Bond Scarcity and the Term Structure

RUGGERO JAPPELLI,* LORIANA PELIZZON,† AND MARTI G. SUBRAHMANYAM‡

April 13, 2023

ABSTRACT

We develop a general quantity-driven equilibrium model that integrates the term structure of interest rates and the repo market to shed light on the effects of bond scarcity on the yield curve. In our model, investors with habitat preferences for special bonds exert price pressure in the secondary market. Arbitrageurs take the opposite side, but finance their positions by rolling over repo contracts, thus increasing the demand for the targeted securities in the repo market and lowering the rates to borrow against these bonds. We characterize in closed form the endogenous dynamic interaction between bond prices and repo rates, and illustrate both the strong localization of supply effects and the resulting portfolio rebalancing of investors. Our calibrated model can quantitatively match US Treasury data and the coexistence and dynamic interactions of general and special bonds in the yield curve. The requisite absence of arbitrage opportunities among bonds with equivalent cash flows obtains since the expected risk-return ratio accounts for the short repo rate, which varies at the instrument level.

JEL classification: 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, 968, 10012, New York, msubrahm@stern.nyu.edu.

We thank Giovanni Dell'Ariccia, Wenxin Du, Darrell Duffie, Ester Faia, Matthias Fleckenstein, Yesol Huh, Sebastian Infante, Urban Jermann, Francis Longstaff, Elisa Luciano, Toomas Laarits, Andrea Modena, Cecilia Parlatore, Pietro Reggiani, Stephen Schaefer, Fabian Smetak, Claudio Tebaldi, Davide Tomio, Bruce Tuckman, Dimitri Vayanos, Ernst-Ludwig von Thadden, Olivier Wang, Geoffery Zheng, and seminar and conference participants at the Bank of Italy, the Federal Reserve Board of Governors, the Leibniz Institute for Financial Research SAFE, the 21st International Conference CREDIT, the 15th edition of the IRMC, NYU Stern, the NYU Stern student seminar, UMass Amherst, 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 ackowledges the Alexander von Humboldt Foundation and the 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 the 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

Recent quantitative easing (QE) programs by central banks have sharply demonstrated the importance of demand factors in fixed income markets. The price of nearly riskless securities delivering known streams of payments rises persistently with large purchases (Bernanke, 2020), posing a challenge to standard bond valuation models in the financial economics literature as well as to Ricardian equivalence theories in the macroeconomic literature. These purchases induce a scarcity of high-quality collateral, which exerts downward pressure on the rates at which the targeted securities trade in the repurchase agreements (repo) market.[1] More generally, special bonds in substantial excess demand trade at economically large premia over and above the price of instruments with equivalent cash flows in the bond market, and secure comparatively lower financing rates in the repo market. Thus, special repo rates are also elastic with respect to the demand for bonds. Nonetheless, the theoretical research thus far designed to explain quantity effects on bond prices abstracts from the bonds' collateral value in the repo market. To fill this gap in the literature, this paper studies the endogenous interactions between the entire term structure of security prices and their respective repo rates in the context of sovereign bond markets. In our model, durable assets such as bonds not only serve as investment opportunities, but also as collateral for loans, in the spirit of Kiyotaki and Moore (1997).

We offer a comprehensive quantity-driven model of the term structure of interest rates that integrates and endogenizes the money market. There is growing recognition that demand and supply forces, particularly QE, affect both bond prices in the bond market (D'Amico and King, 2013; Greenwood and Vayanos, 2010, 2014) and the repo rates secured by bonds in the money market, both in the US (D'Amico et al., 2018) and in the EU (Arrata et al., 2020; Corradin and Maddaloni, 2020). To deliver such empirical regularities, we build a theoretical model yielding a number key findings which, to our knowledge, are new to the literature. Specifically, our model shows in closed form that the pricing of demand forces on the bond market emerges independently of preferences toward risk. The risk-neutral valuation of quantity effects depends on the elasticity of bond supply in the repo market, which measures the marginal cost of closing the arbitrage with a replicating portfolio composed of bonds not targeted by price pressure with otherwise equivalent cash flows. Thus, QE affects the term structure partly *because* it induces repo specialness, defined as the differential repo financing rate granted by bonds in exceptional demand relative to bonds with equivalent cash flows. The demand pressure, unrelated to fundamentals, exerted on the bond market is reflected in the specialness on the repo market by the endogenous response of market participants.

[1] 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 prespecified subsequent trading day. The counterparty enters the reverse side of the trade by buying the collateral on the spot market and stipulating a forward contract to sell it.

Dynamically, bond scarcity, repo specialness, and the term structure feature nontrivial and previously unnoticed interactions with each other. We show that repo specialness strongly influences the term premium, along with the entire yield curve. High levels of repo specialness are evidence of significant cost of carry trades and hedging strategies, both of which are effects countering the reduction in the slope of the yield curve induced by demand forces such as unconventional monetary policy. Thus, even in the absence of risk premia, repo specialness and the general level of interest rates interact with each other through the effect of their (mutual) correlation on the expectations of future rates. Our approach delivers a term structure model derived from first economic principles and featuring strongly localized supply effects, such as kinks. Moreover, we endogenize the short interest rate in the cross section of bonds, allowing for the combined analysis of the bond market and money market rates. From our general equilibrium analysis of securities and money markets, bond scarcity emerges as a powerful policy tool for influencing the yield curve. Repo specialness dampens the duration extraction of QE and therefore reduces the impact of any given quantity of asset purchases on the term premium. On the other hand, bond scarcity increases repo specialness, strengthening the local supply channel of QE.

To derive our results, we build on the Vayanos and Vila (2021) term structure model of the bond market (henceforth, VV). Distinct from their work, we focus on the preferences of investors for specific characteristics. For example, in the US Treasury bond market, securities with the same cash flows might be *on-the-run* or *off-the-run*. Traders prefer the former and bid up their prices.[2] We designate the bonds subject to excess demand as "special." In doing so, we introduce a new dimension in term structure models (TSMs): bonds that share the same tenor might differ in their exposure to exceptional demand forces. 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 borrowing cost of the bond in the repo market, where investors borrow and lend cash using bonds as collateral. The novelty of our paper springs from the latter connection, which gives rise to an endogenous, dynamic relation between the excess demand for a bond on the secondary market and its associated short-term borrowing rate on the repo market. We extend the no-arbitrage restriction between the bond and the repo markets in the insightful static model of Duffie (1996) to derive a rigorous, dynamic, general equilibrium model of money market rates where the short repo rate earned by lending against bonds subject to exceptional demand pressure depends on quantity factors, thus improving our understanding of *both* markets. The expected return-risk ratio on the bond market is consistent with the absence of arbitrage when the

[2]The *on-the-run* versus *off-the-run* definitions do not convey as much in other markets such as most of Europe, Japan, India, and others, because sovereign bonds in those markets are often issued "on tap." Thus, in principle, all bonds can be reissued and the specialness cannot be ascribed to recent issuance. In general, securities go on special when they attract a significant degree of excess demand. Such demand pressure sometimes arises when a bond becomes the *cheapest-to-deliver* in the futures market, or when the issue is labelled as Green or Islamic.

short repo rate varies at the instrument level.

The importance of the repo market and its close connection to the bond market underscores the relevance of our quantity-driven framework where we model both markets in general equilibrium. The repo market is the lifeblood of the financial system, where institutional investors routinely obtain collateralized financing, and its sheer size is enormous – much larger than the bond market itself. The average daily volume of outstanding repo transactions is about \$12 Trillions, roughly 14% of the world's GDP, of which Treasury repo transactions constitute about \$8 Trillions. By contrast, the daily volume in the US Treasury bond market averages around \$0.6 Trillions.[3]

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 which we leverage in our model. General collateral repo agreements are often called "cash-driven" transactions, because their primary purpose is to achieve collateralized financing to provide liquidity. In these transactions, each bond within a certain basket can be delivered as collateral. Thus, general collateral is a set of securities that trade at the same repo rate. On the other hand, 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, for example to cover short-selling commitments, causes buyers in the repo market to accept a lower interest rate in exchange for cash in these "security-driven" transactions. By lowering the attainable financing rate, "special" bonds yield a "repo dividend" (Duffie, 1996) that varies with the tenor of the collateral (D'Amico and Pancost, 2022) and the demand for that *particular* bond. As an illustration, Figure 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.[4] We distinguish between general collateral (GC) and special collateral (SC) transactions, and plot the latter for benchmark time-to-maturity buckets. The SC rates are generally below the GC rates. Further, it is clearly noticeable that the SC rates might vary stochastically across tenors and over time. How do we explain these empirical facts about the repo market? And what are their consequences for the bond market?

From a theoretical viewpoint, existing models of the term structure treat the short rate, such as the repo rate, as unique and inelastic, and thus *exogenous* to quantities. As a result, standard TSMs are silent regarding the impact of excess demand on money market rates. To fill this gap, we present a quantity-driven general equilibrium model of the bond and the repo markets which

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

[4] Italian sovereign bonds include Buoni del Tesoro Poliennali (BTP), Buoni Ordinari del Tesoro (BOT), and Certificati di Credito del Tesoro (CCT). Data from MTS markets at a millisecond precision

endogenizes the short rate, both in the cross section and in the time series. In equilibrium, repo specialness is stochastic, dynamic, and impacted by excess demand in the bond market. When subject to exceptional demand pressure in the bond market, a security becomes overpriced relatively to instruments with equivalent cash flows.[5] The lure of price deviations from economic fundamentals induces term-structure arbitrageurs, such as hedge funds, to reverse in the repo market the position targeted by exceptional pressure on the bond market and sell the security short. Gradually, this behavior raises the demand for high-quality collateral in the repo market, which raises the price of the bond. The outcome of this spiral can be characterized in a unique tractable closed-form solution, whose economic primitives are the exceptional demand (excess trading volume) in the bond market and the elasticity of collateral supply in the repo market.[6] Naturally, demand segmentation in the bond secondary market has its mirror image in the distinction between GC and SC in the repo market. 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.[7]

QE impacts financial markets, and thus the economy at large, by substituting part of the long-term bonds held by investors with cash. The endogenous response of the private sector is often referred to as portfolio balancing, and grouped into local supply and duration risk channels (see Joyce et al., 2012). The former results from preferences for specific bonds, and the latter from falling term or risk premia induced by a reduction in the average duration of the bonds held by investors. In our model, the strong localization of supply effects originates with the obligation of term-structure arbitrageurs to deliver the special bonds sold short, come what may, while the duration risk effect acts through their preferences toward risk. The VV model of the bond market focuses on the changes induced by QE on the risk premium required by term-structure arbitrageurs such as banks, tying to the preferences of these investors both the duration risk and the local supply effects. In our setup, these two channels have starkly different drivers, so that the transmission of purchases might even not be smooth across the yield curve or arise when purchases are predictable, which is relevant since central banks tend to be upfront about QE volumes. Importantly, we find closed-form solutions for the impact of QE on the repo market. Thus, our equilibrium model provides us with the requisite machinery to analyze the combined effects of demand pressure on bond prices and collateralized overnight financing rates, deriving a feasible choice set to

[5]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 (Duffie, 1996), and since those are often the cheapest among the basket of deliverable bonds for the settlement of futures contracts (Merrick Jr et al., 2005). Empirically, this pattern is well documented. Among others, Barclay et al. (2006) using clearing records show that the trading volume as well as the market share of electronic intermediaries decline by about 90% when Treasury securities go *off-the-run*.

[6]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.

[7]We address a long-standing puzzle in the literature. Cornell and Shapiro (1989) are among the first to show the existence of mispricing of what we shall refer to as special bonds with respect other to bonds of similar tenor.

guide the decisions of policymakers. In existing models of QE, it is unclear why the central bank would split large-scale asset purchases into multiple operations which span several months, if not years. Among the novel implications of our theory, we show that policymakers aiming to influence the yield curve while minimizing distortions in the repo market should favor predictable repeated reverse auctions to a one-time operation, *ceteris paribus*. This was generally the practice of the major central banks – including the Fed, the ECB, and the BoJ – during the past decade. Simply put, bond prices are forward-looking expectations of payoffs discounted at the entire stream of future repo rates, but current repo rates only reflect the contemporary stock of available collateral.[8]

A general equilibrium model that integrates the bond and the repo market has certain advantages. For instance, yield curve fitting errors of Treasury securities are widely used by academics, policymakers, and practitioners. To quote an example, in an influential paper Hu et al. (2013) use the dispersion in the Treasury yield curve fitting errors as a measure of pricing noise, which proxies for the shortage of arbitrage capital in the economy. One caveat of considering the Treasury market in isolation from the repo market is that bond mispricing might not be executable, if the borrowing cost of the 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. Importantly, it seems appropriate to drop 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 reaches as much as 50 bps. A calibration of our theory using realistic parameters illustrates quantitatively our main findings. For comparability with previous studies in the literature, we use US Treasury bond data from Gürkaynak et al. (2007). We start from the simplest case where a bond is assumed to remain on special through its entire life-cycle and its specialness features a certain time decay, an assumption that is later on relaxed. 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). We uncover a novel driver of local supply effects, which arise because the specific collateral repo agreements necessary to short-sell a bond subject to demand pressure require the delivery of that particular bond. Thus, the transmission of price pressure might not be smooth across the maturity spectrum. We then turn to the case featuring several levels of specialness accompanying the bond through its life-cycle, empirically relevant to the US market. Through time, *on-the-run-bonds* gradually become *first-off-the-run* and *second-off-the-run*, etc., to finally come to rest in the absorbing status

[8]A Reverse Repurchase Facility as introduced by the Fed might attenuate the bond and repo specialness of securities targeted by QE. Details about this monetary policy instrument are available at https://www.newyorkfed.org/markets/domestic-market-operations/monetary-policy-implementation/repo-reverse-repo-agreements.

of general bonds, as their specialness decreases.[9] Our contribution nests traditional and more recent TSMs as particular cases arising from the specification of a particular pricing kernel. While we build on the Preferred-Habitat Theory, the Expectation Hypothesis and the Liquidity Premium Theory of the term structure are also consistent with our model. Overall, our research agenda 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). Differently, in our theory, price differences are not attributable to specialized or constrained marginal investors but rather stem from the holding cost of arbitrage, i.e., the cost of repeatedly borrowing the position to sell it short, empirically documented by Fontaine and Garcia (2012).

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

2 Literature Review

The Term Structure of Interest Rates The term structure of interest rates describes the relation between the time to maturity and the yield of a bond. After the global financial crisis, unconventional monetary policy has renewed the efforts by researchers to explain the effects of demand pressure on fixed income securities. For instance, D'Amico and King (2013) and Greenwood and Vayanos (2014) document the partial transmission of QE that results from market segmentation. VV provide the analytical structure to harmonize these findings with the received preferred-habitat theory (pioneered by Culbertson, 1957 and Modigliani and Sutch, 1966), which makes the point that participants in bond market differ in their investment horizons. Rather then focusing on preferred maturity habitats, we focus on preferences for special bonds which could arise from liquidity considerations (Pasquariello and Vega, 2009).[10] We propose several extensions to the preferred-habitat model. First, we endogenize special repo rates by allowing arbitrageurs to finance their positions in the repo market. Second, our main results do not depend on the arbitrageurs' risk aversion. The impact of demand on prices arises instead from the execution cost of arbitrage, since

[9] The statement readily generalizes in Section 4.3 to bonds with same tenor and varying degrees of specialness resulting from the respective demand intensities, with the potential to fit any bond and repo rate pair.

[10] Naturally, bonds differ on many other dimensions than with respect to their maturity. For example, Chen et al. (2022) use the constraints prohibiting Islamic financial institutions to invest capital in compliance with Shariah law to identify clientele effects on bond prices and repo rates, and D'Amico et al. (2022) focus on Green premia.

the short rate responds to quantity. Importantly, 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 price differences result from differentials in their demand. However, the comparatively higher price of the sought-after bond is reflected by the appropriate special repo rate.[11] Finally, we introduce imperfect substitutability over maturities in the habitat preferences of investors, which allows us to model the rebalancing induced by QE on their portfolios. Recently, the elegant framework proposed by VV has been extended to the foreign exchange market in Greenwood et al. (2020) 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 above focuses on the effects of demand pressure on the repo market, formalizing the microfoundations of the behavior of arbitrageurs.

The Repo Market Repo contracts are similar to collateralized loans. In a seminal 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 (empirically validated by Jordan and Jordan, 1997) where special repo rates, those significantly below prevailing riskless rates, decrease as arbitrageurs intensify the search for collateral to sell the bond short on the secondary market. Differently from the earlier Duffie paper, which is static in nature, we explore the repo specialness in a dynamic sense both in the time series and in the cross-section of bonds, explaining it as the result of the interaction between demand forces and costly arbitrage. 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. To our knowledge, our paper is the first general equilibrium model formalizing these ideas in a term structure framework where repo specialness arises endogenously.

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 the recent advances in the literature on the heterogeneity in asset demand across investors. In the model of Vayanos and Weill (2008), search costs induce endogenous specialness: i.e., between two assets with identical cash flows, the one where short-sellers concentrate their trades is priced at a premium reflecting a larger

[11] The existence of a replicating portfolio is a sufficient condition for a valuation free of preferences.

pool of buyers.[12] 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, contain extensive descriptions of the institutional aspects of the US and European markets for repurchase agreements. Empirically, D'Amico et al. (2018) document that large-scale asset purchases affect repo specialness through the collateral scarcity channel, whose economically large impact is reflected in the secondary market for sovereign bonds.[13] Among other findings, Pelizzon et al. (2022) reproduce the Hu et al. (2013) "noise" measure in the Eurozone and show that yield curve fitting errors are associated with the scarcity of high-quality collateral induced by QE. Consistently with our model, Graveline and McBrady (2011) and Maddaloni and Roh (2021) show that buy-and-hold investors, such as pension funds and insurance companies, participate in the repo market substantially less than in the secondary market, increasing the scarcity of collateral. He et al. (2022) propose a preferred-habitat model that explains the behavior of Treasury convenience yields at times of crises, 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 and generic (e.g., *on-the-run* and *off-the-run*) securities, giving rise to equilibrium price differences between bonds with identic 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 satisfies a generalized notion of the Sharpe ratio that allows the short financing rate to depend on the characteristics of the collateral.

3 The Model

3.1 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 $i = \{\mathrm{g}, \mathrm{s}\}$, as general as opposed to special bonds. General and special bonds of the same tenor have equivalent cash flows, but their prices might differ because of 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. All stochastic processes are modeled under the equivalent martingale measure defined on

[12] Gârleanu et al. (2021) examine the stock and the securities lending markets when beliefs are heterogeneous.

[13] 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 to more than 50% in 2016, when it peaks at the huge level of 550 bps (`https://www.bis.org/publ/mktc11.pdf`, Graph IV.13.)

the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, and all are adapted to the filtration $(\mathcal{F}_t)_{t\in\mathrm{T}}$.[14] 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.[15]

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

Bonds can be used as collateral to obtain overnight secured financing in the repo market.[16] As is standard, to model 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).[17] 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 discussed, the repo market is segmented. Arbitrageurs with overnight cash on their hands have two distinct riskless options to lend money, either against special or general collateral bonds, at the respective market rates. Namely:

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

2. Reverse the position in the special collateral ($i = s$), which is in elastic 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 forego loan returns to borrow special bonds necessary to meet their pending short-selling commitments. Specialness premia $r_t - r_t^n$ do not result in any arbitrage opportunities, as demonstrate below. However, the supply of special bonds is *elastic* to quantities, and so are SC repo rates. In fact, the amount of special bonds outstanding 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. We endogenize the

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

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

[16] We focus on overnight repo transactions for ease of notation because the modelling of term repos would require an additional index. Empirically, the overnight tenor attracts the dominant volume proportion, by far. 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`.

[17] In the baseline model, we consider unlimited overnight borrowing without default risk, but Section 4.2 discusses borrowing constraints and haircuts. The results hold under re-use of the collateral as long as the passthrough of the rehypothecated collateral is less than 1, as is well understood empirically and need not be discussed here.

difference between the GC and the SC repo rates as a result of demand effect on the bond market, which induces the search for collateral on the repo market, to deliver the stylized facts illustrated by Figure 1.

3.1.1 Preferred-Habitat Investors

Preferred-habitat investors (such as bond market mutual funds) as a group 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.[18] Preferred-habitat investors are not active on the repo market, or perhaps less so than arbitrageurs.[19] We define as "special" those bonds that are targeted by preferred-habitat investors, and index them through $i = s$; to fix ideas, think of *on-the-run* and *first-off-the-run* securities as obvious candidates for "specialness."[20] Conversely, we refer to bonds of all maturities for which the excess demand is permanently zero as "general," and index them through their status $i = g$; for example, *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 which evolves as a 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 exeptional demand risk factors only affect special bonds. The process for demand risk in Equation (4) is autoregressive and mean-reverting.[21] 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 innovations.[22] To express the model in full generality, we allow for demand shocks and GC rate innovations to be correlated with tenor-specific coefficient ρ_n. Under normal market conditions, Equation (3) de-

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

[19] Bond-market mutual funds often target bellwether indices composed of *on-the-run* bonds of selected maturities and have mandates preventing them to achieve 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.

[20] 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-delivery, Green, and Islamic.

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

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

scribes preferences for liquidity and those arising from coordination equilibria among investors. In the context of QE, this formulation captures the purchases of targeted bonds by central banks 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. Later on, in Section 4.3, we show that similar results obtain by generalizing the specification to allow for a broader set of bonds targeted by purchases.

3.1.2 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.[23] For example, arbitrageurs (such as hedge funds) would sell short a bond 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 downwards 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, in equilibrium, arbitrageurs are only active in special bonds. 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 equally 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, general collateral 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.[24] The next expression is the dynamics of arbitrageurs' wealth W_t, where Δ denotes the first difference operator.

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

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

[24] 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-sellers.

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 where 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. The first term on the right hand 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 outright the bond in the spot market, and finance such 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. Differently from 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).[25]

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 that we propose, the demand forces which 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 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 specify exogenously the process for the short rate are suitable to describe the GC repo market, but leave the larger SC segment of the market unmodeled.

[25] 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.

3.1.3 General Bonds, Special Bonds

Two issues of the same tenor may differ in terms of their 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 general bonds as those 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 only reflect the risk of changes in the short rate r_t.[26] 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 exceptional demand risk q_t^n only exerts pressure on the price of bonds targeted by preferred-habitat investors. And indeed, we regularly observe that bonds on special are overpriced with respect to general bonds with identical cash flows in Treasury markets. In Section 3.3 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.

3.2 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 their model to an arbitrary equivalent 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.

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

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

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

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, the innovations in the short rate η_{t+1} and in the demand risk factor υ_{t+1}^{n-1}, in general correlated. In equilibrium, we verify that Equation (8) holds for both the special and the general bonds. However, by market clearing, arbitrageurs' net exposures at the close of business day are only short positions in special bonds.[27] Substituting Equation (8) into the arbitrageurs' wealth dynamics in Equation (6),

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

Each period, arbitrageurs maximize the expected value of 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, yet leaving unrestricted the preferences of arbitrageurs which uniquely pin down such a market price of risk.[28] Replacing Equation (9) into 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 the Jensen's correction term, so that

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

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

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

[27] Empirically, D'Amico et al. (2018) use the repo volume spread, calculated as volume of reverse repo to 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* relative to *off-the-run* Treasury bonds.

[28] As demonstrated in Section 3.4, the specification in Vayanos and Vila 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 (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 the cash with 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. Differently from the classical formulation where 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 the position.

3.2.1 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. By specifying the appropriate market price of risk $\underset{1\times 2}{\lambda(\cdot)}$, several interesting cases arise. In Section 3.4, 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.[29] 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 cross section of overnight special repo rates, r_t^n.

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

Equation (13) compares the textbook equilibrium concept with ours. Since the seminal paper by Vasicek (1977) and the two-factor 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 bonds can be used for collateralized borrowing at various 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 a 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.

[29]For a reference to the Girsanov theorem in discrete time see Föllmer and Schied (2008).

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 a different cost 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.[30] 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. Differently, in our interpretation 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 the market demand segmentation. An econometric test for the relative performance of the two TSMs is described in Section 3.5 below. Here, we focus on the close connection between the bond market and the repo market across the term structure of interest rates, and provide a general solution of the model that endogenizes repo specialness l_t^n, defined as the difference between GC and SC rates conditional on time to maturity:

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

3.2.2 *Affine Representation*

We cast our affine term structure model 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).[31] 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

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

[31] Grasselli and Tebaldi (2008) establish conditions for closed-form bond prices in admissable TSMs.

Menini, 2002). Factors that are more persistent exert a stronger impact on long-term yields.

3.3 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 general and special collateral bonds is instead riskless, since both their prices and repo rates are known.[32] 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 the 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 term structure model. To this end, let us specify as an auxiliary variable the difference between the pricing constants of bonds of different status and same tenor in Equation (7) by defining $D_n = C_n^s - C_n^g$.

Lemma 1. *In equilibrium,*

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

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 the details. *Q.E.D.*

Both the general and the special bonds promise the payment of equivalent cash flows at maturity. Therefore, their relative price, (on the left hand 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 hand side of the equation). Intuitively, absent this equivalence, the arbitrageurs would earn a free lunch by selling short (purchasing outright) the bond overpriced (underpriced) relatively to the other bond and to its own repo rate. Since both the 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 the 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* $e^{(B_n q_t^n + D_n)} = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{n} l_{t+j}^{n-j}}\right]$.

[32] We abstract from search costs in over-the-counter markets (see Duffie et al., 2005; Jankowitsch et al., 2011).

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, both computed under the equivalent martingale measure. Intuitively, the prices of special bonds reflect the 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. These papers are however 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 the 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 the repo specialness to arbitrageurs' demand for the 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. Conversely, the GC repo transaction rates are not sensitive to demand pressure, because any instrument within a basket of bonds can be delivered on the buy-back day. And 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 general collateral market, where each bond is substitutable with others included in the basket of deliverables, features a perfectly inelastic price elasticity to quantity. The special collateral market, where contracts command the delivery of specifically designated bonds, is instead characterized by its positive loan price elasticity of supply, since the amount outstanding 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 too can be thought of as preferred-habitat investors targeting and holding until maturity specific bonds on the cash market, and increasing their specialness in the repo market. For example, the ECB offers the bond purchased during QE operations for lending in its cash-collateralized securities lending facility at lower rates than the prevailing market ones, generating mispricing between instruments with equivalent cash flows (Pelizzon et al., 2022).

Differently from 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 raising 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 be upward sloping (see Duffie, 1996, Figure 3). This point is illustrated in Figure 2, which shows that the supply of the special collateral bonds is linear in its repo rate, with slope $\mathcal{E}^s$.[33] The demand is however inelastic, as arbitrageurs have the commitment to deliver the specific bond. With rightward shifts in the demand curve for SC in the repo market, equilibrium specialness increases as collateral holders require a higher compensation to pledge additional units of the special security. As we later on show, the chart is a general representation of the SC segment of the repo market which 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), that 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

[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 polynomial of second degree would result from a quadratic TSM, and so on for higher order specifications.

autoregressive structure with persistence φ_n.[34] The key insight is that the persistence parameters of these process 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 next outline 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 $\mathbb{Q}$.

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

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

with the initial condition

$$A_1 = 1, B_1 = -\mathcal{E}^i, C_1^i = 0.$$

Proof. Appendix B demonstrates the statement. Intuitively, the initial values A_1 and C_1^i 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_n^i 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_t^n measuring exceptional demand only exerts upward pressure on the price of targeted bonds, while not affecting 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 of 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

[34] For generality, we are allowing for tenor-specific parameters in the equation for demand risk. Gradually, these parameters guiding the process for excess demand for the issue change as the 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, e.g., from φ_{10} to φ_9.

our setup does not restrict the collateral value of all securities to be a common exogenous short rate. In fact, the joint modelling 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 generates specialness endogenously.

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, as these instruments are in elastic supply on the repo market.*

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

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

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

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 to reduce some equilibrium price. Since the GC borrowing rate r_t is not sensitive to quantities, 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 (its short rate risk) and generates riskless profits when both bonds reach maturity, thus contradicting the concept of equilibrium that requires market clearing, i.e., finite quantities.

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, i.e., $r_t^n \leq r_t$.

We would like to 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 special collateral is elastic. Our contribution allows us to understand such search for collateral as the reflection of exceptional demand pressure on the bond market. Ultimately, the demand risk factor on the bond market determines specialness in the repo market endogenously through the maximizing behavior of arbitrageurs.

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

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

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

An immediate implication of the previous result is that specialness equals to zero for GC bonds that are in inelastic supply, as we would expect ($\mathcal{E}^g = 0$; see Remark 1).[35] Importantly, the ingredients 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 of 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 a predictable and a random component. Interestingly, this result demonstrates that the elasticity of collateral supply $\mathcal{E}^i$ does not depend of the bond tenor. Regardless of the bond's tenor, repo specialness exactly reflects the excess demand in the bond market, as one would expect from a quantity-driven theory of repo

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

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

rates. Thus, Figure 2 is a general representation of the SC segment of the repo market independent of the tenor of the bond. Notice that Equation (18) is just 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. Summarizing, 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 1, 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.

Lemma 2. *The pricing recursion in Proposition 1 is consistent with the optimality of arbitrageurs, who value bonds 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, as well as 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 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 which are constant in the risk factors.

Note that from arbitrageurs' FOC, 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 hand side of Equation (19), and its special repo rate on the right hand 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 their 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 3.2 characterizes the prices of both the general and the special bonds.

Remark 2. *From the perspective of the arbitrageurs, general and special bonds are in equilibrium equally profitable. 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.*

3.4 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 D derives λ^{VV}, the market price of risk associated to 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. Recall that our general results hold under the risk-neutral measure. To obtain the 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 II 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, the quantity of collateral demanded in the market affects repo specialness together 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, either general or special. Thus, bond prices include a risk compensation as the principal notional 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 $A_1 = 1, B_1 = -\mathcal{E}^i, C_1 = 0$.*

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

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 as well as 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 of the holdings of the private sector. In Section 5.2 we show that our TSM generates strongly localized supply effects, a feature which, to our knowledge, has been 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 the general yield curves.*

Proof. Consider a generic term premium $\lambda_r(\cdot)$, 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 equilibrium,

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

[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 corresponding to our Proposition 1.

which differs from zero, since the holdings of arbitrageurs influence the market price of risk. In the above, we have used Proposition 1 which shows that $B_n = -\mathcal{E}^s \prod^{n-1}(1 + \varphi_n)$, as well as 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, which has a natural interpretation. When bonds are in infinite supply, QE lowers the term premium by inducing the 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 bond 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 the optimal holdings of the arbitrageurs is directly proportional to the persistence of specialness φ_n, which increases the likelihood of large realizations of specialness in the future conditionally 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) who have price semielasticity 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 that their scarcity affects 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 completely unrestricted the functional form of λ_r.

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 a less negative market price of risk λ_r^{VV}, and QE reduces the slope of the yield curve. When we instead allow for bonds to be in finite supply, 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 is 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 allows for, such a hedging strategy might become more costly exactly when it is more necessary, raising term premia. As a consequence, a reduction in specialness, for instance through a securities lending facility (SLF), results in stronger impacts of QE on the reduction of risk premia. As we show in Section 5.2 below, 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 (with each other) on the expectations of future rates. Suppose, for instance, that the central bank announces it will hike its reference rates in the future. Then, the expectations of future special repo rates will also change reflecting the rising expectations of the general collateral rates. The converse is also true, and expectations of special repo rates increases lead the general term structure to become more upward sloping, with the obvious caveat that correlations may also change over time.

3.5 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 (repo specialness) of the asset. Equation (13) is relatively simple to take 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 generic as well as 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 returns using market data and to assess the robustness of the estimates to different frequencies. Similarly, a common approach is to use the variation of returns to proxy for the standard deviation $\hat{s}^n$. Finally, the exercise requires a measure for the risk-free rate $\hat{r}_t$ and one for the tenor-specific overnight special repo rate $\hat{r}_t^n$ 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. From the GC and SC rates, the repo specialness $\hat{l}_t^n$ can be inferred. 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.

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

We leave to future research the task of carrying out a formal econometric test of this specification. Our model suggests that β_2 should be negative to prevent arbitrage opportunities. Intuitively, special bonds should have lower excess returns relatively to general bonds, since the former ones are generating additional cash flows on the repo market.

4 Extensions and Generalizations

4.1 Imperfect Substitutability in the Demand of Preferred-Habitat Investors

In general, preferred-habitat investors who 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 comparatively much lower price, it seems reasonable to think that the company will closely monitor the prices of both securities before implementing their hedging trades. These considerations induce us to generalize the demand specification of the preferred-habitat investors to model the consequences of their rebalancing, empirically documented by Koijen et al. (2021), on financial markets. Consider the following extension of the demand specification of the 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 $n \in [1, 2, \dots, \mathtt{N}]$ and define[40]

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

In Equation (20), special bonds are substitutable among each other. 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 3 corresponds to the case when $\mathcal{A}$ is a positive definite diagonal matrix. If however other maturities are imperfect substitutes, off-diagonal elements of $\mathcal{A}$ are negative as 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 demand cross-price sensitivity decreases as the distance to the main diagonal increases. In general, $\mathcal{A}$ might well be asymmetric.[41]

In order to write down compactly the joint evolution of the autoregressive demand risk factors, 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,

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

Which we can write more compactly as

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

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

[41]To see this, consider one example where 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 10-years time to maturity semgent are instead not (or perhaps, less) willing to shift their demand pressure to the 10-years bonds ($\alpha_{10,9} = 0$) because they are committed by institutional constraints to invest in the long-duration fixed income market composed by bonds having a time to maturity equal to or larger than 10 years.

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

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 Equations (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, $q_t^{\texttt{N}+1} = 0$. 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 $\rho_{i,j}$, 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, conditionally 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 which consist of the pricing recursion coefficients.

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

Equation (23) is a generalization of Equation (7) that allows the prices of special bonds to depend of the whole term structure of demand risk factors. This formulation reflects substitutability across bonds in the demand of preferred-habitat investors. The model in Section 3 corresponds to the case where $\mathcal{A}$ and B are diagonal matrices. For convenience, we assemble the rows of B into the vectors $(B^n)_{n=1}^{\texttt{N}}$ which represent the price sensitivity of bonds with maturity n to the demand risk factors across the whole 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

$$\underset{\mathtt{N}\times 1}{M_t} = A\Delta r_t + B\Delta Q_t + C - A\Big[r_t + (1-\varrho)(\theta - r_t)\Big] - B\Big[\Phi Q_t + \overline{Q}\Big],$$
$$\underset{\mathtt{N}\times\mathtt{N}}{\Sigma}\ \ \underset{\mathtt{N}\times 1}{U} = \Omega V_{t+1} + \sigma_r \eta_{t+1} I_{\mathtt{N}}.$$

Note the parallel between Equation (24) and Equation (8) in the univariate model of Section 3. 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 3, where the demand risk factors are allowed to correlate among each other as well as with the general collateral rate. We let $\tilde{M}_t = M - .5\mathbb{E}_t^{\mathbb{Q}}\Big[U_{t+1}'\Sigma'\Sigma U_{t+1}\Big]$ denote the vector of drifts after accounting for the Jensen's correction terms, and $R_t = \Big[r_t^1 \quad \dots \quad r_t^{\mathtt{N}}\Big]$ 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-hand side, the drift term is different 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 E. On the left-hand side, the specialness of bonds now reflects demand pressure across the whole 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

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

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 shows up already under $\mathbb{Q}$ in the recursion for the pricing coefficients, derived in Appendix E. 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 in VV. Using the market clearing condition,

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

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 from investors targeting their preferred habitat.

4.2 Heterogeneous Arbitrageurs: Haircuts and Borrowing Contstraints

Thus far, the 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, who hold positions $(X_t^{j,n})_{n\in\mathbb{N}}$. Clearly, different business models give rise also to differences in counterparty risk. From the perspective of academics, market participants, and policymakers, haircuts are seen to mitigate 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 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 \$5 times the price of the bond as a 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 contstraints, the maximization programs of the arbitrageurs would incorporate the scarcity of capital and the requirement of each position being 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

[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 default.

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 wealth modified to reflect the foregone 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 martinage 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 foregone returns remunerated at the GC rate times the haircut. Moreover, carry trades are only possible when capital is available (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 the behavior of the arbitrageurs.[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 the 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 which follows Proposition 1 with one exception: the initial condition A_1^j becomes arbitrageur-specific and shifts upward in proportion to the haircut. Namely,

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

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

The interpretation of the above analysis is straightforward. For example, consider a hedge fund investing \$1 Mn in the repo market at the GC rate. Since the fund has a high risk tolerance (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 GC or SC bond in the market for repurchase agreements, it has to pledge more cash. Recall from Section 3.4 that risk aversion affects positively 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 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 even more so at longer horizons. Thus, arbitrage profitability increases in the horizon of the carry trade for agents with lower risk aversion than the prevailing one 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.

4.3 The Degrees of Specialness and the Treasury Auction Cycle

So far, simplicity considerations led us to consider the bond (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 differential demand pressure across them, suggesting a generalization of the status of specialness from a binary to a categorical variable. We thus reindex special bonds through $i \in \mathbf{s} = \{s_{\mathrm{P}}, \ldots, s_1\}$; to fix ideas, think of *on-the-run*, *first-off-the-run* securities, *et cetera*. Without loss of generality, we order the elements in the set s as decreasing in the degree of their specialness. The demand of preferred-habitat investors thus becomes

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

allowing 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 paper 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. As an 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 rollover their exposures into new issues, typically the largest specialness spreads 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. Wathersheds 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 by the auction date, and after two weeks by the issuance date. As an example, the 2-year and 5-year US Treasuries are issued on a monthly auction cycle, and the 10-year and 20-year are 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). Predictability of bond specialness 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. relatively to the US market, in the European repo markets the collateral specialness is substantially more persistent. As an illustration, Figure 3 shows the 1-year volume-weigthed trailing average of the SC transaction 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 during their entire trading life cycle. We further note that the specialness of the bonds with 15 years of maturity at issuance peaks after about 5 years, when the time-to-maturity reaches 10 years, and sharply decays 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 the European sovereign bond markets, in the remainder of the analysis

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

we focus on the US market where, as a result of the regular 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 a bond q_t^n. Thus, the cyclical behavior of repo spreads is guided by the parameters governing exceptional demand pressure in Equation (4). As discussed in Section 5, 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 US and the economic intuition that preferred-habitat investors roll over their position into liquid bonds.

5 Calibration

5.1 Two Yield Curves

The calibration of our model is tantamount to the combined modelling of the general and the special yield curves in the bond market, and of the specialness in the repo market. The exercise is interesting because it allows for the analysis of the effects of counterfactual scenarios determined by conventional monetary policy tools that guide the short rate behavior as well as unconventional instruments that act through demand pressure on the bond and repo markets.[45] We use the simple model structure outlined in Example 2, and refer to well-established contributions in the literature on financial economics.

For comparability with VV, we set $\mathtt{N} = 30$ 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 calibration of 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 the standard deviation of the 1-year yield, respectively. The market price of GC risk λ_r replicates the average 10-year bond yield in the data. To measure $\mathcal{E}^s$, we use the estimate of the impact of bond purchases on their returns conditionally 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 pace φ. We set $\overline{q}_t$ to match the average *on-the-run* repo spread of 19.4 bps documented by D'Amico et al. (2018). This value is approximately similar to the GC repo/T-bill spread of 23.65 bps found by Nagel (2016), albeit more conservative. We tune the persistence parameter φ to the ratio between the average *on-the-run* repo spread 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. A comparison of the current level of the risk factors to their long-run means determines whether the curve is in contango or in backwardation.

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 6 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 4, our model features several salient characteristics. First, as we can see from the top panel, 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 (its price is higher) than that of the general bond. Thus, the yield curve composed by interpolating the prices of special zero coupon bonds lies below the yield curve of general bonds, but their difference shrinks with time to maturity, as demand pressure shocks die out over time. That is intuitive, given that the two curves are generated by rolling over GC and SC rate risk and that special repo rates are generally below general ones. In fact, the vertical distance between the general and the special collateral curve at short residual maturities reflects the elasticity of the repo market supply of special collateral $\mathcal{E}^s$, and at the longer end of the yield curve the persistence φ. The gradually decreasing pattern of bond specialness reminds us of the spread between *on-the-run* and GSW-fitted yields documented in Greenwood et al. (2015), Figure 1. Second, the joint modelling of the general and special yield curves on 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 4, we show the repo rate on GC (in red), that is constant across time to maturity, as well as on SC transactions (in blue). The SC rate is $\mathcal{E}^s \times \overline{q}_t$ times lower than the GC rate, except for the more special 10-year bond, which we use to illustrate local supply effects.

5.2 *Local Supply Effects*

In Figure 4, exceptional demand pressure directed towards the 10-years special bond q_t^{10} is stronger. Targeted demand pressure captures the structural intervention of the central banks through policies such as QE. A central bank can be modeled as a buy-and-hold investor which 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. In the top panel of Figure 4, 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 modelling 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

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

intervention by the central bank is represented in the bottom panel of Figure 4, where the cross-section of special repo rates reaches a trough for the 10-years tenor special collateral that is more aggressively targeted, illustrating the endogeneity of repo rates. Simply put, when some investors exert significant demand pressure raising a bond's price and lowering its yield, arbitrageurs sell the security short, increasing its repo specialness. Since the special collateral is not substitutable 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 4.

This calibration exercise generates several interesting policy implications. To cite just one, consider any two levels of exceptional demand for long-term and for short-term bonds, respectively, that 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: Indeed, if the decay of exceptional demand pressure is fast, the two bonds will be approximately exposed to the same repo dividend. The same repo 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 impact the yield of long-term bonds while minimizing distortions on the repo market. Simply put, prices are forward looking while special repo rates reflect the *contemporary* stock of collateral. As the rate of decay of exceptional demand pressure diminishes, the bond price immediately increases, thus reflecting expectations of its declining future specialness. 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, 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 relative to a one-time operation. This is generally in consonance with the practice of the major central banks, including the ECB, the BoJ, and the Fed during the past decade. As noted, we expect the SLF policies to reduce repo specialness and allow to arbitrage away the resulting kinks in the yield curve, affecting the localization of supply effects.

5.3 Differently Special Bonds

Figure 5 illustrates the general case of our model with varying degrees of bond specialness introduced in Section 4.3. 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. 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 of 19.4 bps, 8.4 bps and 4 bps, respectively. We set special repo rates to these values in our calibration. Correspondingly, we raise the persistence parameter to the cardinality of the set of special bonds $P = 3$, thus setting the value of φ to 1.5 percentage points. Graphically, bonds jump upwardly to the more seasoned yield curve as investors roll over their portfolios to newly issues 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*, to finally come 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.

5.4 Bond Scarcity and the Term Premium

In the previous calibrations, the market price of risk λ was set to a constant, tuned to the average term premium in the data, as discussed. To illustrate the effects of QE on the slope of the yield curve, we allow the market price of risk to depend on the holdings of arbitrageurs and show the resulting term structure in the top panel of Figure 6. We use our baseline calibration and tune the market price of risk to Christensen and Rudebusch (2012), who estimate a reduction of the 10-year premium of 29 bps using data from eight QE announcements in the US Treasury market. For clarity, we mute local supply effects. We use the same scale of previous calibrations, so as to allow for a better comparison with the previous figures. All else being equal, when holdings data enter the risk premium, asset purchases reduce the slope of the yield curve. In the bottom panel of Figure 6, we carry a counterfactual exercise which holds fixed the quantity of purchases and increases their effect on repo specialness, widening the vertical distance between the GC and the SC yield curves. The resulting higher repo specialness dampens the reduction of the term premium induced by QE, highlighting the dynamic interactions between bond scarcity, repo specialness, and the term structure of interest rates.

6 Conclusion

Empirical fixed income markets research in the last two decades has documented systematic patterns in the spread between general bonds and special bonds that are difficult to explain in the context of uncertainty in the short rate dynamics. The extant literature has lacked a coherent theory to reconcile this evidence with existing models of the term structure of the interest rates. In this paper, 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 whole 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 have, however, abstracted from credit risk and market liquidity considerations, which may give rise to additional effects. Our model was implemented without taking a particular stance on investor preferences, for ease of comparison with other techniques in the literature. At the same time, we illustrate how our general formulation nests preference-based approaches as special cases.

The theory that we have presented in this paper has two attractive features. First, we have provided a unified framework that connects the secondary market for (nearly) risk-free bonds, e.g., Treasury bonds, with the repo market for collateralized financing. Policymakers could use our model to asses the combined effects of exceptional demand pressure, such as quantitative easing 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. As a specific application, this structure has allowed us to fit jointly the *on-the-run* and *off-the-run* yield curves in the US in a manner that is consistent with the absence of arbitrage opportunities. In other sovereign bond markets such as within the Eurozone, specialness may arise due to futures market microstructure issues (e.g., cheapest-to-deliver bonds) and search costs. Third, we have characterized the many dynamic and multifaceted interconnections 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 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 Treasury auctions that account for cyclicality in specialness, enable 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. This article has discussed the demand pressure for special issues that have the same cash flows as benchmark securities; applications could focus on Green or Islamic bonds premia. Future research could generalize the method that we have proposed to multifactor or quadratic term structure models from the theory side and test its preditcions empirically.

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 hand side of Lemma 1 to its expected leaded value. By taking logs on both sides of the expression and rearranging terms, 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 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 FOC 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 (1 factor,) and demand ($\mathtt{N}$ factors, one for each tenor) risk. Specifically,

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

E Equilibrium with Imperfect Demand Substitutability

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

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

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

$$
\begin{aligned}
-\log b^{n+1}_t(i) &= -\log E^{\mathbb{Q}}_t\Big[b^n_{t+1}(i)\Big] \\
&= \mathbb{E}^{\mathbb{Q}}_t\Big[-\log b^n_{t+1}(i)\Big] - .5\text{Var}^{\mathbb{Q}}_t\Big[-\log b^n_{t+1}(i)\Big], \\
A_{n+1} r_t + B^{n+1} Q_t + C^i_{n+1} &= A_n\Big[\varrho r_t + (1-\varrho)\theta\Big] + B^n\Big[\Phi Q_t + \overline{Q}\Big] + C^i_n \\
&- .5\Big[A_n^2\sigma_r^2 + \sum_{i=1}^{\mathtt{N}} B^2_{n,i}\langle \Omega^\top, \Omega\rangle_{n,i} + 2A_n \sum_{i=1}^{\mathtt{N}} B_{n,i}\rho_{n,i}\Big].
\end{aligned}
$$

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.

$$
\begin{aligned}
A_{n+1} &= \varrho A_n + 1, \\
B^{n+1} &= B^n \Phi - H^{n+1,i}, \\
C^i_{n+1} &= C^i_n + A_n(1-\varrho)\theta - .5A_n^2\sigma_r^2 + B^n\overline{Q} - .5\Big[\sum_{i=1}^{\mathtt{N}} B^2_{n,i}\langle \Omega^\top, \Omega\rangle_{n,i} + 2A_n \sum_{i=1}^{\mathtt{N}} B_{n,i}\rho_{n,i}\Big].
\end{aligned} \tag{30}
$$

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

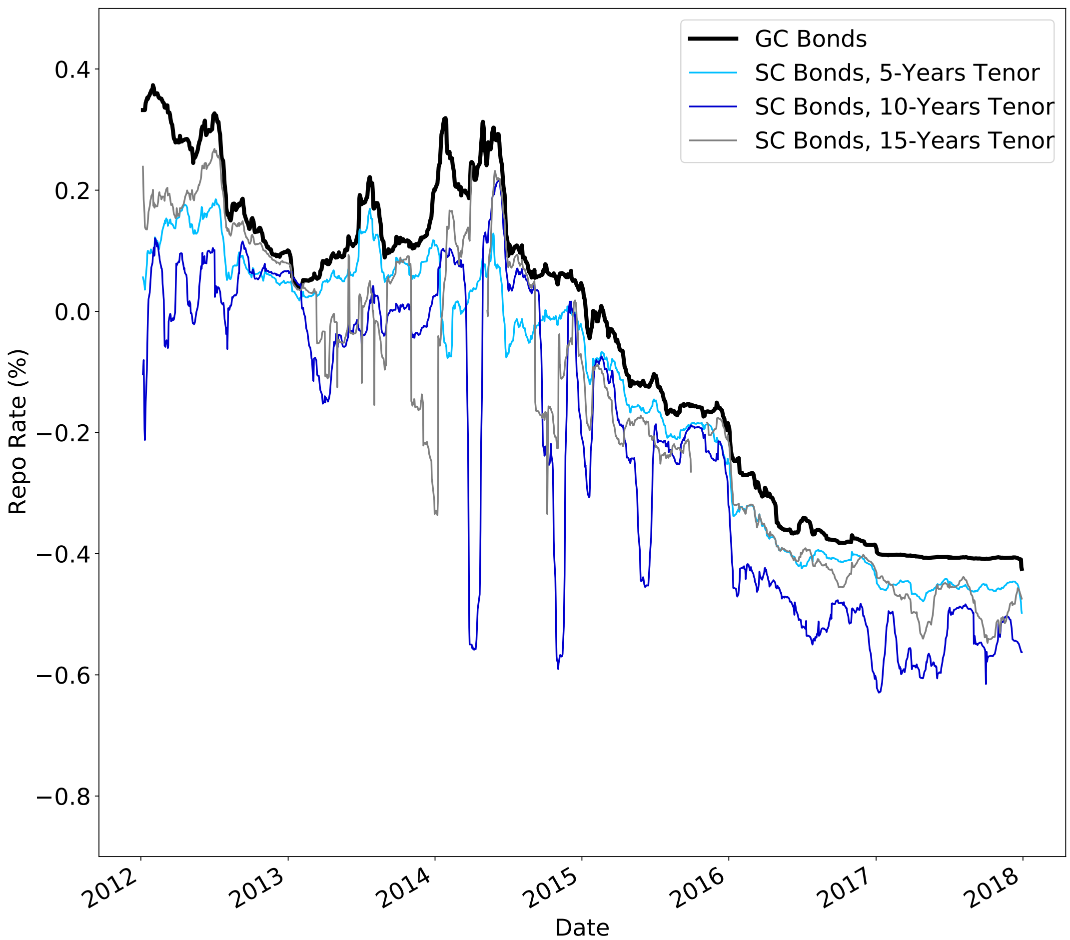

FIGURE 1: **General and special repo rates for Italian Treasury bonds**. This figure shows the volume-weighted monthly trailing average of the daily rates on tick-by-tick repo transactions collateralized by Italian Treasury bonds (most notably Buoni del Tesoro Poliennali, Buoni Ordinari del Tesoro, and 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 between general collateral (GC) and special collateral (SC) transactions, the latter shown for benchmark time-to-maturity buckets.

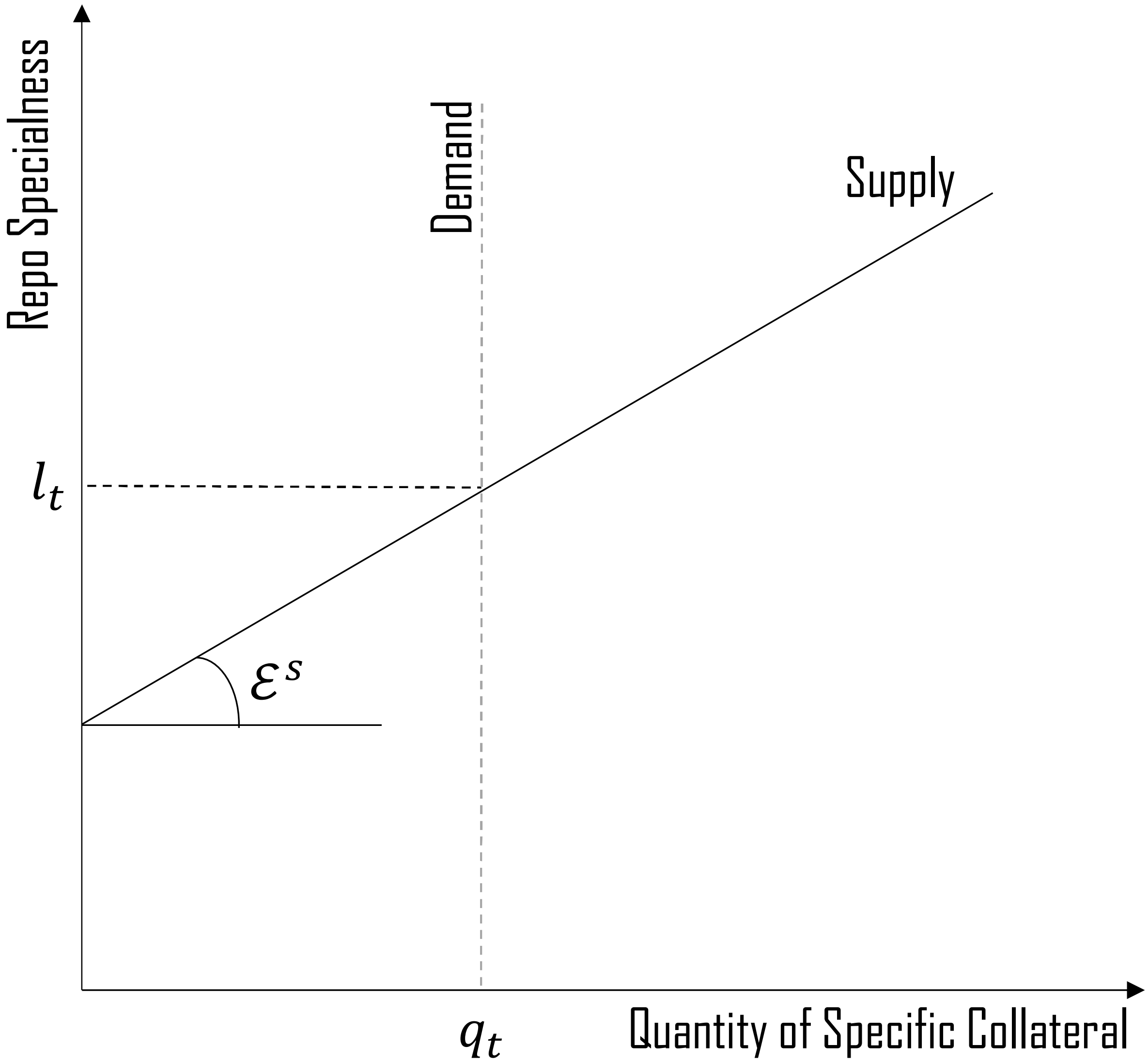

FIGURE 2: **Demand and supply of special collateral**. This figure illustrates the functioning of the market for repurchase agreements collateralized by special bonds. The horizontal axis shows demand pressure on the cash market and the vertical axis repo specialness. The supply curve is upward sloping.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

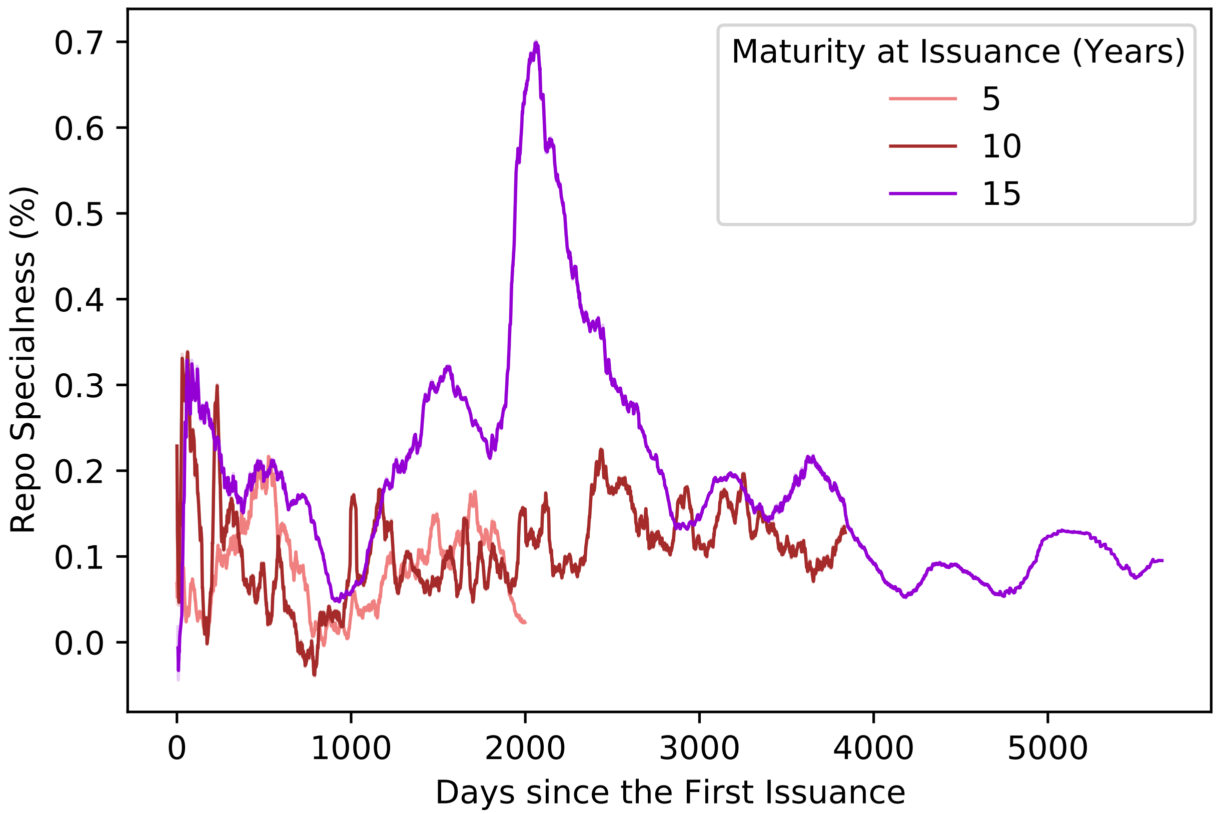

FIGURE 3: **Repo Specialness of Italian Treasury bonds**. This figure shows the volume-weighted 6-months trailing average of the daily rates on tick-by-tick repo transactions collateralized by Italian Treasury bonds (most notably Buoni del Tesoro Poliennali, Buoni Ordinari del Tesoro, and Certificati 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.

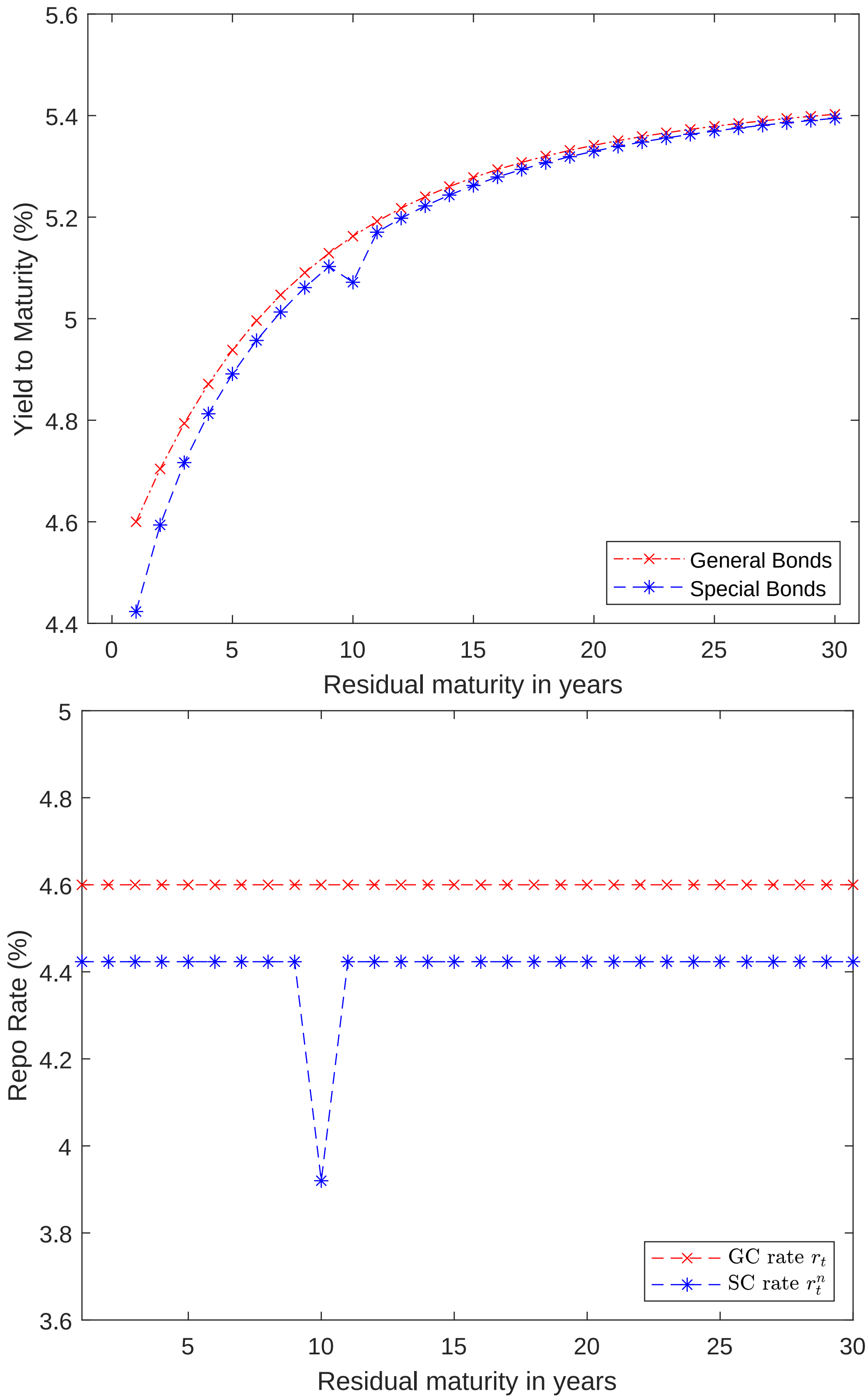

FIGURE 4: **Yield curves and repo rates**. The top panel of the figure shows the term structure of interest rates. The bottom panel shows general and special overnight repo rates plotted against collateral tenor. Rates are expressed on a per annum basis. The curves in red show the general bonds, not exposed to demand pressure. The curves in blue show special bonds targeted by exceptional demand pressure. Table I presents the calibration.

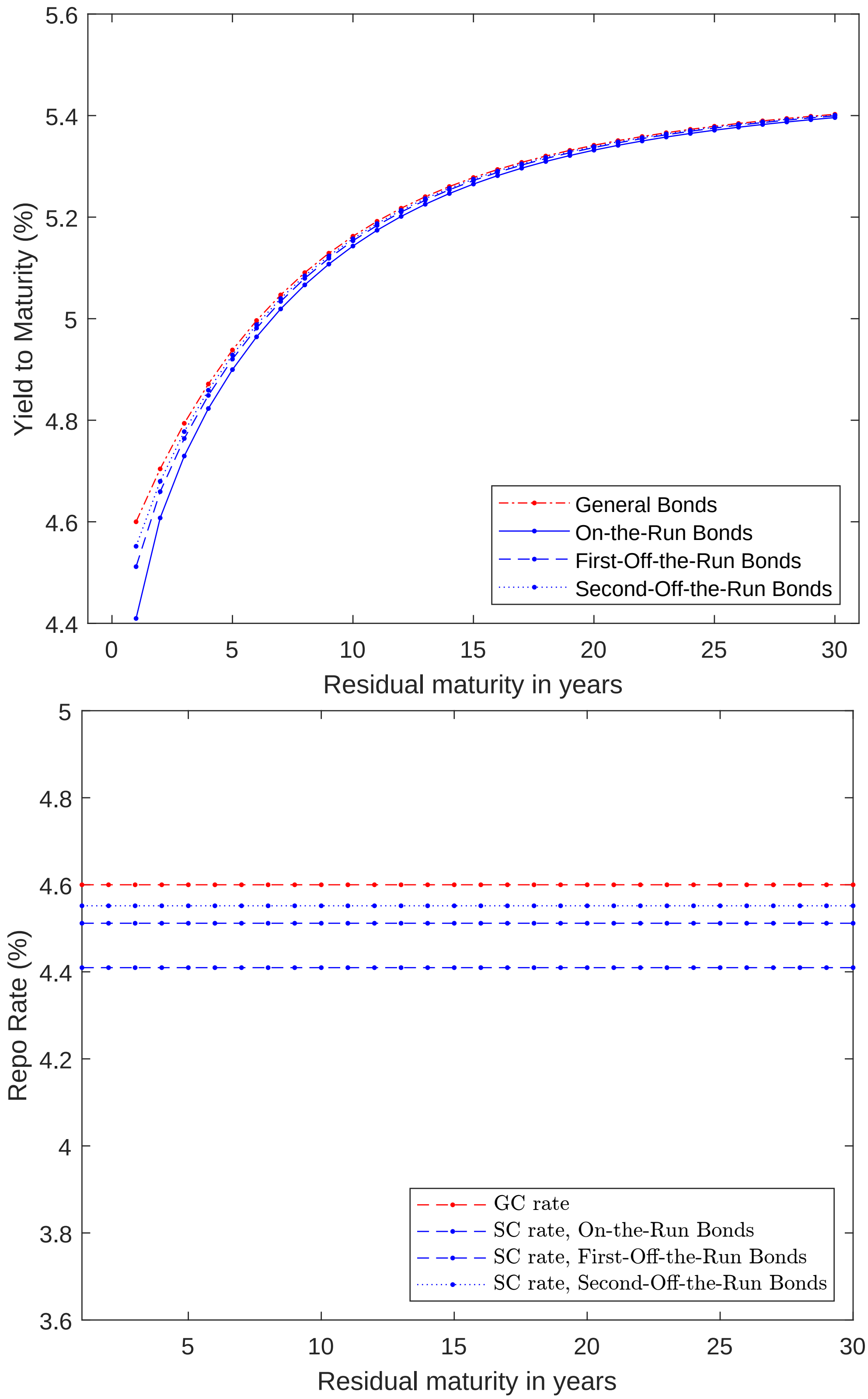

FIGURE 5: **Degrees of Specialness**. The top panel of the figure shows the term structure of interest rates. The bottom panel shows general and special overnight repo rates plotted against collateral tenor. Rates are expressed on a per annum basis. The curves in red show the general bonds, not exposed to demand pressure. The curves in blue show special bonds differently targeted by exceptional demand pressure. Section 5.3 presents the calibration.

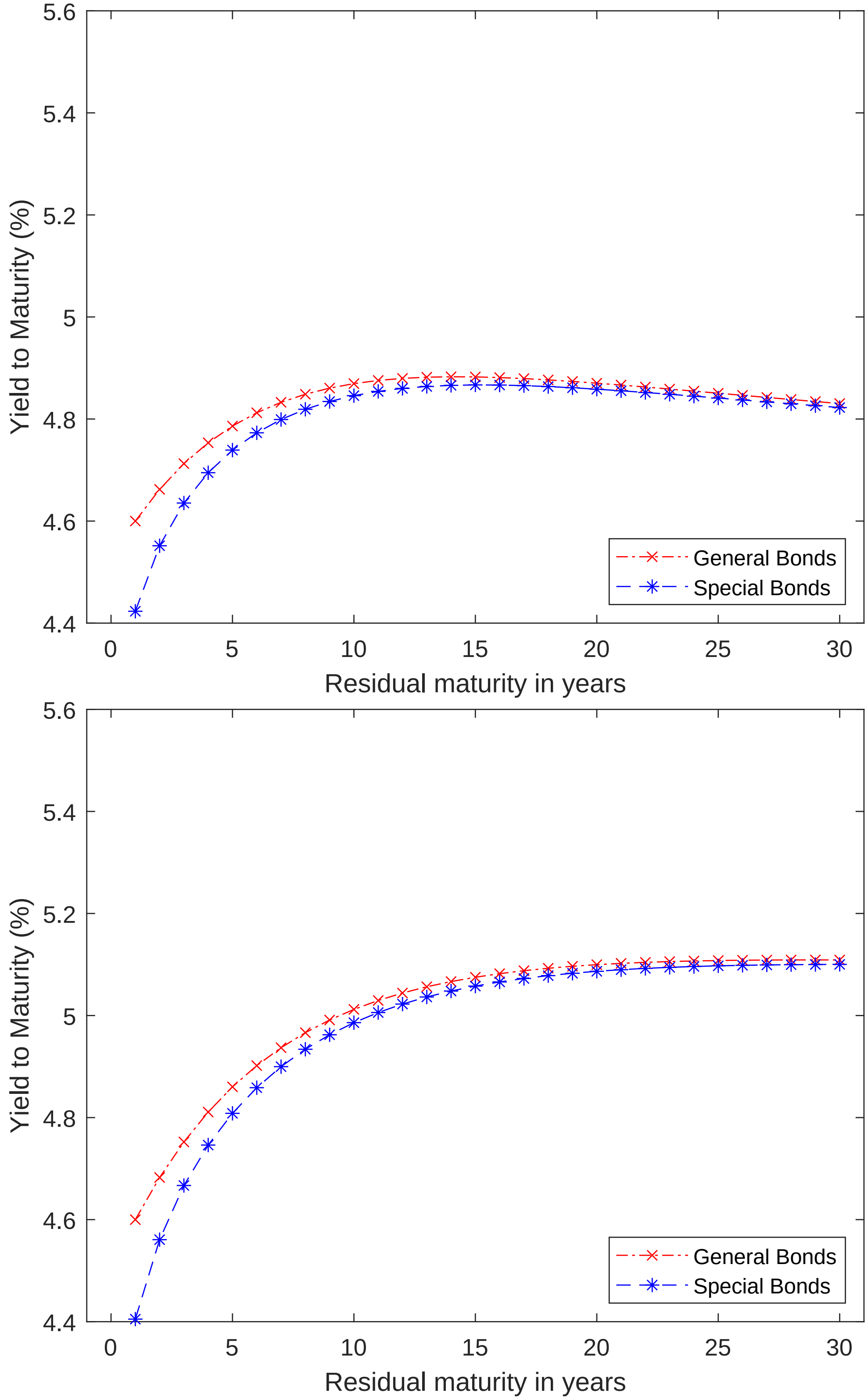

FIGURE 6: **Bond Scarcity and the Term Premium**. The top panel of the figure shows the term structure of interest rates tuned to the compression in the term premium induced by QE. The bottom panel shows a counterfactual exercise which holds fixed the quantity of purchases and increases their effect on repo specialness. Rates are expressed on a per annum basis. The curves in red show the general bonds, not exposed to demand pressure. The curves in blue show special bonds differently targeted by exceptional demand pressure. Section 5.4 presents the calibration.

TABLE I: **Calibration**

Yield Curve Calibration on 1985 - 2020 Data			
Parameter	Value	Source	Data and Moment
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

Exceptional Demand Pressure and Local Supply Effects			
Parameter	Value	Source	Data and Moment
$\mathcal{E}^s$ 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

TABLE II: **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 seminal paper by Vasicek (1977) develops the equilibrium consistent with the absence of arbitrage. The two factor model by Brennan and Schwartz (1979) derive 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.

References

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.

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.

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.

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., 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. URL https://doi.org/10.1111/j.1468-0297.2012.02550.x.

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., J. Klausmann, and N. A. Pancost. 2022. The Benchmark Greenium. Working Paper.

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.

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.

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.

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

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

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.

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. 2020. A quantity-driven theory of term premia and exchange rates. Working Paper.

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.

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.

Joyce, M., D. Miles, A. Scott, and D. Vayanos. 2012. Quantitative easing and unconventional monetary policy–an introduction. *The Economic Journal* 122:F271–F288.

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* pp. 473–489.

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.