialness increases because collateral holders require greater compensation to pledge additional units of the special security. As we show below, the chart in Figure 2 is a general representation of the SC segment of the repo market that holds independently of the tenor of the bond.

Essentially, Equation (16) shows the existence of a mapping between demand pressure on the bond on the secondary market and its specialness on the repo market, characterizing the differential price of nearly maturing special and general securities. The extent to which a bond is special on the repo market is a function of its demand pressure on the bond market and of the elasticity of repo supply $\mathcal{E}^i$ in Equation (17), which yields the initial condition for the iterative pricing of demand risk. Our next result solves for the term structure of bond prices in closed form and verifies the conjecture formulated in Equation (7). To this end, we exploit the recursive structure of the problem. From the Vasicek stochastic process in Equation (2), we know that the persistence of the GC rate is ϱ. Likewise, the stochastic process for exceptional demand in Equation (4) has an autoregressive structure with persistence φ_n.[34] The key insight is that the persistence parameters of these processes determine the equilibrium pricing of the respective risk factors, since a long position can be replicated by a series of short-term investments at the GC rate for general bonds and at the SC rate for special bonds. The equilibrium we outline next is consistent both with the expectation hypothesis and the liquidity premium theory of the term structure, since we have left risk premia unrestricted by specifying the model under the equivalent martingale measure $\mathbb{Q}$.

[33] The first order relation between specialness and demand risk that captures a linear SC supply curve results from the affine specification and can be generalized to higher orders. For example, a second-degree polynomial would result from a quadratic TSM, and so on for higher-order specifications.

[34] For generality, we are allowing for tenor-specific parameters in the equation for demand risk. Gradually, these parameters guiding the process of excess demand for the issue change as time to maturity diminishes. The repo specialness of the bond reflects the term structure of preferred-habitat demand; the parameters of the process guiding excess demand change with bond tenor, for example, from φ_{10} to φ_9.

Proposition 1. *The coefficients in the affine pricing Equation (7) obey the recursion*

$$\begin{cases} A_{n+1} &= 1 + \varrho A_n \\ B_{n+1} &= -\mathcal{E}^i + \varphi_n B_n \\ C^i_{n+1} &= C^i_n + A_n(1-\varrho)\theta - 0.5A_n^2\sigma_r^2 + B_n(1-\varphi_n)\kappa_n - 0.5B_n^2\sigma_{q,n}^2 - A_nB_n\rho_n \end{cases}$$

with the initial condition

$$A_1 = 1, B_1 = -\mathcal{E}^i, C^i_1 = 0.$$

Proof. Appendix B demonstrates the statement. Intuitively, the initial values A_1 and C^i_1 result from Equation (1), which, coupled with the absence of arbitrage, requires the yield to maturity of a general bond over a one-period interval to equal the GC rate. The initial condition for B_1 follows from Lemma 1, which forces the repo rate of transactions collateralized by bond issues on special to reflect exceptional demand pressure in proportion to the elasticity of collateral supply. *Q.E.D.*

Recall that A_n captures the compensation for bearing duration risk measured through the interest rate on the GC, common to both general and special bonds, while B_n prices demand pressure and only affects the valuation of special bonds. The coefficient C^i_n soaks up the average discount factor conditional on the tenor and the status. Proposition 1 shows that the sequences $(A_n)_{n\in\mathbb{N}}$ and $(B_n)_{n\in\mathbb{N}}$ are convergent if the persistence parameters ϱ and φ_n are below one in absolute value. Market segmentation arises in equilibrium, as the risk factor q^n_t measuring exceptional demand only exerts upward pressure on the price of targeted bonds and does not affect the price of general bonds.

The finding in Proposition 1 is novel because securities with identical cash flows would have the same price in all the earlier TSMs. Instead, this result shows that in equilibrium price differences arise for bonds targeted by demand pressure, *ceteris paribus*. The key insight is that our setup does not restrict the collateral value of all securities to a common exogenous short rate. In fact, the joint modeling of the general and special yield curves that is consistent with the absence of arbitrage requires a generalization of the canonical TSM to account for the collateral value of bonds in the market for collateralized financing. As a sample application, our model is the first among TSMs to address the *on-the-run/off-the-run* bond spread (Krishnamurthy, 2002) in an equilibrium framework that is consistent with the notion of no arbitrage and endogenously generates specialness.

The recursion for the B_n coefficients in Proposition 1 is parametrized by $\mathcal{E}^i$, which captures the elasticity of collateral supply; namely, the sensitivity of the repo rate on transactions backed by collateral maturing overnight to demand pressure. We nest more traditional models as special cases which obtain by setting $\mathcal{E}^i = 0$, a case corresponding to TSMs where there is no pricing of exceptional demand pressure, the lending rate is exogenous, and the collateral is general. Let us further clarify this point.

Remark 1. *The B_n coefficients are a sequence of zeros for GC bonds, as their repo supply is inelastic. Conversely, for SC bonds the B_n coefficients assume negative values, leading to higher bond prices because these instruments are in elastic supply on the repo market.*

$$\begin{cases} \mathcal{E}^i = 0 \Longleftrightarrow B_n = 0 \quad \forall n & i = g, \\ \mathcal{E}^i > 0 \Longleftrightarrow B_n < 0 \quad \forall n & i = s. \end{cases}$$

As a consequence, C_n is also a function of bond status. More specifically, the difference between C_n^s and C_n^g that we have referred to as D_n behaves as follows:

$$D_n = B_n(1 - \varphi_n)\kappa_n - 0.5 B_n^2 \sigma_{q,n}^2 - A_n B_n \rho_n, \qquad D_1 = 0.$$

Remark 1 is quite intuitive: The B_n coefficients switch off to zero for GC bonds, which are not subject to demand pressure and symmetrically are in inelastic supply on the repo market. Bonds on special are overpriced relative to those that are not subject to demand pressure. We provide a simple sign characterization: $B_n \leq 0 \quad \forall\, n$. The economic reasoning is as follows. Assume by contradiction $B_n > 0$ for some tenor n that would occur if net demand pressure were reduced to some equilibrium price. Since the GC borrowing rate r_t is not sensitive to quantity, arbitrageurs would want to buy an infinite amount of the relatively underpriced special issue and short sell the general one in order to create a portfolio that achieves a perfect hedge against the financing costs of the position (i.e., its short rate risk) and generates riskless profits when both bonds reach maturity, thus contradicting the concept (i.e., finite quantities) of equilibrium that requires market clearing. Remark 1 shows that the effect of demand pressure on bond prices is nonnegative because $B_n \leq 0$, which maps to the well-known result that repo rates are lower for issues on special that guarantee cheaper cash equivalence, since $\mathcal{E}^s > 0$. In general, we prefer not to rule out the unlikely event of negative specialness that could result from selling pressure. However, unless the demand factor q_t^n is negative, SC repo rates are below the GC rate; that is, $r_t^n \leq r_t$.

We employ the closed-form results in Lemma 1 and Proposition 1 above to endogenize repo specialness of bonds with arbitrary tenor. What determines specialness in our model is the behavior of arbitrageurs. When a bond is overpriced because it is exposed to exceptional demand pressure, term structure arbitrageurs reverse the bond in the repo market to sell it short, accepting the risk of rolling over reverse repo contracts until the position is closed or the bond matures, whichever comes first. Repo specialness increases in the short selling behavior of arbitrageurs because the supply of SC is elastic. Our contribution allows us to understand this search for collateral as the reflection of exceptional demand pressure on the bond market. Ultimately, the demand risk factor on the bond market endogenously determines specialness in the repo market through the maximizing behavior of arbitrageurs.

Proposition 2. *Equilibrium specialness is affine in demand pressure:*

$$l_t^n(q_t^n) = \mathcal{E}^i q_t^n = \mathcal{E}^i \underbrace{\left(\varphi_n q_{t-1}^{n+1} + (1-\varphi_n)\kappa_n \right)}_{\textit{Predictable from time } t-1} + \mathcal{E}^i \underbrace{\left(\sigma_{q,n} \upsilon_t^n \right)}_{\textit{Innovation at time } t}. \tag{18}$$

Proof. See Appendix C. *Q.E.D.*

An immediate implication of the previous result is that specialness equals zero for GC bonds that are in inelastic supply, as we would expect ($\mathcal{E}^g = 0$; see Remark 1).[35] Importantly, the elements in Proposition 2 are empirically observable and can be estimated from repo quantities and prices, since no risk compensation is involved.[36] Repo specialness is composed of a predictable component, the foreseeable excess demand for collateral, and a stochastic component, the innovation in the demand for collateral. The first term in Equation (18) is the sum of the unconditional mean of the excess demand and its previous realization, weighted on the persistence of the process. The second term measures the effect of the current demand innovation υ_t^n on the repo specialness l_t^n of the bond. As in D'Amico and Pancost (2022), specialness has both a predictable and a random component. Interestingly, this result demonstrates that the elasticity of collateral supply $\mathcal{E}^i$ does not depend on bond tenor. Regardless of the bond's tenor, repo specialness precisely reflects the excess demand in the bond market, as one would expect from a quantity-driven theory of repo rates. Thus, Figure 2 is a general representation of the SC segment of the repo market independent of the tenor of the bond. Note that Equation (18) is simply Equation (4) multiplied by $\mathcal{E}^i$. The same forces leading to price pressure on the secondary bond markets are those that generate repo specialness. To sum up, a targeted demand shock υ_t^n increases the bond log prices by a factor of $-B_n = \mathcal{E}^s \prod^{n-1}(1+\varphi_n)$ and their repo spreads by a factor of $\mathcal{E}^s$. Thus, when the persistence parameters φ_n are below one,

[35] Remark 1 further shows that $D_1 = 0$, thus ensuring consistency between Equation (16) and Proposition 2.

[36] Repo rates result from the combination of a spot and a forward agreement and thus must be known at time t.

the quantity effects are stronger on the repo market than on the bond market. Bond prices are forward-looking and reflect the expected flow of future repo rates, whose dependence on the current shock dies out over time, while repo rates simply reflect the contemporary stock of collateral. We direct the interested reader to Appendix D for a discussion of the supply and demand curves for collateral in the repo market.

Lemma 2. *The pricing recursion in Proposition 1 is consistent with the optimality of arbitrageurs, who value bonds by taking into account their financing rate on the repo market.*

Proof. By replacing the expressions for the expected bond log price variation given by Equation (11) into the risk-adjusted optimality condition of the arbitrageurs $\mu_t^n = r_t^n$ in Equation (12), we obtain

$$\begin{aligned} & r_t \Delta A_n + q_t^n \Delta B_n + \Delta C_n^i - A_{n-1}(1-\varrho)(\theta - r_t) - B_{n-1}(1-\varphi_n)(\kappa_n - q_t^n) \\ & - 0.5 A_{n-1}^2 \sigma_r^2 - 0.5 B_{n-1}^2 \sigma_{q,n-1}^2 - A_{n-1} B_{n-1} \rho_{n-1} = r_t^n = r_t - l_t^n = r_t - \mathcal{E}^i q_t^n, \end{aligned} \tag{19}$$

with the second equivalence coming from the definition of the n-th special repo rate in Equation (14) and the third following from Proposition 2. Proposition 1 states the unique solution of Equation (19), which can be obtained by isolating all terms in each of the risk factors and those free of the risk factors, and requiring coefficients within each group to add up to zero. Arbitrageurs' behavior is perfectly consistent with Proposition 1, which states that the same recursion that would obtain by solving the difference equation. *Q.E.D.*

Difference Equation (19) must hold for all possible values of the risk factors r_t and q_t^n. From the latter representation, we immediately note the initial conditions for the recursion of the coefficients: The coefficients A_n on r_t must start from the value of 1. The series of B_n coefficients on the demand risk factor q_t^n starts from the initial condition $-\mathcal{E}^i$, the price elasticity of the bond on the repo market, which sets our contribution apart from previous TSMs by allowing bonds with equivalent cash flows to trade at different prices, even under the risk-neutral measure. The C_n^i sequence starts from zero, and adds up the terms that are constant in the risk factors.

Note that from arbitrageurs' first order condition, specialness is indeterminate (as in Duffie, 1996, Proposition 6), leaving unidentified the B_n coefficients that capture the price impact of demand pressure. In fact, specialness affects both the price of the bond on the left side of Equation (19) and its special repo rate on the right side. However, the initial condition for B_1 is set by Lemma 1. Our TSM framework can estimate the risk-adjusted demand-induced *counterfactual* prices of any bond by using readily available data on the elasticity of repo supply, without technical assumptions such as staggered settlement, by exploiting the richness of the Duffie and Kan (1996) representation paired with the breakthrough of the VV TSM. Let us close the model by verifying that the equilibrium concept presented in Section 4.2 characterizes the prices of both general and special bonds.

Remark 2. *From the arbitrageurs' perspective, general and special bonds are equally profitable in equilibrium. The optimality condition for special bonds achieved by setting $i = s$ in Equation (19) folds into the optimality condition for general bonds, which results from the same Equation evaluated at $i = g$, by using Remark 1.*

4.4 Bond Scarcity and the Term Premium

The market price of risk governs the slope of the yield curve; for instance, more negative values result in a steeper yield curve.[37] Consider the following examples, which arise as particular cases in our setup:

$$\lambda^{\mathrm{RN}} = \underline{0}, \quad \lambda^{\mathrm{V}} = \lambda(t, r), \quad \lambda^{\mathrm{VV}} = -a \sum_{m \in \mathtt{N}} \mathbb{E}_t^{\mathbb{P}} \left[U_{t+1}^n X_t^m \sigma^n U_{t+1}^m \right].$$

λ^{RN} corresponds to risk neutrality. The celebrated Vasicek (1977) paper derives the equilibrium under general conditions and no demand uncertainty, which we achieve in our model by using the market price of risk λ^{V} and setting all $\sigma_{q,n}$ to zero. Furthermore, when all demand innovations υ_t^n are perfectly correlated, our equilibrium model reduces to Brennan and Schwartz (1979). Appendix E derives λ^{VV}, the market price of risk associated with the n-th bond in the VV model expressed in discrete time with $1 + \mathtt{N}$ factors, where a denotes the risk aversion of arbitrageurs, which rationalizes the underreaction of long rates to short- rate shocks. Naturally, the aforementioned setup achieves the same result in its discrete-time version. Recall that our general results hold under the risk-neutral measure. To obtain equilibrium under the $\mathbb{P}$ measure, it suffices to apply a Girsanov transformation to Equation (12) by using the preferred specification for the market price of risk. Table OA.1 in the Appendix compares our theory with benchmark TSMs. The following example provides a closed-form solution for bond prices under the physical probability measure.

[37] Repo specialness, the spread between GC and SC rates, does not vary with the market price of risk. Repo rates are determined at the inception of the contract and involve no risk. As discussed above, the quantity of collateral demanded in the market affects repo specialness, along with the elasticity of collateral supply. Thus, repo rates simply reflect the contemporary stock of collateral on the market. Conversely, bonds are forward-looking expectations of the relevant future repo rates, whether general or special. Thus, bond prices include a risk compensation because the notional principal is discounted at the entire stream of future repo rates. Proposition 3 discusses the relation between repo specialness and the term premium.

Example 2. *Suppose there is only one demand risk factor, $\upsilon_t^n = \upsilon_t \, \forall n$, and assume for simplicity that it is independent of the short rate.*[38] *Then, the pricing coefficients under the physical measure $\mathbb{P}$ are given by the following recursion, with $A_1 = 1, B_1 = -\mathcal{E}^i, C_1 = 0$:*

$$\begin{cases} A_{n+1} & = 1 + \varrho A_n, \\ B_{n+1} & = B_1 + \varphi_n B_n, \\ C_{n+1}^i & = C_n^i + A_n(1-\varrho)\theta - 0.5A_n^2\sigma_r^2 + B_n(1-\varphi_n)\kappa_n - 0.5B_n^2\sigma_{q,n}^2 \\ & \quad +0.5\left[\lambda_r^2 - (\lambda_r - A_n\sigma_r)^2 + \lambda_q^2 - (\lambda_q - B_n\sigma_{q,n})^2\right], \end{cases}$$

where λ_r and λ_q are the prices of the short rate and demand risk factors, respectively.[39]

The literature on QE suggests that asset purchases affect the term structure by influencing the risk premium and by inducing local supply effects (see, e.g., D'Amico et al., 2012). These findings call for the specification of a term premium that depends on the holdings of the private sector. In Section 6.2, we show that our TSM generates strongly localized supply effects, a feature that, to our knowledge, is absent from the previous literature. The next result clarifies that repo specialness affects the term premium in *any* specification of risk premia featuring portfolio holdings.

Proposition 3. *Suppose that the holdings of the arbitrageurs X_t^n affect the market price of risk. Then, the repo specialness l_t^n affects the term premium of both the special and general yield curves.*

Proof. Consider a generic term premium $\lambda_r(\cdot)$, a differentiable function of the holdings of the arbitrageurs X_t^n. From Equation (3), $Z_t^n(g) = 0$. Furthermore, from Equation (5), $X_t^n(i) = Z_t^n(i)$. Thus, we focus on the holdings $X_t^n(s)$ of special bonds, without loss of generality. In

[38] An excellent reference for discrete-time affine models with independent factors is Backus et al. (1998).

[39] In models where the market price of risk is free from equilibrium quantities (e.g., Vasicek, 1977; Brennan and Schwartz, 1979), no further step is required, and bond prices follow from Equation (7). In VV, the market price of risk itself depends on the pricing coefficients through the market clearing exposures of arbitrageurs. This example, with two independent factors, corresponds to Lemma A.2 in VV, where closed-form solutions are available for the limiting case of infinite risk aversion and risk neutrality, the latter of which corresponds to our Proposition 1.

equilibrium,

$$\frac{\partial \lambda_r(\cdot)}{\partial l_t^n} = \frac{\partial \lambda_r}{\partial X_t^n(s)} \frac{\partial X_t^n(s)}{\partial l_t^n} = -\frac{\partial \lambda_r}{\partial X_t^n(s)} \frac{\partial Z_t^n(s)}{\partial l_t^n} = -\frac{\partial \lambda_r}{\partial X_t^n(s)} \frac{\partial [q_t^n - \alpha^n (A_n r_t + B_n q_t^n + C_n^s)]}{\partial l_t^n}$$
$$= -\frac{\partial \lambda_r}{\partial X_t^n(s)} \frac{\partial [\frac{l_t^n}{\mathcal{E}^s} - \alpha^n (A_n r_t + B_n \frac{l_t^n}{\mathcal{E}^s} + C_n^s)]}{\partial l_t^n} = -\frac{\partial \lambda_r}{\partial X_t^n(s)} \frac{1 - \alpha^n B_n}{\mathcal{E}^s}$$
$$= -\frac{\partial \lambda_r}{\partial X_t^n(s)} \left[\frac{1}{\mathcal{E}^s} + \alpha^n \prod^{n-1} (1 + \varphi_n) \right],$$

which differs from zero, since the holdings of arbitrageurs influence the market price of risk. Above, we have used Proposition 1, which shows that $B_n = -\mathcal{E}^s \prod^{n-1}(1+\varphi_n)$, and Proposition 2, which states that $l_t^n = \mathcal{E}^s q_t^n$. In most applications, $\frac{\partial \lambda_r}{\partial X_t^n(s)} < 0$, since the arbitrageurs demand higher compensation in the form of a term premium when engaging in quantitatively larger carry trades. *Q.E.D.*

We view this result as one of our key contributions and one that has a natural interpretation. When bonds are in infinite supply, QE lowers the term premium by inducing arbitrageurs to increase their short selling activity. However, when bonds are scarce, special repo rates arises from the combination of finite supply and excess demand, and act in the *opposite* direction, raising term premia. Specialness is the cost of carry trade arbitrage positions hedged against interest rate risk. With higher specialness, arbitrageurs will want to scale down their positions, *ceteris paribus*. To induce them to roll over large quantities of carry trade positions, the risk premium must rise. From the closed-form solution above, this effect is stronger when bonds have a lower elasticity of supply on the repo market $\mathcal{E}^s$. As illustrated in Figure 2, any given level of repo specialness maps to higher bond quantities in equilibrium when the collateral is in less elastic supply. Moreover, the effect of specialness on arbitrageurs' optimal holdings is directly proportional to the persistence of specialness φ_n, which increases the likelihood of large realizations of specialness in the future, conditional on current values. On the demand side, higher specialness raises bond valuations, reducing the bidding pressure of preferred-habitat investors (including those other than the central bank) that have price semi-elasticity α^n. These three channels work in the same direction, and special repo rates increase the term premium through their combined effect. Special bonds are commonly used to hedge interest rate risk, and it is natural for their scarcity to affect the entire term structure. Proposition 3 shows that a significant amount of repo specialness influences the term premium for *all* bonds whenever the market price of risk depends on the portfolio of the investors, leaving the functional form of λ_r completely unrestricted.

As a concrete example, we have adapted to our general model the market price of interest rate risk specified by VV:

$$\lambda_r^{\mathrm{VV}} = -a \sum_{m \in \mathtt{N}} X_t^m (A_{m-1}\sigma_r + B_{m-1}\rho_m).$$

Recall that by the market clearing condition, X_{t+1}^n is negative, at least with buying pressure. Thus, in the model of Vayanos and Vila (2021), a larger magnitude of arbitrageurs' exposures results in a more positive or less negative market price of risk λ_r^{VV}, while QE reduces the slope of the yield curve. When we instead allow for the supply of bonds to be finite, specialness reduces the optimal holdings of arbitrageurs and dampens the decrease in the slope of the yield curve. The term premium required by arbitrageurs increases with repo specialness for two clear reasons. First, higher repo specialness is tantamount to higher costs of arbitrage. As shown in Proposition 3, any increase in repo specialness affects the holdings of the private sector by a factor of $\frac{\partial X_t^n(s)}{\partial l_t^n} = -\left[\frac{1}{\mathcal{E}^s} + \alpha^n \prod^{n-1}(1+\varphi_n)\right]$, as we endogenously derived in equilibrium. Second, it is intuitive that the repo specialness of, say, m-tenor bonds commands a compensation for its correlation ρ_m with the general interest rate risk. Many fixed-income trading desks use special bonds, which enjoy superior liquidity, in order to hedge their duration risk. If repo specialness is correlated with the interest rate, as our model permits, such a hedging strategy might become more costly precisely when it becomes more necessary, raising term premia. As a consequence, a reduction in specialness, for instance through an SLF, results in stronger impacts of QE on the reduction of risk premia. As we show in Section 6.2, the SLF policy also controls the localization of the supply effects of QE.

The above analysis shows that even in the absence of risk aversion, repo specialness and the general level of interest rates interact with each other through the effect of their correlation on the expectations of future rates. Suppose, for instance, that a central bank announces it will hike its reference rates in the future. Then, the expectations of future special repo rates will also change to reflect the rising expectations of GC rates. The converse is also true, and expectations of special repo rate increases lead the general term structure to become more upward sloping, with the obvious caveat that correlations may also change over time.

4.5 Testable Predictions

Perhaps the most interesting testable prediction of our theory is a preference-free asset pricing equation that generalizes the classical term structure equilibrium equation. Based on the notion of arbitrage, we point out that the excess return to risk ratio should be constant in the cross-section of nearly risk-free bond returns, but only after taking into account the convenience yield (that is, the repo specialness) of the asset. Equation (13) is relatively simple to apply to the data. To test its empirical counterpart, we require a panel of nearly riskless bonds that consists of observations of their secondary market and repo quotes. The data should include both generic and special bonds with the same tenor n.

A formal empirical analysis is beyond the scope of this paper, but we can sketch the necessary steps. It is natural to estimate the (Jensen-adjusted) drift term of each bond $\hat{m}_t^n$ as the period-to-period bond return using market data and to assess the robustness of the estimates to different frequencies. Similarly, a common approach is to use variation in returns to proxy for the standard deviation $\hat{s}^n$. Finally, the exercise requires a measure for the risk-free rate r_t and one for the tenor-specific overnight special repo rate r_t^n. One can compute both by using volume-weighted averages of GC rates and SC repo market rates, grouping bonds by their tenor, and use time fixed effects to soak up the adjustment in the market price of risk. The repo specialness l_t^n can be inferred from the GC and SC rates. Next, the following simple panel linear regression model could test whether the proposed equilibrium TSM reasonably improves on the canonical specification (Vasicek, 1977; Brennan and Schwartz, 1979) by accounting for bond-specific short rates.

$$\frac{\hat{m}_t^n}{\hat{s}^n} = \text{Time FE} + \beta_1 \frac{r_t}{\hat{s}^n} + \beta_2 \frac{l_t^n}{\hat{s}^n} + \text{error term}$$

Our model suggests that β_2 should be negative to prevent arbitrage opportunities. Intuitively, special bonds should have lower excess returns relative to general bonds, since the former generate additional cash flows on the repo market. We caution the reader that while this preliminary analysis may be useful, a formal test of the above requires more sophisticated specifications to account for the simultaneous determination of bond prices and repo specialness.

Moreover, Proposition 3 can be tested by regressing the term premium on the average specialness $\bar{l}_t$ after controlling for variables in Ξ_t:

$$y_t^{10}(g) - y_t^2(g) = \gamma_0 + \gamma_1 \Xi_t + \gamma_2 \bar{l}_t + \text{error term}$$

Our model suggests that γ_2 should be positive, since high levels of repo specialness reduce the short selling behavior of term structure arbitrageurs and their required compensation for risk. We leave to future research the task of carrying out a formal econometric test of these specifications. Clearly, the term premium should be estimated by dropping highly special securities from the pool of high-quality bonds used to fit the yield curve – a practice currently followed by the Fed but not by the ECB, even though the specialness of German bunds routinely reaches as much as 50 bps.

5 Extensions and Generalizations

5.1 Imperfect Substitutability in the Demand of Preferred-Habitat Investors

In general, preferred-habitat investors that aim to match the duration of their liabilities by using the most liquid issue of a certain bond may also consider special bonds featuring a similar but not identical time to maturity for which terms may be more attractive, trading off prices against maturity proximity to their respective demand shocks. For example, suppose an insurance company wishes to hedge the interest rate risk of its 10-year liabilities. Ignoring coupon effects, one way to achieve immunization in the bond market is by targeting the most liquid maturity-matched bond issue. However, if a bond with a residual maturity of $9\frac{3}{4}$ years has a much lower price, it is reasonable to think that the company will closely monitor the prices of both securities before implementing its hedging trades. These considerations induce us to generalize the demand specification of the preferred-habitat investors to model the consequences of their rebalancing on financial markets, as empirically documented by Koijen et al. (2021). Consider the extension of the demand specification of preferred-habitat investors,

$$Z_t(i) = \begin{cases} Q_t - \mathcal{A}\mathcal{B}_t(i) & i = s, \\ \underline{0} & i = g, \end{cases} \tag{20}$$

where we consider a set of discrete tenors $n \in [1, 2, \ldots, \mathtt{N}]$ and define[40]

$$\underset{\mathtt{N}\times 1}{Z_t(i)} = \begin{bmatrix} Z_t^1(i) \\ \vdots \\ Z_t^{\mathtt{N}}(i) \end{bmatrix}, \quad \underset{\mathtt{N}\times 1}{Q_t} = \begin{bmatrix} q_t^1 \\ \vdots \\ q_t^{\mathtt{N}} \end{bmatrix}, \quad \underset{\mathtt{N}\times\mathtt{N}}{\mathcal{A}} = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,\mathtt{N}} \\ \vdots & \ddots & \vdots \\ \alpha_{\mathtt{N},1} & \cdots & \alpha_{\mathtt{N},\mathtt{N}} \end{bmatrix}, \quad \underset{\mathtt{N}\times 1}{\mathcal{B}_t(i)} = \begin{bmatrix} \log b_t^1(i) \\ \vdots \\ \log b_t^{\mathtt{N}}(i) \end{bmatrix}.$$

In Equation (20), special bonds are substitutable with one another. The vectors $Z_t(i), Q_t$, and $\mathcal{B}_t(i)$ stack vertically excess demands functions, demand shocks, and the log prices of special bonds of each tenor, respectively. The matrix $\mathcal{A}$ consists of the excess demand semi-elasticities to prices across tenors $\alpha_{n,m}$, representing the change in the quantity demanded of the special bond m resulting from the percentage change in the price of the special bond with time to maturity n. The baseline model in Section 4 corresponds to the case when $\mathcal{A}$ is a positive definite

[40]Without loss of generality, as discrete indexes can capture any frequency interval, e.g., monthly, yearly, etc.

diagonal matrix. If, however, other maturities are imperfect substitutes, off-diagonal elements of $\mathcal{A}$ are negative because demand increases non-linearly in the price of bonds of different maturities (i.e., linearly in their log price), so that the marginal rate of substitution between pairs of maturities varies along the demand curve. It is reasonable (but not necessary) to assume that the cross-price sensitivity of demand decreases as the distance to the main diagonal increases. In general, $\mathcal{A}$ might well be asymmetric.[41]

In order to write the joint evolution of the autoregressive demand risk factors compactly, recall that there is no previous demand for newly issued bonds of the longest maturity (by construction). Expressing the Vasicek processes from Equation (4) jointly,

$$\begin{cases} q^1_{t+1} & = \phi_1 q^2_t + (1-\phi_1)\kappa_1 + \sigma_{q,1}\upsilon^1_{t+1}, \\ q^2_{t+1} & = \phi_2 q^3_t + (1-\phi_2)\kappa_2 + \sigma_{q,2}\upsilon^2_{t+1}, \\ \vdots & \\ q^{\mathtt{N}}_{t+1} & = \phi_{\mathtt{N}} \underbrace{q^{\mathtt{N}+1}_t}_{=0} + (1-\phi_{\mathtt{N}})\kappa_{\mathtt{N}} + \sigma_{q,\mathtt{N}}\upsilon^{\mathtt{N}}_{t+1}, \end{cases} \tag{21}$$

which we can write more compactly as

$$Q_{t+1} = \Phi Q_t + \overline{Q} + \Omega V_{t+1}, \tag{22}$$

$$\underset{\mathtt{N}\times\mathtt{N}}{\Phi} = \begin{bmatrix} 0 & \varphi_1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \ddots & 0 & \varphi_{\mathtt{N}-1} \\ 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \underset{\mathtt{N}\times 1}{\overline{Q}} = \begin{bmatrix} (1-\varphi_1)\kappa_1 \\ \vdots \\ (1-\varphi_{\mathtt{N}})\kappa_{\mathtt{N}} \end{bmatrix}, \quad \underset{\mathtt{N}\times\mathtt{N}}{\Omega} = \begin{bmatrix} \sigma_{q,1} & \cdots & \rho_{1,n} \\ \vdots & \ddots & \vdots \\ \rho_{\mathtt{N},1} & \cdots & \sigma_{q,\mathtt{N}} \end{bmatrix}, \quad \underset{\mathtt{N}\times 1}{V} = \begin{bmatrix} \upsilon^1_{t+1} \\ \vdots \\ \upsilon^{\mathtt{N}}_{t+1} \end{bmatrix}.$$

The matrix Φ displays the persistence parameters ϕ_n on the superdiagonal, and 0 elsewhere. The autoregressive representation in Equation (22) is natural, as illustrated by the system of Equation (21), noting that the process for demand risk factors evolves by replacing the time subscript t with $t+1$ and the tenor superscript n with $n-1$. By construction, there are no previous demand shocks on special issues of the longest maturity, $q^{\mathtt{N}+1}_t = 0$. To state the model in full generality, we allow the innovations in the preferred-habitat demand for maturity j to covary with those for other maturities and denote the respective correlation coefficients via $\rho_{i,j}$,

[41] To see this, consider an example in which preferred-habitat investors targeting the 9-year tenor bond are willing to substitute with a bond with 10 years to maturity ($\alpha_{9,10} < 0$), but preferred-habitat investors populating the segment with 10 years to maturity are instead not (or perhaps less) willing to shift their demand pressure to the 10-year bonds ($\alpha_{10,9} = 0$) because they are committed by institutional constraints to invest in the long-duration fixed-income market that is composed of bonds with a time to maturity equal to or longer than 10 years.

represented in the off-diagonal elements of Ω. Let us conjecture by analogy with the scalar case that the vector of price processes is affine in the short rate and, conditional on the bond status, in the vector of demand shocks:

$$-\mathcal{B}_t(i) = \begin{cases} Ar_t + BQ_t + C & i = s, \\ Ar_t + C & i = g, \end{cases} \tag{23}$$

where A, B, and C are matrices that consist of the pricing recursion coefficients.

$$\underset{\mathtt{N}\times 1}{A} = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}, \qquad \underset{\mathtt{N}\times\mathtt{N}}{B} = \begin{bmatrix} B_{1,1} & \cdots & B_{1,\mathtt{N}} \\ \vdots & \ddots & \vdots \\ B_{\mathtt{N},1} & \cdots & B_{\mathtt{N},\mathtt{N}} \end{bmatrix} = \begin{bmatrix} B^1 \\ \vdots \\ B^{\mathtt{N}} \end{bmatrix}, \qquad \underset{\mathtt{N}\times 1}{C} = \begin{bmatrix} C_1^i \\ \vdots \\ C_{\mathtt{N}^i} \end{bmatrix}.$$

Equation (23) is a generalization of Equation (7) that allows the prices of special bonds to depend on the entire term structure of demand risk factors. This formulation reflects substitutability across bonds in the demand of preferred-habitat investors. The model in Section 4 corresponds to the case where $\mathcal{A}$ and B are diagonal matrices. For convenience, we assemble the rows of B into the vectors $(B^n)_{n=1}^{\mathtt{N}}$, which represent the price sensitivity of bonds with maturity n to the demand risk factors across the entire term structure. By the market clearing condition in Equation (5), arbitrageurs are only active in special bonds to smooth away price differences induced by exceptional demand pressure. We next solve their maximization program and drop the bond status $i = s$ for clarity of notation. Note that first-differencing the vector of log prices $\mathcal{B}_{t+1}$ amounts to computing the vector of one-period bond returns, whose law of motion is

$$\Delta\mathcal{B}_{t+1} = M_t - \Sigma U_{t+1}, \tag{24}$$

where

$$\underset{\mathtt{N}\times 1}{M_t} = A\Delta r_t + B\Delta Q_t + C - A\Big[r_t + (1-\varrho)(\theta - r_t)\Big] - B\Big[\Phi Q_t + \overline{Q}\Big],$$

$$\underset{\mathtt{N}\times\mathtt{N}}{\Sigma} \quad \underset{\mathtt{N}\times 1}{U} = \Omega V_{t+1} + \sigma_r \eta_{t+1} I_{\mathtt{N}}.$$

Note the parallel between Equation (24) and Equation (8) in the univariate model of Section 4. The term M_t is simply a vector that stacks vertically all the predictable changes m_t^n in the log price of bonds with tenor n. Similarly, the matrix product ΣU_{t+1} is the multidimensional version of the vector product $\sigma^n U_{t+1}^n$ in Section 4, where the demand risk factors are allowed to correlate with one another and with the GC rate. We let $\tilde{M}_t = M - 0.5\mathbb{E}_t^{\mathbb{Q}}\left[U'_{t+1}\Sigma'\Sigma U_{t+1}\right]$ denote the vector of drifts after accounting for Jensen's correction terms, and $R_t = \left[r_t^1 \quad \ldots \quad r_t^{\mathtt{N}}\right]$ represent the vector of special repo rates. The first order condition for the optimality of the arbitrageurs' problem with respect to special bonds expressed under the $\mathbb{Q}$ measure reads

$$\tilde{M}_t = R_t. \tag{25}$$

Equation (25) is the natural extension of Equation (12). However, both sides of the equilibrium reflect the generalization of the demand function of preferred-maturity investors. On the left side, the drift term differs from the baseline case since the coefficients in the affine pricing Equation (23) satisfy an extension of the recursion in Proposition 1, where substitutability across bonds affects the coefficients already under $\mathbb{Q}$, as demonstrated in Appendix F. On the left side, the specialness of bonds now reflects demand pressure across the entire term structure of interest rates, because arbitrageurs take the opposite side of preferred-habitat investors that can substitute across maturities. Specifically, the repo specialness of the m-period maturity bond now reacts to demand pressure on every other bond, and its gradient with respect to the term structure of demand pressure (which could be measured by the volume in the bond market of special issues in excess of that of general issues) is given by

$$H^{m,i} = [\eta_1^{m,i} \qquad \ldots \qquad \eta_{\mathtt{N}}^{m,i}], \qquad \eta_n^{m,s} = -\frac{\partial l_t^{m,s}}{\partial q_t^n}, \qquad \eta_n^{m,g} = 0. \tag{26}$$

The vector $H^{m,i}$ from Equation (26) generalizes the elasticity of the supply of collateral in the repo market $\mathcal{E}^i$ of Equation (17) and already appears under $\mathbb{Q}$ in the recursion for the pricing coefficients derived in Appendix F. On the other hand, the semi-elasticity of substitution parameters in $\mathcal{A}$ affects bond prices under the physical measure $\mathbb{P}$ if quantities enter the market price of risk, as is the case in VV. Using the market clearing condition,

$$\begin{aligned}
\lambda &= a\mathbb{E}_t^{\mathbb{P}}\left[\Sigma U_{t+1} X_t \Sigma^\top U_{t+1}^\top\right] \\
&= a\mathbb{E}_t^{\mathbb{P}}\left[\Sigma U_{t+1}\Big(Q_t - \mathcal{A}\mathcal{B}_t\Big)\Sigma^\top U_{t+1}^\top\right] \\
&= a\mathbb{E}_t^{\mathbb{P}}\left[\Sigma U_{t+1}\Big(Q_t + \mathcal{A}\big[Ar_t + BQ_t + C\big]\Big)\Sigma^\top U_{t+1}^\top\right].
\end{aligned}$$

The market price of risk decreases with substitutability across varieties because the off-diagonal elements of $\mathcal{A}$ are negative when other maturity segments are regarded as imperfect substitutes for investors targeting their preferred habitat.

5.2 Heterogeneous Arbitrageurs: Haircuts and Borrowing Constraints

The existing literature has considered term structure arbitrageurs as a homogeneous group, abstracting from important differences amongst them. For instance, hedge funds are aggressive investors, while broker dealers have a relatively higher risk aversion. Consider a mass one of mean-variance arbitrageurs indexed by j, with varying degrees of risk aversion a^j and levels of wealth W_t^j, that hold positions $(X_t^{j,n})_{n\in\mathtt{N}}$. Clearly, different business models also give rise to differences in counterparty risk. From the perspective of academics, market participants, and policymakers, haircuts are viewed as mitigating such counterparty risk. The term structure literature focuses on risk-free bonds, for which we can abstract from the default of the issuer and focus on counterparty risk. On the other hand, repo haircuts are on average larger with higher borrower and lender credit and funding liquidity risk (Martin et al., 2014), because both parties could default and because both are typically interested in rolling over the transaction.[42] This motivates us to consider a counterparty-specific haircut h^j applied to GC and SC repo positions as decreasing in the risk aversion of the j-th term structure arbitrageur. For instance, $h^j = 0.05$ means that the j-th investor must pledge five times the price of the bond as collateral in order to obtain \$100 of repo financing.

For greater generality, we consider borrowing constraints requiring arbitrageurs to have "skin in the game" and back the haircuts of their positions with their own wealth. In the presence of haircuts and borrowing constraints, the maximization programs of arbitrageurs would incorporate the scarcity of capital and the requirement that each position be backed by the commitment of a certain haircut of a bond's market value instead of generating returns at the GC repo rate r_t. Let us denote through ν_j the multiplier associated with the non-negativity constraint on the wealth of the j-th arbitrageur, whose problem is

$$\max_{\{X_t^{j,n}\}_{n\in\mathtt{N}}} \mathbb{E}_t^{\mathbb{P}}\left[\Delta W_{t+1}^j\right] - \frac{a^j}{2}\mathbb{V}_t^{\mathbb{P}}\left[\Delta W_{t+1}^j\right] + \nu_j\left[W_t^j - h^j\sum_{n\in\mathtt{N}} X_t^{j,n}\right], \tag{27}$$

$$\Delta W_{t+1}^j = \left(W_t^j - h^j\sum_{n\in\mathtt{N}} X_t^{j,n}\right) r_t + \sum_{n\in\mathtt{N}} X_t^{j,n}\left(\log\frac{b_{t+1}^{n-1}}{b_t^n} - r_t^n\right). \tag{28}$$

Equation (27) is the objective function with borrowing constraints, specified under $\mathbb{P}$ in order to express the idiosyncratic degree of risk aversion a^j. Equation (28) is the law of motion of

[42] There are several instances of "fails" in the repo market, but these are mostly cases where the repo contract is simply rolled over for another day or two, rather than a case of true default.

wealth modified to reflect the forgone returns remunerated at the GC repo rate and proportional to the haircut locked up by each long position in the special bonds; namely, the opportunity cost $h^j \sum_{n\in \mathtt{N}} X_t^{j,n} r_t$. Let us solve the problem under the equivalent martingale measure. The Kuhn-Tucker first order conditions for an interior optimum are

$$\mu_t^n - r_t^n - h^j r_t = 0, \qquad \forall \qquad t, n,$$
$$\nu_j \left[W_t^j - h^j \sum_{n\in \mathtt{N}} X_t^{j,n} \right] = 0, \quad \nu_j \geq 0.$$

This generalization nests the baseline Equation (13) when the haircut h^j and the borrowing constraint ν^j are equal to zero. Intuitively, the risk-adjusted expected return of a bond over and above its special repo rate must now be equal to the cost of the position in terms of forgone returns remunerated at the GC rate times the haircut. Moreover, carry trades are only possible when capital is available (i.e., the borrowing constraint is respected). The optimization program thus sheds light on the effect of holding costs (Pontiff, 1996) and capital constraints (Gromb and Vayanos, 2018) on arbitrageur behavior.[43]

We outline a comparative statics analysis, leaving a formal treatment of the issue as a suggestion for future research. Borrowing constraints lead to a "gambling for resurrection" effect. With limited liability, it might be optimal to increase the risk profile as wealth shrinks, since short selling frees up cash on the spot. Interestingly, haircuts generate clientele effects on the supply side. Each term structure arbitrageur faces an effective yield curve that follows Proposition 1, with one exception: the initial condition A_1^j becomes arbitrageur-specific and shifts upward in proportion to the haircut. That is,

$$A_1^j = 1 + h^j.$$

The interpretation of the above analysis is straightforward. For example, consider a hedge fund investing \$1 million in the repo market at the GC rate. Since the fund has a high risk tolerance (i.e., a low risk aversion a^j), a large haircut applies to the transaction, and at time t the fund receives $1 + h^j$ units of the GC bond in exchange for cash. Thus, when haircuts h^j are larger, the reward for cash is higher for a given GC rate r_t. Conversely, if the fund wishes to reverse a GC or SC bond in the market for repurchase agreements, it has to pledge more cash. Recall from Section 4.4 that risk aversion positively affects the average slope of the yield curve. As a result, market participants attaching a low penalty a^j to the variance of future wealth specialize

[43] Regulatory constraints in the spirit of Du et al. (2018) are achieved by appropriately redefining Equation (27).

in arbitraging away price differences on longer-maturity bonds.

On the one hand, less risk-averse arbitrageurs must pledge a relatively large amount of cash h^j for each bond they short. On the other hand, the market compensation for the rollover risk is higher than the one they would require and is even substantial at longer horizons. Thus, arbitrage profitability increases in the horizon of the carry trade for agents with lower risk aversion than the one that prevails in the market. In our example, broker dealers specialize in term structure arbitrage at the short end of the yield curve, where low credit risk grants a comparative advantage, and hedge funds at the long end of the yield curve. In summary, preferred-habitat investors are by no means specific to the demand side of the market. Arbitrageurs are also heterogeneous in their business models, which affects their carry trades through their choice sets and preferences.

5.3 *The Degrees of Specialness and the Treasury Auction Cycle*

So far, simplicity considerations have led us to consider the bonds (volume-weighted) average specialness for a given tenor. However, bonds with the same time to maturity often trade at different degrees of specialness due to the differential demand pressure across them, suggesting a generalization of the status of specialness from a binary to a categorical variable. We thus re-index special bonds through $i \in \mathbf{s} = \{s_{\mathrm{P}}, \ldots, s_1\}$; to clarify this idea, think of *on-the-run* securities, *first-off-the-run* securities, and so on. Without loss of generality, we order the elements in the set s as decreasing in their degree of specialness. The demand of preferred-habitat investors thus becomes

$$Z_t^n(i) = \begin{cases} q_t^n - \alpha^n \log b_t^n(i) & i \in \mathbf{s}, \\ 0 & i = g, \end{cases}$$

and allows for varying degrees of demand risk across differentially special bonds,

$$q_{t+1}^n(s_{\mathrm{p}}) = \varphi_n q_t^{n+1}(s_{\mathrm{p}+1}) + (1-\varphi_n)\kappa_n + \sigma_{q,n}\upsilon_{t+1}^n(s_{\mathrm{p}}). \tag{29}$$

Equation (29) models the gradual convergence of demand pressure to zero as the bond matures. For instance, excess demand for the *on-the-run* bond (indexed by s_{P}) transitions with persistence φ_n to buying pressure in the next period, when the same bond becomes *first-off-the-run* (indexed by $s_{\mathrm{P}-1}$), and so forth. Naturally, the *on-the-run* bond has the highest specialness, $l_t^n(q_t^n(s_{\mathrm{P}}))$. Each of the results derived in the present study naturally extends to the case where the full distribution of bond prices and special repo rates is endogenized to reflect different demand pressures on the secondary bond market for bonds of a given tenor. For example, the above discussion is relevant to the US Treasury auction cycle.

The US government generally issues Treasury bonds at a pre-announced frequency. As market participants roll over their exposures into new issues, the largest specialness spreads typically arise between two auctions. For instance, Krishnamurthy (2002) documents the systematic convergence of the repo spread tied to the 30-year Treasury bond over successive issuances. Watersheds in the auction cycle are the announcement date on which forward contracts on the new bond are initiated, often referred to as "when-issued" trading, followed after about one week later by the auction date, and two weeks later by the issuance date. For example, the 2-year and 5-year US Treasuries are issued on a monthly auction cycle, and the 10-year and 20-year notes are issued on a quarterly cycle.[44] Within each cycle, regular "retaps" provide additional amounts of a previously issued security in many sovereign bond markets. Thus, specialness premia also exhibit a strong cyclicality, because the auction frequency is generally regular and predictable. However, repo specialness is not confined to *on-the-run* bonds. Typically, specialness gradually decreases over the life cycle of the bond as the security becomes *first-off-the-run*, *second-off-the-run*, and so forth (see, e.g., Tuckman and Serrat, 2022). The predictability of bond specialness described above extends to cheapest to deliver bonds for futures contracts in European markets (Buraschi and Menini, 2002), especially when bonds are issued on a retaps basis; that is, increasing the amount outstanding of already issued bonds. Relative to the US market, collateral specialness is substantially more persistent in European repo markets. As an illustration, Figure OA.2 shows the one-year volume-weighted trailing average of SC transactions collateralized by Italian treasury bonds, grouped by different maturities, as a function of the number of days passed since the bonds were first issued. From the chart, we see that the repo "specialness" of Italian bonds with original maturities of 5, 10, and 15 years can be detected throughout their entire trading life cycle. We further note that the specialness of bonds with 15 years of maturity at issuance peaks after about 5 years, when the time to maturity reaches 10 years, and decays sharply thereafter. However, these aggregate patterns are influenced by retaps and market conditions. Even though repo specialness has a stronger persistence – and a larger impact on bond prices – in European sovereign bond markets, in the remainder of the analysis we focus on the US market where, as a result of the regular US Treasury auction cycle, it is more readily interpretable. We illustrate the yields and repo rates corresponding to differentially special bonds in the calibration below. From Proposition 2, in the equilibrium of our model, specialness l_t^n is proportional to the excess demand for bond q_t^n.

[44]See https://www.treasurydirect.gov/auctions/general-auction-timing/ for additional details.

Thus, the cyclical behavior of repo spreads is guided by the parameters governing exceptional demand pressure in Equation (4). As discussed in Section 6, a low persistence of demand innovations φ_n and a long-run mean $\kappa_n = 0$ are consistent with the strong cyclicality of special repo rates in the United States and the economic intuition that preferred-habitat investors roll over their positions into liquid bonds.

6 Calibration

6.1 Two Yield Curves

The calibration of our model is tantamount to the combined modeling of the general and special yield curves in the bond market and of the specialness in the repo market. The aim of this calibration is to highlight the effects of counterfactual scenarios determined by conventional monetary policy tools that guide short rate behavior and the use of unconventional instruments through QE, which act through demand pressure on the bond and repo markets.[45] We use the simple model structure outlined in Example 2 above and refer to well-established contributions in the literature on financial economics.

For comparability with VV, we set the maturity, $\mathtt{N} = 30$, and use publicly available 1985–2020 US Treasury data from Gürkaynak et al. (2007) (GSW). It is worth emphasizing that the latter data set excludes bonds targeted by exceptional demand pressure, thus fitting well with our purpose of calibrating the general yield curve. We express all rates on a per annum basis. We take a standard value for the long-run mean θ from He and Milbradt (2014) and specify ϱ and σ_r to match the autocorrelation and standard deviation of the one-year yield, respectively. The market price of GC bond risk λ_r in this calibration is considered constant and equal to 0.42, replicating the average 10-year bond yield in the GSW data. To measure $\mathcal{E}^s$, we use the estimated impact of bond purchases on their returns conditional on other characteristics in D'Amico and King (2013). To model demand risk, we use a homogeneous level of excess demand $\overline{q}_t$ for the special bond across tenors which reverts to zero at the speed φ.

We set $\overline{q}_t$ to 26 bps to match the average *on-the-run* repo spread of 19.4 bps documented by D'Amico et al. (2018). This value approximates the GC repo/T-bill spread of 23.65 bps found by Nagel (2016), although it is more conservative, and far lower than the repo spread of around 40 bps observed in the German bund repo market. We tune the persistence parameter φ to the

[45] In affine TSMs, the persistence parameters define the curvature of the yield curve, and the relative importance of shocks is more pronounced at shorter maturities, as current realizations of stationary risk factors are relatively more informative for the near future.

ratio between the average *on-the-run* repo spread to the average repo spread of *second-off-the-run* and older bonds on special of 4.88 bps in D'Amico et al. (2018). Thus, the half-life of $\overline{q}_t$ is six months. To illustrate local supply effects in our model, q_t^{10} reproduces the 10-year special bond price residual from the GSW model estimates in D'Amico et al. (2018). We explain these choices in detail in Table I.

As shown in Figure 3, our model features several salient characteristics. First, as the top panel shows, two yield curves – general and special – co-exist simultaneously. For each tenor, the yield to maturity of the special bond exposed to demand pressure is lower (i.e., its price is higher) than that of the general bond. Thus, the yield curve constructed by interpolating the prices of SC zero coupon bonds lies below the yield curve of GC bonds, but their difference shrinks with time to maturity as demand pressure shocks decline over time. That is intuitive, given that the two curves are generated by rolling over GC and SC rate risk and that SC repo rates are generally below GC ones. In fact, the vertical distance between the GC and the SC curve at short residual maturities reflects the elasticity of the repo market supply of SC $\mathcal{E}^s$ and the persistence φ at the longer end of the yield curve. The gradually decreasing pattern of bond specialness recalls the spread between *on-the-run* and GSW-fitted yields documented in Figure 1 in Greenwood et al. (2015). Second, the joint modeling of the GC and SC yield curves in the bond market is only possible in the context of our theory, because we account for differentials in the special repo rates induced by these bonds. In the bottom panel of Figure 3, we show the repo rate for GCs in red, which is assumed to be constant across time to maturity, and for SC transactions in blue. That is, the SC rate captures the average special repo rates across all the transactions of special bonds with that maturity. As demonstrated in Proposition 2, the SC rate is $\mathcal{E}^s\overline{q}_t$, defining the specialness as lower than the GC rate, except for the most special 10-year bond, since we use the variable q_t^{10} to illustrate local supply effects, as described in subsection 6.2. In subsection 6.4, we then relax the assumption of constant risk aversion of the arbitrageurs in the tradition of Vasicek and instead assume, as in VV, that the market price of risk and hence the term premium depend on the arbitrageurs' holdings.

6.2 *Local Supply Effects*

In Figure 3, exceptional demand pressure directed toward the 10-year maturity special bond, q_t^{10}, is stronger. This targeted demand pressure may capture the structural intervention of central banks through policies such as QE. A central bank can be modeled as a buy-and-hold investor that exerts extraordinary purchasing pressure on the market for nearly riskless sovereign bonds with particular tenors.[46] Targeted net excess demand may also reflect institutional constraints on investors, the reopening of a Treasury auction, or short squeezes, as diverse positions may induce a spike in valuations in otherwise common value settings (Nyborg and Strebulaev, 2003). In the top panel of Figure 3, excess demand induces a proportional kink in the yield curve (as noted, among others, by Gürkaynak et al., 2007, in Figure 4). Thus, from a modeling perspective, the flexibility of our framework allows for nonmonotonicity and bridges the gap between equilibrium models of the term structure of interest rates and econometric interpolation techniques (in the spirit of Nelson and Siegel, 1987). The mirror image of the intervention by the central bank is represented in the bottom panel of Figure 3, where the cross-section of special repo rates reaches a trough for the 10-year tenor SC that is more aggressively targeted, illustrating the endogeneity of repo rates. Simply put, when some investors exert significant demand pressure that raises a bond's price and lowers its yield, arbitrageurs borrow the bond in the SC market to respond to the large demand for this bond created by preferred-habitat investors, thus increasing its repo specialness. Since SC cannot be replaced with similar bonds on the repo market, the net supply effects on both prices and special repo rates are strongly localized. Introducing substitutability in the habitat preferences of buy-and-hold investors would gradually smooth local supply effects across the yield curve, as demonstrated in Section 5.

This calibration exercise generates several interesting and important policy implications. To cite just one, consider any two levels of exceptional demand for long- and short-term bonds, respectively, which both have the same effect on special repo rates. Then, the demand pressure at the short end of the yield curve has a larger effect on bond yields. The intuition is straightforward: If the decay of exceptional demand pressure is rapid, the bonds at the two maturities will be exposed to approximately the same repo dividend, although that dividend is of course discounted more heavily at the long end of the yield curve. Perhaps a more subtle remark is that policymakers can fine-tune the persistence of their asset purchases to be impactful for the yield of long-term bonds, while minimizing distortions on the repo market. Simply put, prices have a forward-looking outlook while special repo rates reflect the *existing* stock of collateral. As the rate of decay of exceptional demand pressure diminishes, the bond price immediately increases, thus reflecting expectations that its future specialness will decline. On the other hand, what matters for the degree of collateral specialness is the quantity of bonds available on the repo market at each point in time. Thus, by fixing the overall amount purchased of a bond and the effect of the purchase on its yield, predictable repeated reverse auctions smooth the distortions in the repo market across intervention dates when compared to a one-time operation. This is generally consonant with the practice of the major central banks, including the ECB,

[46] For instance, the Fed reports its Treasury portfolio holdings by tenor in its system open market account.

the Bank of Japan, and the Fed, over the past decade.[47] As noted, we expect the SLF policies to reduce repo specialness and allow the resulting kinks to be arbitraged away in the yield curve, affecting the localization of supply effects and inducing market forces to smooth them.

6.3 Differently Special Bonds

Figure OA.3 illustrates the general case of our model with varying degrees of bond specialness introduced in Section 5.3. For simplicity, we mute local supply effects across the term structure and keep the repo specialness constant across maturities. The calibration again follows Table I. However, rather than collapsing special repo rates onto their average, we now allow for differences in the distribution of bonds of the same maturity. D'Amico and Pancost (2022) document that the average specialness of *on-the-run*, *first-off-the-run*, and *second-on-the-run* and older bonds is 19.4, 8.4, and 4.0 bps, respectively. We set special repo rates to these values in our calibration. Correspondingly, we raise the persistence parameter to account for the number of classes of differently special bonds $P = 3$, thus setting the value of $\varphi^3 at 1.5$ percentage points. Graphically, bond yields jump upward to the more seasoned yield curve as investors roll over their portfolios to include new issues to replace older bonds as they mature, and demand pressure gradually dies out over time. Through time, *on-the-run-bonds* gradually become *first-off-the-run* and *second-off-the-run* bonds, finally coming to rest in the absorbing status of general bonds as their yield increases and their special repo rate decreases. This dynamic process accompanies the convergence of each security toward maturity.

6.4 Bond Scarcity and the Term Premium

The second main result of our model is that repo specialness generated by QE in the repo market impairs the effect that QE has on the term premium of the yield curve. To illustrate the effects of QE on the slope of the yield curve, the received wisdom suggests that the market price of risk λ should depend on the asset purchases, as argued by VV. We show the resulting term structure in Figure 4. In order to model QE, it appears reasonable to consider an economy at the zero lower bound (ZLB), as shown in Panel A. We start from a level of the term premium of $y^{10} - y^2 = 0.39$ in the general term structure of interest rates ($i = g$), delivered by the parameters in Table I when the economy is at the ZLB. When the short interest rate is in

[47] From the FAQ on the Public Sector Purchase Program available on the ECB website: "The need to preserve smooth market functioning calls for the necessary amount of purchases at yields below the Deposit Facility Rate [special bonds] to be distributed over time, rather than abruptly changing the sectors of the yield curve where asset purchases take place." Thus, the ECB distributes its bond purchases to smooth distortions, as we argue it should. We are not aware of other models of QE and the term structure that generate this striking pattern.

the proximity of this zero bound, the central bank may influence the yield curve by purchasing assets. In Panel B, we show the effect of asset purchases on the yield curve in the Vayanos and Vila (2021) framework. To this end, we specify the values of arbitrageurs' risk aversion as $a = 4.5$ and the demand elasticity of habitat investors as $\alpha = 6.21$, in line with the literature. In this scenario, the central bank reduces the term premium to 0.24, compressing it by around 38%. However, since collateral is infinitely available and $\mathcal{E} = 0$, asset purchases do not influence money markets. This lack of an effect on the short-term interest rate appears to be at odds with the consensus view in the empirical literature discussed in the stylized facts above. In our model, we thus incorporate the effects of asset purchases on money markets. In Panel C, we again set $\mathcal{E}$ at 0.68, as we have in previous calibrations to match the evidence in D'Amico and King (2013) on US markets, while keeping everything else fixed. Bonds directly purchased by the central bank become special and trade at lower yields. By inducing specialness, asset purchases become less effective in influencing the general term structure of interest rates, inducing a term premium of 0.27, which is lower than the baseline but substantially higher than the case with exogenous money market. *Ceteris paribus*, in counterfactual scenarios where collateral is scarce, asset purchases thus achieve a lower reduction of the term premium. In Panel D, we set $\mathcal{E}$ at 0.78 to sketch an illustration of EU markets, where bonds issued on tap reach comparatively higher levels of repo specialness. This calibration may help model the findings in Arrata et al. (2020); those authors estimate that purchasing 1% of a bond outstanding is associated with a decline in its repo rate of 0.78 bps. We kept the same values for other parameters to allow for a better comparison with the previous panels. QE induces even higher levels of specialness, as measured by the vertical distance between the term structure of general and special bonds. In Panel D, the term premium is 0.29 and QE achieves a reduction with respect to the baseline 0.39 value of around 25%. In general, limits to arbitrage arising from money markets may substantially impair the transmission of QE to the term premium, as demonstrated in Proposition 3. For clarity of exposition, we have muted local supply effects.

We draw two main lessons by endogenizing money markets in these calibrations. First, by inducing repo specialness, QE generates a vertical distance between the GC and the SC yield curves. Second, the resulting repo specialness dampens the effect of QE on the term premium, highlighting the dynamic interactions between bond scarcity, repo specialness, and the term structure of interest rates. The key mechanism that drives our results is that SC rates largely reflect bond scarcity, a limit to arbitrage that may substantially prevent fixed-income intermediaries from entering aggressive short positions over the long term. Substantial levels of special repo rates thus induce marginal agents to hold more conservative positions, reducing their exposure to interest rate changes and their need for greater compensation for bearing negative duration risk in the form of a term premium, given their short positions on bonds that are scarce. As an important insight, this counterfactual result highlights that dysfunctional money markets impair the transmission of unconventional monetary policy, consistent with the market data presented in Figure 1.

7 Conclusion

Empirical fixed-income market research in the last two decades has documented systematic patterns in the spread between general and special bonds that are difficult to explain in the context of uncertainty in short-rate dynamics. The existing literature lacks a coherent theory to reconcile this evidence with existing models of the term structure of interest rates. In the present study, we have proposed an endogenous explanation for special repo rates based on the short selling behavior of term structure arbitrageurs. We have done so by characterizing the equilibrium relation between bond prices and repo specialness across the entire term structure of interest rates. The preferred-habitat approach that we have used gives rise to equilibrium price differences between bonds with identical cash flows that are reflected in their respective repo spreads. Our derived equilibrium concept accounts for the collateral value of the bonds in the market for repurchase agreements, both general and special. We draw three main lessons by endogenizing money markets. First, by inducing repo specialness, quantitative easing generates a vertical distance between the general collateral and special collateral yield curves. Second, by requiring the delivery of specific securities, SC money markets reconcile the quantitative discipline imposed by the absence of arbitrage opportunities with the presence of strongly localized supply effects of QE on the term structure of interest rates. Third, the repo specialness induced by QE also dampens the effect of QE on the term premium, highlighting the dynamic interactions between bond scarcity, repo specialness, and the term structure of interest rates. The key mechanism that drives this result is that SC rates largely reflect bond scarcity, which is a cost for arbitrageurs and therefore a limit to arbitrage that may substantially prevent fixed-income intermediaries from entering aggressive short positions over the long term. Thus, lower levels of special repo rates may induce marginal agents to hold more conservative positions, reducing their exposure to interest rate changes and the compensation they require for bearing duration risk in the form of a term premium. In the present study, we have relaxed the standard assumption of the uniqueness of the instantaneous interest rate by proposing an endogenous market for the risk-free asset whose supply is elastic to quantities. (We have, however, abstracted from credit risk and market liquidity considerations, which may give rise to additional effects.) For ease of comparison with other techniques in the literature, our model was implemented without taking a particular stance on investor preferences. At the same time, we illustrate how our general formulation nests preference-based approaches as special cases.

The theory that we have presented has two especially attractive features. First, we have provided a unified framework that connects the secondary market for (nearly) risk-free bonds, such as US Treasury bonds, with the repo market for collateralized financing. Policymakers could use our model to assess the combined effects of exceptional demand pressure, such as QE or tapering, on the secondary market for government bonds and on the repo market for collateralized financing. Second, we have developed a generalized term structure equilibrium concept that accounts for the collateral value of bonds. Our framework is attractive for applied researchers, who may exploit exogenous shocks in both the bond and repo markets rather than considering these markets in isolation. Third, we have characterized the many dynamic and multifaceted connections between bond scarcity, repo specialness, and the term structure of interest rates. We have derived our results in closed form so as to perform comparative statics experiments and derive testable predictions and illustrated them through quantitative calibrations on bond and money markets. We have then proposed three simple extensions of our model to consider regular US Treasury auctions that account for cyclicality in specialness, enabling us to derive the equilibrium effects of heterogeneous arbitrageurs through haircuts and borrowing constraints and to examine the equilibrium effects of substitutability between bonds in the demand of preferred-habitat investors. The present study has discussed the demand pressure for special issues that have the same cash flows as benchmark securities; applications could focus on Green or Islamic bond premia. The structure we derive suggests that by estimating the yield curve on both general and special bonds together, a common practice, may result in a distorted fit that no longer sends reliable signals about impending economic conditions. Future research could generalize the method that we have proposed to multifactor or quadratic term structure models from the theory side and test its predictions empirically. Overall, the paper shows that dysfunctional money markets, in the context of the large expansion in the role of their collateral-driven segment, can substantially impair the transmission of unconventional monetary policy across the yield curve. Hence, these effects on both the bond and repo markets should be considered jointly in the conduct of monetary policy.

A Proof of Lemma 1

By substituting Equation (7) into the affine representation in Equation (15), we obtain the price of general and special bonds, since $q_t^n = 0$ for general bonds whose status is $i = g$:

$$
\begin{aligned}
b_t^n(g) &= \mathbb{E}_t^{\mathbb{Q}}\big[e^{-\sum_{j=0}^{n} r_{t+j}}\big] = e^{(-A_n r_t - C_n^g)},\\
b_t^n(s) &= \mathbb{E}_t^{\mathbb{Q}}\big[e^{-\sum_{j=0}^{n} r_{t+j}^{n-j}}\big] = e^{(-A_n r_t - B_n q_t^n - C_n^s)}.
\end{aligned}
$$

Lemma 1 results after taking the ratio of the price of the general bond $b_t^n(g)$ to the price of the special bond $b_t^n(s)$ by noting that $r_t^n = r_t - l_t^n$ and $D_n = C_n^s - C_n^g$. *Q.E.D.*

B Proof of Proposition 1

By definition of the equivalent martingale measure,

$$
b_t^{n+1}(i) = E_t^{\mathbb{Q}}\big[b_{t+1}^n(i)\big].
$$

From Equation (7), the $-\log$ price of the n-th tenor bond at $t+1$ and its expectation and variance are, respectively,

$$
\begin{aligned}
&-\log b_{t+1}^n(i) = A_n r_{t+1} + B_n q_{t+1}^n + C_n^i\\
&= A_n\big[\varrho r_t + (1-\varrho)\theta + \sigma_r \eta_{t+1}\big] + B_n\big[\varphi_n q_t^{n+1} + (1-\varphi_n)\kappa_n + \sigma_{q,n}\upsilon_{t+1}^n\big] + C_n^i,
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}\big[-\log b_{t+1}^n(i)\big] &= A_n \mathbb{E}_t^{\mathbb{Q}}\big[r_{t+1}\big] + B_n \mathbb{E}_t^{\mathbb{Q}}\big[q_{t+1}^n\big] + C_n^i\\
&= A_n\big[\varrho r_t + (1-\varrho)\theta\big] + B_n\big[\varphi_n q_t^{n+1} + (1-\varphi_n)\kappa_n\big] + C_n^i,\\
\text{Var}_t^{\mathbb{Q}}\big[-\log b_{t+1}^n(i)\big] &= A_n^2 \text{Var}_t^{\mathbb{Q}}\big[r_{t+1}\big] + B_n^2 \text{Var}_t^{\mathbb{Q}}\big[q_{t+1}^n\big] + 2A_n B_n \text{Cov}_t^{\mathbb{Q}}\big[r_{t+1}, q_{t+1}^n\big]\\
&= A_n^2\sigma_r^2 + B_n^2\sigma_{q,n}^2 + 2A_n B_n \rho_n.
\end{aligned}
$$

Since the shocks are Gaussian, we can use the properties of the log-normal distribution:

$$
\begin{aligned}
-\log b_t^{n+1}(i) &= -\log E_t^{\mathbb{Q}}\big[b_{t+1}^n(i)\big]\\
&= \mathbb{E}_t^{\mathbb{Q}}\big[-\log b_{t+1}^n(i)\big] - 0.5\text{Var}_t^{\mathbb{Q}}\big[-\log b_{t+1}^n(i)\big]\\
A_{n+1} r_t + B_{n+1} q_t^{n+1} + C_{n+1}^i &= A_n\big[\varrho r_t + (1-\varrho)\theta\big] + B_n\big[\varphi_n q_t^{n+1} + (1-\varphi_n)\kappa_n\big] + C_n^i\\
&\quad - 0.5\big[A_n^2\sigma_r^2 + B_n^2\sigma_{q,n}^2 + 2A_n B_n \rho_n\big].
\end{aligned}
$$

By matching coefficients (Backus et al., 1998, Section 4), we obtain the desired recursions. As for the initial conditions, from Equation (1) we know that $A_1 = 1$ and $C_1^i = 0$, by the absence of arbitrage between the investment in the general bond and at the GC rate, and from Lemma 1 that $B_1 = -\mathcal{E}^i$. *Q.E.D.*

C Proof of Proposition 2

By virtue of Lemma 1, we have

$$
\begin{aligned}
e^{B_n q_t^n + D_n} &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{n} r_{t+j}}\right]\mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{n} r_{t+j}^{n-j}}\right]^{-1} \\
&= e^{-l_t^n}\mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=1}^{n} r_{t+j}}\right]\mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=1}^{n} r_{t+j}^{n-j}}\right]^{-1} \\
&= e^{-l_t^n}\mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{n-1} r_{t+1+j}}\right]\mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=1}^{n-1} r_{t+1+j}^{n-1-j}}\right]^{-1} \\
&= e^{-l_t^n}\mathbb{E}_t^{\mathbb{Q}}\left\{\mathbb{E}_{t+1}^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{n-1} r_{t+1+j}}\right]\right\}\mathbb{E}_t^{\mathbb{Q}}\left\{\mathbb{E}_{t+1}^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{n-1} r_{t+1+j}^{n-1-j}}\right]^{-1}\right\} \\
&= e^{-l_t^n}\mathbb{E}_t^{\mathbb{Q}}\left[e^{(-A_{n-1}r_{t+1}-C_{n-1}^g)}\right]\mathbb{E}_t^{\mathbb{Q}}\left[e^{(-A_{n-1}r_{t+1}-B_{n-1}q_{t+1}^{n-1}-C_{n-1}^s)}\right]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{e^{\left(-l_t^n - A_{n-1}\left[\varrho r_t+(1-\varrho)\theta-0.5\sigma_r^2\right]-C_{n-1}^g\right)}}{e^{\left(-A_{n-1}\left[\varrho r_t+(1-\varrho)\theta-0.5\sigma_r^2\right]-B_{n-1}\left[\varphi_{n-1}q_t^n+(1-\varphi_{n-1})\kappa_{n-1}-0.5\sigma_{q,n-1}^2-A_{n-1}\rho_{n-1}\right]-C_{n-1}^s\right)}} \\
&= e^{\left(-l_t^n+B_{n-1}\left[\varphi_{n-1}q_t^n+(1-\varphi_{n-1})\kappa_{n-1}-0.5\sigma_{q,n-1}^2\right]+D_{n-1}-A_{n-1}B_{n-1}\rho_{n-1}\right)} \\
&= e^{\left(-l_t^n - A_{n-1}B_{n-1}\rho_{n-1}\right)}\mathbb{E}_t^{\mathbb{Q}}\left[e^{(B_{n-1}q_{t+1}^{n-1}+D_{n-1})}\right].
\end{aligned}
$$

The second equivalence results from the definition of specialness in Equation (14), the fourth follows by the law of iterated expectations, and the fifth by the Laplace representation of bond prices in Equation (15). We then express the expected values, after accounting for Jensen's terms. We have related the left side of Lemma 1 to its expected leaded value. By taking logs on both sides of the expression and rearranging terms, it follows that

$$
\begin{aligned}
l_t^n &= D_{n-1} + B_{n-1}((1-\varphi_{n-1})\kappa_{n-1} - 0.5B_{n-1}\sigma_{q,n-1}^2 - A_{n-1}\rho_{n-1}) - D_n \\
&\quad + (\varphi_{n-1}B_{n-1} - B_n)q_t^n \\
&= \mathcal{E}^i q_t^n.
\end{aligned}
$$

The latter equivalence results from the recursions in Proposition 1 and Remark 1. *Q.E.D.*

D Repo Market Clearing

Let the demand for special repo by arbitrageurs be D and denote by S the collateral supplied by preferred-habitat investors on the repo market. From the market clearing condition on the bond market, arbitrageurs' demand on the repo market is $D = Z(s) = q_t^n(1-\alpha^n B_n)-\alpha^n[A_n r_t+C_n]$. As previously demonstrated, in equilibrium the aggregate relation $l_t^n = \mathcal{E}^i q_t^n$ must hold in order to prevent any arbitrage opportunities. We use the demand curve and the equilibrium specialness to pin down the supply curve S. The repo market is completely characterized by the following supply and demand curves:

$$\begin{cases} D & = -\alpha^n[A_n r_t + C_n] + (1-\alpha^n B_n)q_t^n, \\ S & = -\alpha^n[A_n r_t + C_n] + \frac{1}{\mathcal{E}^s}(1-\alpha^n B_n)l_t^n. \end{cases}$$

The economic intuitions are as follows. First, observe that both the demand and supply of SC repo agreements react by the same amount to any change in the general level of interest rates. To see this more clearly, consider a decrease in demand for the special bond induced by a higher level of interest rates – namely, any increase in $\alpha^n[A_n r_t + C_n]$ that occurs as preferred-habitat investors demand less of the bond on the cash market and thus offer less of the bond on the repo market. This reduction in supply is exactly offset by the reduced demand for reverse repo agreements by arbitrageurs, who must symmetrically reduce their exposure for the bond market to clear. Meanwhile, quantity shocks matter for the determination of equilibrium specialness. Consider an unexpected increase in q_t^n, which prompts an increase in demand for the bond of magnitude $q_t^n(1-\alpha^n B_n)$, corresponding to an equal increase in the demand for reverse repo agreements. The preferred-habitat investors are only willing to supply the SC on the repo market at a specialness premium l_t^n, which reflects their private valuation of the security. Thus, both the bond and repo markets clear. For example, suppose a central bank rolls out a QE program. Bonds on the cash market become scarcer, and arbitrageurs sell them short to fill the gap with respect to their interest rate risk, thus increasing the demand on the repo market. If the central bank does not offer purchases on the SLFs, the supply will slope upwards as investors with private valuations for the bonds require greater compensation to forgo the security for one period. SLFs can parallel the QE program and reduce the elasticity of repo supply. Importantly, the elasticity parameter $\mathcal{E}^s$ can be estimated from market data.

E Risk Adjustment

Under the physical measure $\mathbb{P}$, VV mean-variance arbitrageurs optimize

$$
\begin{aligned}
&\max_{\{X_t^n\}_{n\in\mathtt{N}}} \mathbb{E}_t^{\mathbb{P}}\Big[\Delta W_{t+1}\Big] - \frac{a}{2}\mathbb{V}_t^{\mathbb{P}}\Big[\Delta W_{t+1}\Big] \\
&= \max_{\{X_t^n\}_{n\in\mathtt{N}}} W_t r_t + \sum_{n\in\mathtt{N}} X_t^n\Big(\mu_t^n - r_t^n\Big) - \frac{a}{2}\mathbb{E}_t^{\mathbb{P}}\Bigg[\Big(\sum_{n\in\mathtt{N}} X_t^n \sigma^n U_{t+1}^n\Big)^2\Bigg] \\
&= \max_{\{X_t^n\}_{n\in\mathtt{N}}} W_t r_t + \sum_{n\in\mathtt{N}} X_t^n\Big(\mu_t^n - r_t^n\Big) - \frac{a}{2}\sum_{n\in\mathtt{N}} X_t^n\sigma^n\Bigg(\sum_{m\in\mathtt{N}}\mathbb{E}_t^{\mathbb{P}}\Big[U_{t+1}^n X_t^m \sigma^m U_{t+1}^m\Big]\Bigg).
\end{aligned}
$$

The first order condition with respect to a position in the n-th tenor bond on special is

$$
\begin{aligned}
\mu_t^n - r_t^n &= \sigma^n a\Bigg(\sum_{m\in\mathtt{N}}\mathbb{E}_t^{\mathbb{P}}\Big[U_{t+1}^n X_t^m \sigma^m U_{t+1}^m\Big]\Bigg) \\
&= \sigma^n a\Bigg(\sum_{m\in\mathtt{N}}\mathbb{E}_t^{\mathbb{P}}\Big[[\eta_{t+1} \quad \upsilon_{t+1}^n]' X_t^m [A_{m-1}\sigma_r \quad B_{m-1}\sigma_{q,m}][\eta_{t+1} \quad \upsilon_{t+1}^m]'\Big]\Bigg) \\
&= [A_{n-1}\sigma_r \quad B_{n-1}\sigma_{q,n}]a\Bigg(\sum_{m\in\mathtt{N}}\begin{bmatrix} X_t^m(A_{m-1}\sigma_r + B_{m-1}\rho_m) \\ X_t^m(A_{m-1}\rho_m + B_{m-1}\sigma_{q,m})\end{bmatrix}\Bigg) \\
&= -A_{n-1}\sigma_r\lambda_r^{\mathrm{VV}} - B_{n-1}\sigma_{q,n}\lambda_q^{\mathrm{VV}}
\end{aligned}
$$

which decomposes the market price of risk into the compensation for short-rate (one factor) and demand ($\mathtt{N}$ factors, one for each tenor) risk. Specifically,

$$
\begin{aligned}
\lambda_r^{\mathrm{VV}} &= -a\sum_{m\in\mathtt{N}}[X_t^m(A_{m-1}\sigma_r + B_{m-1}\rho_m)], \\
\lambda_q^{\mathrm{VV}} &= -a\sum_{m\in\mathtt{N}}[X_t^m(A_{m-1}\rho_m + B_{m-1}\sigma_{q,m})].
\end{aligned}
$$

F Equilibrium with Imperfect Demand Substitutability

Mutatis mutandis, we can apply the same steps as in Appendix C. From Equation (23),

$$\begin{aligned}
-\log b^n_{t+1}(i) &= A_n r_{t+1} + B^n Q_{t+1} + C^i_n \\
&= A_n\Big[\varrho r_t + (1-\varrho)\theta + \sigma_r \eta_{t+1}\Big] + B^n\Big[\Phi Q_t + \overline{Q} + \Omega V_{t+1}\Big] + C^i_n, \\
\mathbb{E}^{\mathbb{Q}}_t\Big[-\log b^n_{t+1}(i)\Big] &= A_n \mathbb{E}^{\mathbb{Q}}_t\Big[r_{t+1}\Big] + B^n \mathbb{E}^{\mathbb{Q}}_t\Big[Q_{t+1}\Big] + C^i_n \\
&= A_n\Big[\varrho r_t + (1-\varrho)\theta\Big] + B^n\Big[\Phi Q_t + \overline{Q}\Big] + C^i_n, \\
\text{Var}^{\mathbb{Q}}_t\Big[-\log b^n_{t+1}(i)\Big] &= A^2_n \text{Var}^{\mathbb{Q}}_t\Big[r_{t+1}\Big] + \text{Var}^{\mathbb{Q}}_t\Big[B^n \Omega V_{t+1}\Big] \\
&\quad + 2A_n \sum_{i=1}^{\text{N}} B_{n,i}\, \text{Cov}^{\mathbb{Q}}_t\Big[\eta_{t+1}, \upsilon^i_{t+1}\Big] \\
&= A^2_n \sigma^2_r + \sum_{i=1}^{\text{N}} B^2_{n,i} \langle \Omega^\top, \Omega\rangle_{n,i} + 2A_n \sum_{i=1}^{\text{N}} B_{n,i}\rho_{n,i}.
\end{aligned}$$

Since the shocks are Gaussian, we use the properties of the multivariate log-normal distribution.

$$\begin{aligned}
-\log b^{n+1}_t(i) &= -\log E^{\mathbb{Q}}_t\Big[b^n_{t+1}(i)\Big] \\
&= \mathbb{E}^{\mathbb{Q}}_t\Big[-\log b^n_{t+1}(i)\Big] - 0.5\text{Var}^{\mathbb{Q}}_t\Big[-\log b^n_{t+1}(i)\Big], \\
A_{n+1} r_t + B^{n+1} Q_t + C^i_{n+1} &= A_n\Big[\varrho r_t + (1-\varrho)\theta\Big] + B^n\Big[\Phi Q_t + \overline{Q}\Big] + C^i_n \\
&\quad - 0.5\Big[A^2_n \sigma^2_r + \sum_{i=1}^{\text{N}} B^2_{n,i}\langle \Omega^\top, \Omega\rangle_{n,i} + 2A_n \sum_{i=1}^{\text{N}} B_{n,i}\rho_{n,i}\Big].
\end{aligned}$$

By matching coefficients, we obtain their growth rate. As for the initial conditions, from Equation (1) we have $A_1 = 1$ and $C_1 = 0$, and $B^1 = -H^{1,i}$ by a straightforward extension of Lemma 1:

$$\begin{aligned}
A_{n+1} &= \varrho A_n + 1, \\
B^{n+1} &= B^n \Phi - H^{n+1,i}, \qquad (30) \\
C^i_{n+1} &= C^i_n + A_n(1-\varrho)\theta - 0.5A^2_n\sigma^2_r + B^n\overline{Q} - 0.5\Big[\sum_{i=1}^{\text{N}} B^2_{n,i}\langle\Omega^\top, \Omega\rangle_{n,i} + 2A_n \sum_{i=1}^{\text{N}} B_{n,i}\rho_{n,i}\Big].
\end{aligned}$$

Equation (30) is a recursion for the n-th row of the B matrix. *Q.E.D.*

References

Altavilla, C., G. Carboni, and R. Motto. 2021. Asset purchase programmes and financial markets: lessons from the euro area. *International Journal of Central Banking* 17:1–48.

Arrata, W., B. Nguyen, I. Rahmouni-Rousseau, and M. Vari. 2020. The scarcity effect of QE on repo rates: Evidence from the euro area. *Journal of Financial Economics* 137:837–856.

Backus, D., S. Foresi, and C. Telmer. 1998. Discrete-time models of bond pricing. Working Paper.

Ballensiefen, B., A. Ranaldo, and H. Winterberg. 2023. Money Market Disconnect. *The Review of Financial Studies* forthcoming.

Banerjee, S., and J. J. Graveline. 2013. The cost of short-selling liquid securities. *The Journal of Finance* 68:637–664.

Barclay, M. J., T. Hendershott, and K. Kotz. 2006. Automation versus intermediation: Evidence from Treasuries going off the run. *The Journal of Finance* 61:2395–2414.

Berardi, A., R. Brown, and S. Schaefer. 2021. Bond risk premia: The information in (really) long term rates. Working Paper.

Bernanke, B. 2020. The new tools of monetary policy. *American Economic Review* 110:943–83.

Brennan, M., and E. Schwartz. 1979. A continuous time approach to the pricing of bonds. *Journal of Banking & Finance* 3:133–155.

Buraschi, A., and D. Menini. 2002. Liquidity risk and specialness. *Journal of Financial Economics* 64:243–284.

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

Chen, M., J. Cherian, Z. Li, Y. Shao, and M. G. Subrahmanyam. 2022. Clientele effect in sovereign bonds: Evidence from Islamic Sukuk bonds in Malaysia. *Critical Finance Review* 11:677–745.

Cherian, J., E. Jacquier, and R. Jarrow. 2004. A model of the convenience yields in on-the-run treasuries. *Review of Derivatives Research* 7:79–97.

Christensen, J. H., and G. D. Rudebusch. 2012. The response of interest rates to US and UK quantitative easing. *The Economic Journal* 122:F385–F414.

Copeland, A., A. Martin, and M. Walker. 2014. Repo runs: Evidence from the tri-party repo market. *The Journal of Finance* 69:2343–2380.

Cornell, B., and A. Shapiro. 1989. The mispricing of US Treasury bonds: A case study. *The Review of Financial Studies* 2:297–310.

Corradin, S., and A. Maddaloni. 2020. The importance of being special: Repo markets during the crisis. *Journal of Financial Economics* 137:392–429.

Costain, J., G. Nuño, and C. Thomas. 2022. The term structure of interest rates in a heterogeneous monetary union. Working Paper.

Cox, J., J. Ingersoll, and S. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53:385–407.

Culbertson, J. 1957. The term structure of interest rates. *The Quarterly Journal of Economics* 71:485–517.

Dai, Q., and K. Singleton. 2003. Term structure dynamics in theory and reality. *The Review of Financial Studies* 16:631–678.

D'Amico, S., J. Klausmann, and N. A. Pancost. 2022. The benchmark greenium. Working Paper.

Du, W., B. M. Hébert, and W. Li. 2022. Intermediary balance sheets and the treasury yield curve. Working Paper.

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

Duffie, D. 1996. Special repo rates. *The Journal of Finance* 51:493–526.

Duffie, D., N. Gârleanu, and L. Pedersen. 2005. Over-the-counter markets. *Econometrica* 73:1815–1847.

Duffie, D., and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6:379–406.

D'Amico, S., W. English, D. López-Salido, and E. Nelson. 2012. The Federal Reserve's large-scale asset purchase programmes: Rationale and effects. *The Economic Journal* 122:F415–F446.

D'Amico, S., R. Fan, and Y. Kitsul. 2018. The scarcity value of treasury collateral: Repo-market effects of security-specific supply and demand factors. *Journal of Financial and Quantitative Analysis* 53:2103–2129.

D'Amico, S., and T. King. 2013. Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply. *Journal of Financial Economics* 108:425–448.

D'Amico, S., and N. Pancost. 2022. Special repo rates and the cross-section of bond prices: The role of the special collateral risk premium. *Review of Finance* 26:117–162.

Fama, E., and R. Bliss. 1987. The information in long-maturity forward rates. *The American Economic Review* 77:680–692.

Fisher, M. 2002. Special repo rates: An introduction. *Economic Review-Federal Reserve Bank of Atlanta* 87:27–44.

Fleckenstein, M., and F. Longstaff. 2020. Renting balance sheet space: Intermediary balance sheet rental costs and the valuation of derivatives. *The Review of Financial Studies* 33:5051–5091.

Fleckenstein, M., and F. A. Longstaff. 2021. Treasury richness. Working Paper.

Fontaine, J.-S., and R. Garcia. 2012. Bond liquidity premia. *The Review of Financial Studies* 25:1207–1254.

Föllmer, H., and A. Schied. 2008. *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter.

Gabaix, X., A. Krishnamurthy, and O. Vigneron. 2007. Limits of arbitrage: Theory and evidence from the mortgage-backed securities market. *The Journal of Finance* 62:557–595.

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2018. The financial market effects of the Federal Reserve's large-scale asset purchases. *International Journal of Central Banking* 7:45–52.

Gârleanu, N., S. Panageas, and G. Zheng. 2021. A long and a short leg make for a wobbly equilibrium. Working Paper.

Gourinchas, P.-O., W. Ray, and D. Vayanos. 2022. A preferred-habitat model of term premia, exchange rates, and monetary policy spillovers. Working Paper.

Grasselli, M., and C. Tebaldi. 2008. Solvable affine term structure models. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 18:135–153.

Graveline, J., and M. McBrady. 2011. Who makes on-the-run Treasuries special? *Journal of Financial Intermediation* 20:620–632.

Greenwood, R., S. Hanson, and J. Stein. 2015. A comparative-advantage approach to government debt maturity. *The Journal of Finance* 70:1683–1722.

Greenwood, R., S. Hanson, J. Stein, and A. Sunderam. 2023. A quantity-driven theory of term premia and exchange rates. *Quarterly Journal of Economics* forthcoming.

Greenwood, R., and D. Vayanos. 2010. Price pressure in the government bond market. *American Economic Review* 100:585–90.

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

Gromb, D., and D. Vayanos. 2018. The dynamics of financially constrained arbitrage. *The Journal of Finance* 73:1713–1750.

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

Hanson, S., A. Malkhozov, and G. Venter. 2022. Demand–supply imbalance risk and long-term swap spreads. Working Paper.

He, Z., and K. Milbradt. 2014. Endogenous liquidity and defaultable bonds. *Econometrica* 82:1443–1508.

He, Z., S. Nagel, and Z. Song. 2022. Treasury inconvenience yields during the covid-19 crisis. *Journal of Financial Economics* 143:57–79.

Heath, D., R. Jarrow, and A. Morton. 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60:77–105.

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

Jankowitsch, R., A. Nashikkar, and M. Subrahmanyam. 2011. Price dispersion in OTC markets: A new measure of liquidity. *Journal of Banking & Finance* 35:343–357.

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

Jordan, B., and S. Jordan. 1997. Special repo rates: An empirical analysis. *The Journal of Finance* 52:2051–2072.

Kiyotaki, N., and J. Moore. 1997. Credit cycles. *Journal of Political Economy* 105:211–248.

Koijen, R. S., F. Koulischer, B. Nguyen, and M. Yogo. 2021. Inspecting the mechanism of quantitative easing in the euro area. *Journal of Financial Economics* 140:1–20.

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

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

Maddaloni, A., and H. Roh. 2021. Not equally special: Collateralised trading of non-banks. Working Paper.

Mancini, L., A. Ranaldo, and J. Wrampelmeyer. 2016. The euro interbank repo market. *The Review of Financial Studies* 29:1747–1779.

Martin, A., D. Skeie, and E.-L. von Thadden. 2014. Repo runs. *The Review of Financial Studies* 27:957–989.

Merrick Jr, J., N. Naik, and P. Yadav. 2005. Strategic trading behavior and price distortion in a manipulated market: anatomy of a squeeze. *Journal of Financial Economics* 77:171–218.

Modigliani, F., and R. Sutch. 1966. Innovations in interest rate policy. *The American Economic Review* 56:178–197.

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

Nelson, C., and A. Siegel. 1987. Parsimonious modeling of yield curves. *Journal of Business* 60:473–489.

Nyborg, K. G., and I. A. Strebulaev. 2003. Multiple unit auctions and short squeezes. *The Review of Financial Studies* 17:545–580.

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

Pelizzon, L., M. Subrahmanyam, and D. Tomio. 2022. Central bank-driven mispricing. Working Paper.

Pontiff, J. 1996. Costly arbitrage: Evidence from closed-end funds. *The Quarterly Journal of Economics* 111:1135–1151.

Tuckman, B., and A. Serrat. 2022. *Fixed income securities: tools for today's markets*. John Wiley & Sons.

Tuckman, B., and J.-L. Vila. 1992. Arbitrage with holding costs: A utility-based approach. *The Journal of Finance* 47:1283–1302.

Van Binsbergen, J., W. Diamond, and M. Grotteria. 2022. Risk-free interest rates. *Journal of Financial Economics* 143:1–29.

Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5:177–188.

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

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

TABLE I: **Calibration**

Parameter	Value	Source	Data and Moment
Yield Curve Calibration on 1985–2020 Data			
θ Long-run mean of r_t	0.0200	He and Milbradt (2014)	Table I Risk-free rate, long-run mean
ϱ Persistence of r_t	0.9	Gürkaynak et al. (2007) data	Autocorrelation of 1-year yields Equal to 0.9
σ_r Standard deviation of r_t	0.0115	Gürkaynak et al. (2007) data	Volatility of 1-year yields Equal to 2.63
λ_r Market price of GC risk	0.42	Gürkaynak et al. (2007) data	Average of 10-year yields Equal to 0.0517

Parameter	Value	Source	Data and Moment
Exceptional Demand Pressure and Local Supply Effects			
$\mathcal{E}^s$ Slope of special collateral supply $\frac{\partial l_t^1}{\partial q_t^n}$	0.68	D'Amico and King (2013)	Table VII Purchases conditional impact on returns
$\overline{q}_t$ Level of excess demand for the Special bonds	0.0026	D'Amico et al. (2018)	Table I Average general/special Repo spread equal to 19.4 bps
q_t^{10} Level of excess demand for the 10-years tenor special bond	0.0100	D'Amico et al. (2018)	Table I Average price residual of 10-year Special bonds equal to 53 bps of par
φ Persistence of Excess demand pressure	0.25	D'Amico et al. (2018)	Table I Average new to old special bonds Repo spread ratio equal to 0.25

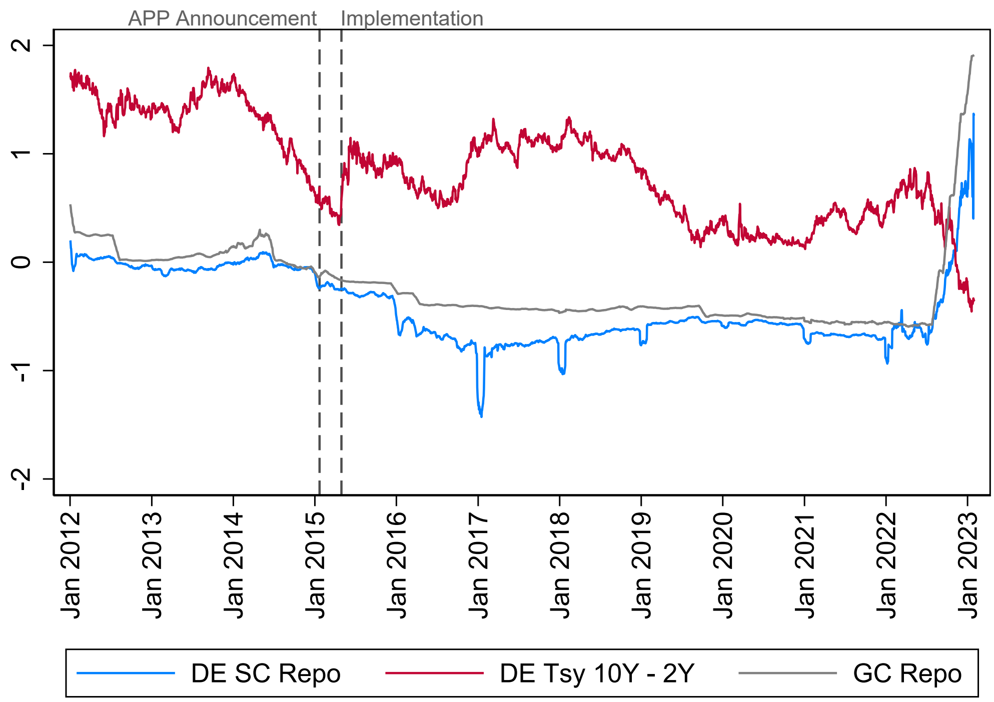

FIGURE 1: **Bond Scarcity and the Term Premium.** The figure shows the unscaled series of the term premium of the German Treasury yield curve, measured as the difference between benchmark 10-year and 2-year yields sourced from Bloomberg, along with the one-month trailing value-weighted average of special and general collateral repo rates on German government bonds, as recorded by MTS, sampled at millisecond resolution and aggregated at daily frequency.

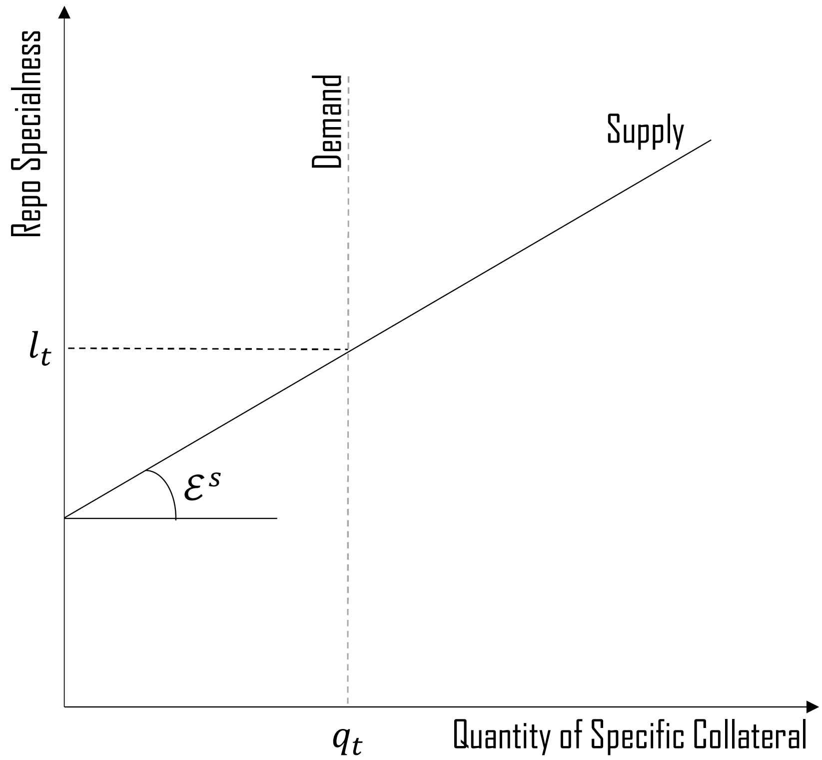

FIGURE 2: **Demand and supply of special collateral.** This figure illustrates the functioning of the market for repurchase agreements collateralized by special bonds. The horizontal axis shows demand pressure on the cash market, and the vertical axis shows repo specialness. The supply curve is upward sloping. The demand curve is flat because of the commitment of short sellers to deliver the specific issue. The supply, by contrast, is elastic, as holders of special collateral bonds require greater compensation to pledge additional units of the security on the repo market.

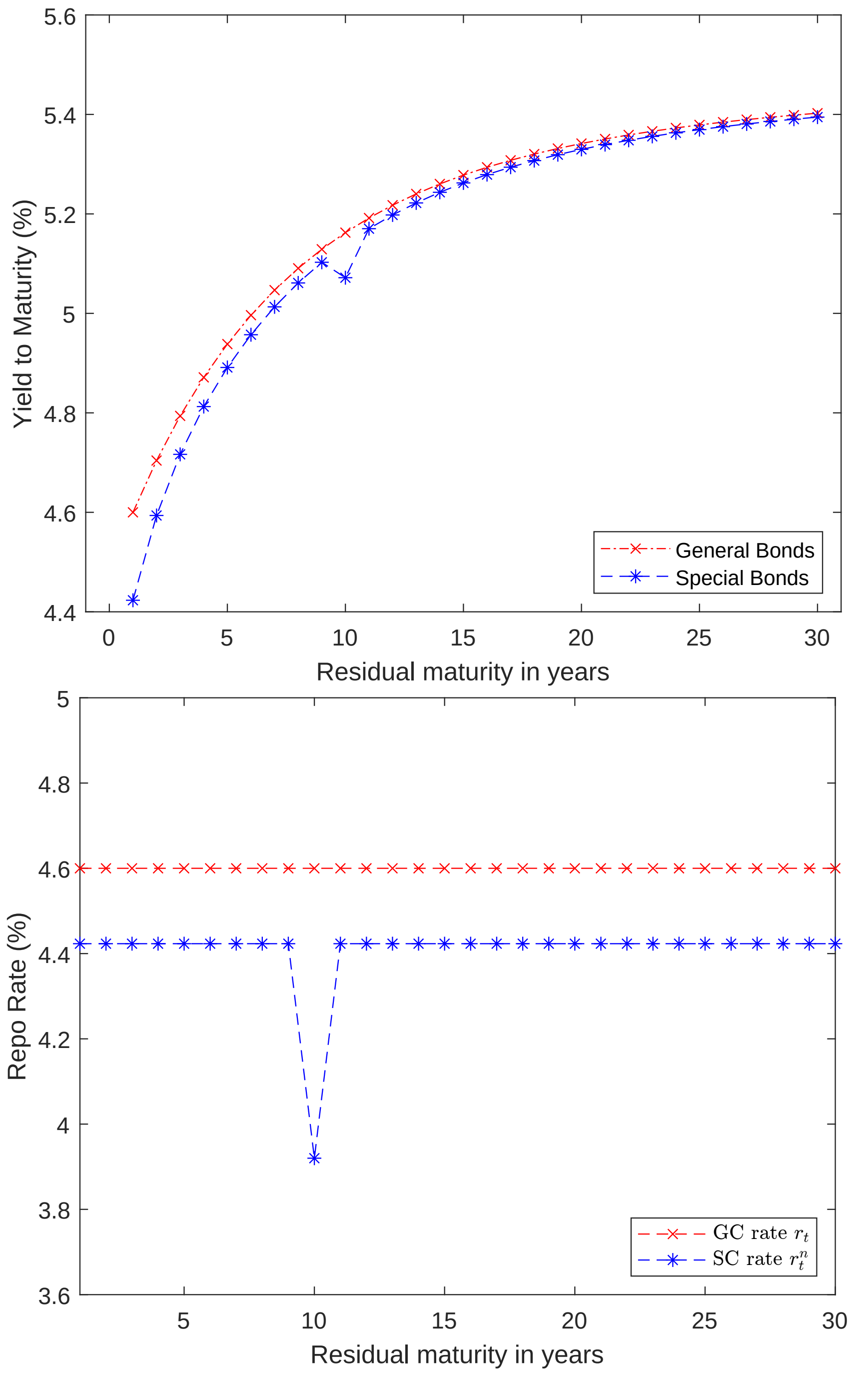

FIGURE 3: **Yield curves and repo rates.** The top panel shows the term structure of interest rates. The bottom panel shows general and special overnight repo rates plotted against collateral tenor. Rates are expressed on a per annum basis. The curves in red show the general bonds, which are not exposed to demand pressure. The curves in blue show special bonds, which are targeted by exceptional demand pressure. Table I presents the calibration.

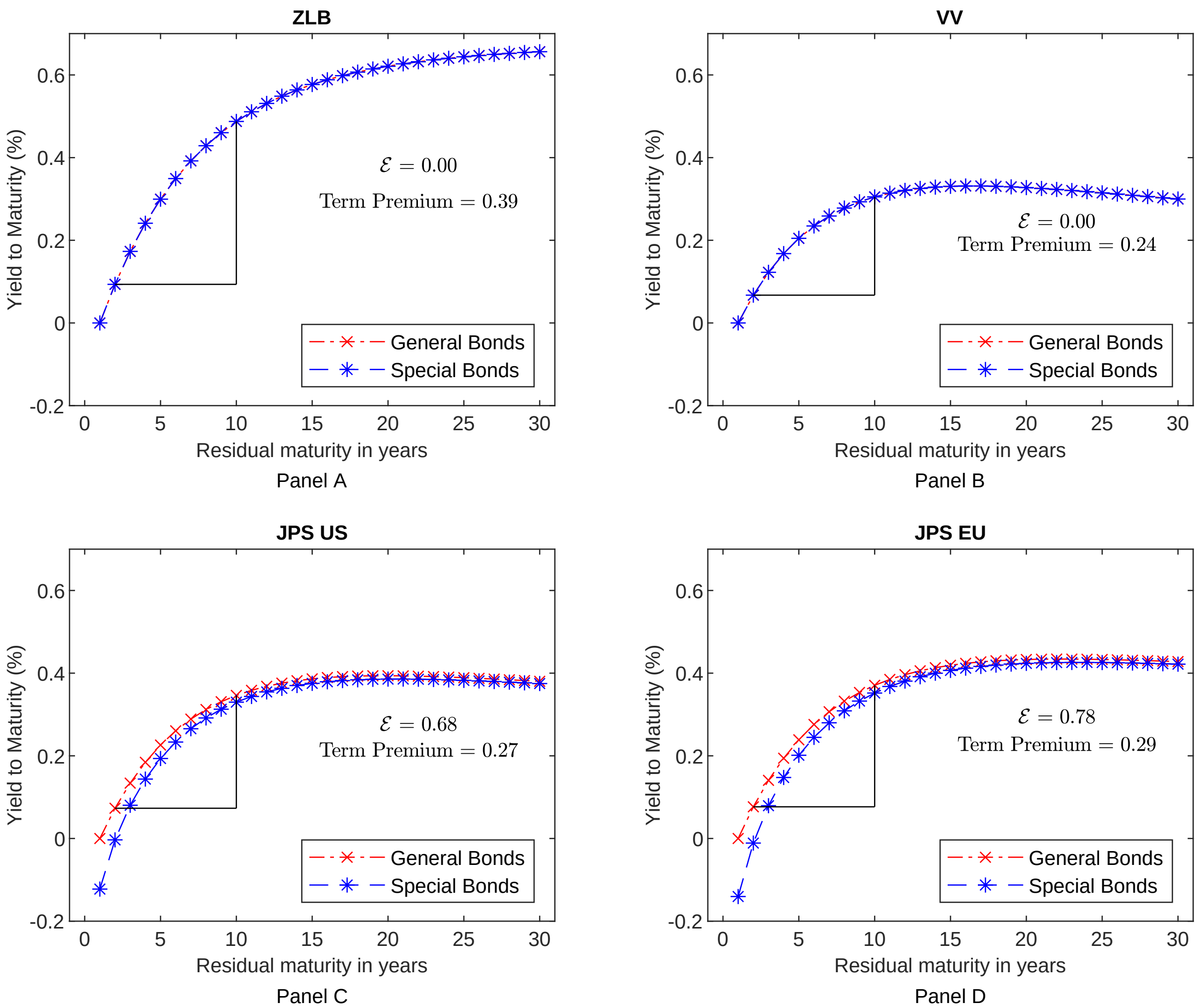

FIGURE 4: **Bond Scarcity and the Term Premium.** Panel A of the figure shows the term structure of interest rates at the zero lower bound (ZLB). Panel B shows the effect of quantitative easing (QE) in the Vayanos and Vila (VV) model. Panels C and D present our calibrations (JPS), where we relax the assumption that the collateral is in infinite supply, *ceteris paribus*. Panel C presents the calibration of our model to the US market, while Panel D shows its calibration to the EU market. Rates are expressed on a per annum basis. The curves in red show the general bonds, which are not exposed to demand pressure. The curves in blue show the special bonds differently targeted by exceptional demand pressure. Section 6.4 presents the calibration.

Internet Appendix

TABLE OA.1: **Models Comparison**

	Factors Number	Market Price of Risk	Short Rate	Equilibrium Segmentation	Substitutability in Bond Demand
Vasicek	1	$\lambda(t, r)$	Time series $r_{t+1} = \varrho r_t + (1 - \varrho)\theta + \sigma_r \eta_{t+1}$	No	Yes, perfect
Brennan and Schwartz	2	$\lambda(t, r)$	Time series $r_{t+1} = \varrho r_t + (1 - \varrho)\theta + \sigma_r \eta_{t+1}$	No	Yes, perfect
Vayanos and Vila	1 + K	$\lambda(a, X_t^n, \Sigma_n, U^n)$	Time series $r_{t+1} = \varrho r_t + (1 - \varrho)\theta + \sigma_r \eta_{t+1}$	No	No
Jappelli, Pelizzon, and Subrahmanyam	1 + N	Arbitrary	Time series and cross-section $r_{t+1} = \varrho r_t + (1 - \varrho)\theta + \sigma_r \eta_{t+1}$ $r_t^n = r_t - l_t^n$	Yes	Yes, imperfect

Notes: the foundational paper by Vasicek (1977) develops the equilibrium consistent with the absence of arbitrage. The two-factor model by Brennan and Schwartz (1979) derives the term structure from the instantaneous rate of return on a short and a long bond. More recently, Vayanos and Vila (2021) focus on the effects of demand pressure on the term structure of interest rates. Our paper connects the insights from the previous literature by deriving an arbitrage-consistent, preferred-habitat explanation of the cross-section of instantaneous bond returns.

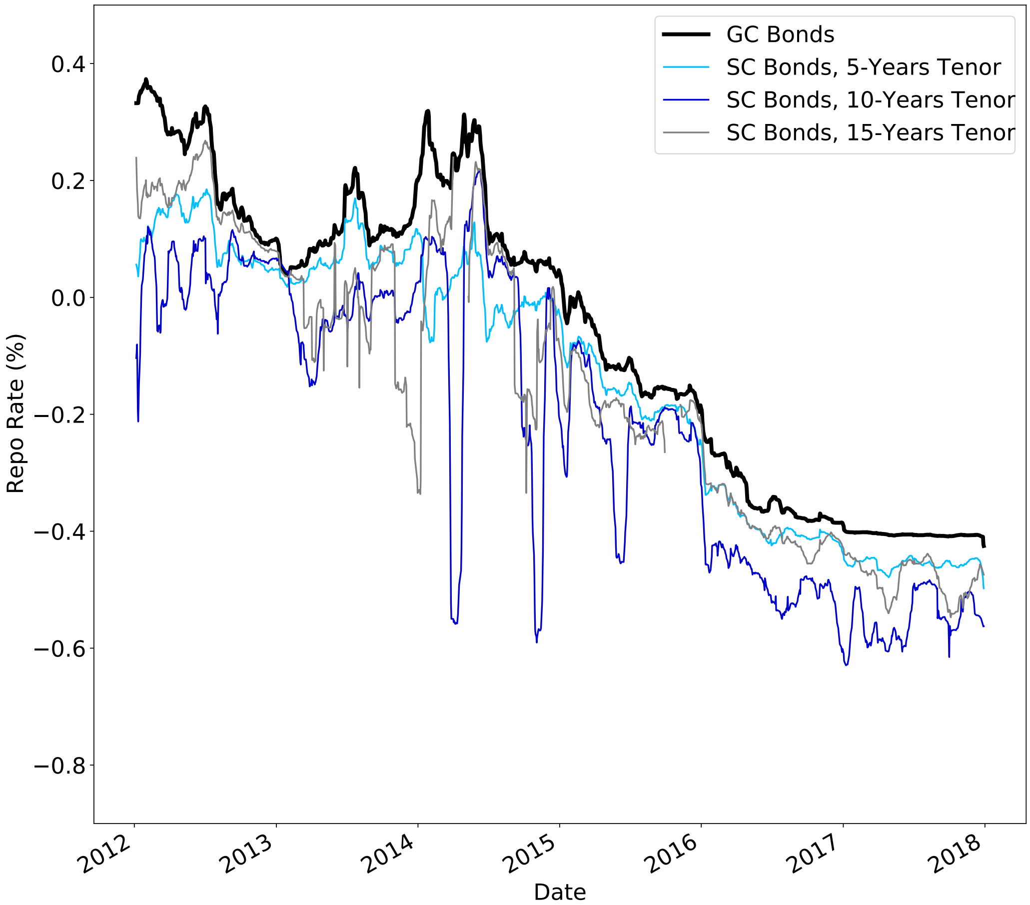

FIGURE OA.1: **General and Special Repo Rates for Italian Treasury Bonds**. This figure shows the volume-weighted monthly trailing average of the daily rates on tick-by-tick repo transactions collateralized by Italian treasury bonds (most notably Buoni del Tesoro Poliennali, Buoni Ordinari del Tesoro, and Certificati di Credito del Tesoro), as recorded by MTS markets from 2012 to 2018. Daily rates are the volume-weighted average of intraday rates. Each trading day, repo transactions for 22 trading days are averaged. We distinguish between general collateral (GC) and special collateral (SC) transactions; the latter are shown for benchmark time-to-maturity buckets.

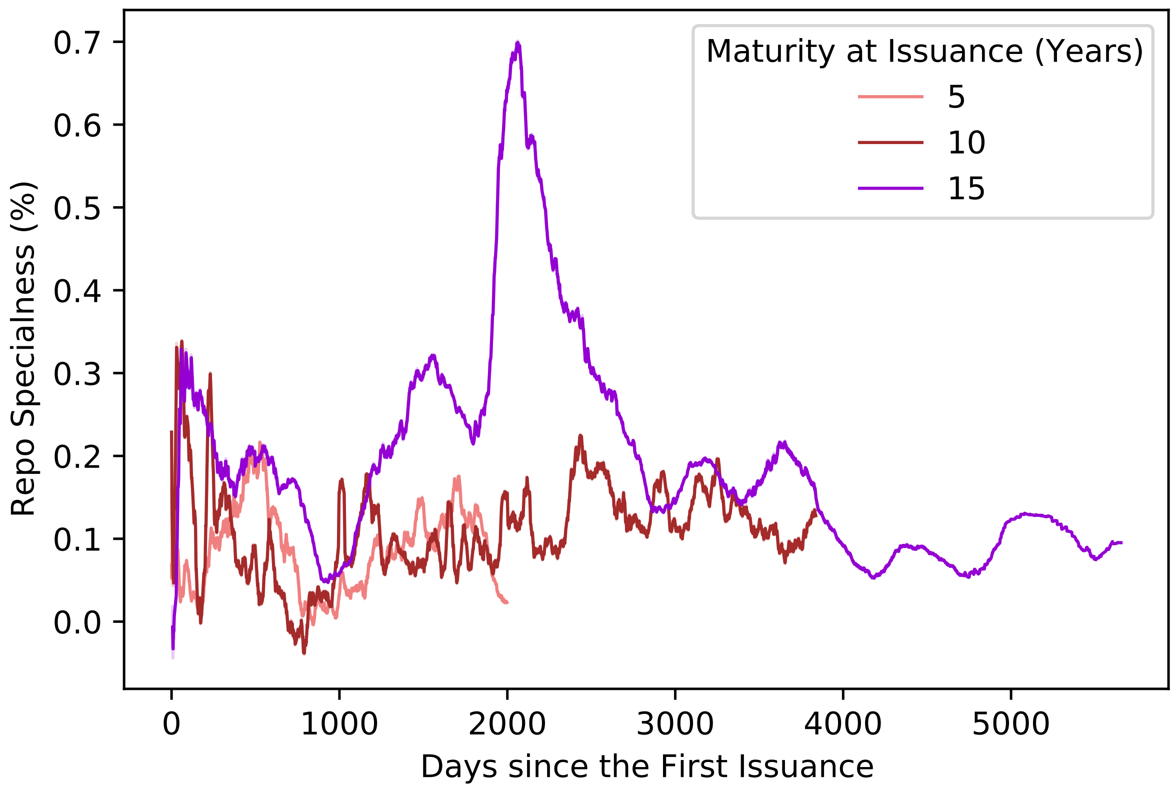

FIGURE OA.2: **Repo Specialness of Italian Treasury Bonds**. This figure shows the volume-weighted six-month trailing average of the daily rates on tick-by-tick repo transactions collateralized by Italian treasury bonds (most notably Buoni del Tesoro Poliennali, Buoni Ordinari del Tesoro, and Certificati di Credito del Tesoro), as recorded by MTS markets from 2012 to 2018. Daily rates are the volume-weighted average of intraday rates. Each day, repo transactions for 365 days are averaged. We distinguish between three benchmark bond maturities at issuance.

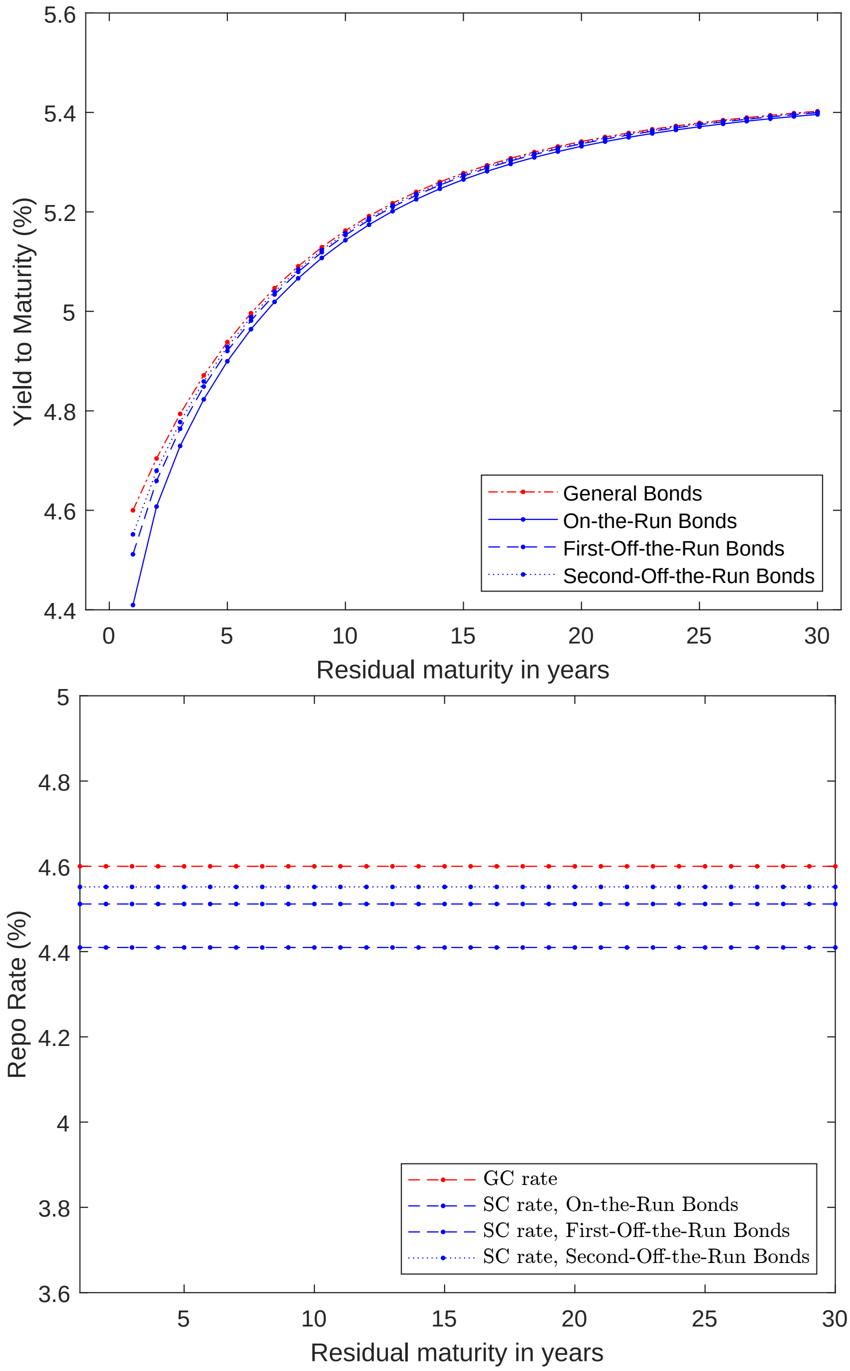

FIGURE OA.3: **Degrees of Specialness.** The top panel shows the term structure of interest rates. The bottom panel shows general and special overnight repo rates plotted against collateral tenor. Rates are expressed on a per annum basis. The curves in red show the general bonds, which are not exposed to demand pressure. The curves in blue show special bonds, which are differently targeted by exceptional demand pressure. Section 6.3 presents the calibration.