NBER WORKING PAPER SERIES

GRANULAR TREASURY DEMAND WITH ARBITRAGEURS

Kristy A.E. Jansen
Wenhao Li
Lukas Schmid

Working Paper 33243
http://www.nber.org/papers/w33243

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2024

We thank Viral Acharya (discussant), Daniel Andrei (discussant), Adrien d'Avernas (discussant), Zefeng Chen, Greg Duffee, Darrell Duffie, William Diamond, Valentin Haddad, Ben Hebert, Francois Gourio, Erica Jiang, Ralph Koijen, Arvind Krishnamurthy, Moritz Lenel (discussant), Martin Lettau, Jane Li, Karen Lewis, Tom Sargent, Alexi Savov, Selale Tuzel, Quentin Vandeweyer (discussant), Annette Vissing-Jorgensen, Robert Richmond, Olivier Wang, Xingtan Zhang, Geoffery Zheng, and seminar and conference participants at the SF Fed, BI-SHoF Conference on Asset Pricing and Financial Econometrics, Junior Valuation Workshop at Wharton, Zurich Quantitative Macroeconomics Workshop, UBC Summer Finance Conference, USC, Chicago Fed, Stanford SITE, Stanford Junior Macro-Finance Conference, SAIF, PBCSF Tsinghua, CKGSB, Peking University Guanghua, NBER Financial Market Frictions and Systemic Risks Conference, UIUC, NYU, CMU Tepper-LAEF Conference, Princeton Macro-Finance Conference, UNSW, ANU, University of Sydney, University of Technology Sydney, UCLA Macro-Finance Lunch, JHU Carey Finance Conference, and Berkeley Haas for very useful comments and discussions. We gratefully acknowledge financial support from NBER Financial Market Frictions and Systemic Risks initiative. We thank Winston Chen for excellent research assistance. Views expressed are those of the authors and do not necessarily reflect official positions of DNB or the Eurosystem. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Granular Treasury Demand with Arbitrageurs
Kristy A.E. Jansen, Wenhao Li, and Lukas Schmid
NBER Working Paper No. 33243
December 2024
JEL No. G12

ABSTRACT

We construct a novel dataset of sector-level U.S. Treasury holdings, covering the majority of the market. Using this dataset, we estimate maturity-specific demand functions and elasticities of different investors and the Fed, and integrate them into a dynamic equilibrium model of the Treasury market with risk-averse arbitrageurs. Quantifying the model reveals that (1) strong arbitrage leads to an elastic Treasury market and a steeply downward-sloping term structure of market elasticity; (2) monetary tightening raises term premia due to arbitrageurs interacting with investors exhibiting high cross-elasticities; (3) QE has limited impact unless the Fed credibly commits to sustained balance sheet expansion.

Kristy A.E. Jansen
University of Southern California
Marshall School of Business
Los Angeles, CA 90089
kjansen@marshall.usc.edu

Lukas Schmid
Marshall School of Business
University of Southern California
and CEPR
lukas@marshall.usc.edu

Wenhao Li
University of Southern California
Marshall School of Business
Los Angeles, CA 90089
liwenhao@marshall.usc.edu

1. Introduction

The U.S. Treasury market is a cornerstone of the financial system, shaping monetary and fiscal policy, influencing global investment flows, and serving as a benchmark for financial instruments. Most recently, the U.S. Treasury market has taken center stage in the swift policy responses to the global pandemic in 2020, when the Federal Reserve (Fed) aggressively purchased long-term bonds under quantitative easing (QE) policies, to counteract outflows from foreign investors and mutual funds. The impact of such interventions critically depends on how bond investors accommodate sales or purchases, given their risk capacity and mandates. Against this backdrop, we ask: What is the role of arbitrageurs in the Treasury market? Do investors cross substitute Treasuries across maturities? How elastic is the Treasury market? What makes the Fed's interest-rate and QE policies effective?

In this paper, we address these questions by proposing an equilibrium model of the U.S. Treasury market that draws on important insights from two influential literatures: the burgeoning literature on demand-based asset pricing pioneered by Koijen and Yogo (2019), and the preferred habitat view of the term structure of interest rates in Vayanos and Vila (2021). We extend the concept of "preferred-habitat investors" in Vayanos and Vila (2021) to what we refer to as "granular-demand investors", whose asset demand can be flexibly estimated from Treasury holdings data, similar to Koijen and Yogo (2019), and we show that these approaches are naturally integrated. In our context, we rely on a novel granular dataset on Treasury portfolio holdings covering most of the market, including commercial banks, insurance companies, foreign investors, among others, as well as the Fed. These demand estimates can be naturally embedded into a dynamic equilibrium model featuring risk-averse arbitrageurs clearing the market. By empirically linking arbitrageurs to hedge funds and dealers and their Treasury holdings, the model enables us to quantitatively evaluate the role of arbitrageurs and how the Treasury market adjusts in equilibrium to demand shocks, macroeconomic changes, and conventional and unconventional monetary policies.

Our analysis reveals three main findings. First, the Treasury market is elastic because arbitrageurs exhibit low estimated risk aversion that significantly weakens demand impact. The strength of arbitrage is heterogeneous, stronger at shorter maturities because of lower risks, leading to a downward-sloping term structure of market elasticity. Second, term premia rise in response to a monetary policy tightening, since granular-demand investors exhibit high estimated cross-elasticities and rebalance towards higher yielding short-term Treasuries and reduce long-term bond positions accordingly, forcing arbitrageurs to absorb more risks and increase term premia. This is in sharp contrast to Vayanos and Vila (2021) and rationalizes the widely documented excess sensitivity of long rates to monetary policy shocks. Third, the effects of Fed purchases on bond yields are weak unless the Fed credibly commits to a persistent expansion of its balance sheet, thereby rationalizing a slow unwinding of the Fed's unconventional monetary policies.

As a first step, we create a rich novel dataset on Treasury holdings at the maturity bucket level across a wide range of institutions, such as insurance companies, mutual funds, broker-dealers, foreign investors, and the Fed, among others. Our dataset covers close to 80% of the total Treasury amount outstanding at any given point in time over the 2011Q4-2022Q4 period. We classify granular-demand investors as commercial banks, insurance companies and pension funds (ICPFs), money market funds (MMFs), mutual funds, foreign officials, and foreign private investors, while the arbitrageurs in our setting are the broker-dealers and hedge funds, mainly for two reasons: First, as shown by Hanson and Stein (2015) and Du et al. (2023b), broker-dealers and hedge funds behave as the opposite of yield-seeking investors, accommodating flows from the rest of the market. Second, broker-dealers and hedge funds generally have better access to a wider range of trading instruments and platforms, allowing them to deploy sophisticated arbitrage strategies[1].

Building on the instrument in Koijen and Yogo (2020) and Fang et al. (2022), we identify own and cross-maturity yield sensitivities at the investor level based on our panel dataset of Treasury holdings across maturity buckets and time. While we consider the Fed separately, its demand aligns with granular-demand investors, increasing long-term holdings when yields are high and reducing them during monetary tightenings. This is consistent with the policy objective of lowering long-term yields during QE and maintaining overall consistency during monetary tightenings. Moreover, across sectors, maturity preference is prevalent: MMFs hold a large amount of the total Treasury outstanding with maturities below one year, while insurance companies and pension funds have a greater demand for longer maturities.

Our empirical approach adds to the extant demand-based asset pricing literature in three ways. First, we incorporate cross elasticities in our demand estimation. Treasuries of different maturities are substitutes in providing liquidity to investors, and notably, our estimates point to significant cross substitution. Second, our demand functions are dynamic in nature, and capture dependence on macroeconomic conditions such as inflation, GDP gap, and total Debt/GDP. Third, we exploit information from both the cross section and time series, consistent with Gabaix and Koijen (2024) and Haddad et al. (2024a) that time-series information is needed to quantify the aggregate market elasticity.

We then embed the estimated demand functions into an equilibrium model of the Treasury market, with granular-demand investors, the Fed, and risk-averse arbitrageurs, in the spirit of Vayanos and Vila (2021). We add to this class of models, first, by allowing for cross-substitution by

[1]For example, broker-dealers and hedge funds take significantly negative net positions in Treasuries at certain periods, while we do not observe negative Treasury positions for other investors.

non-arbitrageurs.[2] Cross substitution generates a positive reaction of term premia to a monetary policy tightening, in contrast to the negative reaction in Vayanos and Vila (2021).[3] Second, we include a monetary-policy rule that depends on macroeconomic dynamics rather than treating the short-term interest rate as exogenous. Third, we incorporate latent outside assets held by the arbitrageurs, adding the aspect of realism that prices of risk are not entirely driven by arbitrageur's Treasury portfolios. We let the data inform us how outside-asset risk exposure interacts with Treasury pricing.

To get intuition, we first analyze a simplified version of the model that we can solve analytically, and then proceed to estimate the full model by minimizing fitting errors on the dynamics of the yield curve. Importantly, we identify the critical arbitrageur risk aversion parameter by matching hedge funds' and dealers' average Treasury holdings. Based on our estimates, we decompose Treasury yield fluctuations into driving forces. On the one hand, we find that short-term yields are mainly affected by monetary policy rates, but macroeconomic fluctuations and latent demand shocks become increasingly important at longer maturities, as arbitrageurs price in their exposure when absorbing these risks. On the other hand, we find that yield fluctuations are mostly driven by banks, foreign officials, and foreign mutual funds.

Quantitatively, we find that a $1 billion dollar extra latent demand of the overall Treasury market increases total Treasury valuation by $0.23 billion, indicating a multiplier of 0.23, in sharp contrast to the multiplier of 5 in the equity market (Gabaix and Koijen 2021) and 3.5 at the rating-level corporate bond market (Chaudhary et al. 2022). Intriguingly, in a counterfactual that excludes arbitrageurs from the Treasury market, the Treasury-market multiplier becomes larger than that of equity and corporate bonds. Therefore, the Treasury market is elastic in the presence of arbitrageurs who are readily stepping in to absorb demand imbalances. This arbitrage force is stronger at the shorter end of the maturity spectrum but becomes weaker at longer maturities due to larger risks, leading to a downward-sloping term structure of market elasticity.

We use our estimated model as a laboratory to examine conventional and unconventional monetary policies that involve interventions in the Treasury market. Regarding monetary policy shocks, our model predicts higher risk premia in response to a monetary tightening, consistent with the literature that empirically identifies term premium responses to monetary policy shocks (Hanson and Stein 2015; Bauer et al. 2023). In our setting, in view of a significant cross elasticity revealed by the data, non-arbitrageurs reduce holdings of long-term Treasuries and force arbitrageurs to hold more long-term Treasuries and charge a higher risk premium. Notably, when cross elasticities are

[2]See Chaudhary et al. (2022) for estimates of cross elasticities in the corporate bond market, and An and Huber (2024) for estimations of cross elasticities in the currency market.

[3]Kekre et al. (2024) introduces arbitrageur's wealth effect to Vayanos and Vila (2021) and also generates a positive reaction of the term premium to monetary policy tightening. We do not incorporate such a channel since the nonlinearity due to wealth effects causes numerical challenges that are beyond our paper.

excluded in our quantitative model, the outcome is reversed, highlighting that cross elasticities are crucial for resolving the puzzle in Vayanos and Vila (2021).

Regarding QE, our model suggests that if bond purchases are expected to be transient, they have little impact on Treasury yields. However, the response of Treasury yields becomes much more prominent if investors expect QE to represent a permanent demand shift of the Fed. Our model thus suggests that the impact of Fed purchases on bond yields is weak unless the Fed credibly commits to a persistent expansion of its balance sheet. This quantitative finding based on a granular analysis is consistent with the theoretical predictions in Greenwood et al. (2015), and gives guidance on the implementation of quantitative tightening (QT).

Related Literature

Our paper contributes to a growing literature that analyzes granular asset demand in fixed-income markets, building on the seminal work by Koijen and Yogo (2019). Specifically, Bretscher et al. (2024), Chaudhary et al. (2022), Siani (2022), and Darmouni et al. (2022) apply demand systems to corporate bond markets, Fang et al. (2022) to global government bond markets, Koijen et al. (2021) to the euro area government bond market, Jansen (2024) to the Dutch government bond market, and Jiang et al. (2022) to international bond and currency markets. Allen et al. (2020) analyze the demand of T-bill auctions and find that auction format matters for portfolio allocations. Doerr et al. (2023) and Stein and Wallen (2023) provide a granular analysis of the demand of money-market funds for near-money assets. Closest to ours, Eren et al. (2023) apply a demand system to the overall U.S. Treasury market using Flows of Funds data. Consistent with their study, we also find that investment funds and banks are more price elastic than ICPFs and foreign officials within the U.S. Treasury market. We contribute to this literature by using granular data on U.S. Treasury holdings by different institutions, including the Fed, and estimating cross-elasticities.

Furthermore, our paper is related to the preferred habitat view of the term structure of interest rates, e.g., Culbertson (1957), Modigliani and Sutch (1966), Guibaud et al. (2013), Greenwood and Vayanos (2014), and Vayanos and Vila (2021). Recent papers have started to build a tighter connection between data and theory. Droste et al. (2021) identify demand shocks from Treasury auctions and calibrate the model in a New Keynesian framework to study the impact of QE. Hanson et al. (2024) quantify the demand and supply shocks in the interest-rate swap market. Khetan et al. (2023) leverage more detailed data on interest-rate swaps and find a high level of segmentation. Bahaj et al. (2023) utilize transaction-level data on UK inflation swaps to quantify a model of inflation risks. Our contribution is to build and estimate a quantitative version of Vayanos and Vila (2021) that accounts for empirically estimated demand functions and actual arbitrageurs' Treasury holdings. Our results also contribute to a growing literature that quantifies the impact of QE (Krishnamurthy and Vissing-Jorgensen 2011; d'Amico et al. 2012; Swanson 2021; Selgrad 2023; Jiang and Sun 2024; Haddad et al. 2024b).

Our estimates of investor demand are consistent with the hypothesis of "yield-oriented investors" in Hanson and Stein (2015). We both theoretically and quantitatively confirm the rationale in Hanson and Stein (2015) that cross-substitution drives the positive term premium response to monetary policy tightening. This also addresses a broad literature that shows that risk premia overall rise with monetary policy tightening (Bernanke and Kuttner 2005; Gertler and Karadi 2015; Bekaert et al. 2013; Kekre et al. 2024).

Our paper is also related to the recent literature on the specialty of U.S. government debt. Krishnamurthy and Vissing-Jorgensen (2012) show that there is a downward-sloping aggregate demand curve for the convenience provided by Treasuries. The literature shows that Treasury convenience yield is closely connected to financial crises (Del Negro et al. 2017; Li 2024), monetary policy (Nagel 2016; Drechsler et al. 2018; Diamond and Van Tassel 2021), exchange rates (Jiang et al. 2021), inflation (Cieslak et al. 2024), pricing of stocks (Di Tella et al. 2023), hedging properties of Treasuries (Brunnermeier et al. 2024; Acharya and Laarits 2023), banking (Diamond 2020; Li et al. 2023; Krishnamurthy and Li 2023), financial regulation (Payne et al. 2022; Payne and Szőke 2024), and government debt valuation (Jiang et al. 2024b,a). We contribute to the above literature by unpacking the demand for Treasuries and sources of demand variations across investors.

Finally, arbitrageurs are important in our analysis, in the same spirit as a growing literature that focuses on financial intermediaries (He and Krishnamurthy 2013; Adrian et al. 2014; He et al. 2017; Du et al. 2018; Wallen 2020; Jermann 2020; Haddad and Muir 2021; Fang and Liu 2021; Kargar 2021; Favara et al. 2022; Du et al. 2023a; Diamond et al. 2024; An and Huber 2024). Haddad and Sraer (2020) show that banks' interest income gap significantly predicts Treasury returns. d'Avernas and Vandeweyer (2023) and d'Avernas et al. (2023) provide theories of how different types of intermediaries together with the central bank affect the Treasury market dynamics. Duffie et al. (2023) uses dealer-level data on Treasury holdings to show that dealer balance sheet utilization is important for Treasury pricing. Du et al. (2023b) quantitatively show that balance sheet frictions of intermediaries are important in pricing Treasuries. A key contribution relative to this literature is that we cover the majority of the Treasury market beyond intermediaries, and explicitly link the pricing kernel with intermediation activities in the Treasury market.

2. Data

One of our contributions is the construction of a novel granular dataset of U.S. Treasury holdings at the sector level, capturing the majority of the market. Indeed, our dataset covers all major institutional holders of U.S. Treasuries, including banks, the Federal Reserve, primary dealers and hedge funds, money market and mutual funds, ETFs, and foreign official and private investors. We next describe these data sources, the construction of our dataset, and stylized facts about U.S. Treasury holders.

2.1. Treasury Holdings Data Sources

The Flow of Funds (FoFs) is the standard data source for extant research regarding investors in U.S. Treasuries (e.g., Krishnamurthy and Vissing-Jorgensen (2007), Eren et al. (2023)). While the FoFs provides information about Treasury holdings at the investor sector level, the holdings are aggregated across all maturities and thus does not allow for a more granular analysis regarding, for example, the term structure of interest rates or the cross-substitution across maturities. To address these limitations, we compile a richer and more detailed dataset by leveraging multiple data sources to obtain U.S. Treasury holdings with the highest level of granularity available. Table 1 summarizes our primary data sources, with further details provided in Appendix A.1.

Table 1. **Data sources**

This table provides a summary of the different data sources that we use in this paper.

Investor Type	**Data Source**	**Frequency**	**Period**	**Detail**
Banks	CALL Reports	Quarterly	1976Q1-2022Q4	Maturity bucket
Fed	Federal Reserve	Weekly	2003W1-2022W52	Security
Primary Dealers	Federal Reserve	Weekly	1998W5-2022W52	Maturity bucket
Hedge Funds	Form PF SEC	Quarterly	2011Q4-2022Q4	Aggregate
Insurers and Pension Funds	eMAXX	Quarterly	2010Q1-2022Q4	Security
Money Market Funds	IMoneyNet	Monthly	2011M8-2022M12	Security
	Flow of Funds	Quarterly	1993Q1-2022Q4	Aggregate
Mutual Funds	Morningstar	Monthly/Quarterly	2000M1-2022M12	Security
ETFs	ETF Global	Daily/Monthly	2012M1-2022M12	Security
Foreign Official and Private	Public TIC	Quarterly	2011Q4-2022Q4	T-bill/non T-bill

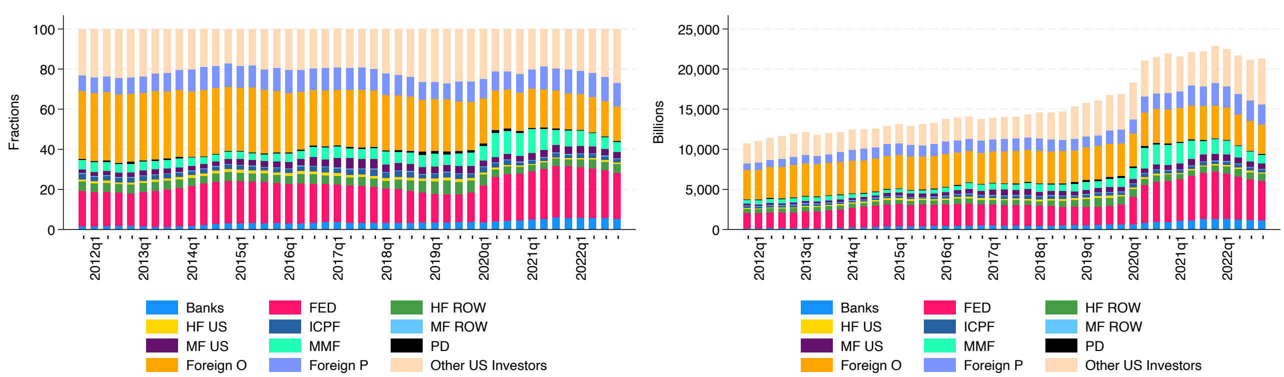
2.2. Data Aggregation

The reporting frequency and granularity differ across the data sources. In constructing our final dataset, we thus need to make choices to aggregate the data. In particular, to maintain consistency across datasets, we analyze data at the quarterly frequency from 2011Q4 to 2022Q4. We then group Treasuries into three maturity buckets. We denote remaining time to maturity as τ and divide Treasuries into three maturity buckets: $\tau < 1Y, 1Y \leq \tau < 5Y, \tau \geq 5Y$, and denote these maturity buckets as $m \in \{1,2,3\}$. The choice for these three maturity buckets is motivated by two considerations: First, this division reflects commonality across portfolio holdings data availability for different sectors. Second, as we show later, we need sufficient cross-sectional variation across maturity buckets to apply our instrument, and using more than three buckets complicates identification due to a reduction in variation across buckets. Finally, for stationarity, we scale all quantities by the ratio of potential GDP at the end of our sample period over the potential GDP at that particular quarter. We provide details on the data aggregation process in Appendix A.2. In our analysis, we also rely on macroeconomic dynamics, and we provide details on the macro variables in Appendix A.3.

2.3. Stylized Facts about Treasury Holdings

Figure 1. **Holdings of U.S. Treasuries by Investor Type**

Panel (a) plots the fraction of total U.S. Treasury outstanding (TAO) held by each investor type over time. Panel (b) plots the corresponding market values (billions). Sectors are U.S. banks (Banks), Federal Reserve (FED), hedge funds outside the U.S. (HF ROW), U.S. hedge funds (HF US), U.S. insurance companies and pension funds (ICPF), mutual funds outside the U.S. (MF ROW), U.S. mutual funds (MF US), U.S. money market funds (MMF US), U.S. and foreign primary dealers (PD), foreign official (Foreign O), and foreign private (Foreign P), and other U.S. investors (Other U.S. Investors). Other U.S. Investors is defined as the total U.S. Treasuries' outstanding minus the holdings of all the other sectors. We report market values and the quarterly sample period is 2011Q4-2022Q4.



(a) % Total Holdings of TAO

(b) Total Holdings (billions)

Figure 1 shows the dollar values and the fraction of total outstanding of U.S. Treasuries held by each investor type from 2011Q4 to 2022Q4. On average, our dataset contains 78% of the holders of U.S. Treasuries. Based on FoF data, the remaining 22% consists of U.S. households (11%), pension funds (5%), local governments (4%), and non-financial corporations (2%).

In Figure 2, we plot maturity-bucket level Treasury holdings of each investor type over the same period. The figure reveals several notable facts. First, MMFs are only active in maturity bucket 1 and hold between 10% to 35% of short-term Treasuries outstanding. Second, at the other end of the spectrum, ICPFs barely hold short-term Treasuries but hold around 5% of the Treasuries with maturities beyond 5 years. Third, the Fed holds substantially more of the intermediate and long-term bonds outstanding as opposed to short-term bonds. Fourth, mutual funds hold few short-term bonds, but are equally spread among maturity buckets 2 and 3. Fifth, only primary dealers and hedge funds have negative holdings in certain periods. Finally, foreign official holdings significantly declined, mainly in the short and medium-maturity buckets.

Table 2. **Marginal Holders U.S. Treasuries**

Panel (a) reports the marginal holders of U.S. Treasuries that we obtain by regressing percentage changes in holdings as a fraction of total outstanding (TAO) on the contemporaneous percentage changes in TAO. Panel (b) reports the average fraction of TAO held by each sector over our sample period. We report results for both the aggregate and by maturity bucket. Sectors are U.S. banks (Banks), Federal Reserve (FED), hedge funds outside the U.S. (HF ROW), U.S. hedge funds (HF US), U.S. insurance companies and pension funds (ICPF), mutual funds outside the U.S. (MF ROW), U.S. mutual funds (MF US), U.S. money market funds (MMF US), U.S. and foreign primary dealers (PD), foreign official (Foreign O), and foreign private (Foreign P), and other U.S. investors (Other U.S. Investors). The numbers are in percentage points and the quarterly sample period is from 2011Q4 to 2022Q4. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Panel (a): Marginal Holders (% of outstanding)

	Banks	**Fed**	**HF ROW**	**HF US**	**ICPF**	**MF ROW**	**MF US**	**MMF**	**PD**	**Other US**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Aggregate	4.4***	39.5***	-1.9	-0.5	1.1	0.1	1.9	30.2***	2.0*	7.4	6.3**	9.5***
$\tau < 1Y$	2.5***	8.0***	12.7***	3.1***	0.3	0.0	0.7*	51.0***	3.4***	8.0***	8.4***	1.8
$1Y \leq \tau < 5Y$	7.7**	34.5**	-24.7	-8.6*	-4.5	0.2	9.5*		0.9	33.0	36.1***	15.9
$\tau \geq 5Y$	1.7	38.2***	15.7*	4.0*	3.4*	0.5*	6.0**		3.5*	22.4*	-1.7	6.4*

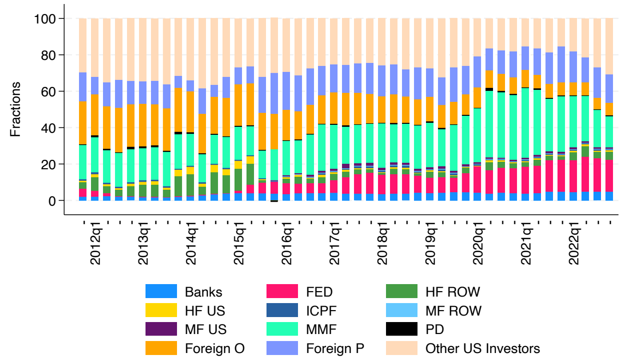
Panel (b): Average Holders (% of outstanding)

	Banks	**Fed**	**HF ROW**	**HF US**	**ICPF**	**MF ROW**	**MF US**	**MMF**	**PD**	**Other US**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Aggregate	3.4	19.6	4.4	1.2	2.3	0.3	3.2	5.9	0.7	21.7	27.2	10.2
$\tau < 1Y$	3.7	8.0	4.3	1.2	0.8	0.1	0.7	22.6	0.6	27.2	16.1	14.8
$1Y \leq \tau < 5Y$	3.3	18.9	5.6	1.4	2.4	0.3	3.4		0.6	12.1	46.5	5.5
$\tau \geq 5Y$	3.4	28.6	3.2	0.9	3.3	0.4	4.7		0.9	29.3	13.1	12.3

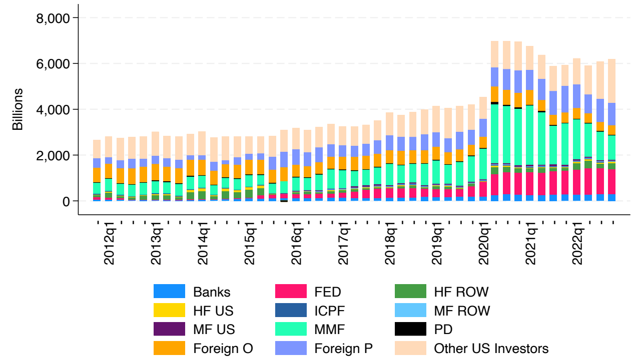
In Table 2, we further examine which investors absorb the additional debt when supply increases, or put differently, we investigate who are the marginal holders of U.S. Treasuries. To that end, similar to Fang et al. (2022), but focusing on maturity buckets, we decompose the marginal holders of Treasuries. For each maturity bucket m and sector ι, we regress changes in holdings on

Figure 2. **Holdings of U.S. Treasuries by Maturity Bucket**

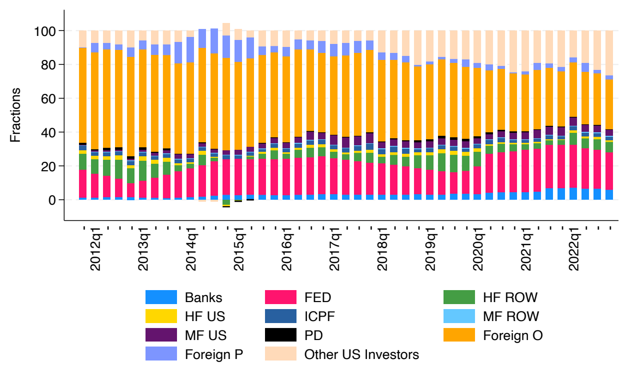
Left panels display the fraction of total U.S. Treasury outstanding (TAO) held by each investor type by maturity buckets. Right panels plot the corresponding market values (billions). Sectors are U.S. banks (Banks), Federal Reserve (FED), hedge funds outside the U.S. (HF ROW), U.S. hedge funds (HF US), U.S. insurance companies and pension funds (ICPF), mutual funds outside the U.S. (MF ROW), U.S. mutual funds (MF US), U.S. money market funds (MMF US), U.S. and foreign primary dealers (PD), foreign official (Foreign O), and foreign private (Foreign P), and other U.S. investors (Other U.S. Investors). Other U.S. Investors is defined as the total U.S. Treasuries' outstanding minus the holdings of all the other sectors. We report market values and sample period is quarterly from 2011Q4 to 2022Q4.



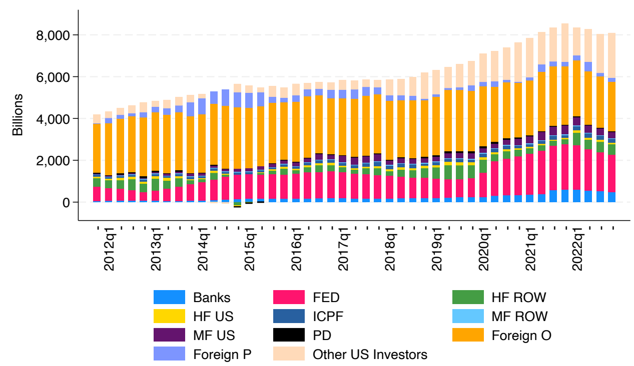
(a) $\tau < 1Y$: % Holdings as of TAO



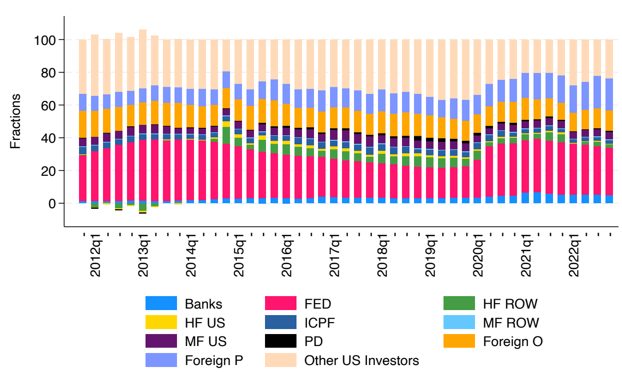
(b) $\tau < 1Y$: Total Holdings (billions)



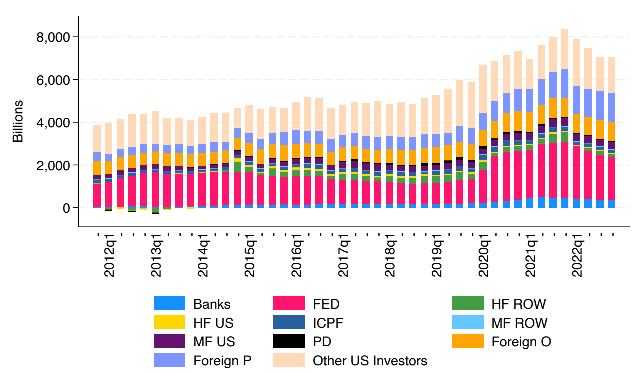
(c) $1 \leq \tau < 5Y$: % Holdings of TAO



(d) $1 \leq \tau < 5Y$: Total Holdings (billions)



(e) $\tau \geq 5Y$: % Holdings of TAO



(f) $\tau \geq 5Y$: Total Holdings (billions) changes in the total supply of debt:

$$\frac{Z_t^\iota(m) - Z_{t-1}^\iota(m)}{S_{t-1}(m)} = a^\iota(m) + b^\iota(m)\frac{S_t(m) - S_{t-1}(m)}{S_{t-1}(m)} + \varepsilon_t^\iota(m), \quad \forall \iota \tag{1}$$

where $Z_t^\iota(m)$ equals the total Treasury market value holdings of sector ι in maturity bucket m at time t (in billions) and $S_t(m)$ the total market value supply of Treasuries in maturity bucket m at time t (in billions). The accounting identity in Equation (1) implies that the sum across sectors must add up to the total so that $\sum_\iota \beta^\iota(m) = 100\%$ for all m. We also aggregate the holdings of each sector across maturities and estimate the total marginal contribution of each sector.

Panel (a) of Table 2 shows the results. Notably, the largest absorbers of U.S. Treasuries are the Fed and MMFs. In the aggregate, they absorb 39.5% and 30.2% of the U.S. debt, respectively, whereby the Fed mainly absorbs long-term Treasuries and MMFs short-term Treasuries. These percentages do not simply reflect proportional expansions to average holdings. Indeed, Panel (b) shows that, on average, the Fed and MMFs only hold 19.6% and 5.9% of the total debt outstanding, respectively. On the other hand, foreign officials only absorb 6.3% of additional U.S. debt, while their average holdings are substantially larger at 27.2%.

3. Empirical Results

Our data reveal substantial heterogeneity in Treasury holdings across sectors. In this section, we first set up a stylized model that can account for such heterogeneity to guide our empirical analysis. The model suggests a distinction of sectors into what we refer to as "granular-demand investors" and arbitrageurs, and the model nests both investor classes as special cases. We will embed them in a rich equilibrium model of the Treasury market in Section 4.

In the context of granular-demand investors, who we plausibly associate with banks, insurance companies and pension funds, mutual funds, money market funds, foreign investors, and other U.S. investors, the model suggests implementing a demand analysis much in the spirit of Koijen and Yogo (2019). This demand-based approach aptly and flexibly captures the rich heterogeneity of institutional patterns. For example, money market funds cannot hold any Treasuries beyond one-year maturity due to regulatory requirements. Pension funds and insurance companies naturally have preferences for long-maturity Treasuries because of long-duration liabilities (Sharpe and Tint 1990; Campbell and Viceira 2002). Banks have a special demand for Treasuries due to various liquidity-based regulations such as the liquidity coverage ratio.

In contrast, arbitrageurs, who we classify as broker-dealers and hedge funds following the literature (Hanson and Stein 2015; Du et al. 2023b), hold significant gross short positions for arbitrage purposes and are much less subject to regulatory or institutional constraints. In the absence of such non-pecuniary attributes, our model reduces to a portfolio optimization with limits to risk-bearing capacity. Intriguingly, according to the model, implementing the same demand-based regressions for arbitrageurs will typically generate misleading results. Therefore, rather than applying a reduced-form regression to uncover "demand functions" for arbitrageurs, we instead use arbitrageurs' positions to structurally discipline the parameter governing risk-bearing capacity in our full model (see Section 4).

3.1. Towards an Empirical Model of Treasury Demand

To guide our empirical analysis, we start with a simple model of investor demand for U.S. Treasuries. We index investor groups by ι and denote their portfolio holdings of maturity $\tau \in \{1, \cdots, N\}$ as $Z_t^\iota(\tau)$ and stack all maturities into a vector Z_t^ι. We denote the return on a Treasury with maturity τ as $R_{t+1}^{(\tau)}$, and the risk-free as r_t. We allow for flexible beliefs and denote the beliefs of sector ι as $\mathbb{E}^\iota$ in expectations, and $\mathbb{V}^\iota$ in covariances. For the sake of realism, we accommodate that investors' portfolios extend beyond Treasuries and denote the non-Treasury holdings as $\tilde{Z}_t^\iota$ and the associated returns as $\tilde{R}_{t+1}^\iota$.

We model the optimization problem of investor ι with wealth W_t^ι as

$$\begin{aligned} &\max_{Z_t^\iota, \tilde{Z}_t^\iota} \mathbb{E}_t^\iota[W_{t+1}^\iota] - \frac{\gamma^\iota}{2}\mathbb{V}_t^\iota(W_{t+1}^\iota) + \underbrace{V^\iota(Z_t^\iota)}_{\text{non-pecuniary}} \\ &W_{t+1}^\iota = W_t^\iota(1+r_t) + \underbrace{\sum_{\tau=1}^{N} Z_t^\iota(\tau)(R_{t+1}^{(\tau)} - r_t)}_{\text{Treasury returns}} + \underbrace{\tilde{Z}_t^\iota(\tilde{R}_{t+1}^\iota - r_t)}_{\text{outside portfolio return}}, \end{aligned} \tag{2}$$

where the objective function includes a non-pecuniary component that captures the special attributes of U.S. Treasuries, such as liquidity or safety, as in Krishnamurthy and Vissing-Jorgensen (2012). The non-pecuniary term can also reflect balance sheet costs of holding cash securities, such as the supplementary leverage regulation on banks. Similarly, the term can represent an inconvenience for certain Treasuries, such as that of short-term Treasuries for pension funds or insurance companies. For tractability, we assume that the derivative of V^ι is affine in the portfolio choice Z_t^ι,

$$\frac{\partial V^\iota(Z_t^\iota)}{\partial Z_t^\iota} = \bar{V}_0^\iota - \bar{V}^\iota Z_t^\iota. \tag{3}$$

In the budget equation, the "outside portfolio" can capture institutional features such as the long-duration liabilities of pension funds and insurance companies. Moreover, we allow for heterogeneous risk-aversion γ^ι. Institutions such as money-market funds can be approximated as agents with extremely high γ^ι and thus unable to bear the risks of long-term bonds.

Denote the aggregate states of the economy as the vector β_t, and the vector of Treasury yields as $y_t = \left(y_t^{(1)}, y_t^{(2)}, \cdots, y_t^{(N)}\right)'$. We allow for flexible beliefs about asset returns,

$$\mathbb{E}^\iota[R_{t+1}^{(\tau)} - r_t] = \mu^\iota(\tau) \cdot \beta_t + \phi^\iota(\tau) \cdot y_t, \tag{4}$$

where the dependence on yields could reflect the heuristic inference regarding how the yield curve predicts expected returns (Fama and Bliss 1987), or "reaching for yield" (Hanson and Stein 2015).

Solving for (2), we obtain the first-order condition for $Z_t^\iota(\tau)$,

$$\mu^\iota(\tau) \cdot \beta_t + \phi^\iota(\tau) \cdot y_t + V_\tau'(Z_t^\iota) = \gamma^\iota \left(\mathbb{V}^\iota(R_{t+1}^{(\tau)}, R_{t+1}) Z_t^\iota + \mathbb{V}^\iota(R_{t+1}^{(\tau)}, \tilde{R}_{t+1}^\iota) \tilde{Z}_t^\iota\right), \tag{5}$$

where we denote the vector of returns as $R_{t+1} = (R_{t+1}(1), R_{t+1}(2), \cdots, R_{t+1}(N))'$. Stacking all the values of $\tau \in \{1, \cdots, N\}$ in (5) and using the assumption in (3), we obtain:

$$Z_t^\iota = \left(\mathbb{V}^\iota(R_{t+1}, R_{t+1}) + \frac{1}{\gamma^\iota}\bar{V}^\iota\right)^{-1} \left(\frac{1}{\gamma^\iota}(\mu^\iota \beta_t + \phi^\iota y_t + \bar{V}_0^\iota) - \mathbb{V}^\iota(R_{t+1}, \tilde{R}_{t+1}^\iota)\tilde{Z}_t^\iota\right), \tag{6}$$

where we define the coefficient matrices $\mu^\iota = (\mu^\iota(1), \cdots, \mu^\iota(N))'$, $\phi^\iota = (\phi^\iota(1), \cdots, \phi^\iota(N))'$. We assume that the outside portfolio covariance term $\mathbb{V}^\iota(R_{t+1}, \tilde{R}_{t+1}^\iota)\tilde{Z}_t^\iota$ can be decomposed into linear dependence on aggregate states β_t plus "noise", in the same spirit as market microstructure models (Kyle 1985; De Long et al. 1990). The noise term can reflect sector-level idiosyncratic risks, such as pension-specific regulation changes, or erroneous stochastic beliefs as in De Long et al. (1990). We lump the noise term with the inverse of the matrix in (6) as a normally distributed vector u_t^ι, and can thus express the solution as:

$$Z_t^\iota = \hat{b}_0^\iota + B_1^\iota y_t + B_2^\iota \beta_t + u_t^\iota. \tag{7}$$

In the absence of restrictions on the belief parameters μ^ι and ϕ^ι, the model can thus span the entire space of affine functions over β_t and y_t plus a normally-distributed noise term. As a result, the solution of the optimization problem in (2) can be flexibly represented in the empirically tractable form (7). In fact, the model suggests that demand-based regressions using (7) are a promising approach to understanding investor behavior.

We note that the model allows for a relevant role for cross-elasticities in that $Z_t^\iota(\tau)$ may depend on $y_t(\tau')$ for $\tau' \neq \tau$, i.e., B_1^ι having non-zero off-diagonal elements. Intuitively, this allows investors to rebalance their portfolios towards higher-yielding maturity buckets. Indeed, certain investors such as insurance companies, mutual funds, and banks, are "yield-seeking" and are attentive to the relative yields among fixed-income securities (Becker and Ivashina 2015; Hanson and Stein 2015; Choi and Kronlund 2018). Importantly, within the context of the model, such cross-elasticities are not a reflection of arbitrage activity, but rather of yield-oriented beliefs as captured in (4). Indeed, in the equilibrium model that we set up in Section 4, equilibrium returns only depend on the aggregate states of the model, including the aggregate state vector β_t and the aggregation of demand imbalances, so that a rational arbitrageur infers all elements of ϕ^{ι} as zero and its demand does not directly respond to yields. Rather, arbitrageurs consistently infer Treasury prices according to the market equilibrium. Accordingly, a demand estimation based on (7) is inadequate for arbitrageurs, so we explicitly model them as enforcing arbitrage within the equilibrium model of Section 4.

The expression for Treasury holdings in (6) also makes it clear how the model accommodates that the optimal Treasury portfolio depends on other assets through the outside portfolio. Indeed, the portfolio depends on other assets if other assets' risk exposure comoves with Treasuries. To the extent that the state vector captures risks priced in other assets, innovations to these variables may transmit to Treasury demand fluctuations. For example, including the credit spread in the state vector allows for credit market shocks to be reflected in Treasury demand. In such a way, the model also captures substitution between corporate bonds and Treasuries.

Finally, we discuss the Federal Reserve's Treasury demand. Clearly, the Fed is not a profit-maximizing institution. The Fed's demand is driven by its policy decisions, for example, to reduce long-term interest rates through its QE program. We find it useful to describe the Fed's Treasury demand also in the form of (7), as a flexible way of capturing its policy objective.

3.2. Empirical Methodology

Inspired by our model specified in Section 3.1, we estimate granular-demand investor ι's demand for U.S. Treasuries according to (7). For practicality, we have two slight modifications. First, we group Treasuries into three maturity buckets, consistent with the empirical aggregation of Treasury holdings, and we denote a maturity bucket as $m \in \{1,2,3\}$. Second, we add bond characteristics in the demand as a control, although those bond characteristics will not be directly modeled. In particular, we implement the following regression:

$$Z_t^{\iota}(m) = \theta_0^{\iota} + b_1^{\iota} y_t(m) + b_2^{\iota} y_t(-m) + (b_3^{\iota})' \mathbf{x}_t(m) + (b_4^{\iota})' \mathbf{Macro}_t + u_t^{\iota}(m), \tag{8}$$

where $y_t(m)$ is the yield for maturity bucket m, $y_t(-m)$ equals the weighted-average yield of the other maturity buckets, $\mathbf{x}_t(m)$ is a vector of value-weighted bond characteristics for maturity bucket m: coupon, maturity bucket fixed effects, bid-ask spread, and $\mathbf{Macro}_t$ equals a set of macro variables, including GDP gap, debt/GDP, core inflation, and credit spread. We residualize the coupon and the bid-ask spread with respect to the maturity fixed effects to address multicollinearity issues and ensure that maturity preferences are not confounded with either of these two characteristics. This residualization also makes sure that demand loadings on bond features do not systematically drive demand. We provide summary statistics for this set of variables in Table A1 and the correlation table in Table A2.

We focus on the dollar value of holdings rather than portfolio weights, because dynamics in total portfolio demand are crucial for the term structure of interest rates – modeling only portfolio weights is not sufficient. For example, inflows into money-market mutual funds will cause extra demand for short-maturity Treasuries, but their below-one-year Treasury portfolio weight is 100% and does not capture such fluctuations. Moreover, we use market values rather than face values because our model in Section 4 indicates that market values are the relevant signals for investors, so our specification in (8) has a direct mapping to our dynamic quantitative model.[4]

Different from Koijen and Yogo (2019), but following our model in Section 4.2, we include "other yield" to capture cross substitution across the maturity structure. We find that if we only include own yield but not other yield in our analysis, we would uncover a coefficient on own yield that is downward biased. The reason is that own yield and other yield are correlated, while demand increases if own yield goes up, but decreases when other yield goes up. Hence, when not accounting for other yield, b_1^l picks up both the positive and negative effect, leading to a coefficient that is biased towards zero.[5]

In our specification, we assume that the macro variables are exogenous to investors, as in Fang et al. (2022) and Koijen and Yogo (2020). That is, investor (latent) demand does not contemporaneously affect macro variables. In addition, we also assume that bond characteristics, except for yields, are exogenous to latent demand. This assumption is the basis for the construction of our instrumental variables.

We could estimate the demand system specified in Equation (8) by GMM if it satisfies the moment condition:

$$\mathbb{E}[u_t^l(m)|y_t(m), y_t(-m), \mathbf{x}_t(m), \mathbf{Macro}_t] = 0. \tag{9}$$

The concern with this moment condition is that the error term may not be orthogonal to yields. For instance, if sectors have a large demand for Treasuries unrelated to bond characteristics or macro variables, then this latent demand is likely to also suppress the yield. As such, we need an instrument for bond yields.

Instrument. We build on the instrument used in Koijen and Yogo (2020) and Fang et al. (2022)

[4]However, our empirical estimates are similar if we replace all market values with face values of Treasury holdings.

[5]Table A7 shows that the coefficient on own yield is attenuated closer to zero when not accounting for other yield.

and use the following three step procedure. First, we estimate demand for each investor type as in Equation (8), but excluding the yield. We then in a second step extract the predicted values $\hat{Z}_t^\iota(m)$. We also follow step (1) and (2) for the nominal value of Treasury supply at each maturity bucket, whereby we regress it on the FFR and macro variables, consistent with the specification of our US Treasury model introduced in Section 4. In a third step, we impose market clearing and extract the imposed yield that sets the implied demand equal to the implied market value of supply:

$$\sum_\iota \hat{Z}_t^\iota(m) = \frac{\hat{S}_t(m)}{(1+\tilde{y}_t(m))^{\tau(m)}}, \tag{10}$$

where $\hat{S}_t(m)$ is the predicted nominal value of supply for maturity bucket m, and $\tau(m)$ the corresponding maturity. We take $\tau(m)$ as the average bond duration for maturity bucket m. We then extract pseudo yield $\tilde{y}_t(m)$ that clears the market at each point in time t and use it as an instrument for the actual yield $y_t(m)$: $I_t(m) = \tilde{y}_t(m)$. We apply the same logic to the value-weighted yield of the other buckets, for which the instrument equals the value-weighted pseudo yield for the other maturity buckets: $I_t(-m) = \tilde{y}_t(-m)$.

In summary, the idea behind the instrument is that the pseudo yield isolates the component of the yield that is driven by bond characteristics and macro variables. This instrument satisfies the exclusion restriction under the identifying assumption that bond characteristics and macro variables are exogenous to investor latent demand, as we assume throughout. Moreover, we rely on the nonlinear relationship between pseudo yields and bond characteristics as well as macro variables. The reason is that in the case of a linear relationship, the pseudo yields would be perfectly collinear with bond characteristics and macro variables in the second stage (8). This assumption of nonlinearity is satisfied because of the convexity effect of compounding interest as in (10), and empirically relevant because of a strong first stage. In Appendix B, we give a stylized example to further clarify these arguments.[6] More formally, we can weaken moment condition (9) to:

$$\mathbb{E}[u_t^\iota(m)|I_t(m), I_t(-m), \mathbf{x}_t(m), \mathbf{Macro}_t] = 0. \tag{11}$$

The first stage estimates of the demand system are summarized in Table A4. The corresponding Kleibergen-Paap statistic to test for weak instruments is 11.13, above the threshold of 10 for rejecting weak instruments (Stock and Yogo 2005).[7]

[6] We thank Quentin Vandeweyer for discussing our paper and providing these stylized examples.

[7] The first stage is the same for all sectors, except MMFs, for which the statistic equals 4.27. The reason is that MMFs do not invest in maturities beyond 1 year, so the instrument cannot exploit heterogeneity across maturities and we should interpret their result with care.

3.3. Demand Functions of Granular-Demand Investors

Table 3. **Demand System Results - IV**

This table shows the IV estimates of our demand system specified in Equation (8). The dependent variable is the market value ($ billion) of U.S. Treasuries held by sector ι in maturity bucket m at time t, adjusted by the ratio of GDP potential at the end of our sample period over the value at current quarter. The endogenous variables are: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m. We instrument own and other yield using pseudo yields specified in Section 3.2. Additional variables include Coupon Rate, Bid-Ask Spread, maturity bucket indicators, Credit Spread, Debt/GDP, Credit Spread, GDP Gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to maturity fixed effects. For explanations of sector abbreviations, refer to the notes of Table 2. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Banks	**ICPF**	**MF ROW**	**MF US**	**MMF**	**Other U.S.**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	63.850**	3.833	6.934*	137.258***	436.596*	172.272	-33.849	32.697
	[26.277]	[11.461]	[3.716]	[47.699]	[236.128]	[199.313]	[115.257]	[94.669]
$y_t(-m)$	-72.167**	-1.247	-3.663	-152.400***	-611.375*	-17.813	-94.278	-42.745
	[28.676]	[13.518]	[4.025]	[53.939]	[367.663]	[257.566]	[154.463]	[125.330]
Coupon Rate	-148.638***	3.053	-4.817	-137.838**	55.752	182.530	-480.953**	-315.103*
	[35.111]	[18.189]	[4.853]	[61.177]	[545.299]	[319.718]	[191.041]	[180.040]
Bid-Ask Spread	7.730	18.664***	3.059**	12.692	136.693	109.723	-102.377**	-65.497
	[7.921]	[4.472]	[1.206]	[16.243]	[140.086]	[76.916]	[46.128]	[56.216]
$\mathbb{1}\{1Y \leq \tau < 5\}$	56.159***	148.746***	12.952***	189.591***		-427.082***	2923.108***	-346.709***
	[15.057]	[4.427]	[2.132]	[26.569]		[122.524]	[91.434]	[83.651]
$\mathbb{1}\{\tau \geq 5\}$	-68.055	182.999***	9.623	36.298		451.302	148.771	44.390
	[47.867]	[20.885]	[7.022]	[91.367]		[413.365]	[226.244]	[186.195]
Credit Spread	15.144	-12.095	0.784	-37.701	-512.281**	286.080	95.977	-30.513
	[20.288]	[13.631]	[2.489]	[40.149]	[202.541]	[185.470]	[90.280]	[130.369]
Debt/GDP	648.082***	-7.771	41.743***	-18.509	5592.173***	2142.833**	-1806.284***	651.782
	[79.844]	[48.167]	[10.595]	[135.214]	[1277.801]	[919.753]	[572.490]	[536.095]
GDP Gap	11.000***	-4.501**	1.424***	12.121**	-75.617***	-9.814	-10.512	8.537
	[3.708]	[1.885]	[0.460]	[5.146]	[21.914]	[29.890]	[17.207]	[17.759]
Core Inflation	16.814**	-0.440	-2.254***	-3.223	59.070	-13.744	-74.315*	3.339
	[6.870]	[3.300]	[0.854]	[11.134]	[95.780]	[49.601]	[40.866]	[33.921]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap Statistic (*first stage*)	11.13	11.13	11.13	11.13	4.27	11.13	11.13	11.13

Table 3 shows the results using the IV methodology outlined in the previous section.[8] We find that all investors have downward sloping demand curves, except for the foreign official sector, although its coefficient is insignificant and small in economic magnitude. That is, granular-demand investors demand more U.S. Treasuries of maturity bucket m when the yield (price) is high (low). In addition, investors load negatively on the yield of other maturity buckets, meaning that their demand for maturity bucket m decreases when the yields of other buckets are high. Generally, we find that other elasticity is slightly higher than own elasticity, but the order of magnitude

[8] The results of the OLS estimates are in Appendix A3.

between the coefficients is similar. This is consistent with the findings in Chaudhary et al. (2022). They find a ratio between cross-elasticity and own-elasticity of close to 1 at the CUSIP level and for portfolios at the rating $\times$ quarter-to-maturity level for corporate bonds, the latter aggregation closely resembling ours. This ratio implies that own and cross-elasticity have the same magnitude, but with opposite sign. Additionally, Table A5 shows that the coefficients on own and other yield are qualitatively similar when not controlling for macro variables. Finally, Table A6 reveals that our results are robust to a specification where pseudo yields are inferred solely from coupon, maturity, GDP gap, and Debt/GDP, while omitting bid-ask spread, credit spread, and core inflation. This finding underscores the robustness of our IV to excluding bond characteristics and macro variables that are potentially endogenous and thereby violating the exclusion restriction of the IV.

Moving to the bond characteristics, ICPFs, and foreign MFs have a higher demand for Treasuries when the bid-ask spreads are high; that is, when Treasuries are less liquid,[9] while foreign official investors reduce their demand at that time. This suggests heterogeneous liquidity preferences across investors. Furthermore, ICPFs have a high demand for long-term Treasuries, while foreign officials have a high preference for medium-term bonds, highlighting the importance of heterogeneity in maturity preferences across investors. By means of the investment mandates of MMFs, they only operate in the shortest maturity bucket. Moving to the macro variables, we find that banks, MFs U.S., and MFs ROW increase their demand for Treasuries when the GDP gap is high, while MMFs and ICPFs reduce their demand. Foreign investors reduce their demand for Treasuries when core inflation is high, while banks increase their demand. Finally, we find that Banks, MF ROW, MMFs, and Other U.S. Investors increase demand for Treasuries when debt/GDP is high, while foreign officials heavily reduce their demand in response to a rise in the U.S. debt burden, consistent with the trends described in Table 2.

By looking at holdings in market values, Table 3 does not allow a comparison among the price elasticities across investor types. As such, we scale the holdings for each sector by the average holding of that sector, across buckets and time. Figure 3a plots the coefficients on own and other yield for each investor type. Interestingly, mutual funds and MMFs appear to be most price elastic, followed by banks[10]. ICPFs and foreign official investors are the least price elastic. Although our estimated own elasticities appear large for certain sectors, these sectors tend to be small relative to the total amount outstanding (e.g., the most elastic U.S. mutual fund sector holds only 4.7% of the market). In addition, we also uncover significant cross elasticity, which reduces the equilibrium

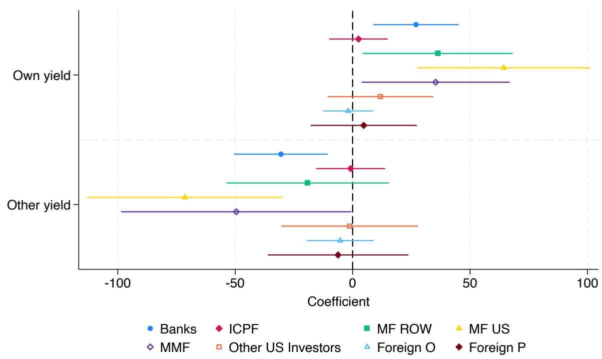
[9] Bretscher et al. (2024) find that ICPFs' corporate bond demand has a positive loading on the bid-ask spread. For instance, ICPFs may prefer illiquid assets to keep their solvency positions appearing more stable. However, they find that MFs prefer liquid bonds in the *cross*-section. This finding does not necessarily contradict our result that picks up a preference for liquidity in the *time*-series by removing maturity fixed effects. Our finding should thus be interpreted as foreign MFs having a higher demand for Treasuries when market liquidity declines.

[10] Eren et al. (2023) also find that banks and investment funds are more price elastic.

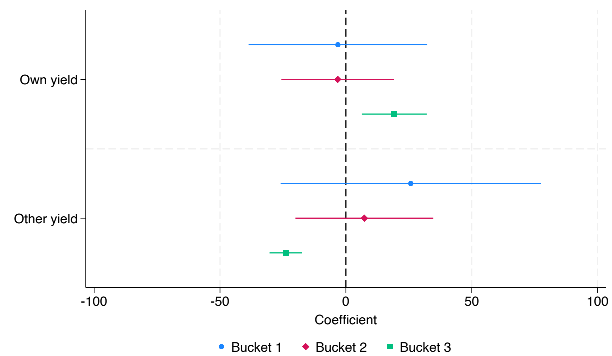
market elasticity because investors tend to cross substitute rather than absorb quantities in net. Importantly, as we illustrate in Section 5.3, the equilibrium market elasticity in the presence of arbitrageurs is significantly different from the value-weighted elasticity of granular-demand investors.

Figure 3. **Yield Elasticities by Investor Type**

Panel (a) plots the coefficients on own and other yield for different granular-demand investors, scaling holdings for each sector by the average holding across time and maturity buckets for that sector to allow for comparison of coefficients across investor types. A coefficient of 50 implies that for a one percentage point increase in yield, the demand goes up by 50%. For explanations of sector abbreviations, refer to the notes of Table 2. Panel (b) shows the yield sensitivities for the Federal Reserve by maturity bucket, whereby we scale the holdings in each bucket by the time-series average holding in that bucket. We use market values scaled by GDP potential and the quarterly sample period is 2011Q4-2022Q4.



(a) Granular-Demand Investors



(b) Federal Reserve

3.4. Demand Functions of the Fed

For the Fed, we estimate their demand curves separately for each maturity bucket. The reason is that the Fed implements unconventional monetary policies mainly via long-term Treasuries. We should, therefore, expect the Fed to respond to yields for its long-term Treasury holdings, but not for its short- and medium-term Treasury holdings. In contrast, we do not have a strong prior that granular-demand investors have significantly different responses to yields across maturities.

Table 4 summarizes the results. Interestingly, in the long-term bucket, the Fed behaves similarly to granular-demand investors: the Fed increases its long-term Treasury holding when the long-term yield is high, while reduces its holding when the short-term yield is high. This revealed behavior is consistent with the Fed's policy goals. Specifically, QE aims to lower long-term yields, prompting the Fed to expand its balance sheet when long-term yields rise. Moreover, the Fed aligns its conventional and unconventional monetary policies by simultaneously increasing the short-

Table 4. **Demand System Results - Fed**

This table shows the IV estimates of our demand system specified in Equation (8) for the Fed. The dependent variable is the market value ($ billion) of U.S. Treasuries held by the Fed in each maturity bucket m at time t, adjusted by the ratio of GDP potential at the end of our sample period over the value at current quarter. The endogenous variables are: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m. We instrument own and other yield using pseudo yields specified in Section 3.2. Additional variables include Coupon Rate, Bid-Ask Spread, Credit Spread, Debt/GDP, Credit Spread, GDP Gap, and Core Inflation. Column (1) shows the results for $\tau < 1Y$, Column (2) for $1Y \leq \tau < 5$, and Column (3) for $\tau \geq 5$. The quarterly sample period is from 2011Q4 to 2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	$\mathbb{1}\{\tau < 1Y\}$	$\mathbb{1}\{1Y \leq \tau < 5Y\}$	$\mathbb{1}\{\tau \geq 5Y\}$
	(1)	(2)	(3)
$y_t(m)$	-14.733	-49.318	385.678**
	[100.514]	[208.133]	[157.594]
$y_t(-m)$	120.213	112.178	-478.703***
	[146.510]	[254.479]	[79.222]
Coupon Rate	-35.947	-2557.515***	246.631
	[186.162]	[256.424]	[248.683]
Bid-Ask Spread	203.700***	102.781	-177.449***
	[59.059]	[75.504]	[65.788]
Credit Spread	24.368	206.475	-231.120*
	[82.169]	[138.053]	[137.150]
Debt/GDP	3643.632***	429.732	4649.721***
	[398.422]	[564.090]	[1020.458]
GDP Gap	-6.980	-16.387	-49.862**
	[7.078]	[14.768]	[22.644]
Core Inflation	46.812	-61.166	155.350***
	[40.232]	[40.724]	[29.301]
Observations	45	45	45
Kleibergen-Paap Statistic (*first stage*)	4.27	9.58	14.67

term interest rate and reducing long-term Treasury holdings, resulting in a negative cross-elasticity in its long-term Treasury demand. Despite the significant price elasticity in long-term Treasury holdings, the Fed's medium and short-term Treasury holdings are not responsive to Treasury yields, consistent with the focus of QE/QT on long-term securities.

Additionally, we find that the Fed reduces demand for long-term bonds when the GDP gap is high, indicating less need to support the economy via QE when the economy is doing well. Moreover, the Fed significantly expands its Treasury holdings in all maturity buckets when Debt/GDP is higher, indicating prominent fiscal accommodations by the Fed.

Figure 3b shows the relative yield sensitivities of the Fed across maturity buckets. Clearly, the Fed's short- and medium-term Treasury holdings are price inelastic, while its long-term Treasury holdings exhibit significant price elasticity, comparable in magnitude to that of banks.

4. An Equilibrium Model of the Treasury Market

The previous section revealed three key findings. First, granular-demand investors and the Fed have downward-sloping demand curves. Second, their total demand exhibits strong cross substitution. Third, the Fed's demand for short- and medium-maturity is not significantly affected by Treasury yields, while its long-maturity Treasury demand increases with long-maturity Treasury yield but decreases with short-term yields, consistent with its policy objectives.

Building on these empirical results, this section develops a model where strategic arbitrageurs interact with granular-demand investors and the Fed in the Treasury market, in the spirit of Vayanos and Vila (2021). We capture Treasury demand of granular-demand investors and the Fed using demand functions, motivated by the model in Section 3.1, while we explicitly model arbitrageurs using a stripped-off version of the model in Section 3.1 that reflects pure arbitrage. After we set up the model, we provide a simplified version that allows us to derive analytical results to obtain intuition regarding the fundamental mechanisms. Finally, we estimate the full model from the data.

To capture the rich economics in the Treasury market, we deviate from Vayanos and Vila (2021) mainly in three aspects. First, we incorporate cross-substitution in investor demand, a critical feature that generates realistic term premium responses to monetary policy shocks. Second, we include a monetary-policy rule that depends on macroeconomic dynamics, rather than treating the short-term interest rate as exogenous, allowing us to quantify the magnitude of monetary policy shocks. Third, we account for latent outside assets held by arbitrageurs, adding the element of realism that prices of risks are not entirely driven by arbitrageurs' Treasury portfolios.

4.1. Model Setup

The model is discrete-time and infinite-horizon. There are four types of agents in the economy: a competitive arbitrageur sector, the Fed, a set of granular-demand investors, and the government. We only explicitly model the strategic decisions by arbitrageurs while we capture the behavior of other agents by policy rules that map directly to our estimated demand functions. We model the Treasury market explicitly by market clearing. Economic dynamics are driven by macroeconomic shocks, monetary policy shocks, and demand shocks.

Consider zero-coupon bonds of maturities $\tau \in \{1, 2, \cdots, N\}$ that all pay a face value of 1 at maturity. Denote by $P_t^{(\tau)}$ and $y_t^{(\tau)}$, respectively the time-t price and yield of the bond with maturity τ. We use "prime" to denote the transpose of vectors and matrices, and all vectors are column vectors. Define the log price vector as

$$p_t = \left(\log(P_t^{(1)}), \log(P_t^{(2)}), \cdots, \log(P_t^{(N)})\right)'. \tag{12}$$

For simplicity, we denote the yield of a one-period bond as r_t, defined as $r_t = -\log(P_t^{(1)})$.

We consider r_t as directly controlled by monetary policy. All other bond yields and prices are endogenously determined in equilibrium. Denote the total return from holding a Treasury of maturity τ as

$$R_{t+1}^{(\tau)} = \frac{P_{t+1}^{(\tau-1)} - P_t^{(\tau)}}{P_t^{(\tau)}}. \tag{13}$$

Accordingly, the total return of one-period Treasury is $R_{t+1} = R_{t+1}^{(1)} = \exp(r_t) - 1 \approx r_t$.

The dynamics of the economy is driven by a K-dimensional vector of macro factors,

$$\beta_t = (\beta_{1,t}, \beta_{2,t}, \cdots, \beta_{K,t})', \tag{14}$$

which follows a VAR(1) process,

$$\beta_{t+1} = \bar{\beta} + \Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}. \tag{15}$$

In the above expression, ε_{t+1} is a K-dimensional vector that follows an i.i.d. standard normal distribution, and Φ is a matrix that determines the long-run dynamics.

We interpret the vector β_t as macro states of the economy that drive the monetary policy stance in equilibrium and also expectations regarding future economic states. Monetary policy depends on contemporaneous economic variables,

$$r_{t+1} = \bar{r} + \phi_r'(\beta_{t+1} - \bar{\beta}) + \rho_r r_t + \sigma_r \varepsilon_{t+1}^r, \tag{16}$$

where ρ_r captures monetary policy inertia, as discussed, for example, in Clarida et al. (2000), and ε_{t+1}^r reflects monetary policy shocks. We assume that monetary policy shocks ε_{t+1}^r are independent from ε_{t+1}, i.e., monetary policy shocks are not subsumed by public information on macro dynamics.

Denote the set of institutions excluding arbitrageurs as $\mathscr{I}$. Sector-ι's ($\iota \in \mathscr{I}$) demand for bonds with maturity $\tau \in \{1, \cdots, N\}$ follows the functional form in (7) of Section 3.1,

$$Z_t^\iota(\tau) = \theta_0^\iota(\tau) - \alpha^\iota(\tau)' p_t - \theta^\iota(\tau)'\beta_t + u_t^\iota(\tau), \tag{17}$$

where we use log prices instead of yields for consistency with Vayanos and Vila (2021), but the two are equivalent. The parameter vector $\alpha^\iota(\tau)$ loads on the whole log-price vector p_t and reflects not only the demand sensitivity to the price of maturity τ itself but also sensitivities to prices of other maturities $\tau' \neq \tau$, capturing cross elasticities. We lump the demand for bonds from granular-demand investors and the Fed together, and refer to it either as the "non-arbitrageur demand", defined as

$$Z_t(\tau) = \sum_{\iota \in \mathscr{I}} Z_t^{\iota}(\tau). \tag{18}$$

Accordingly, we define $\theta_0(\tau)$, $\alpha(\tau)$, $\theta(\tau)$, and $u_t(\tau)$ as the sums of corresponding values from each sector $\iota \in \mathscr{I}$. We use column vector forms to express our setup in a more convenient and compact notation. In vector form, we can write (18) as

$$Z_t = \theta_0 - \alpha p_t - \theta \beta_t + u_t, \tag{19}$$

where $\theta_0 = (\theta_0(1), \theta_0(2), \cdots, \theta_0(N))'$ is an N-dimensional vector, $\alpha = (\alpha(1), \alpha(2), \cdots, \alpha(N))'$ an $N \times K$ matrix, and $\theta = (\theta(1), \theta(2), \cdots, \theta(N))'$ an $N \times K$ matrix. The unobservable, maturity-specific latent demand shock, $u_t = (u_t(1), u_t(2), \cdots, u_t(N))'$, reflects the non-systematic component of demand shocks. We assume that u_t is i.i.d., with mean zero and covariance matrix Σ^u.

On the supply side, we assume that the government issues Treasuries depending on macroeconomic conditions and the monetary policy rate. Accordingly, we specify the aggregate value of government bond supply, or, more precisely, the supply to the public market, i.e. marketable Treasury securities, as

$$S_t(\tau) = \bar{S}(\tau) + \zeta(\tau)'\beta_t + \zeta_r(\tau) r_t, \tag{20}$$

or, in vector form, as

$$S_t = \bar{S} + \zeta \beta_t + \zeta_r r_t, \tag{21}$$

where $\zeta = (\zeta(1), \zeta(2), \cdots, \zeta(N))'$ is an $N \times K$ matrix. We can interpret Equation (21) as coming from a budget equation of the government, where Treasury supply needs to adjust to the need for government financing, which in turn is driven by macroeconomic conditions and the prevailing interest rate. Therefore, our model implicitly captures fiscal dynamics of the government.

We model a representative arbitrageur as a special version of the generic problem in (2). In particular, we shut off the non-pecuniary term to reflect pure arbitrage, and assume rational expectations. We denote arbitrageur positions in Treasuries of maturity τ as $X_t(\tau)$, and the outside asset position as $\tilde{X}_t$.

We view modeling outside assets as adding an important element of realism to models in the spirit of Vayanos and Vila (2021), since arbitrageurs' risk-bearing capacity in the Treasury market plausibly depends on their positions in other markets. We will estimate arbitrageurs' outside asset risk exposure with a revealed preference approach.

Accordingly, arbitrageurs' wealth dynamics evolve as

$$W_{t+1} = W_t(1+R_t) + \sum_{\tau=2}^{N} X_t(\tau)(R_{t+1}^{(\tau)} - R_t) + \tilde{X}_t(\tilde{R}_{t+1} - R_t). \tag{22}$$

We assume that the return of the outside asset is normally distributed and depends on the state of the economy, in that

$$\tilde{R}_{t+1} = \tilde{\phi}'\beta_t + \tilde{\phi}_r r_t + \tilde{\sigma}'\varepsilon_{t+1} + \tilde{\sigma}_r'\varepsilon_{t+1}^r, \tag{23}$$

where $\tilde{\phi}$ is a $K \times 1$ vector, $\tilde{\phi}_r$ is a scalar, $\tilde{\sigma}$ is a $K \times 1$ vector, and $\tilde{\sigma}_r'$ is a scalar.

The objective of arbitrageurs is to maximize a mean-variance utility,

$$\max_{\{X_t^\tau\}_\tau, \tilde{X}_t} E_t[W_{t+1}] - \frac{\gamma}{2} Var_t(W_{t+1}), \tag{24}$$

subject to the wealth dynamics specified in (22).

Finally, for each maturity τ, there is a market-clearing condition,

$$Z_t(\tau) + X_t(\tau) = S_t(\tau). \tag{25}$$

We conjecture that there is an affine equilibrium in the form of

$$p_t = A\beta_t + A_r r_t + A_u u_t + C, \tag{26}$$

where $A = (A(1), A(2), \cdots, A(N))'$ is an $N \times K$ matrix, $A_r = (A_r(1), A_r(2), \cdots, A_r(N))'$ is an $N \times 1$ vector, $A_u = (A_u(1), A_u(2), \cdots, A_u(N))'$ is an $N \times N$ matrix, $C = (C(1), C(2), \cdots, C(N))'$ is an $N \times 1$ vector.

4.2. A Simplified Version with Analytical Solutions

To gain intuition regarding the mechanisms at play in the model, we analyze a simplified version of the model in this subsection. In particular, we assume $N = 2$, so that there are only two maturities for consideration that represent "short" and "long". We assume that the granular-demand investor demand has a simple structure with the matrix capturing the demand response to price (see Equation (19)), given as

$$\alpha = \begin{pmatrix} a & -b/2 \\ -b & a/2 \end{pmatrix}. \tag{27}$$

Since $p(2) = -2y(2)$, the long-term and short-term Treasury demand responses to long-term yield are a and $-b$, so the matrix of demand responses to yields is symmetric. We assume that both

a and b are positive, so that Treasury demand increases in its own yield, but decreases in the other-maturity yield, which is the case for the aggregate granular-demand investor demand as we uncovered in Section 3.

We set $K = 1$ so that the macro factor β_t is only one dimensional, and we interpret this single-dimension factor as "supply" factor that drives the total debt supply. We also set $\phi_r = 0$ so that monetary policy process does not depend on the macro factor, and $\bar{r} = 0$ for simplicity. We further set $\zeta_r = 0$ so that debt supply is

$$S_t^{(\tau)} = \bar{S}^{(\tau)} + \zeta(\tau)\beta_t, \tag{28}$$

for $\tau = \{1, 2\}$. We impose a regularity condition that $\zeta(2) > -\theta(2)$ so that any supply expansion does not automatically get overshadowed by the expansion of demand in response to such supply expansion. Finally, for simplicity, we shut off all outside portfolio exposure by setting $\tilde{X}_t = 0$.

Using the first order conditions and the market clearing condition, we find the following unique equilibrium solution for log prices,

$$\begin{aligned} p_t^{(1)} &= -r_t, \\ p_t^{(2)} &= -\frac{1+\rho_r+\gamma\sigma_r^2 b}{1+\frac{a}{2}\gamma\sigma_r^2} r_t - \frac{\gamma\sigma_r^2(\zeta(2)+\theta(2))}{1+\frac{a}{2}\gamma\sigma_r^2}\beta_t + \frac{\gamma\sigma_r^2}{1+\frac{a}{2}\gamma\sigma_r^2} u_t(2) + \frac{\frac{1}{2}-\gamma\bar{S}^{(2)}+\gamma\theta_0(2)}{\frac{1}{\sigma_r^2}+\frac{a}{2}\gamma}, \end{aligned} \tag{29}$$

where the first equation reflects that the short rate is given by the monetary policy stance, and the second equation comes from arbitrageurs accommodating the imbalance between Treasury supply and non-arbitrageur demand subject to risk aversion. Detailed derivations are provided in Appendix D.2, which also contains proofs of all the following propositions in this section.

Using Equation (29), we summarize the drivers of Treasury price variation in the following proposition.

Proposition 1 (Decomposition of Treasury Pricing)**.** *Monetary policy rate r_t plays a dominant role for short-maturity Treasuries, while macro shocks and latent demand shocks become more important for long-maturity Treasuries.*

Proposition 1 is an intuitive result by simply observing Equation (29). The more general message is that the relative importance of macro factors and latent demand increases as the maturity of Treasuries increase, because the arbitrage force gets weaker at longer maturities. Furthermore, Proposition 1 also implies that a demand shock, either latent demand or permanent demand, has a larger price impact if it comes from longer maturities, because shorter-maturity demand shocks are better accommodated by arbitrageurs given that arbitraging short-maturity Treasuries involves lower risks. Taking the limit, the one-period arbitrage is perfect and the short rate is not affected by any demand shock.

Next, we analyze how arbitrageurs' risk aversion γ affects Treasury pricing.

Proposition 2 (Impact of Arbitrageurs' Risk Aversion)**.** *For long-term Treasuries, a higher arbitrageur risk aversion γ increases the magnitude of the Treasury price sensitivity to the macro factor β_t, latent demand u_t, and permanent demand $\theta_0(2)$.*

Proposition 2 states that higher arbitrageur risk aversion makes Treasury prices more sensitive to many sources of variations in the model, which is intuitive given that arbitrageurs accommodate order imbalances subject to risk aversion.

It is useful to consider two extreme cases. In the first case, we take $\gamma \to \infty$, so that arbitrageurs "drop out" from the market. Then the long-term Treasury price becomes

$$p_t^{(2)} = -\frac{2b}{a} r_t - \frac{2}{a}(\zeta(2) + \theta(2))\beta_t + \frac{2}{a} u_t(2) + \frac{2}{a}(\theta_0(2) - \bar{S}^{(2)}). \tag{30}$$

This is a case where Treasury prices are entirely driven by supply and demand absent from strategic arbitrage. Therefore, there is no distinction among a temporary latent demand shock u_t, a permanent demand shock $\theta_0(2)$, or a supply shock $\zeta(2)\beta_t$ – all of them share the same price impact. Moreover, the short-term rate r_t has an impact on the long-term Treasury price $p_t^{(2)}$ only if the cross substitution b is different from zero.

In the second case, we take $\gamma \to 0$, so that arbitrageurs are risk neutral and arbitrage to the full extent, leading to

$$p_t^{(2)} = -(1+\rho_r) r_t + \frac{1}{2}\sigma_r^2, \tag{31}$$

which is the log Treasury price under the expectations hypothesis (the second term is the Jensen's term after taking logs). Intuitively, the current short rate is r_t and in expectation the next period short rate is $\rho_r r_t$, leading to a log price of $-(1+\rho_r)r_t$ plus a convexity adjustment.

According to the non-arbitrageur demand in (19) and the simplified elasticity matrix in (27) and solution in (29), we obtain non-arbitrageur holdings as

$$Z_t^{(2)} = \frac{\theta_0(2)\frac{1}{\sigma_r^2} - \frac{1}{4}a + \frac{a}{2}\gamma \bar{S}^{(2)}}{\frac{1}{\sigma_r^2} + \frac{a}{2}\gamma} + \frac{\frac{a}{2}(1+\rho_r) - b}{1 + \frac{a}{2}\gamma\sigma_r^2} r_t + \frac{\frac{a}{2}\gamma\sigma_r^2(\zeta(2)+\theta(2))}{1+\frac{a}{2}\gamma\sigma_r^2}\beta_t + \frac{1}{1+\frac{a}{2}\gamma\sigma_r^2} u_t(2). \tag{32}$$

Expression (32) reflects equilibrium non-arbitrageur demand adjustments from two sources: one directly reflects demand dependence on the macro factor β_t and latent demand u_t absent a Treasury price effect, and the other reflects the response to the Treasury price change, both from the own yield and the other yield, due to cross elasticity. We note that if $\gamma \to \infty$, the above expression converges to $Z_t^{(2)} \to \bar{S}^{(2)} + \zeta(2)\beta_t$, i.e., the total debt supply in this simplified model as in (28). In this extreme case, the arbitrageurs' holdings of long-term Treasury become zero. Generally, arbitrageur holdings are $X^{(2)} = S_t^{(2)} - Z_t^{(2)}$, which in this simplified model is

$$X_t^{(2)} = \frac{\frac{1}{\sigma_r^2}\bar{S}^{(2)} + \frac{1}{4}a - \theta_0(2)\frac{1}{\sigma_r^2}}{\frac{1}{\sigma_r^2} + \frac{a}{2}\gamma} - \frac{\frac{a}{2}(1+\rho_r) - b}{1 + \frac{a}{2}\gamma\sigma_r^2} r_t + \frac{\zeta(2) + \theta(2)}{1 + \frac{a}{2}\gamma\sigma_r^2}\beta_t - \frac{1}{1 + \frac{a}{2}\gamma\sigma_r^2} u_t(2). \tag{33}$$

We next discuss how the yield curve responds to monetary policy.

Proposition 3 (Monetary Policy and Risk Premium)**.** *If $2b/a > 1 + \rho_r$ (strong cross elasticity), a positive monetary policy shock increases the term premium and causes over-reaction of long-term yields relative to the expectation hypothesis. On the other hand, if $2b/a < 1 + \rho_r$ (weak cross elasticity), we obtain the opposite result and there is under-reaction of long-term yields.*

We note that Proposition 3 sharply contrasts with the typical results in Vayanos and Vila (2021) type of models without cross elasticity. As Proposition 2 of Vayanos and Vila (2021) shows, there is under reaction of long-term yields relative to the expectations hypothesis. The basic intuition is that when the monetary policy rate rises, long-term Treasuries are cheaper due to the expectations effect, which induces non-arbitrageur investors to hold more of them and reduces the amount arbitrageurs absorb, therefore reducing the risk premium of long-term Treasuries and dampening the yield increase in the first place. When there is strong cross substitution, however, there is another force at work in that non-arbitrageur investors tend to reduce long-term Treasury holdings when short-term rate is higher, which then forces the arbitrageurs to increase their long-term Treasury holdings (see Equation (33)). This counteracts the first force and may cause the yield to be even higher than according to the expectations hypothesis. The proposition provides a sharp characterization of the conditions under which this new force dominates the first one.

In Section 3, we show that for most sectors, the cross elasticity is of a similar order of magnitude as the own elasticity. After aggregating all the sectors, we find that $2b/a = 2.3$ across maturities, while $\rho_r = 0.78$, so the strong cross elasticity is supported in the data. As a result, Proposition 3 suggests overreaction of long-term yield relative to the expectations hypothesis, which is consistent with the literature (Bekaert et al. 2013; Hanson and Stein 2015; Gertler and Karadi 2015; Kekre et al. 2024). Kekre et al. (2024) generates overreaction by introducing wealth effects for arbitrageurs, while we achieve the same result by allowing for cross elasticities.

Apart from traditional monetary policy, unconventional monetary policy can also be analyzed within the framework. We interpret QE as a demand shift, i.e., a higher $\theta_0(2)$.

Proposition 4 (QE and Treasury Pricing)**.** *QE increases Treasury prices and reduces Treasury yields.*

Proposition 4 indicates the pivotal role of Fed's demand in the Treasury market. With a persistent QE in place (higher $\theta_0(2)$), the Fed permanently increases Treasury prices and lowers Treasury yields.

We note that in this two-maturity model, there is no difference between a temporary demand shock $u_t(2)$ and a permanent demand shock $\theta_0(2)$, since after one period, the two-period bond becomes one period and the price is fully determined by the monetary policy rate. In the full model, we expect the effect to be stronger for permanent shocks since they have a stronger impact on the pricing kernel of arbitrageurs, and we will examine this hypothesis quantitatively using the full model.

Finally, we want to caution readers that although Propositions 2 to 4 provide very sharp characterizations regarding the roles of cross elasticities, price responses, and arbitrageur positions, these predictions are obtained under a drastic simplification of the full model. In the richer full model, we consider more than two maturities, so the risk premium on the macro factors β_t will be priced into long-term Treasuries, and the demand elasticity matrix α is more complicated than the one in Equation (27). More importantly, the full model accounts for arbitrageurs' outside portfolio which is affected by all the important factors including r_t and β_t, so predictions about how the short-rate r_t and macro factors β_t affect the Treasury yield curve are more complicated than the simple predictions in this section. Nevertheless, we believe this simple model still provides useful intuition that guides and helps us interpret our quantitative analysis in the following sections.

4.3. Solving and Estimating the Model

As noted, we conjecture an affine solution of the model of the form (26). Given this conjecture, we solve for the mean-variance problem in (24) and derive arbitrageurs' first-order conditions for Treasury holdings. For tractability, we make a simplifying assumption that the idiosyncratic latent demand shocks are not priced and do not carry a risk premium. This is a typical result in most asset pricing models. It is important to note that this assumption does not imply no price impact by latent demand shocks, since u_t can still directly affect prices via demand pressure.

Define the expected return on Treasuries of maturity τ as $\mu_t^{(\tau)} \equiv E_t[R_{t+1}^{(\tau)}]$, where $R_{t+1}^{(\tau)} = \exp(r_{t+1}^{(\tau)}) - 1 \approx r_{t+1}^{(\tau)} + \frac{1}{2}Var_t[r_{t+1}^{(\tau)}]$. The approximation becomes exact when we take a continuous-time approach[11]. The log return can be further expressed as $r_{t+1}^{(\tau)} = p_{t+1}^{(\tau-1)} - p_t^{(\tau)}$ and expanded using (15) and (26).

[11] Refer to Greenwood et al. (2023) for a more detailed discussion.

Next, solving the optimization problem (24), we get the first-order conditions

$$\mu_t^{(\tau)} - r_t = \hat{A}(\tau-1)' \underbrace{\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\Sigma\hat{A}(\hat{\tau}-1)X_t(\hat{\tau})\right) + \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t\right)}_{\lambda_{\beta,t}} + A_r(\tau-1)' \underbrace{\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\sigma_r^2 A_r(\hat{\tau}-1)X_t(\hat{\tau})\right) + \sigma_r\tilde{\sigma}_r\tilde{X}_t\right)}_{\lambda_{r,t}}, \tag{34}$$

where $\lambda_{\beta,t}$ is the price of risk of macroeconomic shocks and $\lambda_{\lambda,t}$ is the price of risk of monetary policy shocks, and $\hat{A}(\tau-1)$ is the risk exposure to macro factors given in Appendix D.1 (see Equation (A8)). For the Treasury price exposure to macroeconomic shocks, $\hat{A}(\tau-1)$, the expected return $\mu_t^{(\tau)} - r_t$ needs to provide compensation, and the compensation per unit of exposure is reflected by $\lambda_{\beta,t}$. Similarly, the exposure of the Treasury price to interest-rate risks, $A_r(\tau-1)$, requires compensation as reflected by $\lambda_{r,t}$.

Moreover, Equation (34) implies that the price of risk in this model is also affected by the "outside asset" position $\tilde{X}_t$, and its risk exposure. Note that we do not have sufficient degrees of freedom to pin down all parameters related to the dynamics of the outside asset. Instead, we assume that they can be spanned by β_t and r_t, so that

$$\begin{aligned} \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t &= \Psi\beta_t + \Lambda r_t + \psi \\ \sigma_r\tilde{\sigma}_r\tilde{X}_t &= \Psi_r\beta_t + \Lambda_r r_t + \psi_r, \end{aligned} \tag{35}$$

where Ψ is a $K \times K$ matrix, ψ is a $K \times 1$ vector, Ψ_r is a $1 \times K$ vector, and ψ_r is a scalar. These extra parameters need to be estimated together in the full model.

Next, we solve for X_t^τ using the market clearing Equation (25) and replace $Z_t(\tau)$ with (18), $S_t(\tau)$ with (20), thereby pinning down the equilibrium arbitrageur holdings as[12]

$$X_t(\tau) = \left(\bar{S}(\tau) + \zeta(\tau)'\beta_t + \zeta_r(\tau)'r_t\right) - \left(\theta_0(\tau) - \alpha(\tau)'p_t - \theta(\tau)'\beta_t + u_t(\tau)\right). \tag{36}$$

Expanding the expected return $\mu_t^{(\tau)}$ and plugging the equilibrium arbitrageur holdings X_t^τ of (36) into the pricing equation (34), we obtain an equilibrium condition that we rewrite purely in terms of β_t, r_t, and u_t. Because the equation holds for all values of these variables, the coefficients in front of them must all be matched, and so does the intercept term. Matching the coefficients, we arrive at a set of iterative equations for the coefficient matrices A, A_r, and A_u, as well as the vector C. These equations are given in Appendix D.1. The common structure of these iterative expressions is that they are all related to the granular-demand price elasticity $\alpha(\tau)$ and arbitrageur risk aversion γ. Therefore, both granular-demand function and arbitrageur risk aversion are central in driving the

[12]Note that this is an equilibrium result, not a "demand function". As discussed in Section 3.1, the demand of rational arbitrageurs does not explicitly depend on yields.

pricing of Treasury securities.

To solve and estimate the model, we denote the actual Treasury yields as $y_t^o(\tau)$ for maturity τ, and the model-implied Treasury yields as $y_t(\tau)$. Let h be the steady-state arbitrageurs' holdings of long-term (above 1Y) Treasuries as a fraction of the total long-term Treasuries outstanding. We find that h is very sensitive to the parameter γ. We denote the corresponding data moment as h^o, which is about 6% for Treasuries in our main sample from 2011 to 2022. Then we estimate the remaining parameters through the following problem:

$$\min_{\{\gamma, A_u, \Psi, \Psi_r, \Lambda, \Lambda_r, \psi, \psi_r\}} \mathbb{E}\left[M \cdot (h - h^o)^2 + \sum_t \sum_\tau (y_t(\tau) - y_t^o(\tau))^2\right], \tag{37}$$

subject to the equilibrium restrictions that discipline these parameters, detailed in Appendix D.3. We pick M to be sufficiently large so that the average intermediary Treasury holding is matched well. In our implementation, we pick $M = 1000000$. Further increasing M does not change the results.

4.4. Estimation Results

In line with our empirical analysis and motivated in Appendix A.3, we choose the macro state vector as $\beta_t =$ (credit spread, GDP gap, core inflation, debt/GDP). Adding additional macroeconomic variables does not significantly increase the explanatory power of the model for Treasury yield dynamics but could introduce over-fitting problems, so we choose this set of four macro variables. We estimate a VAR of the form (15) using the same sample period as in our main empirical analysis. We find that core inflation and debt/GDP are both highly persistent. Nevertheless, the maximum absolute value of the eigenvalue is 0.89, so macro variables converge to their long-run average.

To fit the monetary policy rule, we have to rely on a longer time period, because monetary policy rate does not exhibit much variation during our main sample period. In particular, we use the post-Volcker period (1990 to 2024) excluding the zero lower bound (ZLB) period (2008-2015). We start from 1990 because that is when the Fed gained credibility in its fight against inflation. The coefficients on GDP gap and inflation have the same signs as in the classical Taylor rule (Taylor 1993). Moreover, there is a high level of monetary policy inertia reflected by a coefficient of 0.78 on the lagged policy rate. This dependence on the lagged policy rate generates an impact of the monetary policy rate on long-term yields from the expectations effect and is critical to understanding how the yield curve responds to monetary policy shocks ε_{t+1}^r.

We estimate problem (37) on our main data sample from 2011Q4 to 2022Q4. Since we take expectations, the latent demand component in (26) will drop out in the objective function. In Figure

A4 of Appendix E.1, we show that the model-implied expected yields (equation (26) with $u_t = 0$) can fit the term structure reasonably well, both across maturities and over time. We further show that incorporating information on demand u_t leads to model predictions even closer to the data, consistent with the idea that demand shocks matter in the Treasury market.

The resulting absolute risk-aversion parameter is $\gamma = 0.03$. As shown in later sections, this is a low level of risk aversion that leads to an elastic Treasury market. The novelty of our approach is that we only rely on quantities to pin down arbitrageurs' risk aversion. As a result, the model can generate realistic quantity allocations across sectors and build tight linkages between quantities and prices. Our approach relies on granular data that allow us to distinguish arbitrageurs explicitly from granular-demand investors.

5. Dissecting the Treasury Market

In this section, we put our estimated model to work, and illustrate its basic mechanics and implications by dissecting the Treasury market. In particular, we decompose Treasury yields into different driving forces, quantify the impact of arbitrageur risk aversion, evaluate the aggregate elasticity of the Treasury market, and show the term structure of market elasticity. A key advantage of our approach is that from our granular demand data, we explicitly recover sector-level latent demand and their contribution to total demand, in contrast to the extant literature that relies on latent factors (Ang and Piazzesi 2003; Bikbov and Chernov 2010; Joslin et al. 2014).

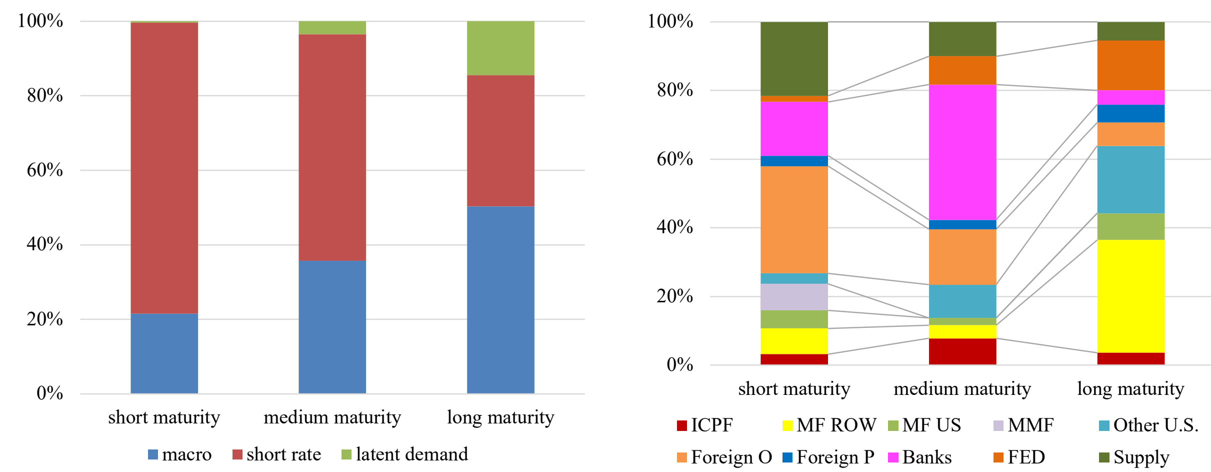
5.1. Decomposing Treasury Pricing

We start by taking guidance from the model to decompose Treasury prices and their variation into their underlying drivers. Indeed, by virtue of Equation (26), we obtain a decomposition of yields into macroeconomic states, monetary policy rate, as well as latent demand, while from Equations (17) and (20), we obtain a decomposition of yield changes into sector-specific demand shocks and supply shocks. Indeed, we define $\Delta Z_t^{\iota} = -\theta(\tau)\Sigma^{1/2}\varepsilon_t + u_t$ as the demand shock of sector ι, and $\Delta S_t = \zeta'\varepsilon_t + \zeta_r\sigma_r\varepsilon_t^r$ as a supply shock in the model. While the first decomposition reveals how state variables and latent demand drive the overall variations in Treasury yields, the second one focuses on sector heterogeneity and the unpredictable components of supply and demand. In both of them, we express the contribution of each variable as the Shapley value of R^2, which is calculated as the marginal contribution of each variable to the R^2 among all possible sets of combinations of dependent variables.

Regarding the first decomposition, Panel (a) in Figure 4 shows that the relative contribution of economic forces varies across the term structure, in line with the model predictions in Proposition

Figure 4. **Decomposition of Treasury Yield Variation.**

In this figure, we decompose Treasury yield variations. In Panel (a), we show the relative contribution of macroeconomic factors, FFR, and latent demand to the variation in Treasury yields, using the relative magnitude of their Shapley values of R^2, which is calculated as the average marginal contribution of each variable to the R^2 among all possible sets of dependent variable combinations. In Panel (b), we focus on how these different factors take effect through the supply and demand forces in the model, by regressing one-quarter difference in Treasury yields of different maturity buckets on aggregate supply shocks and sector-level demand shocks, which are unpredictable components driven by latent demand shocks and macro shocks.



(a) Decomposition into macroeconomic, short rate, and latent demand variations.

(b) Contribution of sector-level demand and aggregate supply shocks.

1. For short-maturity Treasuries, monetary policy plays the dominant role, explaining the vast majority of variation, while macro variables play a secondary role. We note that latent demand almost has no explanatory power for short-maturity Treasury yields. As the maturity increases, the relative importance of the FFR declines while the relative importance of both macro variables and latent demand shocks expand. In particular, for long-maturity Treasuries, macroeconomic variables can explain about half of the variation in yields.

As shown in Figure 4 Panel (b), the second decomposition varies across the maturity structure of Treasuries. Although banks' holdings are very small compared to the entire market (3.4% as shown in Table 2), they play a sizable role in transmitting shocks to the Treasury market, especially in the medium-maturity bucket (1$\sim$5 years). The foreign official sector contributes significant shocks to both short- and medium-maturity buckets, while foreign mutual fund demand shocks significantly contribute to yield variations in the long maturity bucket. Importantly, we find that a sector's contribution to yield variation can substantially differ from its average holdings, as that contribution predominantly depends on how actively a sector responds to shocks.

5.2. Arbitrageur Risk Aversion

In our model with arbitrageurs, the response of Treasury yields to shocks is not only shaped by granular-investors' individual demand elasticities, but, critically, also by arbitrageur risk aversion, as shown by Proposition 2. To provide quantitative guidance regarding the effects of γ, we find it illuminating to compare the baseline case with estimated risk aversion to an extreme case where we send $\gamma \to \infty$ and thus exclude the arbitrageurs. In the latter case ($\gamma \to \infty$), market clearing implies

$$p_t = \alpha^{-1} \left((\theta_0 - \theta \beta_t + u_t) - \left(\bar{S} + \zeta \beta_t + \zeta_r r_t \right) \right). \tag{38}$$

Thus, absent arbitrageurs, the equilibrium price response to a demand shock is simply α^{-1}. On the other hand, with arbitrageurs, the equilibrium demand elasticity also depends on arbitrageur risk-aversion γ, the volatility of macroeconomic shocks Σ, monetary policy uncertainty σ_r and inertia ρ_r, and the persistence of macroeconomic dynamics Φ. In this case, therefore, the equilibrium demand elasticity may significantly differ from the estimated granular-demand investors' demand elasticities.

In Table 5, we illustrate the equilibrium Treasury price response (in %) at each maturity bucket to a latent demand shocks that is 1% of outstanding in a specific maturity bucket. This is a granular version of price multiplier as in Gabaix and Koijen (2021). In Panel (a), we report the multiplier in the full model with arbitrageurs for the three maturity buckets we consider in our empirical analysis, using the average duration as the representing maturity. We find that for a given demand shock, the price response at longer maturities is much stronger. For example, the response of the long-maturity Treasury price is 13 times larger than the short-maturity Treasury price response, given a shock to short-maturity demand. Moreover, shocks to demand for longer-maturity Treasuries are more powerful, reflected by larger multipliers associated with shocks to longer maturities.

In Panel (b), we effectively remove arbitrageurs by setting $\gamma = \infty$, and examine the corresponding price multipliers, obtained from scaling (38) with the outstanding amount in each maturity bucket. We find that in this case, price multipliers are generally one to two orders of magnitude larger than in the baseline case. Clearly, without arbitrageurs, the price impact on T-bills is too large for a world in which the Fed tightly controls the money market. With arbitrageurs, the Fed controls the monetary policy rate by actively accommodating any demand shocks in the one-period (one-quarter) Treasury market, and arbitrageurs propagate these dynamics through the term structure, with weakening price effects at longer maturities.

The force of arbitrage is illustrated in Panel (c) of Table 5, where we report the ratio of the price impacts in the case without arbitrageurs and the baseline case. On average, the price impact

Table 5. **Impact of Latent Demand Shocks on Treasury Prices with and without Arbitrageurs.**

We illustrate the impact of latent demand shocks with and without arbitrageurs. In panels (a) and (b), a value of 1 at row i and column j implies that 1% extra latent demand of maturity bucket i increases the price at maturity j by 1%. Panel (c) shows the ratio of the corresponding cells in Panel (b) over Panel (a).

Panel (a): With Arbitrageur		price change (%) of	
	short maturity	medium maturity	long maturity
shock on short maturity	0.001	0.006	0.013
shock on medium maturity	0.008	0.057	0.139
shock on long maturity	0.015	0.111	0.323
Panel (b): Without Arbitrageur			
shock on short maturity	0.344	1.292	7.108
shock on medium maturity	2.287	7.508	44.102
shock on long maturity	0.707	2.478	12.878
Panel (c): Price Impact Ratio (Panel (b)/Panel (a))			
shock on short maturity	432.899	230.737	531.800
shock on medium maturity	288.740	132.617	316.304
shock on long maturity	46.904	22.229	39.884

in the case without arbitrageur is more than 100 times the one with arbitrageurs. In Appendix E.2, we also show the impact of permanent demand shocks with and without arbitrageurs and reach a similar conclusion.

5.3. How Elastic is the Treasury Market?

In view of the price responses of each maturity bucket in Table 5, we can estimate the aggregate Treasury market multiplier, which is the percentage valuation change of the entire Treasury market for a demand shock to the overall Treasury market that is worth 1% of total Treasury value.

Using long-run average values of total Treasury supply in each maturity bucket as weights, we convert numbers in Table 5 into a total market multiplier, which is 0.23. This implies that for a \$100 billion dollar demand shock on the whole Treasury market, the total Treasury valuation increases by \$23 billion dollars. On the other hand, the multiplier is 0.78 for a representative permanent demand shock[13]. In contrast, Chaudhary et al. (2022) report a multiplier for the corporate bond market of 3.5, while Gabaix and Koijen (2021) find a multiplier of 5 for the stock market. As a result, the equilibrium price impact in the Treasury market is significantly weaker, which suggests

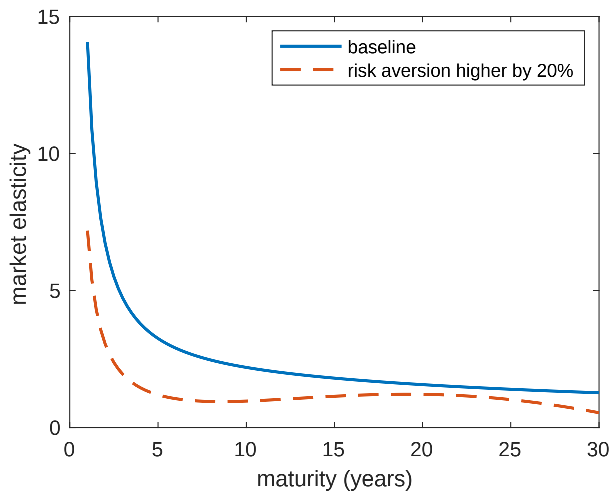
[13] See Appendix E.2 for more details on calculating market multipliers for both latent and permanent demand shocks.

that the Treasury market is, in fact, quite elastic.

Notably, however, once we remove arbitrageurs by setting $\gamma = \infty$, the estimated Treasury market multiplier increases to 26.82, substantially larger than the multipliers in corporate bond and equity markets. This illustrates the importance of accounting for arbitrageurs when computing aggregate price elasticities. Intuitively, with low estimated risk aversion, arbitrageurs aggressively trade and thus dampen the effect of demand shocks.

Figure 5. **The Term Structure of Market Elasticity**

This figure illustrates how market elasticity differs with the maturity of the demand shock. In particular, we define the market elasticity of maturity τ as the inverse of *total market multiplier at maturity* τ, which is the percentage change in total Treasury valuation in response to a change in Treasury demand at maturity τ equal to 1% of total Treasury outstanding.



We also estimate multipliers for each maturity and find notable heterogeneity among them. We define the market elasticity at maturity τ as the inverse of total market multiplier at maturity τ, which is the percentage change in total Treasury valuation in response to a change in Treasury demand at maturity τ equal to 1% of total Treasury value. Figure 5 reveals a sharply downward-sloping term structure of the market elasticity with a one-year elasticity of around 14, while falling to 2 at the 10-year maturity. Intuitively, this steep decline reflects the higher risk exposure arbitrageurs have to bear when absorbing shocks to the demand of longer-maturity Treasuries. It is then natural to conjecture that the slope of the term structure of market elasticity is sensitive to arbitrageur risk aversion. Indeed, when we increase risk aversion by 20%, the slope is muted and no longer monotonic, as the dotted line in Figure 5 shows. In this case, arbitrage at the shorter end of the maturity spectrum becomes harder, while demand factors become more important and heterogeneously affect the term structure.

5.4. Treasury Market Elasticity under the Microscope

The aggregate Treasury market elasticity masks heterogeneous responses and elasticities of different granular-demand investors as well as arbitrageurs. The granular nature of our setting allows us to trace the equilibrium response of the entire Treasury market down to portfolio adjustments at the sector level. We illustrate such rebalancing in response to selling pressure and distress in the Treasury market. We do so by considering both a temporary or a permanent negative demand shock, and examining the responses of all sectors, including that of the Fed.

We use our granular sector-level demand functions to examine how each sector adjusts to latent demand shocks at the average state in the model. We start with the equilibrium with a zero latent demand shock at the steady state, so that the benchmark price is $\bar{p} = A\bar{\beta} + A_r\bar{r} + C$. Next, we introduce a demand shock Δu to the model, and then trace out how each sector absorbs this shock. The new price will be $p = A\bar{\beta} + A_r\bar{r} + A_u\Delta u + C$, with corresponding change $\Delta p = p - \bar{p} = A_u\Delta u$. Each sector ι, including the Fed, absorbs this shock according to the demand specifications in equation (17), namely $\Delta Z^{\iota} = -\sum_{\iota \in \mathscr{I}} \alpha^{(\iota)} A_u \Delta u$, where we exclude the shock Δu itself and only consider adjustments due to price changes. For permanent demand shocks, we need to solve the equilibrium coefficients $\{A, A_r, A_u, C\}$ again under a different level of demand θ_0 and obtain a new price p', which then allows us to calculate the demand change $\Delta Z^{\iota} = -\sum_{\iota \in \mathscr{I}} \alpha^{(\iota)}(p' - p)$.

In Table 6, we show how different sectors rebalance their portfolios in response to a \$100 billion sell-off in the long-maturity Treasury bucket. Panel (a) examines the impact of a temporary shock, while Panel (b) focuses on a permanent shock. The Table reveals that the Fed's balance sheet remains largely unchanged in response to a temporary demand shock, but adjusts significantly to a permanent shock. This contrast reflects the model's property that temporary demand shocks have a much smaller price impact compared to permanent shocks (also see Figure 7). Moreover, it captures the reality that as a central bank, the Fed prioritizes long-term policy goals and avoids reacting to short-term market fluctuations. Specifically, in response to a permanent shock in long-term Treasuries, the Fed increases its long-term Treasury holdings by \$42.6 billion, absorbing 42.6% of the total shock and emerging as the primary stabilizer of the long-term Treasury market during persistent demand shifts.

On the other hand, for a temporary demand shift, arbitrageurs are the only sector that responds significantly. Comparing Panels (a) and (b), we observe that the arbitrageur sector expands its balance sheet substantially in both cases, but for different reasons. In Panel (a), arbitrageurs are willing to hold significantly more long-term Treasuries to accommodate the temporary demand shock because they have lower impact on the risks that they bear. In contrast, Panel B reflects a permanent demand shock, leading to a significant increase in the risk premium and thus the long-term yield (also see Figure 7). In response, the Fed, U.S. mutual funds, and other U.S. investors,

Table 6. **Model-Implied Sector-Level Portfolio Adjustment to Demand Shocks.**

This table illustrates how each sector adjusts their portfolio positions in response to a $100 billion sell shock (temporary as in Panel (a) and permanent as in Panel (b)) of Treasuries at maturity bucket 3 ($\tau \geq 5$Y), excluding the sell shock in the first place.

Panel (a): A $100 billion Temporary Sell Shock of $\tau > 5$ Treasuries

Sector	$\tau < 1$	$1 \leq \tau < 5$	$\tau \geq 5$	Total Change
Banks	-0.9	0.7	-1.0	-1.2
ICPF	0.1	0.2	0.1	0.4
MF ROW	0.1	0.3	0.1	0.5
MF U.S.	-1.9	1.6	-1.9	-2.2
MMF	-24.6	0.0	0.0	-24.6
Other U.S.	7.2	10.1	7.3	24.5
Foreign Official	-6.8	-6.5	-6.9	-20.2
Foreign Private	-0.8	0.1	-0.8	-1.5
Fed	6.5	1.8	-9.3	-1.0
Arbitrageurs	21.1	-8.2	112.4	125.2

Panel (b): A $100 billion Permanent Sell Shock of $\tau > 5$ Treasuries

Sector	$\tau < 1$	$1 \leq \tau < 5$	$\tau \geq 5$	Total Change
Banks	-6.7	-3.3	7.3	-2.7
ICPF	0.0	0.2	0.6	0.7
MF ROW	-0.2	0.1	1.0	0.9
MF U.S.	-14.0	-6.9	15.8	-5.1
MMF	-102.0	0.0	0.0	-102.0
Other U.S.	3.0	11.2	29.7	43.9
Foreign Official	-12.2	-13.4	-11.4	-37.0
Foreign Private	-4.1	-2.4	3.4	-3.1
Fed	15.8	8.2	42.6	66.6
Arbitrageurs	120.5	6.3	11.0	137.8

significantly increase their long-term Treasury holdings. Moreover, because of a strong cross substitution, U.S. mutual funds, foreign investors, banks, and money market funds, all significantly reduce their short-term Treasury holdings.[14] Arbitrageurs fill this gap by significantly expanding their short-term Treasury holdings.

Notably, in both scenarios that we consider in Panels (a) and (b) of Table 6, we find that the foreign official sector exacerbates the original negative demand shock. Note that our estimation

[14] Note that the estimated own and cross elasticity for MMFs have to be interpreted with care because of a rather weak first stage. If the true cross elasticity is lower than our estimates reveal, then MMFs would sell less short-term Treasuries in response to a sell shock, and therefore, simultaneously, arbitrageurs would absorb less short-term Treasuries.

is at quarterly frequency, so this reveals that not only in crisis times such as COVID-19, but also in normal times, the foreign official sector amplifies Treasury yield fluctuations in response to demand shocks.

6. Conventional and Unconventional Monetary Policies

In this section, we use the model to analyze monetary policy shocks and quantitative easing.

6.1. The Impact of Conventional Monetary Policy

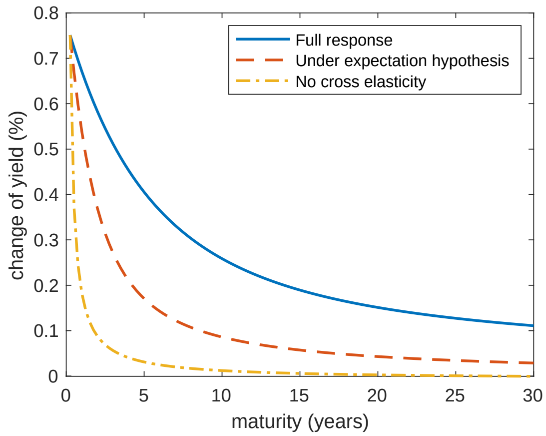
To examine conventional monetary policy, we consider the impact of a one-standard deviation shock to monetary policy, $\varepsilon_r = 1$, at the steady state, which translates into a positive shock to the short rate by 0.75%. In Figure 6, we illustrate how the term structure responds to an increase in the monetary policy rate. The left panel illustrates the response of the yield curve. In the absence of changes in risk premia, the expected future short rate change is the same as the expectation hypothesis, $E_t[\Delta r_{t+h}] = \sigma_r \rho_r^h$, and the expectation component of the yield curve change for maturity τ is $\frac{1}{\tau}\sum_{h=0}^{\tau-1} \sigma_r \rho_t^h$. While, as shown in panel (a) of Figure 6, the expectation component declines quickly over maturity, approaching zero at around a 15-year maturity, the full response in the model strongly reacts to the monetary policy shock even at a 30-year maturity. Accordingly, their difference, i.e., the risk premium or term premium, positively responds to a monetary policy shock, as shown in panel (b).

Although this positive response of the term premium to monetary policy shocks and the "excessive reaction" of long-term yields in our model are well-documented empirically, many models in the literature struggle to rationalize them. While in models with perfect arbitrage (i.e., zero risk aversion) the expectation hypothesis holds and the term premium should not respond, in models of market segmentation (e.g., no arbitrageurs), short-rate shocks do not significantly change long term rates, leading to a negative response of the term premium. In models with risk-averse arbitrageurs and preferred-habitat investors in the spirit of Vayanos and Vila (2021), a higher policy rate typically reduces the term premium, because it lowers Treasury prices and thus boosts non-arbitrageur demand, so that arbitrageurs reduce their Treasury holding and thus command a lower price of risk.

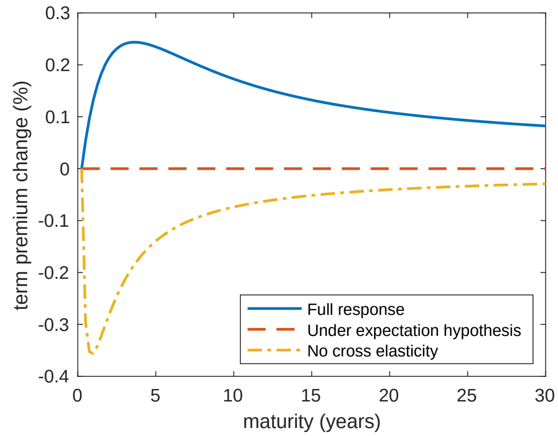
Our model rationalizes the evidence through the presence of cross substitution in investors' demand functions. Intuitively, with the estimated cross-substitution, granular-demand investors tend to rebalance their portfolios towards higher yielding short-term bonds after a positive monetary policy shock, thereby alleviating the demand for long-term bonds in view of falling prices and

Figure 6. **Contemporaneous Yield Curve Response to a Monetary Policy Shock.**

This figure illustrates the impact of a one standard deviation monetary policy shock ($\varepsilon_t^r = 1$) under three different models: the full model, the model with risk-neutral arbitrageurs (the expectation hypothesis), and the model without cross elasticities. The left panel illustrates yield curve responses. The right panel illustrates the response of the term premium, which is the risk premium component of yields.



(a) Yield curve response



(b) Term premium response

leaving a higher share of them for arbitrageurs to absorb. Indeed, under stringent assumptions, Proposition 3 shows that with high cross-substitution term premia rise in response to positive monetary policy shocks. Consistent with this intuition, Figure 6 also shows that when we exclude cross elasticities and re-solve the model, the yield curve under-reacts relative to the expectation hypothesis, aligning with the baseline result in Vayanos and Vila (2021). This suggests that accurately capturing cross elasticities in investor demand is essential for understanding the term structure response to monetary policy shocks[15].

Our mechanism critically depends on granular-demand investors' portfolio adjustments and rebalancing in response to monetary policy shocks. Our granular model does not only allow to put these under the microscope, but it also takes into account strategic debt issuance by the government in response. Table 7 illustrates such supply and demand adjustments in response to the monetary policy shock. As total debt supply is driven by the contemporaneous macro variable Debt/GDP, the government supplies more short-maturity bonds and reduces long-term bond supply, with a total net change of Treasuries outstanding of zero. In response, almost all investors rebalance towards higher-yielding short-term Treasuries reflecting cross-substitution, especially the highly elastic MMFs and other U.S. investors. Consistent with overall monetary tightening, the Fed aggressively sells long-term Treasuries, reflecting the large and negative loading on the short-term interest rate (see Table 4), in fact beyond the shrinking of the corresponding supply. Importantly,

[15]Our rationale based on cross elasticities in investors' demand is consistent with Hanson and Stein (2015) who consider yield-oriented investors comparing long- and short-term yields when making long-term bond investments.

Table 7. **Portfolio Adjustment to Monetary Policy Shocks.**

This table illustrates portfolio adjustments to a one standard deviation shock to monetary policy, which is a 0.75% increase in one-period rate. We also report the model-implied change of Treasury supply in response to this shock. All units are in billions of dollars.

Sector	$\tau < 1$	$1 \leq \tau < 5$	$\tau \geq 5$	Total Change
Banks	12.3	1.4	-21.6	-7.8
ICPF	2.1	1.5	0.5	4.0
MF ROW	3.1	2.1	0.1	5.3
MF U.S.	27.6	4.4	-44.7	-12.8
MMF	13.5	0.0	0.0	13.5
Other U.S.	109.2	84.4	44.1	237.7
Foreign Official	-63.5	-61.0	-65.2	-189.7
Foreign Private	3.8	-1.9	-14.4	-12.5
Fed	44.3	21.8	-170.2	-104.0
Arbitrageurs	67.6	-110.3	109.1	66.3
Total Supply	220.0	-57.7	-162.3	0.0

arbitrageurs absorb most of the net increase in long-term Treasuries so that they command higher risk premia. Notably, the foreign official sector sells Treasuries across all maturities, perhaps reflecting an incentive to defend their local currency in time of a stronger dollar or to provide liquidity to their domestic markets in face of capital outflows.

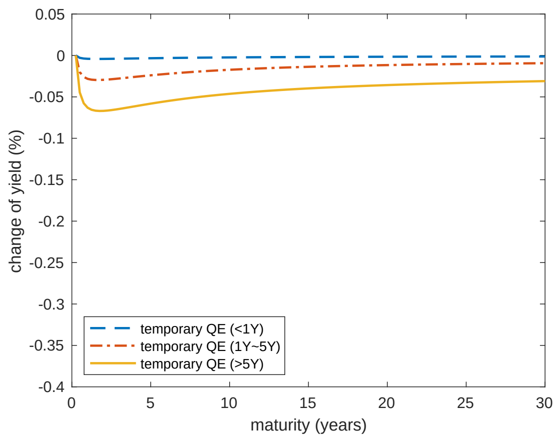
6.2. The Impact of Quantitative Easing

Through the lens of our model, we can think of quantitative easing (QE) policies as changes in the Fed's demand, either through a temporary or a permanent change. We distinguish between transient QE, modeled as an increase in u_t, and permanent QE, which amounts to an increase in θ_0. Proposition 4 suggests that QE increases Treasury prices and decreases Treasury yields.

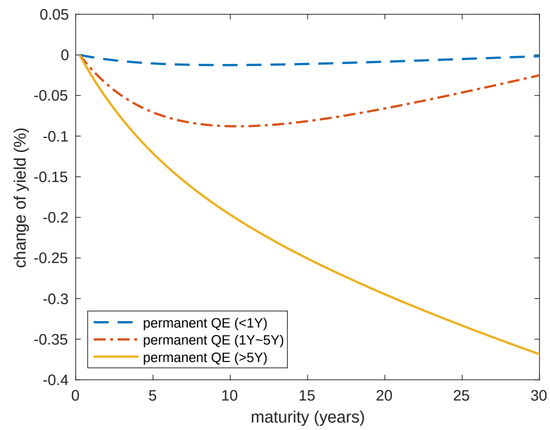
Figure 7 shows the quantitative impact of both transient QE and permanent QE in the full model. In both cases, we consider the steady state yield as the baseline scenario and introduce a $100 billion extra demand on each of the three maturity buckets, respectively, and show the change of yields in response to the demand shock. Panel (a) of Figure 7 shows that transient QE on short-maturity Treasuries has little effect on the yield curve, as dealers elastically arbitrage between short-maturity Treasuries and the one-period rate controlled by monetary policy. As maturity increases, the yield curve becomes more reactive, as arbitrageurs are more reluctant to bear the extra risks involved in absorbing long-maturity Treasuries. It is thus natural that the Fed usually purchases long-term Treasuries in QE programs. These patterns are quantitatively significantly amplified in the case of permanent QE, as panel (b) of Figure 7 illustrates. Moreover,

Figure 7. **Impact of QE Shocks on Treasury Yields.**

This figure illustrates how a $100 billion QE shock on different maturity buckets, either temporary (left panel, increasing latent demand u_t) or permanent (right panel, increasing permanent demand θ_0), affects Treasury yields. For dollar values, we use the stationary model unit as described in Section 4.



(a) Temporary QE Policy



(b) Permanent QE Policy

there is a strong localization effect, in that QE on a specific maturity bucket affects that maturity-bucket yield more strongly than others. As in Vayanos and Vila (2021) and Greenwood et al. (2023), with multiple sources of risks affecting different maturities differentially, arbitrageurs do not aggressively trade against a permanent demand shock, causing a localization of the price impact.

To compare the model-implied results with empirical studies regarding the impact of QE, we have to consider details of the QE implementation. First, the duration of QE purchases ranges between 3 to 10 years, so the average effect is in between our maturity buckets 2 and 3, i.e., between the solid orange line and dotted red line in Figure 7. Second, the expected duration of the QE purchase is between one quarter (panel (a) of Figure 7) and permanent (panel (b) of Figure 7). As a rough approximation, using the average value of bucket 2 and 3, our model implies that the impact of a $100 billion purchase generates yield declines ranging from 3 to 14 bps in 10-year Treasuries, depending on the expected persistence of QE. This is in a similar order of magnitude as the 4.5 bps reported in Gulati and Smith (2022), who survey the extant literature, including Krishnamurthy and Vissing-Jorgensen (2011) and Swanson (2011). Our model highlights that the effectiveness of QE critically depends on how credibly the Fed can signal its commitment to a sustained expansion of its balance sheet, perhaps through Forward Guidance, so that investors perceive it as permanent.

7. Conclusion

In this paper, we estimate an equilibrium model of the U.S. Treasury market to dissect Treasury pricing and understand the impact of conventional and unconventional monetary policies. Our model nests granular-demand investors, whose Treasury demand can be flexibly estimated from a novel dataset on granular Treasury holdings, in the spirit of Koijen and Yogo (2019), risk-averse arbitrageurs, who absorb demand imbalances as in Vayanos and Vila (2021), and the Fed.

Our quantitative analysis reveals an elastic Treasury market and a downward-sloping term structure of market elasticity, as arbitrageurs readily absorb demand imbalances, especially at the short end of the maturity spectrum. Moreover, it rationalizes rising risk premia in response to monetary tightening, as due to cross-substitution arbitrageurs have to increase their long-term Treasury holdings. Finally, the effectiveness of QE is significantly driven by the expected persistence of the Fed's interventions.

We view our paper as the building block of a framework for combining novel data with equilibrium demand-based models to shed light on important macro-finance questions in the government bond market. Future research can build on our approach and incorporate this demand view of Treasury pricing into macroeconomic models to study the macro implications of government bond demand.

References

Acharya, V. V. and Laarits, T. (2023). When do treasuries earn the convenience yield?: A hedging perspective.

Adrian, T., Etula, E., and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, 69(6):2557–2596.

Allen, J., Kastl, J., and Wittwer, M. (2020). Estimating demand systems for treasuries.

An, Y. and Huber, A. (2024). Intermediary elasticity. *Available at SSRN 4825151*.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, 50(4):745–787.

Bahaj, S., Czech, R., Ding, S., and Reis, R. (2023). The market for inflation risk.

Bauer, M. D., Bernanke, B. S., and Milstein, E. (2023). Risk appetite and the risk-taking channel of monetary policy. *Journal of Economic Perspectives*, 37(1):77–100.

Becker, B. and Ivashina, V. (2015). Reaching for yield in the bond market. *The Journal of Finance*, 70(5):1863–1902.

Bekaert, G., Hoerova, M., and Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788.

Bernanke, B. S. and Kuttner, K. N. (2005). What explains the stock market's reaction to federal reserve policy? *The Journal of finance*, 60(3):1221–1257.

Bikbov, R. and Chernov, M. (2010). No-arbitrage macroeconomic determinants of the yield curve. *Journal of Econometrics*, 159(1):166–182.

Bretscher, L., Schmid, L., Sen, I., and Sharma, V. (2024). Institutional corporate bond pricing. *Review of Financial Studies*. Forthcoming.

Brunnermeier, M. K., Merkel, S., and Sannikov, Y. (2024). Safe assets. *Journal of Political Economy*.

Campbell, J. and Viceira, L. (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press.

Chaudhary, M., Fu, Z., and Li, J. (2022). Corporate bond multipliers: Substitutes matter. *Available at SSRN 4292266*.

Choi, J. and Kronlund, M. (2018). Reaching for yield in corporate bond mutual funds. *The Review of Financial Studies*, 31(5):1930–1965.

Cieslak, A., Li, W., and Pflueger, C. (2024). Inflation and treasury convenience.

Clarida, R., Gali, J., and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly Journal of Economics*, 115(1):147–180.

Culbertson, J. (1957). The term structure of interest rates. *The Quarterly Journal of Economics*, 71(4):485–517.

d'Amico, S., English, W., López-Salido, D., and Nelson, E. (2012). The federal reserve's large-scale asset purchase programmes: rationale and effects. *The Economic Journal*, 122(564):F415–F446.

Darmouni, O., Siani, K., and Xiao, K. (2022). Nonbank fragility in credit markets: Evidence from a two-layer asset demand system. *Available at SSRN 4288695*.

d'Avernas, A., Petersen, D., and Vandeweyer, Q. (2023). The central bank's balance sheet and treasury market disruptions.

d'Avernas, A. and Vandeweyer, Q. (2023). Treasury bill shortages and the pricing of short-term assets. *The Journal of Finance*.

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

Del Negro, M., Eggertsson, G., Ferrero, A., and Kiyotaki, N. (2017). The great escape? a quantitative evaluation of the fed's liquidity facilities. *American Economic Review*, 107(3):824–857.

Di Tella, S., Hébert, B., Kurlat, P., and Wang, Q. (2023). The zero-beta rate.

Diamond, W. (2020). Safety transformation and the structure of the financial system. *The Journal of Finance*, 75(6):2973–3012.

Diamond, W., Jiang, Z., and Ma, Y. (2024). The reserve supply channel of unconventional monetary policy. *Journal of Financial Economics*, 159:103887.

Diamond, W. and Van Tassel, P. (2021). Risk-free rates and convenience yields around the world. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Doerr, S., Eren, E., and Malamud, S. (2023). Money market funds and the pricing of near-money assets. *Swiss Finance Institute Research Paper*, (23-04).

Drechsler, I., Savov, A., and Schnabl, P. (2018). A model of monetary policy and risk premia. *The Journal of Finance*, 73(1):317–373.

Droste, M., Gorodnichenko, Y., and Ray, W. (2021). Unbundling quantitative easing: Taking a cue from treasury auctions. *Manuscript, UC Berkeley*.

Du, W., Hébert, B., and Huber, A. W. (2023a). Are intermediary constraints priced? *The Review of Financial Studies*, 36(4):1464–1507.

Du, W., Hébert, B., and Li, W. (2023b). Intermediary balance sheets and the treasury yield curve. *Journal of Financial Economics*, 150(3):103722.

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

Duffie, D., Fleming, M. J., Keane, F., Nelson, C., Shachar, O., and Van Tassel, P. (2023). Dealer capacity and us treasury market functionality.

Eren, E., Schrimpf, A., and Xia, F. D. (2023). The demand for government debt. *Available at SSRN 4466154*.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, pages 680–692.

Fang, X., Hardy, B., and Lewis, K. K. (2022). Who holds sovereign debt and why it matters.

Fang, X. and Liu, Y. (2021). Volatility, intermediaries, and exchange rates. *Journal of Financial Economics*, 141(1):217–233.

Favara, G., Infante, S., and Rezende, M. (2022). Leverage regulations and treasury market participation: Evidence from credit line drawdowns. *Working Paper*.

Gabaix, X. and Koijen, R. S. (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis.

Gabaix, X. and Koijen, R. S. (2024). Granular instrumental variables. *Journal of Political Economy*, 132(7):000–000.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American economic review*, 102(4):1692–1720.

Greenwood, R., Hanson, S., and Vayanos, D. (2015). Forward guidance in the yield curve: Short rates versus bond supply. *National Bureau of Economic Research*.

Greenwood, R., Hanson, S., and Vayanos, D. (2023). Supply and demand and the term structure of interest rates.

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

Guibaud, S., Nosbusch, Y., and Vayanos, D. (2013). Bond market clienteles, the yield curve, and the optimal maturity structure of government debt. *The Review of Financial Studies*, 26(8):1914–1961.

Gulati, C. and Smith, A. L. (2022). The evolving role of the fed's balance sheet: Effects and challenges. *Economic Review (01612387)*, 107(4).

Haddad, V., He, Z., Huebner, P., Kondor, P., and Loualiche, E. (2024a). Causal inference for asset pricing.

Haddad, V., Moreira, A., and Muir, T. (2024b). Asset purchase rules: How QE transformed the bond market.

Haddad, V. and Muir, T. (2021). Do intermediaries matter for aggregate asset prices? *The Journal of Finance*, 76(6):2719–2761.

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

Hanson, S. G., Malkhozov, A., and Venter, G. (2024). Demand-and-supply imbalance risk and long-term swap spreads. *Journal of Financial Economics*, 154:103814.

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

He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103:732–770.

Jansen, K. A. (2024). Long-term investors, demand shifts, and yields. *Review of Financial Studies*. Forthcoming.

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

Jiang, W., Sargent, T. J., Wang, N., and Yang, J. (2024a). A p theory of government debt and taxes. *The Journal of Finance*.

Jiang, Z., Krishnamurthy, A., and Lustig, H. (2021). Foreign safe asset demand and the dollar exchange rate. *The Journal of Finance*, 76(3):1049–1089.

Jiang, Z., Lustig, H., Van Nieuwerburgh, S., and Xiaolan, M. Z. (2024b). The us public debt valuation puzzle. *Econometrica*, 92(4):1309–1347.

Jiang, Z., Richmond, R. J., and Zhang, T. (2022). A portfolio approach to global imbalances.

Jiang, Z. and Sun, J. (2024). Quantitative tightening with slow-moving capital.

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

Kargar, M. (2021). Heterogeneous intermediary asset pricing. *Journal of Financial Economics*, 141(2):505–532.

Kekre, R., Lenel, M., and Mainardi, F. (2024). Monetary policy, segmentation, and the term structure.

Khetan, U., Li, J., Neamtu, I., and Sen, I. (2023). The market for sharing interest rate risk: Quantities and asset prices.

Koijen, R. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.

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

Koijen, R. S. and Yogo, M. (2020). Exchange rates and asset prices in a global demand system. *National Bureau of Economic Research*.

Krishnamurthy, A. and Li, W. (2023). The demand for money, near-money, and treasury bonds. *The Review of Financial Studies*, 36(5):2091–2130.

Krishnamurthy, A. and Muir, T. (2017). How credit cycles across a financial crisis.

Krishnamurthy, A. and Vissing-Jorgensen, A. (2007). The demand for treasury debt.

Krishnamurthy, A. and Vissing-Jorgensen, A. (2011). The effects of quantitative easing on interest rates: channels and implications for policy.

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

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Li, W. (2024). Public liquidity and financial crises. *American Economic Journal: Macroeconomics*.

Li, W., Ma, Y., and Zhao, Y. (2023). The passthrough of treasury supply to bank deposit funding. *Columbia Business School Research Paper, USC Marshall School of Business Research Paper*.

Maggiori, M., Neiman, B., and Schreger, J. (2020). International currencies and capital allocation. *Journal of Political Economy*, 128(6):2019–2066.

Modigliani, F. and Sutch, R. (1966). Innovations in interest rate policy. *The American Economic Review*, 56(1/2):178–197.

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

Payne, J. and Szőke, B. (2024). Convenience yields and financial repression.

Payne, J., Szőke, B., Hall, G., and Sargent, T. J. (2022). Costs of financing us federal debt under a gold standard: 1791-1933.

Selgrad, J. (2023). Testing the portfolio rebalancing channel of quantitative easing.

Sharpe, W. F. and Tint, L. G. (1990). Liabilities— A new approach. *The Journal of Portfolio Management*, 16(2):5–10.

Siani, K. (2022). Raising bond capital in segmented markets. *Available at SSRN 4239841*.

Stein, J. C. and Wallen, J. (2023). The imperfect intermediation of money-like assets.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 5. Cambridge: Cambridge Univ. Press.

Swanson, E. T. (2011). Let's twist again: a high-frequency event-study analysis of operation twist and its implications for qe2. *Brookings Papers on Economic Activity*, 2011(1):151–188.

Swanson, E. T. (2021). Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, 118:32–53.

Tabova, A. M. and Warnock, F. E. (2021). Foreign investors and us treasuries.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy*, volume 39, pages 195–214. Elsevier.

Vayanos, D. and Vila, J.-L. (2021). A preferred-habitat model of the term structure of interest rates. *Econometrica*, 89(1):77–112.

Wallen, J. (2020). *Markups to financial intermediation in foreign exchange markets*. Stanford University.

Internet Appendix of
"Granular Treasury Demand with Arbitrageurs"

Kristy A.E. Jansen Wenhao Li Lukas Schmid

A. Data Sources and Aggregation

This appendix details the various data sources used to construct our dataset of granular U.S. Treasury holdings and explains how these datasets are merged. Specifically, in A.1 we report the data sources of U.S. Treasury holders, in A.2 we discuss the process of merging datasets of Treasury holdings, and in A.3 we provide data sources for macro variables.

A.1. Treasury Holders

A. Banks - CALL Reports

Banks are major investors in the U.S. Treasury market. We obtain banks' holdings of U.S. Treasuries at the maturity bucket level from CALL reports. CALL reports are regulatory filings required for all U.S. banks and include detailed information on a bank's assets, liabilities, income, and expenses. The CALL reports are filed on a quarterly basis and cover the period from the first quarter of 1976 to the end of 2022. Banks report their aggregate U.S. Treasury holdings and their holdings in different maturity buckets of U.S. Treasuries and U.S. Agency bonds combined. The maturity buckets are: $\tau < 3M$, $3M \leq \tau < 1Y$, $1Y \leq \tau < 3Y$, $3Y \leq \tau < 5Y$, $5Y \leq \tau < 15Y$, $\tau \geq 15Y$. To obtain their allocation to U.S. Treasuries for different maturities, we assume that the fraction of Treasuries versus Agency bonds is fixed across maturities at a given point in time. Hence, at each point in time, we multiply the total maturity bucket holdings by the fraction of Treasuries relative to the sum of Treasuries and Agency bonds.

B. Fed - Federal Reserve

In the aftermath of the Great Financial Crisis, the Federal Reserve has become a major player in the U.S. Treasury market. The Federal Reserve System Open Market Account (SOMA) reports security holdings that are acquired through open market operations by the Fed. These data are obtained through the website of the Federal Reserve Bank of New York.[1] The holdings are at the security (CUSIP) level and reported on a weekly basis since the start of 2003.

[1] `https://www.newyorkfed.org/markets/soma-holdings`

C. Primary Dealers - Federal Reserve

To maintain transparency of U.S. and foreign primary dealers trading activities, their total weekly positions are made available through the website of the Federal Reserve Bank of New York.[2] Primary dealers report their holdings for conventional maturity buckets since early 1998. However, the specific maturity buckets reported change over time. The time frames with the same reporting standards are: January 1998 to June 2001, July 2001 to March 2013, April 2013 to December 2014, January 2015 to December 2021, and from January 2022 onward. Generally, more recent data reports finer maturity buckets. To be consistent across time, we treat July 2001 to March 2013 as the baseline and aggregate the maturity buckets of subsequent periods to match that of this time frame. The final maturity buckets are: T-bills, Treasuries with $\tau \leq 3Y$, $3Y < \tau \leq 6Y$, $6Y < \tau \leq 11Y$, and $\tau > 11Y$.

D. Hedge Funds - Form PF

We obtain aggregate U.S. and foreign hedge fund Treasury positions from Form PF that hedge funds file with the SEC.[3] As of 2011Q4, hedge funds must file Form PF if they are registered or are required to register with the SEC, manage private funds, and have at least $150 million in total assets. The Fed reports the totals separately for domestic and foreign hedge funds. We only observe the aggregate Treasury positions, so we rely on the maturity distribution obtained from primary dealers to infer the maturity bucket holdings. That is, we multiply the maturity bucket weights of primary dealers with the aggregate hedge fund Treasury positions at each point in time to obtain maturity bucket specific hedge fund holdings. The reason we rely on primary dealers to infer the maturity distribution is twofold. First, we define both as arbitrageurs, consistent with the literature (Du et al. 2023b; Vayanos and Vila 2021). Second, corroborating the idea that both hedge funds and primary dealers act as arbitrageurs, the aggregate Treasury holdings of primary dealers and hedge fund align closely in that higher aggregate Treasury holdings for primary dealers tend to come with higher holdings for hedge funds (see Figure A1 of the Appendix).

E. Insurers and Pension Funds - eMAXX

eMAXX provides a comprehensive coverage of fixed income holdings of institutional investors at the security (CUSIP) level. The database predominantly covers the holdings of insurance companies, mutual funds, and pension funds (Becker and Ivashina 2015; Bretscher et al. 2024). We

[2]The data and the list of primary dealers that must report can be found here: `https://www.newyorkfed.org/markets/counterparties/primary-dealers-statistics`. Specifically, the Fed allows certain foreign-owned institutions to operate as primary dealers in the U.S. Treasury market if they meet specific criteria.

[3]We thank Moritz Lenel for directing us to this data source.

only use the data on insurance companies and pension funds, and rely on Morningstar for mutual funds. Due to the voluntary nature of reporting by pension funds, the coverage of pension funds in eMAXX is limited, unlike the mandatory reporting by insurance companies. Additionally, we focus on the U.S. eMAXX database, which covers the holdings of North American investors. The holdings data are quarterly and cover the period from the first quarter of 2010 to the end of 2022.

F. Money Market Funds - IMoneyNet and FoFs

IMoneyNet provides a wide coverage of asset holdings (predominately fixed income and cash) by U.S. money market funds (MMFs) at the security (CUSIP) level. We focus on both holdings reported by MMFs domiciled in the U.S. as well as on their offshore holdings. The holdings are reported on a monthly basis since August 2011.

To obtain a larger coverage of the total MMF population, we augment the data with FoFs from the Federal Reserve. Using our security-level database, we verify that on average 99.6% of MMF holdings are in either T-bills or U.S. Treasuries with remaining time to maturity less than 1 year. Hence, we can reasonably assume that MMF Treasury holdings reported in FoF have remaining maturities below 1 year.

G. Mutual Funds - Morningstar

We obtain holdings data on domestic and foreign mutual funds from Morningstar, Inc. The funds report all their positions including stocks, bonds, and cash at the security (CUSIP) level. We focus on both fixed-income and allocation funds. Funds either report monthly or quarterly, and to maintain consistency across the funds and other data sets we use data at quarter ends. Figure A2 reports the aggregate holdings in USD (trillions) over time. These aggregates align closely with the numbers reported in Maggiori et al. (2020).

H. ETFs - ETF Global

We obtain the holdings of U.S. Exchange Traded Funds (ETFs) at the security (CUSIP) level from ETF Global. ETF Global contains extensive coverage of securities held by U.S. ETFs and in our analysis we focus on fixed-income funds. Funds either report daily or monthly, and to maintain consistency with the other datasets we use data at quarter ends. As U.S. ETFs only hold a small fraction of U.S. Treasuries outstanding, we merge them with the U.S. mutual fund sector.

I. Foreign Official and Private - Public TIC

We obtain quarterly U.S. Treasury holdings by foreign investors from the Treasury International Capital Reporting System (TIC). Specifically, we obtain the public TIC Form SLT that exists as of September 2011. As of this date, TIC also provides a breakdown of the total amount held in T-bills versus non T-bills. As of December 2011, TIC also distinguishes between foreign official and foreign private investors. Moreover, to avoid double counting, we subtract from the private foreign Treasury holdings the holdings of foreign mutual funds that we obtain through Morningstar and foreign hedge funds that we obtain through Form PF.

A.2. Data Aggregation

For the data sources in Table 1 that are at the security level, we observe the corresponding CUSIP identifiers that we use to match the holdings data with the CRSP U.S. Treasury Database. The CRSP U.S. Treasury Database contains detailed bond-level information on U.S. Treasuries, including bond yields, prices, bond type, coupon rate, maturity date, issue date, and issuance size. We use the bond prices to convert nominal holdings to market values. For the sectors that report at a more aggregate level (banks, foreign investors, hedge funds, and primary dealers), we use their reported market value holdings directly.

For investors that report at the CUSIP level, including insurers and pension funds, mutual funds, ETFs, money market funds, and the Fed, it is straightforward to divide their holdings in the respective maturity buckets: $\tau < 1Y, 1Y \leq \tau < 5Y, \tau \geq 5Y$. For banks, we aggregate maturity bucket $\tau < 3M$ and $3M \leq \tau < 1Y$ to obtain the first bucket, $1Y \leq \tau < 3Y$ and $3Y \leq \tau < 5Y$ for the second bucket, and $5Y \leq \tau < 15Y$ and $\tau \geq 15Y$ for the third bucket. We follow a similar approach for the primary dealers, whereby we assign T-bills to bucket 1, $\tau \leq 3Y$ and $3Y < \tau \leq 6Y$ to bucket 2, and $6Y < \tau \leq 11Y$ and $\tau > 11Y$ to bucket 3. As motivated earlier, we assume that hedge funds have the same maturity bucket distribution as primary dealers.

For foreign investors, we only observe the fraction that is held in T-bills versus non T-bills. To allocate the foreign holdings to different maturity buckets, we first multiply the T-bill holdings by the inverse of the fraction of the total amount outstanding in maturity bucket 1 of the CRSP universe that is in T-bills, at each point in time. The reason is that on average only 60% of the total amount outstanding in maturity bucket 1 consists of T-bills, while the remaining 40% are bonds and notes with remaining time to maturity below 1 year. This adjustment is meant to more accurately reflect the remaining maturity structure, but our estimations for foreign investors are similar when we assume that T-bills are the only securities held in bucket 1. We then subtract the additional fraction we attribute to maturity bucket 1 from the total non T-bill holdings to compute the total holdings in the remaining maturity buckets. To further determine the fraction in maturity bucket 2 versus 3, we choose the fraction such that the average duration of both the foreign official and foreign private investors' Treasury portfolio is consistent with Tabova and Warnock (2021) at each point in time. To assign the fractions, we take the bond durations of a 6-month, 3-year, and 15-year bond, respectively, as representative bonds for each maturity bucket. However, our main results do not depend on this choice. For instance, the results are qualitatively and quantitatively similar if we choose instead a 10-year or a 20-year bond for the third bucket.

To obtain the residual sector, we subtract the holdings of all investors from the total amount outstanding in each bucket. Since we observe the total foreign investor position, the residual sector consists of U.S. based investors only and hence we will refer to this sector as "Other U.S. Investors".

Finally, in our growing economic environment, portfolio holdings in dollar values will not be stationary. For stationarity, we scale all quantities in our regressions and in the model by the ratio of potential GDP (ticker "NGDPPOT" in FRED, which is nominal potential gross domestic product) at the end of our sample period over the potential GDP at that particular quarter. For example, the ratio of potential GDP in 2022 Q4 to that in 2011 Q4 is 1.6. The dollar value of total debt supply in 2011 Q4 is 10.7 trillion, but we use a scaled value, namely $10.7 * 1.6 = 17.1$ trillion. We use nominal values so that the scaling adjusts for the inflation effect. Moreover, using a GDP adjuster rather than just inflation ensures that we account for the growing scale of the economy. Finally, we use nominal potential GDP rather than nominal GDP to avoid cyclical fluctuations in nominal GDP that causes mechanical correlations among the variables due to the scaling. The underlying assumption is that after accounting for the scaling effect, all quantities are stationary in the fundamental state variables. An alternative scaling is to use a constant exponential growth rate matching the overall economic growth during our sample period, and we find that this approach leads to similar results.

A.3. Macro Data

We complement our dataset with a number of macroeconomic variables that capture relevant drivers of monetary and fiscal policy stances, as well as aggregate economic conditions. Specifically, we obtain four macro variables from the Federal Reserve Economic Data (FRED).

First, we include the GDP gap and core inflation to capture aggregate economic conditions as well as the response of monetary policy to macroeconomic dynamics. They together reflect aggregate demand and supply fluctuations in the economy, and they are also the variables that drive monetary policy in the Taylor rule.

Second, we include the debt/GDP ratio to capture the overall supply and dynamics of government debt. As an indicator of the government's fiscal policy stance, the debt/GDP ratio is plausibly connected to the GDP gap, as well as inflation.

Finally, as an indicator of financial market conditions relevant to the aggregate economy, we include credit spreads, which have been widely shown to predict macroeconomic movements (Gilchrist and Zakrajšek 2012; Krishnamurthy and Muir 2017).

B. Identification of the Instrument

To illustrate the identification of our instrument, we assume a simplified setting of one asset with maturity τ and price $P_t = \frac{1}{(1+y_t)^\tau}$. We also assume one investor and fixed supply S.

Let's assume that the data-generating-process of demand is given by:

$$Z_t = \theta + b_1 y_t + (b_2)'\mathbf{x}_t + (b_3)'\mathbf{Macro}_t + u_t \tag{A1}$$

The instrument is then constructed from a pseudo market clearing $\hat{Z}_t = \frac{S}{(1+y_t)^\tau}$ as:

$$\hat{Z}_t = \hat{\theta} + (\hat{b}_2)'\mathbf{x}_t + (\hat{b}_3)'\mathbf{Macro}_t = \frac{S}{(1+\tilde{y}_t)^\tau} \tag{A2}$$

Solving for $\tilde{y}_t$, we obtain:

$$\tilde{y}_t = \left(\frac{\hat{Z}_t}{S}\right)^{-\frac{1}{\tau}} - 1 = \left(\frac{\hat{\theta} + (\hat{b}_2)'\mathbf{x}_t + (\hat{b}_3)'\mathbf{Macro}_t}{S}\right)^{-\frac{1}{\tau}} - 1 \tag{A3}$$

Plugging back into Equation (A1), we have:

$$Z_t = \theta + b_1 \left(\left(\frac{\hat{\theta} + (\hat{b}_2)'\mathbf{x}_t + (\hat{b}_3)'\mathbf{Macro}_t}{S}\right)^{-\frac{1}{\tau}} - 1\right) + (b_2)'\mathbf{x}_t + (b_3)'\mathbf{Macro}_t + u_t \tag{A4}$$

Hence, the relationship between the pseudo yield $\tilde{y}_t$ and the bond characteristics $\mathbf{x}_t$ and macro variables $\mathbf{Macro}_t$ are not collinear because of the non-linearity that stems from the convexity effect of compounding interest. In our main specification, we also obtain predicted supply based on the FFR and the macro variables. In that case, the denominator of Equation (A4) would also contain the macro variables $\mathbf{Macro}_t$, adding yet another layer of non-linearity.

C. Additional Empirical Analysis

Table A1. **Summary Statistics**

This table provides summary statistics of the main variables of interest: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m, Coupon Rate, Bid-Ask Spread, Credit Spread, Debt/GDP, GDP Gap, and Core Inflation.

	mean	sd	min	max
$y_t(m)$	1.400	1.081	0.041	4.291
$y_t(-m)$	1.469	0.902	0.132	4.289
Coupon Rate	2.039	0.883	0.750	4.158
Bid-Ask Spread	0.046	0.028	0.010	0.096
Credit Spread	0.949	0.233	0.550	1.490
Debt/GDP	0.762	0.095	0.654	0.974
GDP Gap	-1.329	1.910	-9.106	1.846
Core Inflation	2.461	1.322	1.173	6.429

Table A2. **Correlation Table**

This table provides the correlation table of the main variables of interest: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m, Coupon Rate, Bid-Ask Spread, Debt/GDP, Credit Spread, GDP Gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to the maturity fixed effects.

	$y_t(m)$	$y_t(-m)$	Coupon	Bid-Ask Spread	Credit Spread	Debt/GDP	GDP Gap	Inflation
$y_t(m)$	1							
$y_t(-m)$	0.58	1						
Coupon	-0.09	-0.28	1					
Bid-Ask Spread	0.02	-0.03	-0.31	1				
Credit Spread	-0.01	-0.03	0.29	-0.07	1			
Debt/GDP	-0.10	-0.15	-0.57	0.48	-0.14	1		
GDP Gap	0.47	0.55	-0.40	0.24	-0.26	0.17	1	
Inflation	0.40	0.49	-0.49	0.01	-0.01	0.43	0.57	1

Table A3. **Demand System Results - OLS**

This table shows the OLS estimates of our demand system specified in Equation (8). The dependent variable is the market value of US Treasuries held by sector ι in maturity bucket m at time t. The independent variables are: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m, Coupon Rate, Bid-Ask Spread, indicator variable if the holdings are in maturity bucket 2 ($\mathbb{1}\{1Y \leq \tau < 5\}$), indicator variable if the holdings are in maturity bucket 3 ($\mathbb{1}\{\tau \geq 5\}$), Credit Spread, Debt/GDP, GDP Gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to the maturity fixed effects. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Banks	**ICPF**	**MF ROW**	**MF US**	**MMF**	**Residual**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	56.676***	-4.351	3.845**	34.136**	26.371	94.762	-203.031***	-84.390
	[14.325]	[6.016]	[1.574]	[15.409]	[82.959]	[96.325]	[61.419]	[57.919]
$y_t(-m)$	-57.315***	5.909	-0.143	-31.597**	105.763	-29.348	77.017	112.609
	[14.624]	[7.585]	[1.881]	[15.447]	[109.295]	[120.441]	[67.085]	[75.499]
Coupon Rate	-135.101***	10.707	-1.277	-17.017	512.046	191.910	-304.136***	-163.445
	[23.444]	[12.861]	[2.925]	[33.696]	[390.244]	[233.422]	[114.897]	[121.617]
Bid-Ask Spread	7.806	19.847***	3.413***	24.162*	19.679	129.796**	-80.549**	-54.510
	[7.638]	[4.834]	[1.143]	[13.148]	[96.878]	[65.928]	[38.206]	[47.171]
$\mathbb{1}\{1Y \leq \tau < 5\}$	58.001***	151.759***	14.013***	224.702***		-391.160***	2983.216***	-308.538***
	[12.694]	[5.180]	[1.626]	[20.429]		[119.457]	[68.496]	[89.159]
$\mathbb{1}\{\tau \geq 5\}$	-51.544**	197.389***	15.426***	231.527***		551.566***	456.849***	274.318**
	[23.479]	[12.992]	[2.969]	[26.707]		[203.121]	[114.795]	[114.354]
Credit Spread	11.640	-13.700	-0.022	-65.416**	-205.221	290.260	57.080	-66.427
	[18.426]	[12.313]	[2.458]	[32.228]	[151.589]	[191.649]	[84.304]	[124.695]
Debt/GDP	697.984***	-3.425	47.785***	200.882*	7587.435***	1757.942**	-1590.849***	998.612*
	[63.868]	[43.153]	[9.359]	[106.771]	[537.329]	[878.071]	[511.343]	[511.718]
GDP Gap	10.028***	-4.249**	1.406***	11.035***	-69.571***	3.596	-8.555	4.775
	[3.480]	[1.753]	[0.435]	[4.187]	[24.891]	[28.517]	[16.147]	[16.596]
Core Inflation	15.072**	0.410	-2.171***	-1.397	-90.427*	17.296	-63.522*	0.139
	[6.703]	[3.189]	[0.766]	[8.157]	[52.450]	[50.306]	[35.251]	[33.024]
R-squared	0.903	0.914	0.843	0.855	0.946	0.657	0.979	0.501
Observations	135	135	135	135	45	135	135	135

Table A4. **First Stage IV**

This table shows the first stage estimates of the IV methodology specified in Equation (8). The dependent variable in Column (1) is $y_t(m)$, the value-weighted yield of maturity bucket m and in Column (2) is $y_t(-m)$, the value-weighted yield of the other maturity buckets $-m$. We instrument own and other yield using pseudo yields specified in Section 3.2. Additional variables include Coupon Rate, Bid-Ask Spread, indicator variable if the holdings are in maturity bucket 2 ($\mathbb{1}\{1Y \leq \tau < 5\}$), indicator variable if the holdings are in maturity bucket 3 ($\mathbb{1}\{\tau \geq 5\}$), Credit Spread, Debt/GDP, GDP gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to the maturity fixed effects. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	$y_t(m)$	$y_t(-m)$
	(1)	(2)
$I_t(m)$	0.722***	0.437***
	[0.046]	[0.048]
$I_t(-m)$	0.690***	0.876***
	[0.118]	[0.076]
Coupon Rate	-0.281	-0.784***
	[0.191]	[0.157]
Bid-Ask Spread	-0.024	-0.077
	[0.069]	[0.061]
$\mathbb{1}\{1Y \leq \tau < 5\}$	-1.027***	-0.055
	[0.311]	[0.223]
$\mathbb{1}\{\tau \geq 5\}$	-0.042	-0.783***
	[0.253]	[0.199]
Credit Spread	0.833***	0.898***
	[0.202]	[0.171]
Debt/GDP	1.804**	0.752
	[0.730]	[0.662]
GDP Gap	-0.237***	-0.225***
	[0.047]	[0.028]
Core Inflation	0.160***	0.151***
	[0.062]	[0.048]
R-squared	0.890	0.894
Observations	135	135

Table A5. **Demand System Results - IV no macro**

This table shows the IV estimates of our demand system specified in Equation (8). The dependent variable is the market value of US Treasuries held by sector ι in maturity bucket m at time t. The endogenous variables are: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m. We instrument own and other yield using pseudo yields specified in Section 3.2. Additional variables include Coupon Rate, Bid-Ask Spread, indicator variable if the holdings are in maturity bucket 2 ($\mathbb{1}\{1Y \leq \tau < 5\}$), and indicator variable if the holdings are in maturity bucket 3 ($\mathbb{1}\{\tau \geq 5\}$). We orthogonalize the coupon and the bid-ask spread with respect to the maturity fixed effects. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Banks	**ICPF**	**MF ROW**	**MF US**	**MMF**	**Residual**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	146.066***	0.557	8.414**	141.113***	1210.993***	313.912	-293.653**	99.110
	[38.843]	[9.782]	[4.275]	[49.284]	[434.761]	[204.318]	[122.231]	[86.344]
$y_t(-m)$	-161.043***	-5.469	-5.703	-130.824***	-2285.615***	-318.431	199.981	-123.476
	[44.117]	[12.116]	[4.372]	[50.581]	[574.119]	[244.400]	[136.224]	[101.632]
Coupon Rate	-360.163***	7.351	-9.967**	-141.013***	-4860.838***	-248.146	193.798	-497.699***
	[47.643]	[13.938]	[4.534]	[46.457]	[1619.799]	[245.908]	[160.971]	[132.489]
Bid-Ask Spread	22.638	15.807***	5.994***	21.304	1131.961**	161.940*	-118.377**	-46.004
	[16.315]	[4.264]	[1.322]	[14.153]	[446.927]	[84.296]	[49.492]	[58.635]
$\mathbb{1}\{1Y \leq \tau < 5\}$	27.473	150.612***	12.477***	185.848***		-462.763***	3012.506***	-369.049***
	[23.555]	[4.969]	[2.554]	[29.830]		[121.314]	[93.922]	[86.631]
$\mathbb{1}\{\tau \geq 5\}$	-220.327***	185.541***	6.683	40.850		121.980	636.045***	-82.673
	[75.197]	[18.561]	[8.187]	[92.872]		[409.193]	[225.281]	[162.500]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap statistic (*first stage*):	17.92	17.92	17.92	17.92	6.09	17.92	17.92	17.92

Table A6. **Demand System Results - IV alternative pseudo yield**

This table shows the IV estimates of our demand system specified in Equation (8). The dependent variable is the market value of U.S. Treasuries held by sector ι in maturity bucket m at time t, adjusted for GDP potential. The endogenous variables are: $y_t(m)$, which is the value-weighted yield of maturity bucket m, $y_t(-m)$, which is the value-weighted yield of the other maturity buckets excluding maturity bucket m. We instrument own and other yield using pseudo yields specified in Section 3.2, but we leave out the bid-ask spread, credit spread, and core inflation in determining the pseudo yields. Additional control variables include Coupon Rate, Bid-Ask Spread, indicator variable if the holdings are in maturity bucket 2 ($\mathbb{1}\{1Y \leq \tau < 5\}$), indicator variable if the holdings are in maturity bucket 3 ($\mathbb{1}\{\tau \geq 5\}$), Credit Spread, Debt/GDP, Credit Spread, GDP Gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to the maturity fixed effects. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Banks	**ICPF**	**MF ROW**	**MF US**	**MMF**	**Other US Investors**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	80.308***	4.761	6.377**	119.801***	449.737**	145.170	-93.150	26.001
	[24.048]	[9.694]	[3.109]	[38.100]	[223.508]	[175.354]	[93.762]	[88.226]
$y_t(-m)$	-91.837***	-2.831	-3.588	-138.151***	-633.948**	24.210	-22.188	-39.676
	[26.610]	[11.760]	[3.483]	[44.329]	[321.304]	[229.082]	[125.733]	[118.286]
Coupon Rate	-168.230***	1.571	-4.624	-122.320**	42.489	222.458	-409.389**	-311.057*
	[32.920]	[17.577]	[4.390]	[53.242]	[531.712]	[290.088]	[160.986]	[166.095]
Bid-Ask Spread	5.943	18.617***	3.185***	15.328	140.023	111.588	-96.074**	-64.218
	[8.048]	[4.467]	[1.196]	[15.390]	[131.357]	[77.183]	[43.538]	[55.851]
$\mathbb{1}\{1Y \leq \tau < 5\}$	50.591***	148.476***	13.195***	196.112***		-418.811***	2943.054***	-343.984***
	[13.944]	[4.538]	[1.933]	[23.794]		[120.090]	[79.762]	[82.811]
$\mathbb{1}\{\tau \geq 5\}$	-99.390**	181.017***	10.415*	66.530		507.277	262.229	54.898
	[44.827]	[17.588]	[5.903]	[74.048]		[363.952]	[183.827]	[174.674]
Credit Spread	19.663	-11.724	0.776	-40.877	-522.200***	276.286	79.398	-31.145
	[21.765]	[13.775]	[2.446]	[38.098]	[193.865]	[186.482]	[83.831]	[129.543]
Debt/GDP	610.977***	-12.415	39.827***	-14.686	5529.946***	2255.687**	-1666.044***	640.370
	[81.620]	[48.080]	[10.155]	[125.989]	[1007.451]	[916.652]	[554.062]	[540.285]
GDP Gap	11.227***	-4.422**	1.499***	12.799**	-75.848***	-11.527	-11.501	9.130
	[3.792]	[1.910]	[0.477]	[5.149]	[21.716]	[30.398]	[16.927]	[17.973]
Core Inflation	16.635**	-0.312	-2.077**	-1.116	63.786	-16.243	-74.024*	4.843
	[6.597]	[3.322]	[0.835]	[10.110]	[75.793]	[49.773]	[37.837]	[33.332]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap Statistic (*first stage*)	15.43	15.43	15.43	15.43	4.05	15.43	15.43	15.43

Table A7. **Demand System Results - IV no other yield**

This table shows the IV estimates of our demand system specified in Equation (8), excluding other yield. The dependent variable is the market value of US Treasuries held by sector ι in maturity bucket m at time t. The endogenous variable is $y_t(m)$, which is the value-weighted yield of maturity bucket m. We instrument own and other yield using pseudo yields specified in Section 3.2. Additional variables include Coupon Rate, Bid-Ask Spread, indicator variable if the holdings are in maturity bucket 2 ($\mathbb{1}\{1Y \leq \tau < 5\}$), indicator variable if the holdings are in maturity bucket 3 ($\mathbb{1}\{\tau \geq 5\}$), Credit Spread, Debt/GDP, Credit Spread, GDP Gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to the maturity fixed effects. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Banks	**ICPF**	**MF ROW**	**MF US**	**MMF**	**Residual**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	20.646*	3.087	4.741***	46.022**	80.663**	161.608**	-90.289**	7.108
	[12.225]	[4.713]	[1.545]	[18.159]	[39.923]	[66.379]	[39.403]	[36.768]
Coupon Rate	-80.866***	4.224	-1.377	5.280	414.985	199.257	-392.417***	-274.961**
	[20.840]	[13.467]	[2.122]	[34.344]	[407.796]	[195.077]	[91.510]	[109.395]
Bid-Ask Spread	10.123	18.706***	3.180***	17.746	46.508	110.314	-99.250**	-64.080
	[9.395]	[4.557]	[1.183]	[14.364]	[95.683]	[74.943]	[42.695]	[53.563]
$\mathbb{1}\{1Y \leq \tau < 5\}$	68.861***	148.965***	13.597***	216.415***		-423.947***	2939.702***	-339.186***
	[15.583]	[4.457]	[1.794]	[23.204]		[118.893]	[74.681]	[83.942]
$\mathbb{1}\{\tau \geq 5\}$	23.528	184.582***	14.272***	229.698***		473.907***	268.412***	98.634
	[19.173]	[8.277]	[2.710]	[29.847]		[141.663]	[77.168]	[67.383]
Credit Spread	-1.735	-12.387	-0.072	-73.347**	-243.628*	281.914	73.925	-40.510
	[20.063]	[13.295]	[2.560]	[36.008]	[133.309]	[193.586]	[85.946]	[126.429]
Debt/GDP	855.805***	-4.181	52.287***	420.152***	7277.569***	2194.104**	-1534.920***	774.817
	[86.977]	[38.086]	[10.768]	[142.550]	[390.054]	[890.121]	[457.434]	[508.858]
GDP Gap	7.548**	-4.561**	1.249***	4.832	-69.348***	-10.665	-15.021	6.492
	[3.607]	[1.969]	[0.407]	[4.217]	[22.415]	[31.623]	[13.238]	[15.981]
Core Inflation	11.329	-0.535	-2.533***	-14.805	-68.659*	-15.098	-81.479*	0.091
	[8.763]	[3.457]	[0.869]	[9.636]	[38.799]	[52.227]	[43.071]	[35.375]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap statistic (*first stage*):	116.35	116.35	116.35	116.35	618.66	116.35	116.35	116.35

Figure A1. **U.S. Treasury Holdings Hedge Funds versus Primary Dealers**. This graph shows the aggregate holdings of U.S. Treasuries (in billions) by hedge funds (left y-axis) and primary dealers (right y-axis) over time.

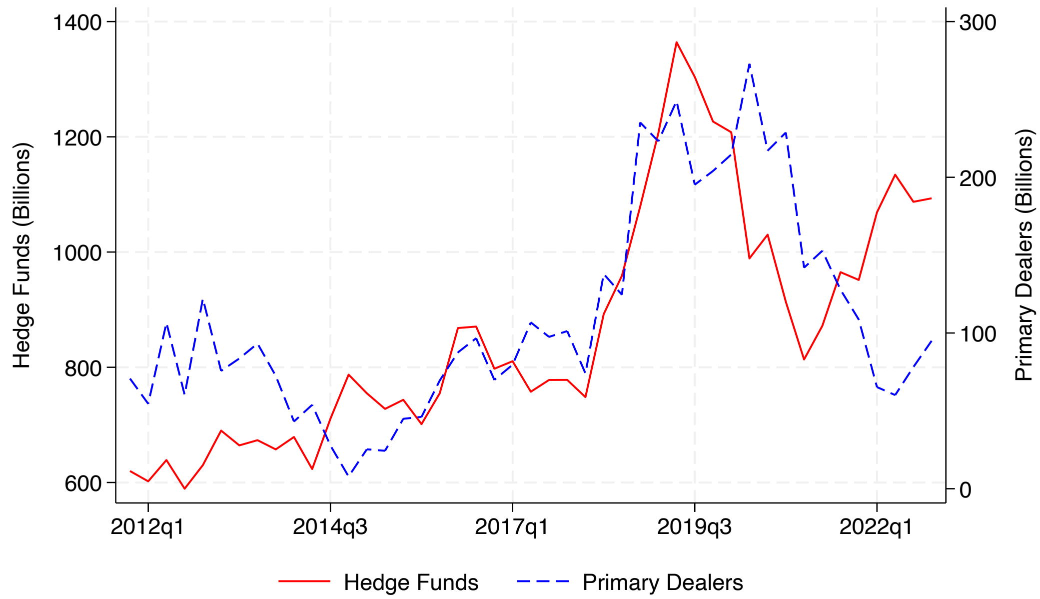
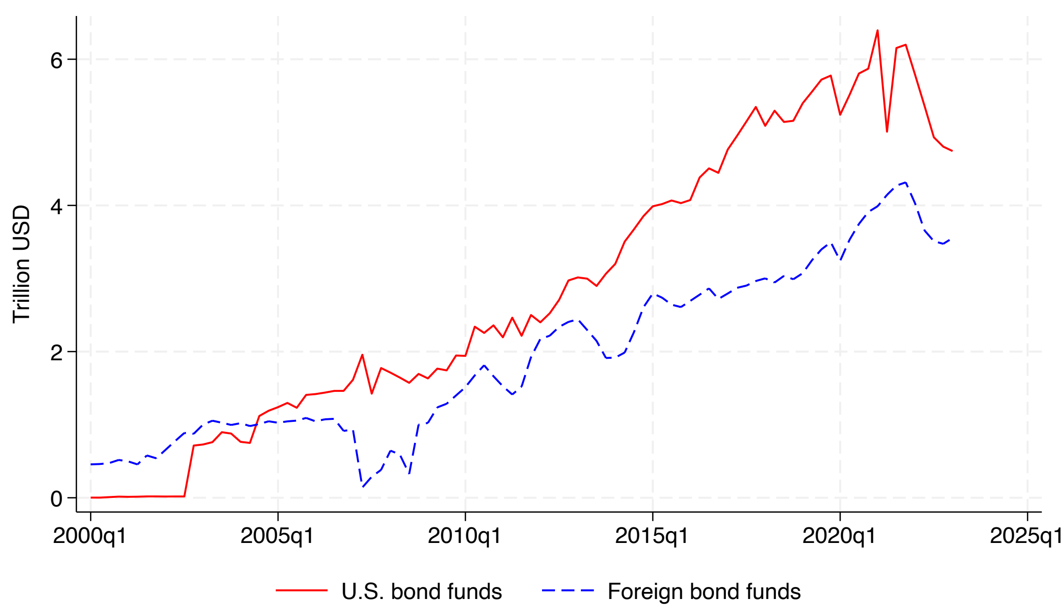


Figure A2. **Morningstar Aggregate Holdings by Domestic and Foreign Bond Funds**. This graph shows the aggregate holdings of US and foreign bond funds in USD (trillions) over time.



D. Model Derivations and Estimation

D.1. Derivations for the Full Model

As noted, we conjecture an affine solution of the model of the form (26). In order to solve the model, we need to pin down the matrices A, A_r, and A_u, as well as the vector C. We next outline the critical steps in the model solution.

We start with the holding return of bonds with maturity τ from t to $t+1$, using (15) and (26),

$$\begin{aligned} r_{t+1}^{(\tau)} &= p_{t+1}^{(\tau-1)} - p_t^{(\tau)} \\ &= A(\tau-1)'\beta_{t+1} + A_r(\tau-1)r_{t+1} - A(\tau)'\cdot\beta_t - A_r(\tau)r_t + A_u(\tau-1)'u_{t+1} - A_u(\tau)'u_t \\ &= A(\tau-1)'(\bar{\beta} + \Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) + A_r(\tau-1)(\bar{r} + \phi_r'(\Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) + \rho_r r_t + \sigma_r \varepsilon_{t+1}^r) \\ &\quad - A(\tau)'\cdot\beta_t - A_r(\tau)r_t + A_u(\tau-1)'u_{t+1} - A_u(\tau)'u_t + C(\tau-1) - C(\tau). \end{aligned} \tag{A5}$$

We can approximate the total holding return as

$$R_{t+1}^{(\tau)} = \exp(r_{t+1}^{(\tau)}) - 1 \approx r_{t+1}^{(\tau)} + \frac{1}{2}Var_t[r_{t+1}^{(\tau)}], \tag{A6}$$

which becomes exact when we take a continuous-time approach. Refer to Greenwood et al. (2023) for a more detailed discussion. Since there is no uncertainty regarding the current short rate, this approximation also leads to $R_{t+1} = R_{t+1}^{(1)} = \exp(r_t) - 1 \approx r_t$.

With (A5) and (A6), we can express the total return as

$$\begin{aligned} R_{t+1}^{(\tau)} &= A(\tau-1)'(\bar{\beta} + \Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) - A(\tau)'\cdot\beta_t + C(\tau-1) - C(\tau) \\ &\quad + \frac{1}{2}\left(A(\tau-1)' + A_r(\tau-1)\phi_r'\right)\Sigma\left(A(\tau-1) + \phi_r A_r(\tau-1)\right) \\ &\quad + A_r(\tau-1)(\bar{r} + \phi_r'(\Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) + \rho_r r_t + \sigma_r\varepsilon_{t+1}^r) - A_r(\tau)r_t \\ &\quad + \frac{1}{2}(A_r(\tau-1)\sigma_r)^2 + A_u(\tau-1)'u_{t+1} - A_u(\tau)'u_t + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1). \end{aligned} \tag{A7}$$

We note that the return $R_{t+1}^{(\tau)}$ in (A7) contains four important components. The first one reflects innovations to the macroeconomic factors β_t. The second one reflects innovations to latent demand u_t. The third one is the innovation to the monetary policy rate r_t. The final components are the Jensen terms for each type of risk, including the macroeconomic shocks, monetary policy shocks, and latent demand shocks.

To simplify expressions, we denote

$$\hat{A}(\tau-1) = A(\tau-1) + \phi_r A_r(\tau-1), \tag{A8}$$

so that $\hat{A}(\tau-1)' = A(\tau-1)' + A_r(\tau-1)\phi_r'$. Therefore, Equation (A7) can be simplified as

$$\begin{aligned} R_{t+1}^{(\tau)} &= A(\tau-1)'(\bar{\beta} + \Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) - A(\tau)'\cdot\beta_t + C(\tau-1) - C(\tau) \\ &+ \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) + A_u(\tau-1)'u_{t+1} - A_u(\tau)'u_t + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) \\ &+ A_r(\tau-1)(\bar{r} + \phi_r'(\Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) + \rho_r r_t + \sigma_r \varepsilon_{t+1}^r) - A_r(\tau) r_t + \frac{1}{2}(A_r(\tau-1)\sigma_r)^2. \end{aligned} \tag{A9}$$

Wealth thus evolves as

$$\begin{aligned} W_{t+1} &= W_t(1+r_t) + \sum_{\tau=2}^{N} X_t^{(\tau)}(R_{t+1}^{(\tau)} - r_t) + \tilde{X}_t(\tilde{R}_{t,t+1} - r_t) \\ &= W_t(1+r_t) + \tilde{X}_t(\tilde{R}_{t,t+1} - r_t) + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) \\ &+ \sum_{\tau=2}^{N} X_t^{(\tau)} \begin{pmatrix} A(\tau-1)'(\bar{\beta} + \Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) - A(\tau)'\cdot\beta_t + \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) \\ + A_u(\tau-1)'u_{t+1} - A_u(\tau)'u_t + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) \\ + A_r(\tau-1)(\bar{r} + \phi_r'(\Phi(\beta_t - \bar{\beta}) + \Sigma^{1/2}\varepsilon_{t+1}) + \rho_r r_t + \sigma_r \varepsilon_{t+1}^r) - A_r(\tau) r_t \\ + C(\tau-1) - C(\tau) + \frac{1}{2}(A_r(\tau-1)\sigma_r)^2 - r_t \end{pmatrix} \\ &= W_t(1+r_t) + \sum_{\tau=2}^{N} X_t^{(\tau)} \begin{pmatrix} A(\tau-1)'\left(\bar{\beta} + \Phi(\beta_t - \bar{\beta})\right) - A(\tau)'\beta_t + \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) \\ - A_u(\tau)'u_t + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) + C(\tau-1) - C(\tau) \\ + A_r(\tau-1)(\bar{r} + \phi_r'\Phi(\beta_t - \bar{\beta}) + \rho_r r_t) - A_r(\tau) r_t + \frac{1}{2}(A_r(\tau-1)\sigma_r)^2 - r_t \end{pmatrix} \\ &+ \left(\sum_{\tau=2}^{N} X_t^{(\tau)}\left(A(\tau-1)'\Sigma^{1/2} + A_r(\tau-1)\phi_r'\Sigma^{1/2}\right) + \tilde{X}_t\tilde{\sigma}'\right)\varepsilon_{t+1} + \left(\sum_{\tau=2}^{N} X_t^{(\tau)} A_r(\tau-1)\sigma_r + \tilde{X}_t\tilde{\sigma}_r'\right)\varepsilon_{t+1}^r \\ &+ \left(\sum_{\tau=2}^{N} X_t^{(\tau)} A_u(\tau-1)'\right)u_{t+1} + \tilde{X}_t(\tilde{\phi}'\beta_t + \tilde{\phi}_r r_t - r_t). \end{aligned} \tag{A10}$$

To simplify notatations, it is convenient to define the expected return on Treasuries of maturity τ

as

$$\mu_t^{(\tau)} = A(\tau-1)'\left(\bar{\beta}+\Phi(\beta_t-\bar{\beta})\right) - A(\tau)'\beta_t + \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) - A_u(\tau)'u_t + C(\tau-1) - C(\tau)$$
$$+\frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) + A_r(\tau-1)(\bar{r}+\phi_r'\Phi(\beta_t-\bar{\beta})+\rho_r r_t) - A_r(\tau)r_t + \frac{1}{2}(A_r(\tau-1)\sigma_r)^2. \tag{A11}$$

In that case, we obtain expected next-period wealth

$$E_t[W_{t+1}] = W_t(1+r_t) + \sum_{\tau=2}^{N} X_t(\tau)\left(\mu_t^{(\tau)} - r_t\right) + \tilde{X}_t(\tilde{\phi}'\beta_t + \tilde{\phi}_r r_t - r_t),$$

and variance of next-period wealth

$$\begin{aligned}
Var_t(W_{t+1}) &= \left(\sum_{\tau=2}^{N} X_t(\tau)\hat{A}(\tau-1)'\Sigma^{1/2} + \tilde{X}_t\tilde{\sigma}'\right)\left(\sum_{\tau=2}^{N} X_t(\tau)\Sigma^{1/2}\hat{A}(\tau-1) + \tilde{X}_t\tilde{\sigma}\right) \\
&+ \left(\sum_{\tau=2}^{N} X_t(\tau)A_r(\tau-1)\sigma_r + \tilde{X}_t\tilde{\sigma}_r'\right)^2 \\
&+ \left(\sum_{\tau=2}^{N} X_t(\tau)A_u(\tau-1)'(\Sigma^u)^{1/2}\right)\left((\Sigma^u)^{1/2}\sum_{\tau=2}^{N} X_t(\tau)A_u(\tau-1)\right) \\
&= \sum_{\tau=2}^{N}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1)(X_t(\tau))^2 + 2\sum_{\hat{\tau}\neq\tau}\hat{A}(\tau-1)'\Sigma\hat{A}(\hat{\tau}-1)X_t(\tau)X_t(\hat{\tau}) \\
&+ 2\sum_{\tau=2}^{N}\hat{A}(\tau-1)'\Sigma^{1/2}\tilde{\sigma}\cdot(X_t(\tau)\tilde{X}_t) + \tilde{\sigma}'\tilde{\sigma}(\tilde{X}_t)^2 + \left(\sum_{\tau=2}^{N} X_t(\tau)A_r(\tau-1)\sigma_r + \tilde{X}_t\tilde{\sigma}_r'\right)^2 \\
&+ \sum_{\tau=2}^{N} A_u(\tau-1)'\Sigma^u A_u(\tau-1)(X_t(\tau))^2 + 2\sum_{\hat{\tau}\neq\tau} A_u(\tau-1)'\Sigma^u A_u(\hat{\tau}-1)X_t(\tau)X_t(\hat{\tau}).
\end{aligned}$$

Consequently, we can write the FOC of arbitrageurs in (34) as

$$\begin{aligned}
\mu_t^{(\tau)} - r_t &= \gamma\left(\sum_{\hat{\tau}=2}^{N} \hat{A}(\tau-1)'\Sigma\hat{A}(\hat{\tau}-1)X_t(\hat{\tau}) + \hat{A}(\tau-1)'\Sigma^{1/2}\tilde{\sigma}\tilde{X}_t\right) \\
&+ \gamma\left(\sum_{\hat{\tau}=2}^{N} A_r(\tau-1)\sigma_r^2 A_r(\hat{\tau}-1)X_t(\hat{\tau}) + A_r(\tau-1)'\sigma_r\tilde{\sigma}_r\tilde{X}_t\right) \\
&+ \gamma\left(\sum_{\hat{\tau}=2}^{N} A_u(\tau-1)'\Sigma^u A_u(\hat{\tau}-1)X_t(\hat{\tau})\right) \\
&= \hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\Sigma\hat{A}(\hat{\tau}-1)X_t(\hat{\tau})\right) + \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t\right) \\
&+ A_r(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\sigma_r^2 A_r(\hat{\tau}-1)X_t(\hat{\tau})\right) + \sigma_r\tilde{\sigma}_r\tilde{X}_t\right) \\
&+ A_u(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N} \Sigma^u A_u(\hat{\tau}-1)X_t(\hat{\tau})\right).
\end{aligned} \tag{A12}$$

$$\tilde{\phi}'\beta_t + \tilde{\phi}_r r_t - r_t = \gamma\left(\sum_{\tau=2}^{N} A(\tau-1)'\Sigma^{1/2}\tilde{\sigma}\cdot X_t(\tau) + \tilde{\sigma}'\tilde{\sigma} + \sum_{\tau=2}^{N} A_r(\tau-1)'\sigma_r\tilde{\sigma}_r\cdot X_t(\tau) + (\tilde{\sigma}_r)^2\right). \tag{A13}$$

Defining the prices of risk as

$$\lambda_{\beta,t} = \gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\Sigma\hat{A}(\hat{\tau}-1)X_t(\hat{\tau})\right) + \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t\right), \tag{A14}$$

$$\lambda_{r,t} = \gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\sigma_r^2 A_r(\hat{\tau}-1)X_t(\hat{\tau})\right) + \sigma_r\tilde{\sigma}_r\tilde{X}_t\right), \tag{A15}$$

$$\lambda_{u,t} = \gamma\left(\sum_{\hat{\tau}=2}^{N} \Sigma^u A_u(\hat{\tau}-1)X_t(\hat{\tau})\right). \tag{A16}$$

Using definitions in (A14), (A15), and (A16), and expanding $\mu_t^{(\tau)}$ with (A11), we rewrite arbitrageur FOC in (A12) as

$$\begin{gathered}
A(\tau-1)'\left(\bar{\beta} + \Phi(\beta_t - \bar{\beta})\right) - A(\tau)'\beta_t + \tfrac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) + A_r(\tau-1)(\bar{r} + \phi_r'\Phi(\beta_t - \bar{\beta}) + \rho_r r_t) \\
+ C(\tau-1) - C(\tau) - A_r(\tau)r_t + \tfrac{1}{2}(A_r(\tau-1)\sigma_r)^2 - A_u(\tau)'u_t + \tfrac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) - r_t \\
= \hat{A}(\tau-1)'\lambda_{\beta,t} + A_r(\tau-1)\lambda_{r,t} + A_u(\tau-1)'\lambda_{u,t}.
\end{gathered} \tag{A17}$$

Ultimately, these coefficients are pinned down in equilibrium, that is, when markets clear. The market clearing condition is

$$Z_t(\tau) + X_t(\tau) = S_t(\tau). \tag{A18}$$

for maturity $\tau \in \{1, 2, \cdots, N\}$. As a next step, using expressions for $Z_t(\tau)$ in (18) and $S_t(\tau)$ in (20), we express the equilibrium arbitrageur holdings solved from (A18) as

$$X_t(\tau) = \left(\bar{S}(\tau) + \zeta(\tau)'\beta_t + \zeta_r(\tau)'r_t\right) - \left(\theta_0(\tau) - \alpha(\tau)'p_t - \theta(\tau)'\beta_t + u_t(\tau)\right). \tag{A19}$$

As a result, our model implies that the price of risks $\lambda_{\beta,t}$, $\lambda_{r,t}$, and $\lambda_{u,t}$ all vary over time, and depends on the quantity of Treasury supply $S_t(\tau)$, non-arbitrageur demand $Z_t(\tau)$, as well as the outside portfolio returns $\tilde{X}_t$.

In the main text, we impose the assumption that $A_u(\tau-1)'\lambda_{u,t} \approx 0$, which holds well quantitatively after we estimate the model. The idea is that idiosyncratic latent demand shocks do not affect price of risks. Under this simplification assumption, we plug (A19) into the pricing equation (A17) and expand $\lambda_{\beta,t}$ and $\lambda_{r,t}$ using (A14) and (A15),

$$\begin{aligned}
& A(\tau-1)'\left(\bar{\beta} + \Phi(\beta_t - \bar{\beta})\right) - A(\tau)'\beta_t + \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) + C(\tau-1) - C(\tau) \\
& A_r(\tau-1)(\bar{r} + \phi_r'\Phi(\beta_t - \bar{\beta}) + \rho_r r_t) - A_r(\tau)r_t + \frac{1}{2}(A_r(\tau-1)\sigma_r)^2 - A_u(\tau)'u_t \\
& + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) - r_t \\
= & \hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\Sigma\hat{A}(\hat{\tau}-1)\begin{pmatrix} (\bar{S}(\hat{\tau}) + \zeta(\hat{\tau})'\beta_t + \zeta_r(\tau)'r_t) \\ -(\theta_0(\hat{\tau}) - \alpha(\hat{\tau})'p_t - \theta(\hat{\tau})'\beta_t + u_t(\hat{\tau})) \end{pmatrix}\right) + \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t\right) \\
+ & A_r(\tau-1)\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\sigma_r^2 A_r(\hat{\tau}-1)\begin{pmatrix} (\bar{S}(\hat{\tau}) + \zeta(\hat{\tau})'\beta_t + \zeta_r(\tau)'r_t) \\ -(\theta_0(\hat{\tau}) - \alpha(\hat{\tau})'p_t - \theta(\hat{\tau})'\beta_t + u_t(\hat{\tau})) \end{pmatrix}\right) + \sigma_r\tilde{\sigma}_r\tilde{X}_t\right).
\end{aligned} \tag{A20}$$

With the assumption in (35), and the affine expression of p_t in (26), we rewrite (A20) as

$$
\begin{aligned}
& A(\tau-1)'\left(\bar{\beta}+\Phi(\beta_t-\bar{\beta})\right)-A(\tau)'\beta_t+\frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1)+C(\tau-1)-C(\tau) \\
& +A_r(\tau-1)(\bar{r}+\phi_r'\Phi(\beta_t-\bar{\beta})+\rho_r r_t)-A_r(\tau)r_t+\frac{1}{2}(A_r(\tau-1)\sigma_r)^2-A_u(\tau)'u_t \\
& +\frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1)-r_t \\
=& \hat{A}(\tau-1)'\gamma\left(\begin{array}{c} \sum_{\hat{\tau}=2}^{N}\left(\Sigma\hat{A}(\hat{\tau}-1)\left(\begin{array}{c}(\bar{S}(\hat{\tau})+\zeta(\hat{\tau})'\beta_t+\zeta_r(\tau)'r_t) \\ -(\theta_0(\hat{\tau})-\alpha(\hat{\tau})'(A\beta_t+A_r r_t+A_u u_t+C)-\theta(\hat{\tau})'\beta_t)-u_t(\hat{\tau})\end{array}\right)\right) \\ +\Psi\beta_t+\Lambda r_t+\psi \end{array}\right) \\
& +A_r(\tau-1)\gamma\left(\begin{array}{c} \sum_{\hat{\tau}=2}^{N}\left(\sigma_r^2 A_r(\hat{\tau}-1)\left(\begin{array}{c}(\bar{S}(\hat{\tau})+\zeta(\hat{\tau})'\beta_t+\zeta_r(\tau)'r_t) \\ -(\theta_0(\hat{\tau})-\alpha(\hat{\tau})'(A\beta_t+A_r r_t+A_u u_t+C)-\theta(\hat{\tau})'\beta_t)-u_t(\hat{\tau})\end{array}\right)\right) \\ +\Psi_r\beta_t+\Lambda_r r_t+\psi_r \end{array}\right).
\end{aligned} \tag{A21}
$$

Matching the coefficients on β_t, r_t, u_t, and the constant term, we obtain iteration equations as follows:

$$
\begin{aligned}
& A(\tau-1)'\Phi-A(\tau)'+A_r(\tau-1)\phi_r'\Phi \\
& =\hat{A}(\tau-1)'\underbrace{\gamma\left(\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\left(\zeta(\hat{\tau})'+\alpha(\hat{\tau})'A+\theta(\hat{\tau})'\right)\right)+\Psi\right)}_{\lambda_{\beta,\beta}} \\
& +A_r(\tau-1)\underbrace{\gamma\left(\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2 A_r(\hat{\tau}-1)\left(\zeta(\hat{\tau})'+\alpha(\hat{\tau})'A+\theta(\hat{\tau})'\right)\right)+\Psi_r\right)}_{\lambda_{\beta,r}},
\end{aligned} \tag{A22}
$$

$$
\begin{aligned}
A_r(\tau-1)'\rho_r-A_r(\tau)-1=& \hat{A}(\tau-1)'\underbrace{\gamma\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\left(\zeta_r(\tau)'+\alpha(\hat{\tau})'A_r\right)+\Lambda\right)}_{\lambda_{r,\beta}} \\
& +A_r(\tau-1)'\underbrace{\gamma\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2 A_r(\hat{\tau}-1)\left(\zeta_r(\tau)'+\alpha(\hat{\tau})'A_r\right)+\Lambda_r\right)}_{\lambda_{r,r}},
\end{aligned} \tag{A23}
$$

$$
\begin{aligned}
-A_u(\tau)' = \hat{A}(\tau-1)'\gamma\Sigma &\left(\left(\sum_{\hat{\tau}=2}^{N} \hat{A}(\hat{\tau}-1)\alpha(\hat{\tau})'A_u \right) - \left(0, \hat{A}(1), ..., \hat{A}(N-1)\right) \right) \\
&+ A_r(\tau-1)'\gamma\sigma_r^2 \left(\left(\sum_{\hat{\tau}=2}^{N} A_r(\hat{\tau}-1)\alpha(\hat{\tau})'A_u \right) - \left(0, A_r(1), ..., A_r(N-1)\right) \right),
\end{aligned} \tag{A24}
$$

$$
\begin{aligned}
&A(\tau-1)'(I-\Phi)\bar{\beta} + A_r(\tau-1)(\bar{r} - \phi_r'\Phi\bar{\beta}) + \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) \\
&+ \frac{1}{2}(A_r(\tau-1)\sigma_r)^2 + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) + C(\tau-1) - C(\tau) \\
&= \hat{A}(\tau-1)'\gamma \left(\sum_{\hat{\tau}=2}^{N} \Sigma\hat{A}(\hat{\tau}-1)\left(\bar{S}(\hat{\tau}) - \left(\theta_0(\hat{\tau}) - \alpha(\hat{\tau})'C\right)\right) + \psi \right) \\
&+ A_r(\tau-1)'\gamma \left(\sum_{\hat{\tau}=2}^{N} \sigma_r^2 A_r(\hat{\tau}-1)\left(\bar{S}(\hat{\tau}) - \left(\theta_0(\hat{\tau}) - \alpha(\hat{\tau})'C\right)\right) + \psi_r \right).
\end{aligned} \tag{A25}
$$

D.2. Proofs of Results in the Simple Model

Since the simple model is a special case of the main model, we can use the derivations for the main model to help with proofs in the simple model. In particular, we will rely on the iteration equations in (A22), (A23), (A24), and (A25) in Section D.1 to help derive the simple model.

Derivations of Equilibrium Treasury Prices in Equation (29)

First, we note that due to perfect arbitrage, we must have $p_t^{(1)} = -r_t$, so that $A(1) = 0$, $A_r(1) = -1$, $C(1) = 0$, and $A_u(1)' = (0,0)$. The holding return for 2-period Treasury bond as in (A7) can be simplified as

$$
R_{t+1}^{(2)} = -A(2)\cdot\beta_t + C(1) - C(2) - (\rho_r r_t + \sigma_r \varepsilon_{t+1}^r) - A_r(2)r_t + \frac{1}{2}\sigma_r^2 - A_u(2)'u_t. \tag{A26}
$$

Next, we set $\tau = 2$ in the iteration equation for A_r in (A23), which leads to

$$
\begin{aligned}
A_r(1)\rho_r - A_r(2) - 1 &= A_r(1)\gamma\left(\sigma_r^2 A_r(1)\alpha(2)'A_r\right) \\
-\rho_r - A_r(2) - 1 &= -\gamma\left(-\sigma_r^2(-b, \frac{a}{2})\begin{pmatrix} -1 \\ A_r(2) \end{pmatrix}\right) \\
-\rho_r - A_r(2) - 1 &= \gamma\sigma_r^2\left(b + \frac{a}{2}A_r(2)\right).
\end{aligned}
$$

Therefore,

$$A_r(2) = -\frac{1+\rho_r+\gamma\sigma_r^2 b}{1+\frac{1}{2}\gamma\sigma_r^2 a}.$$

To obtain $A(2)$, we set $\tau = 2$ in the iteration equation for A in (A22),

$$\begin{aligned}
-A(2) &= A_r(1)\gamma\left(\sigma_r^2 A_r(1)(\zeta(2)+\alpha(2)'A+\theta(2))\right) \\
&= -\gamma\left(-\sigma_r^2(\zeta(2)+(-b,\frac{a}{2})\begin{pmatrix} 0 \\ A(2) \end{pmatrix}+\theta(2))\right) \\
&= \gamma\sigma_r^2(\frac{a}{2}A(2)+\zeta(2)+\theta(2)).
\end{aligned}$$

which leads to

$$A(2) = -\frac{\gamma\sigma_r^2(\theta(2)+\zeta(2))}{1+\gamma\sigma_r^2\frac{a}{2}}.$$

Next, we solve for A_u. For $\tau = 2$, equation (A24) leads to

$$-A_u(2)' = -\gamma\sigma_r^2\left(-\alpha(2)'\begin{pmatrix} A_u(1)' \\ A_u(2)' \end{pmatrix} - (0, A_r(1))\right)$$

$$\begin{aligned}
-A_u(2)' &= -\gamma\sigma_r^2\left(-(-b,\frac{a}{2})\begin{pmatrix} A_u(1)' \\ A_u(2)' \end{pmatrix} - (0,-1)\right) \\
A_u(2)' &= \gamma\sigma_r^2\left(-\frac{a}{2}A_u(2)' - (0,-1)\right) \\
A_u(2)' &= \frac{1}{1+\gamma\sigma_r^2\frac{a}{2}}(0,\gamma\sigma_r^2).
\end{aligned}$$

Consequently, we obtain the A_u matrix as

$$A_u = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\gamma\sigma_r^2}{1+\gamma\sigma_r^2\frac{a}{2}} \end{pmatrix}.$$

Then, we solve for $C(2)$ via setting $\tau = 2$ in equation (A25),

$$\frac{1}{2}\sigma_r^2 + C(1) - C(2) = A_r(1)\gamma\left(\sigma_r^2 A_r(1)(\bar{S}^{(2)} - \theta_0(2) + \alpha(2)'C)\right)$$

$$\frac{1}{2}\sigma_r^2 - C(2) = \gamma\sigma_r^2\left(\bar{S}^{(2)} - \theta_0(2) + (-b,\frac{a}{2})\begin{pmatrix} 0 \\ C(2) \end{pmatrix}\right)$$

$$C(2) = \frac{\frac{1}{2}\sigma_r^2 - \gamma\sigma_r^2\bar{S}^{(2)} + \gamma\sigma_r^2\theta_0(2)}{1+\gamma\sigma_r^2\frac{a}{2}}$$
$$= \frac{\frac{1}{2} - \gamma\bar{S}^{(2)} + \gamma\theta_0(2)}{\frac{1}{\sigma_r^2} + \gamma\frac{a}{2}}.$$

Summarizing all the above, we obtain

$$p_t^{(2)} = -\frac{1+\rho_r+\gamma\sigma_r^2 b}{1+\frac{a}{2}\gamma\sigma_r^2}r_t - \frac{\gamma\sigma_r^2(\zeta(2)+\theta(2))}{1+\frac{a}{2}\gamma\sigma_r^2}\beta_t + \frac{\gamma\sigma_r^2}{1+\frac{a}{2}\gamma\sigma_r^2}u_t(2) + \frac{\frac{1}{2}-\gamma\bar{S}^{(2)}+\gamma\theta_0(2)}{\frac{1}{\sigma_r^2}+\frac{a}{2}\gamma},$$

which is identical to equation (29).

Proof of Proposition 1

According to equation (29), $p_t^{(1)}$ is entirely explained by r_t, while $p_t^{(2)}$ are also explained by β_t and $u_t(2)$. As a result, macro shocks and latent demand shocks are more important for long-maturity Treasuries.

Proof of Proposition 2

To prove Proposition 2, we derive three important sensitivities.

$$\frac{\partial p_t^{(2)}}{\partial \beta_t} = -\frac{\gamma\sigma_r^2(\zeta(2)+\theta(2))}{1+\frac{a}{2}\gamma\sigma_r^2}$$
$$\frac{\partial p_t^{(2)}}{\partial u_t} = \frac{\gamma\sigma_r^2}{1+\frac{a}{2}\gamma\sigma_r^2}$$
$$\frac{\partial p_t^{(2)}}{\partial \theta_0(2)} = \frac{\gamma\sigma_r^2}{1+\frac{a}{2}\gamma\sigma_r^2}$$

The magnitudes of these three sensitivities clearly all increase with γ.

Proof of Proposition 3

The expectation component of the long-term Treasury yield is

$$\bar{y}_t^{(2)} = \frac{1+\rho_r}{2}r_t$$

Using $y_t^{(2)} = -p_t^{(2)}/2$ and equation (29), we get the term premium expression

$$\begin{aligned}
& y_t^{(2)} - \bar{y}_t^{(2)} \\
=& \left(\frac{1+\rho_r+\gamma\sigma_r^2 b}{2+a\gamma\sigma_r^2} - \frac{1}{2}(1+\rho_r) \right) r_t + \frac{\gamma\sigma_r^2(\zeta(2)+\theta(2))}{2+a\gamma\sigma_r^2}\beta_t - \frac{\gamma\sigma_r^2}{2+a\gamma\sigma_r^2}u_t(2) - \frac{\frac{1}{2}-\frac{\bar{r}}{\sigma_r^2}-\gamma\bar{S}^{(2)}+\gamma\theta_0(2)}{2+a\gamma\sigma_r^2} \\
=& \frac{b-\frac{1}{2}(1+\rho_r)a}{\frac{2}{\gamma\sigma_r^2}+a} r_t + \frac{\gamma\sigma_r^2(\zeta(2)+\theta(2))}{2+a\gamma\sigma_r^2}\beta_t - \frac{\gamma\sigma_r^2}{2+a\gamma\sigma_r^2}u_t(2) - \frac{\frac{1}{2}-\frac{\bar{r}}{\sigma_r^2}-\gamma\bar{S}^{(2)}+\gamma\theta_0(2)}{2+a\gamma\sigma_r^2}
\end{aligned}$$

As a result,

$$\frac{\partial(y_t^{(2)} - \bar{y}_t^{(2)})}{\partial r_t} = \frac{b-\frac{1}{2}(1+\rho_r)a}{\frac{2}{\gamma\sigma_r^2}+a}$$

while the baseline response according to the expectation hypothesis is

$$\frac{\partial \bar{y}_t^{(2)}}{\partial r_t} = \frac{1+\rho_r}{2} > 0$$

Consequently, the full response is

$$\frac{\partial y_t^{(2)}}{\partial r_t} = \underbrace{\frac{1+\rho_r}{2}}_{\text{expectation hypothesis}} + \underbrace{\frac{b-\frac{1}{2}(1+\rho_r)a}{\frac{2}{\gamma\sigma_r^2}+a}}_{\text{change of term premium}}$$

When $2b > (1+\rho_r)a$, the term premium component is positive, so that the long-term Treasury yield over-reacts to monetary policy shock compared to the expectation hypothesis. When $2b < (1+\rho_r)a$, the term premium component is negative, so that the long-term Treasury yield under-reacts to monetary policy shock compared to the expectation hypothesis.

Proof of Proposition 4

The impact of QE on Treasury price, as reflected by the increase of the permanent demand $\theta_0(2)$, is as follows,

$$\frac{\partial p_t^{(2)}}{\partial \theta_0(2)} = \frac{\gamma\sigma_r^2}{1+\frac{a}{2}\gamma\sigma_r^2} > 0.$$

Therefore, Treasury prices increase with QE, which implies a decrease of Treasury yields.

D.3. Setting Model Parameters

The model is quite flexible accounting for the rich dependence of investor demand on macroeconomic factors and Treasury prices, as well as dynamics in the state variables. In this subsection, we provide details of how we use data to directly inform model parameters.

We take the average duration as the maturity for each maturity bucket, obtaining $\tau_1 = 2$, $\tau_2 = 10$, and $\tau_3 = 42$ (all in quarters). For each maturity bucket, we sum up the coefficients of non-arbitrageur demand in Table 3 and 4. To convert regression results to the model format, we express the demand for each maturity bucket separately, and use the intercept term to capture maturity-bucket fixed effects. We then add the maturity-by-maturity bucket estimates of the Fed to the granular-demand investor demand to obtain total non-arbitrageur demand. For simplicity, our model does not capture characteristic-based demand (i.e., loadings on coupon rate and bid-ask spread), so we take the average of these components and add them to the intercept of non-arbitrageur demand.

Moreover, in the model, the demand is expressed as a function of prices, not yields, so we need to convert the yield sensitivity into price sensitivity, using the chain rule,

$$\frac{\partial Z(\tau)}{\partial p^{\tau}} = \frac{\partial Z(\tau)}{\partial y^{\tau}} \frac{\partial y^{\tau}}{\partial p^{\tau}} = -\frac{1}{\tau} \frac{\partial Z(\tau)}{\partial y^{\tau}} \tag{A27}$$

Second, we estimate the supply dynamics in Equation (21). We implement a linear regression of the Treasury total supply in each maturity bucket and then recover the loadings on macro factors, the short rate, and the intercept $\bar{S}$. Similar to the demand estimation, we concentrate the supply into three maturities that represent the average duration of three maturity buckets. In Figure A3, we illustrate that the model fits the total supply well. The R^2s of all three regressions are above 95%.

Third, we estimate the monetary policy dynamics in (16). We rewrite the monetary policy equation as

$$r_{t+1} = (\bar{r} - \phi_r' \bar{\beta}) + \phi_r' \beta_{t+1} + \rho_r r_t + \sigma_r \varepsilon_{t+1}^r, \tag{A28}$$

where the intercept term is identified as a whole. To fit the monetary policy rule, we have to use a longer time period, because monetary policy rate does not have much variation during our main sample period. In particular, we use the post-Volcker period (1990 to 2024) excluding the zero lower bound (ZLB) period (2008-2015). We start from 1990 because it is when the Fed gained credibility in its fight of inflation. The resulting monetary-policy equation is:

$$\begin{aligned} r_{t+1} = 1.9 &- 1.36 * \text{credit spread}_{t+1} + 0.06 * \text{GDP gap}_{t+1} + 0.22 * \text{core inflation}_{t+1} \\ &- 1.13 * \text{debt/GDP}_{t+1} + 0.78 * r_t + 0.75 * \varepsilon_{t+1}^r \end{aligned} \tag{A29}$$

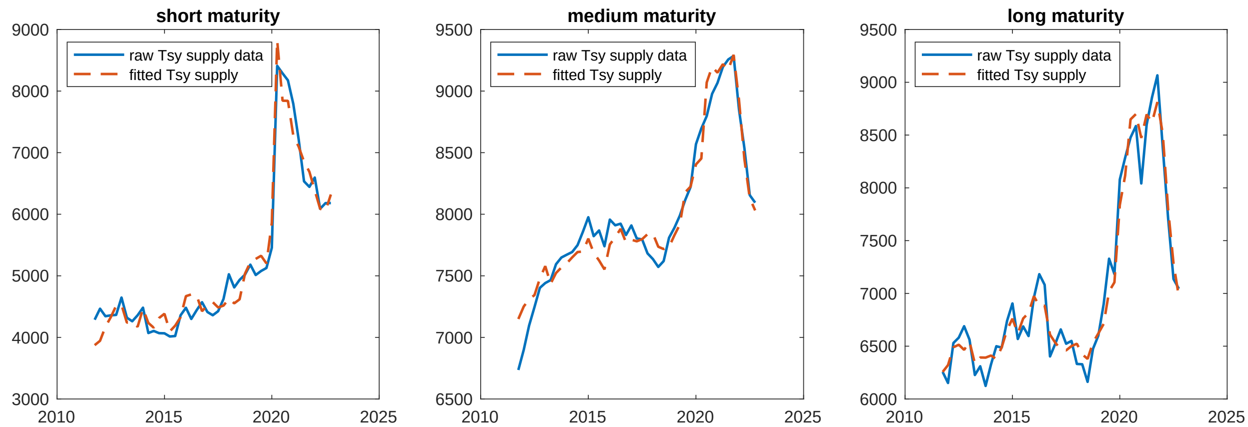


Figure A3. Treasury Supply: Data versus Model Fitting.

Equation (A29) suggests that the Fed lowers the interest rate if credit spread is high, GDP gap (GDP deviation from potential GDP) is low and tightens interest rate if inflation is high. The coefficient on GDP gap and inflation have the same signs as the classical Taylor rule (Taylor 1993) but much smaller coefficients. Moreover, there is moderate amount of monetary policy inertia reflected by the coefficient of 0.78 on lagged policy rate. This dependence on lagged policy rate generates an impact of monetary policy rate on long-term yields from the expectation effect and is critical to understand how yield curve responds to monetary policy shocks ε^r_{t+1}.

Fourth, we estimate the dynamics of macro factors in Equation (15). It is important to get the long-run average of macroeconomic factors correct. Therefore, we take the sample average of macro factors directly as $\bar{\beta}$. Denote the demeaned macro factors as $\hat{\beta}_t$. Then we recover the coefficients with the following regression:

$$\tilde{\beta}_{t+1} = \Phi\tilde{\beta}_t + \Sigma^{1/2}\varepsilon_{t+1}. \tag{A30}$$

Alternatively, we could directly run a linear regression with an intercept to uncover $\bar{\beta}$ and Φ simultaneously. We find that the estimations of Φ are similar between the two approaches, but the simultaneous estimation of $\bar{\beta}$ and Φ gives unreasonable long-run average of macro variables. The matrix Σ is estimated as the covariance matrix of the regression residuals in (A30).

D.4. Model Estimation

According to equation (A22) and (A23), we can reformulate the main estimation problem (37) over the parameter set $\{\gamma, A_u, \lambda_{\beta,\beta}, \lambda_{\beta,r}, \lambda_{r,\beta}, \lambda_{r,r}, \psi, \psi_r\}$, replacing Ψ, Ψ_r, Λ, and Λ_r with $\lambda_{\beta,\beta}, \lambda_{\beta,r}, \lambda_{r,\beta}$, and $\lambda_{r,r}$. This equivalent formulation simplifies the iteration equations in solving the model.

Estimation of the model involves high dimensionality and requires a reasonable initialization of model parameters. Our high-level idea is to iteratively divide the model into smaller problems and initialize the model from a bottom-up approach.

At the first step, we solve for the following simpler optimization problem with unconstrained C to initialize the price of risk matrices $\lambda_{\beta,\beta}, \lambda_{\beta,r}, \lambda_{r,\beta}$, $\lambda_{r,r}$, and the constant term C,

$$\min_{\{\lambda_{\beta,\beta},\lambda_{\beta,r},\lambda_{r,\beta},\lambda_{r,r},C\}} E\left[\sum_t \sum_\tau (y_t(\tau) - y_t^o(\tau))^2\right], \tag{A31}$$

subject to

$$A(\tau)' = A(\tau-1)'\Phi + A_r(\tau-1)\phi_r'\Phi - \hat{A}(\tau-1)'\lambda_{\beta,\beta} - A_r(\tau-1)\lambda_{\beta,r} \tag{A32}$$

$$A_r(\tau) = A_r(\tau-1)\rho_r - 1 - \hat{A}(\tau-1)'\lambda_{r,\beta} - A_r(\tau-1)\lambda_{r,r} \tag{A33}$$

where $\hat{A}(\tau-1)$ is a function of $A(\tau-1)$ and $A_r(\tau-1)$ as defined in (A8), and $y_t(\tau) = -p_t(\tau)/\tau$, where $p_t(\tau)$ satisfies the pricing equation in (26), $p_t = A\beta_t + A_r r_t + A_u u_t + C$. The latent demand term u_t is unobservable with mean 0 and uncorrelated with β_t and r_t. As a result, we can rewrite the objective function as

$$\min_{\{\lambda_{\beta,\beta},\lambda_{\beta,r},\lambda_{r,\beta},\lambda_{r,r},C\}} \sum_t \sum_\tau \left(\frac{A\beta_t + A_r r_t}{\tau} + \frac{C}{\tau} + y_t^o(\tau)\right)^2, \tag{A34}$$

We note that this problem does not explicitly involve arbitrageur risk aversion γ, because that is embedded in the solution of risk premium $\lambda_{\beta,\beta}$, $\lambda_{\beta,r}$, $\lambda_{r,\beta}$, $\lambda_{r,r}$ and the intercept C.

In the estimation, the dimension of β is $K = 4$, and the dimension of r_t is 1. The vector C is 120×1 (quarterly frequency of 30 years gives rise to 120 maturities). Therefore, the total degree of freedom is 5*5+120 = 145. This is a very high dimensional optimization problem. Similar to a typical affine term structure estimation, it is important to find a good initial point for the algorithm. We leverage on an important insight from the affine term structure literature, which is to use regressions to initialize the coefficient matrix. In particular, we start with a linear regression problem:

$$\min_{A,A_r,C} \sum_t \sum_\tau (A\beta_t + A_r r_t + C + \tau y_t^o(\tau))^2,$$

Solving this estimation on A, A_r, and C is equivalent to regress the log-price vector

$$(y_t^o(1), 2y_t^o(2), \cdots, Ny_t^o(N))$$

on β_t and r_t, where C serves as the intercept term.

Next, knowing the values of the matrices A, A_r, we can view the iteration equations in (A32) and (A33) as another set of regressions. Rewriting (A32) and (A33) in a regression form,

$$\begin{aligned}\underbrace{A(\tau)'-A(\tau-1)'\Phi}_{\text{left hand side}} &= \underbrace{A_r(\tau-1)}_{\text{dep var}}(\phi_r'\Phi-\lambda_{\beta,r})-\underbrace{\hat{A}(\tau-1)'}_{\text{dep var}}\lambda_{\beta,\beta},\\ \underbrace{-A_r(\tau)+A_r(\tau-1)\rho_r-1}_{\text{left hand side}} &= \underbrace{\hat{A}(\tau-1)'}_{\text{dep var}}\lambda_{r,\beta}+\underbrace{A_r(\tau-1)}_{\text{dep var}}\lambda_{r,r}.\end{aligned} \tag{A35}$$

where the regression coefficients are $\phi_r'\Phi-\lambda_{\beta,r}$, $\lambda_{\beta,\beta}$, $\lambda_{r,\beta}$, and $\lambda_{r,r}$. Note that ϕ_r and Φ are directly estimated in the data. Consequently, we can use regressions to initialize all of the four price of risk matrices and also the constant term C.

Next, we note that for any given γ, and the solved matrix A, A_r, and $\hat{A}$, we can uniquely pin down the latent-demand impact matrix A_u. Therefore, we can effectively eliminate A_u from the parameters to be estimated. To see that, we denote

$$\begin{aligned}\hat{A}^{\text{shift}} &= \left(0,\hat{A}(1),...,\hat{A}(N-1)\right)'\\ A_r^{\text{shift}} &= (0,A_r(1),...,A_r(N-1))'\end{aligned} \tag{A36}$$

which are "shifts" of the original $\hat{A}$ and A_r matrices. Then we can stack all different τ in equation (A24) to get the matrix equation

$$\begin{aligned}&\left(\hat{A}^{\text{shift}}\gamma\Sigma\sum_{\hat{\tau}=2}^{N}\hat{A}(\hat{\tau}-1)\alpha(\hat{\tau})'+A_r^{\text{shift}}\gamma\sigma_r^2\sum_{\hat{\tau}=2}^{N}A_r(\hat{\tau}-1)\alpha(\hat{\tau})'+I\right)A_u\\ =&\hat{A}^{\text{shift}}\gamma\Sigma(\hat{A}^{\text{shift}})'+A_r^{\text{shift}}\gamma\sigma_r^2(A_r^{\text{shift}})'\end{aligned} \tag{A37}$$

which is simply a linear equation for A_u that can be solved immediately. We initialize γ so that average of A_u is 0.0001, which implies that a 100 billion dollar shock will change the average Treasury market price by 1%.

To initialize the intercepts ψ and ψ_r of arbitrageur's outside portfolios, we implement another "regression" according to equation (A25), by treating $\gamma\hat{A}$ and $\gamma\hat{A}_r$ as the independent variables and ψ and ψ_r as the corresponding coefficients.

With initial values of $\{\gamma, A_u, \lambda_{\beta,\beta}, \lambda_{\beta,r}, \lambda_{r,\beta}, \lambda_{r,r}, \psi, \psi_r\}$, we can estimate all of these parameters in the full optimization problem,

$$\min_{\{\gamma,A_u,\lambda_{\beta,\beta},\lambda_{\beta,r},\lambda_{r,\beta},\lambda_{r,r},\psi,\psi_r\}}\mathbb{E}\left[M\cdot(h-h^o)^2+\sum_t\sum_\tau(y_t(\tau)-y_t^o(\tau))^2\right], \tag{A38}$$

subject to iteration equations in (A32) and (A33) that give rise to $\{A, A_r, \hat{A}\}$, equation (A37) that solves for A_u, equation (A25) that solves for C, predicted yield $y_t(\tau)=-p_t(\tau)/\tau$, where $p_t(\tau)$ is determined by (26), and arbitrageur long-term Treasury holdings as a fraction of total outstanding,

$$h = \frac{\sum_{\tau>4} X_t(\tau)}{\sum_{\tau>4} S_t(\tau)}, \tag{A39}$$

where $\tau > 4$ is for maturities above 4 quarters, and $X_t(\tau)$ and $S_t(\tau)$ are given by equation (36) and (20). Expanding $y_t(\tau) = -p_t(\tau)/\tau$ and using (26), we can further express the estimation problem as

$$\min_{\{\gamma, A_u, \lambda_{\beta,\beta}, \lambda_{\beta,r}, \lambda_{r,\beta}, \lambda_{r,r}, \psi, \psi_r\}} \mathbb{E}\left[M \cdot (h - h^o)^2 + \sum_t \sum_\tau \left(\frac{A\beta_t + A_r r_t + C}{\tau} + y_t^o(\tau)\right)^2\right], \tag{A40}$$

After we estimate problem (A40), we can recover the arbitrageur's outside asset risk loadings Ψ, Ψ_r, Λ, Λ_r, from the definitions of $\lambda_{\beta,r}$, $\lambda_{\beta,\beta}$, $\lambda_{r,\beta}$,and $\lambda_{r,r}$ in equations (A22) and (A23):

$$\Psi = \frac{1}{\gamma}\lambda_{\beta,\beta} - \sum_{\hat{\tau}=2}^{N} \Sigma \hat{A}(\hat{\tau}-1)\left(\zeta(\hat{\tau})' + \alpha(\hat{\tau})'A + \theta(\hat{\tau})'\right)$$

$$\Psi_r = \frac{1}{\gamma}\lambda_{\beta,r} - \left(\sum_{\hat{\tau}=2}^{N} \sigma_r^2 A_r(\hat{\tau}-1)\left(\zeta(\hat{\tau})' + \alpha(\hat{\tau})'A + \theta(\hat{\tau})'\right)\right)$$

$$\Lambda = \frac{1}{\gamma}\lambda_{r,\beta} - \sum_{\hat{\tau}=2}^{N} \Sigma \hat{A}(\hat{\tau}-1)\left(\zeta_r(\tau)' + \alpha(\hat{\tau})'A_r\right)$$

$$\Lambda_r = \frac{1}{\gamma}\lambda_{r,r} - \sum_{\hat{\tau}=2}^{N} \sigma_r^2 A_r(\hat{\tau}-1)\left(\zeta_r(\tau)' + \alpha(\hat{\tau})'A_r\right)$$

To speed up the algorithm, we express the equation for C as a linear equation rather than a iterative procedure. In particular, we rewrite (A25) as

$$\begin{aligned}
&A(\tau-1)'(I-\Phi)\bar{\beta} + A_r(\tau-1)(\bar{r} - \phi_r'\Phi\bar{\beta}) + \frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1) \\
&+ \frac{1}{2}(A_r(\tau-1)\sigma_r)^2 + \frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1) - \hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\left(\bar{S}(\hat{\tau}) - \theta_0(\hat{\tau})\right) + \psi\right) \\
&- A_r(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2 A_r(\hat{\tau}-1)\left(\bar{S}(\hat{\tau}) - \theta_0(\hat{\tau})\right) + \psi_r\right) \\
=&\hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\alpha(\hat{\tau})'\right)C + A_r(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2 A_r(\hat{\tau}-1)\alpha(\hat{\tau})'\right)C + C(\tau) - C(\tau-1)
\end{aligned} \tag{A41}$$

The left-hand side is a single value and denote it as $C_0(\tau)$. Also denote the vector

$$\tilde{A}(\tau)' = \hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\alpha(\hat{\tau})'\right) + A_r(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2 A_r(\hat{\tau}-1)\alpha(\hat{\tau})'\right) + (\mathbf{1}_\tau - \mathbf{1}_{\tau-1})'$$

where $\mathbf{1}_\tau$ is an N-dimensional vector that is one for element τ but zero otherwise. Then equation (A41) can be simplified as

$$C_0(\tau) = \tilde{A}(\tau)'C$$

for all $\tau \in \{2,3,\cdots,N\}$. For $\tau = 1$, we know that $p_t^{(1)} = -r_t$, which implies that $C(1) = 0$. Stacking all of the equations for $\tau \in \{1,2,3,\cdots,N\}$, we get

$$\begin{pmatrix} 0 \\ C_0(2) \\ \vdots \\ C_N(1) \end{pmatrix} = \begin{pmatrix} \mathbf{1}_1' \\ \tilde{A}(2)' \\ \vdots \\ \tilde{A}(N)' \end{pmatrix} C,$$

which is a linear system that can be easily solved.

E. Additional Quantitative Results

E.1. Model Fit and Steady-State Yields and Holdings

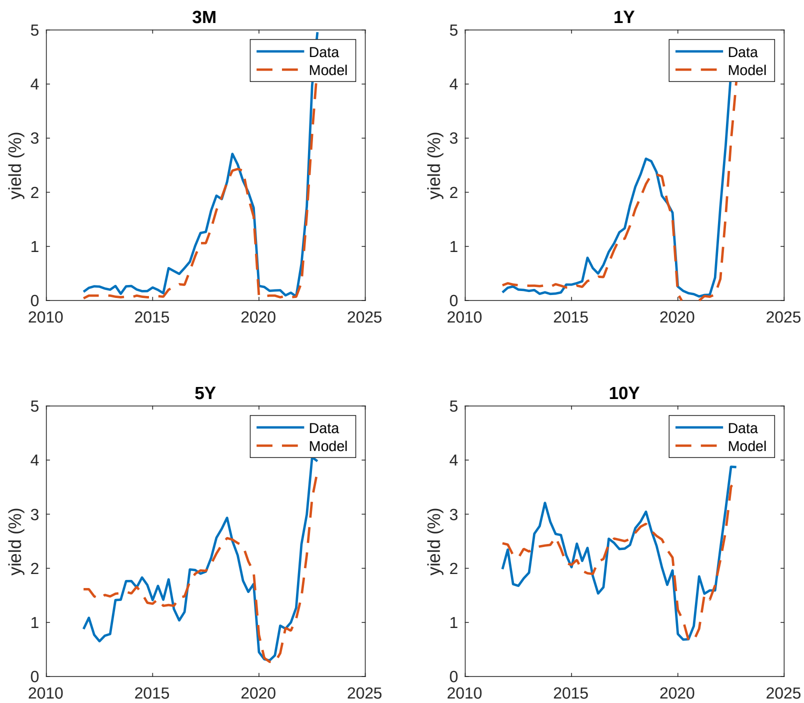
In Figure A4, we show that the model can fit the the term structure reasonably well, both across maturities and over time. Note that these results are achieved by having only fundamental economic variables as state variables, which is much more challenging than a typical affine term structure model that includes factors coming directly from Treasury yields. The equilibrium restrictions in the model impose tight restrictions on how flexibly the model can explain the dynamics of Treasury yields.

In our baseline estimation, we minimize the expected fitting error, so the realizations of u_t do not enter the estimations, as in (A40). According to the model, nevertheless, demand shocks should play a role in explaining the dynamics of the Treasury yield curve. In Figure A5, we compare data with model predictions as in (26) accounting for the impact of latent demand u_t. Contrasting Figure A5 with Figure A4, we find that including u_t in the prediction of yields significantly increases the predictive power, particularly at the long-maturity end. This should be viewed as an out-of-sample test given that we do not use information from u_t to estimate the model (note that average arbitrageur holding h in (37) is not affected by u_t since u_t on average is zero). In particular, the magnitude of fluctuations for long-maturity Treasuries are much closer to the data counterparts after we incorporate the impact of u_t. On the other hand, at the one-year maturity, we find that including u_t causes the predicted yield to fluctuate more than the data. This indicates that there is plausibly stronger arbitrage at short-maturity Treasuries than our mean-variance framework implies, so arbitrageurs play an even bigger role at the short end of the yield curve.

Overall, we find that including u_t improves the predictive power of the model on yield curve dynamics, consistent with the idea that demand shocks matter in the Treasury market.

Figure A4. **Model Fit on the Dynamics of Treasury Yields.**

Model predicted yields are constructed using equation (26) without latent demand (setting $u_t = 0$).



In Figure A6 panel (a), we illustrate the steady-state yield curve. This steady-state yield curve is upward-sloping and mainly reflects the average shape of the yield curve in our model estimation period.

In Figure A6 panel (b), we illustrate the steady-state portfolio allocations across different sectors. The model implies that in the long run, foreign investors are still the largest holder among all groups of investors, and the Fed also plays an important role. Insurance and pension funds are not large holders, but they predominantly hold long-term Treasuries. Finally, as targeted by the calibration, arbitrageurs' longer-term (>1Y) Treasury holding is 6% of the total longer-term Treasuries outstanding.

Figure A5. **Model Fit on the Dynamics of Treasury Yields (including latent demand u).**

Model predicted yields are constructed using equation (26) with latent demand u_t.

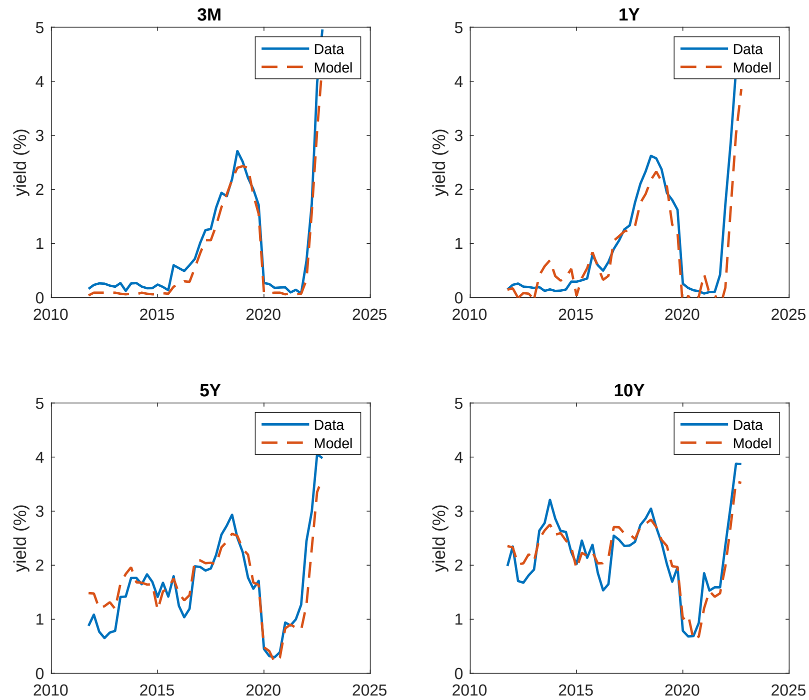
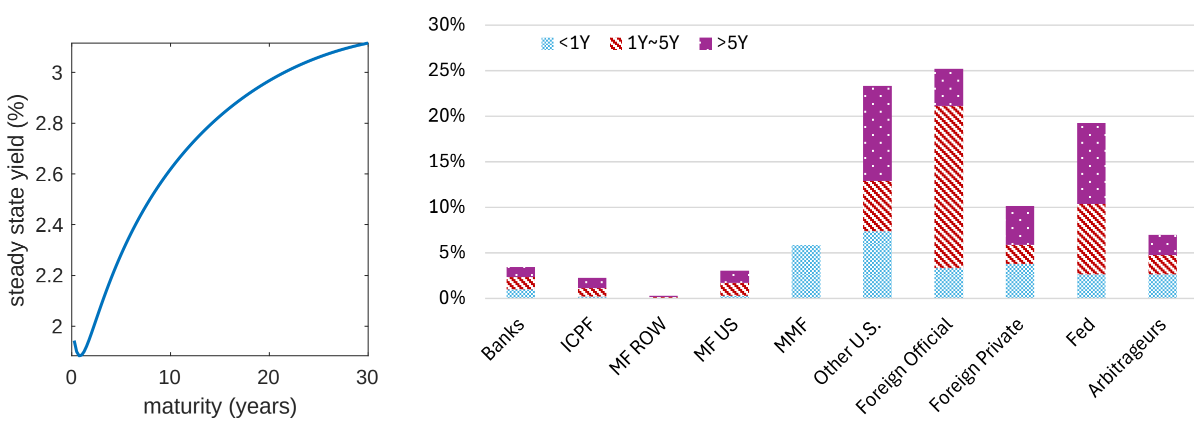


Figure A6. **Steady State.**

This figure illustrates the yield curve and portfolio allocations at the steady state, defined as the state where all shocks are zero. The left panel illustrates the steady-state yield curve. The right panel illustrates the steady-state portfolio holdings (as % over total outstanding) for each group of investors and maturity bucket.



(a) Steady state yield curve

(b) Steady state portfolio allocations

E.2. Calculating Treasury Market Multiplier and Elasticity

The total multiplier of the Treasury market is defined as the % valuation change in the whole Treasury market in response to a "representative demand shock" that is 1% of total Treasury valuation $S^{total} = \sum_\tau S(\tau)$, where $S(\tau)$ is the steady-state outstanding of Treasuries at maturity τ. The representative demand shock reflects the outstanding weight of each maturity bucket. Formally, define the weight vector $\omega = S/S^{total}$. Then the representative demand shock is

$$u = \omega * (S^{total} * 1\%) = S * 1\%.$$

Therefore, the total percentage change in Treasury valuation is

$$100 * \frac{\sum_{\tau'} S(\tau') A_u(\tau', \cdot) u}{S^{total}}.$$

Dividing by the 1% change in total demand is equal to the market multiplier,

$$\mathscr{M} = 100 * \frac{\sum_{\tau'} S(\tau') A_u(\tau', \cdot) u}{S^{total}} / 1\% = 100 * \frac{1}{S^{total}} S' A_u S = 100 * \omega' A_u S. \tag{A42}$$

This market multiplier is closely related to the bucket-level multiplier in Table 5. In particular, the percentage change of price at maturity τ' in response to a 1% extra latent demand of maturity τ is

$$\mathscr{M}(\tau', \tau) = 100 * \frac{A_u(\tau', \tau) * 1\% * S(\tau)}{1\%} = 100 * A_u(\tau', \tau) S(\tau) \tag{A43}$$

As a result, the total market multiplier is a function of maturity-bucket-level multiplier,

$$\mathscr{M} = \sum_{\tau'} \sum_{\tau} \omega(\tau') \mathscr{M}(\tau', \tau) \tag{A44}$$

Using equation (A44) and the values in Table 5 Panel (a), we obtain a total market multiplier of 0.23, i.e., a 1% representative demand shock on the whole Treasury market increases total Treasury valuation by 0.23%. This can also be equivalently stated as a \$1 billion demand shock increasing Treasury valuation by \$0.23 billion. Following the same procedure, we can use Panel (b) of Table 5 to calculate the Treasury market multiplier in the case without arbitrageurs, which leads to a value of 26.82.

Next, we show the aggregate multiplier for permanent demand shocks. In Table A8, we show the price impact of permanent demand shocks in the case with and without arbitrageurs. This table has the same format as our main Table 5. Using Panel (a) of A8, we can calculate the Treasury market multiplier for permanent demand shock as 0.78, which is higher than the case of latent demand shock, because permanent demand shocks significantly change the risk premium.

Panel (b) of Table A8 is identical to Panel (b) of Table 5, because absent from arbitrageurs, latent demand shocks and permanent demand shocks are treated the same by granular-demand investors. Consequently, in the case without arbitrageurs, Treasury market multipliers to permanent demand shock and to latent demand shock are identical.

Table A8. **Impact of Permanent Demand Shocks on Treasury Prices with and without Arbitrageurs.**

We illustrate the impact of permanent demand shocks with and without arbitrageurs. In panels (a) and (b), a value of 1 at row i and column j implies that 1% extra latent demand of maturity bucket i increases the price at maturity j by 1%. Panel (c) shows the ratio of the corresponding cells in Panel (b) over Panel (a).

Panel (a): With Arbitrageur		Price change (%) of	
	short maturity	medium maturity	long maturity
shock on short maturity	0.000	0.003	0.006
shock on medium maturity	0.015	0.109	0.220
shock on long maturity	0.090	0.701	2.130
Panel (b): Without Arbitrageur			
shock on short maturity	0.624	2.349	12.919
shock on medium maturity	2.900	9.520	55.921
shock on long maturity	1.046	3.667	19.061
Panel (c): Price Impact Ratio (Panel (b)/Panel (a))			
shock on short maturity	1335.985	706.168	2069.383
shock on medium maturity	192.915	87.241	254.564
shock on long maturity	11.562	5.234	8.947

Next, we provide details on how to calculate the term structure of market elasticity. According to Section 5.2, the price impact of demand shocks at different maturities is heterogeneous. To capture such heterogeneity, we introduce the concept of market multiplier at maturity τ, which is the percentage change in total Treasury valuation in response to a change of Treasury demand at maturity τ that is expressed as percentage of total Treasury outstanding,

$$\left(\frac{\sum_{\tau'} S(\tau')A_u(\tau',\tau) * (1\% * S^{total})}{S^{total}}\right)/1\%, \tag{A45}$$

Then we define the market elasticity at maturity τ as $\mathscr{E}(\tau)$ as the inverse of the market multiplier at maturity τ in (A45),

$$\mathscr{E}(\tau) = \frac{1}{\sum_{\tau'} S(\tau')A_u(\tau',\tau)} \tag{A46}$$