Granular Treasury Demand with Arbitrageurs

Kristy A. E. Jansen* Wenhao Li† Lukas Schmid‡

July 2025

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) there is a steep downward-sloping term structure of Treasury 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.

Keywords: Treasury demand; financial intermediaries; arbitrage; monetary policy; quantitative easing.

We thank Viral Acharya (discussant), Daniel Andrei (discussant), Adrien d'Avernas (discussant), Yakov Amihud, Zefeng Chen, William Diamond (discusssant), Greg Duffee (discussant), Darrell Duffie, Chuck Fang (discussant), Paul Fontanier, Francois Gourio, Valentin Haddad, Zhiguo He, Ben Hebert, Paul Huebner (discussant), Erica Jiang, Jay Kahn (discussant), Ralph Koijen, Arvind Krishnamurthy, Moritz Lenel (discussant), Anna Pavlova, Martin Lettau, Karen Lewis, Jane Li, Tyler Muir, Walker Ray (discussant), Eric Richert (discussant), Robert Richmond, Tom Sargent, Alexi Savov, Alp Simsek, Adi Sunderam (discussant), Selale Tuzel, Quentin Vandeweyer (discussant), Annette Vissing-Jorgensen, 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, Berkeley Haas, Miami Conference on Fiscal and Monetary Interactions, 5th David Backus Memorial Conference, MFA meetings, NY Fed, Chicago Treasury Market Conference, Yale SOM, NBER Asset Pricing Meeting, Georgetown, Oxford, Warwick, Manchester, Lancaster, QRFE Workshop on Quantitative Finance, STFM conference, UIC Finance Conference, Lake Dishui Finance Conference, Hong Kong International Finance Conference, Fed Conference on Fixed Income Markets, Bank of International Settlements, and LBS Summer Finance Symposium. 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.

*Marshall School of Business, University of Southern California & DNB; kjansen@marshall.usc.edu

†Marshall School of Business, University of Southern California & NBER liwenhao@marshall.usc.edu

‡Marshall School of Business, University of Southern California & CEPR; lukas@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?

To address these questions, we propose a novel framework that integrates granular empirical estimation into a dynamic macro-finance equilibrium model. First, building on insights from demand-based asset pricing (Koijen and Yogo 2019) and the preferred habitat demand theory of the term structure (Vayanos and Vila 2021), we construct detailed investor-specific Treasury demand functions that depend not only on bond characteristics, but also on macroeconomic dynamics. We introduce what we refer to as "granular-demand investors", whose maturity-specific demands and substitution across maturities are flexibly estimated using a novel, comprehensive dataset on Treasury holdings. Second, we embed these demand estimates into a dynamic equilibrium model featuring risk-averse arbitrageurs. These arbitrageurs, empirically associated with hedge funds and broker-dealers, intermediate between different investor sectors, responding endogenously to demand shocks and macroeconomic fluctuations. This micro-to-macro linkage allows us to quantitatively analyze how sector-specific demand shocks propagate through the market, how arbitrageur risk-bearing capacity shapes Treasury market elasticity, and how monetary policy interventions, such as interest rate changes and quantitative easing, transmit through the financial system.

Our analysis reveals three main findings. First, the Treasury market exhibits a steeply downward-sloping term structure of market elasticity, with strong arbitrage forces at the short end because of lower risks but substantially weaker ones at the long end. 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 exhibit behavior opposite to that 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]

Based on insights in Koijen and Yogo (2024) and building on the instrument in Fang et al. (2025), 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 ICPFs have a greater demand for longer maturities.

Our empirical approach advances the demand-based asset pricing literature in three important ways. First, we incorporate cross-maturity elasticities, explicitly estimating how investor demand for Treasuries in one maturity responds to yields in other maturities.[2] This novel feature, enabled by our granular dataset, captures significant substitution effects across maturities overlooked by aggregate demand estimation in Treasury markets. Second, our demand functions are explicitly state-dependent, integrating macroeconomic factors directly into investors' demand, which allows investor demand to vary systematically with inflation, GDP gap, and other macroeconomic conditions. Third, we combine cross-sectional granularity with time-series dynamics in estimating these elasticities. This integrated approach addresses recent insights by Gabaix and Koijen (2024) and Haddad et al. (2024a) on the necessity of time-series variation to quantify aggregate market elasticities.

We then embed the estimated demand functions into an equilibrium model of the Treasury

[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.

[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.

market, with granular demand investors, the Fed, and risk-averse arbitrageurs, in the spirit of Vayanos and Vila (2021). We extend this class of models in three aspects. First, granular-demand investors feature cross-substitution, which 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 arbitrageurs' Treasury portfolios. We let the data inform us how outside-asset risk exposure interacts with Treasury pricing.

To build intuition, we first analyze a simplified, analytically tractable version of our model before estimating the full model by minimizing fitting errors in the dynamics of the yield curve and arbitrageurs' Treasury holdings. Intuitively, risk aversion bridges quantities and prices; thus, jointly targeting yields and quantities enables precise identification of arbitrageurs' risk aversion. Using our estimates, we decompose Treasury yield fluctuations into distinct driving forces. Short-term yields primarily respond to monetary policy rates, whereas macroeconomic fluctuations and latent demand shocks become increasingly influential at longer maturities, reflecting arbitrageurs' risk pricing when absorbing these risks. Additionally, we identify banks, foreign officials, and foreign mutual funds as the main contributors to yield fluctuations arising from macroeconomic and latent demand shocks.

We find a strong downward-sloping term structure of Treasury market elasticity. Demand shocks to long-term Treasury bonds cause much stronger yield curve responses than similar shocks to T-bills. The intuition is that arbitrage is weaker at longer maturities due to larger risks per dollar value of Treasury securities, leading to a downward-sloping term structure of market elasticity. This is consistent with previous studies that find a small impact of T-bill supply shocks on yields (Greenwood et al. 2015c) but much larger impact of QE policies that purchase long-term Treasuries (Krishnamurthy and Vissing-Jorgensen 2012). Averaging across all maturities, the Treasury market is relatively elastic compared to other markets: a $1 billion dollar extra latent demand of the overall Treasury market increases total Treasury valuation by $0.37 billion, indicating a price multiplier of 0.37,[4] 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).

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,

[3]Kekre et al. (2024) introduces arbitrageurs' 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.

[4]Price multiplier is the inverse of market elasticity. This implies that the market elasticity is 2.7.

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 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. (2015a), 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. (2025) 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) and Chaudhary et al. (2024) apply a demand system to the overall U.S. Treasury market using Flows of Funds data. Consistent with their studies, 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 *across* maturity buckets, allowing us to estimate 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. Greenwood et al. (2023a) and Gourinchas et al. (2022) connect preferred-habitat demand with exchange rate dynamics. 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; Fang and Xiao 2024).

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. 2024a,b). 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 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; Siriwardane 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) use 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.

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 both 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), Chaudhary et al. (2024)). While the FoFs provides information about Treasury holdings at the investor sector level, the holdings are aggregated across all maturities and thus limiting the ability for a more granular analysis of, 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 data sources. In constructing our final dataset, we therefore make aggregation choices to ensure consistency. Specifically, we analyze data at a quarterly frequency from 2011Q4–the earliest available period for foreign investors and hedge funds–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 by reducing variation across buckets. To ensure stationarity, we scale all quantities by the ratio of potential GDP at the end of our sample period to the potential GDP at that particular quarter. We provide details on the data aggregation process in Appendix A.2. In our analysis, we also incorporate macroeconomic dynamics, and we provide details on the macro variables in Appendix A.3.

2.3. Stylized Facts about Treasury Holdings

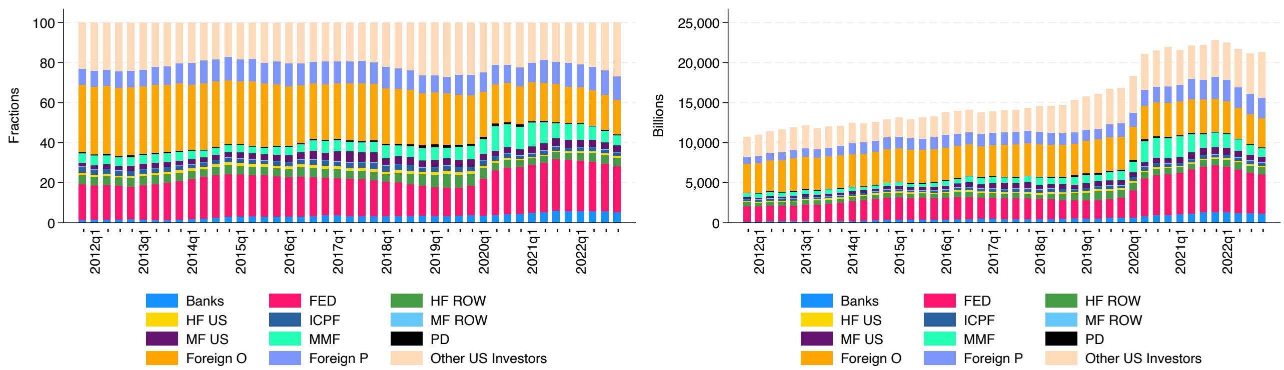
Figure 1 shows the dollar values and the fraction of total outstanding of U.S. Treasuries held by each investor type from 2011Q4 to 2022Q4. On average, our dataset contains 78% of the holders of U.S. Treasuries. Based on FoF data, the remaining 22% consists of U.S. households (11%), pension funds (5%), 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 outstanding short-term Treasuries. 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 report negative holdings across maturity buckets (also see Table 6 in Section 3). Finally, foreign official holdings significantly declined, mainly in the short and medium-maturity buckets.

In Table 2, we further examine which investor types are marginal in trading U.S. Treasuries when supply changes. To that end, similar to Fang et al. (2025), but with a focus on maturity buckets, we decompose the marginal holders of Treasuries. For each maturity bucket m and sector

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)

ι, we regress changes in holdings on changes in the total supply of debt:

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

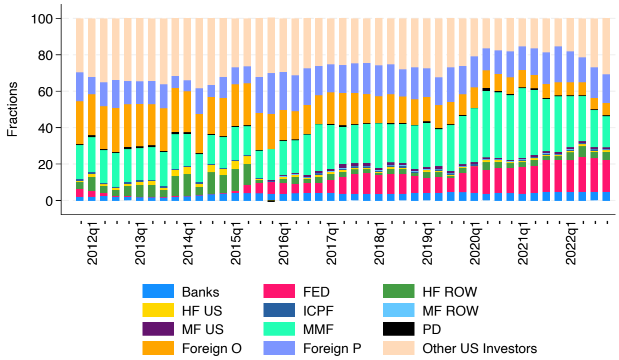
where $Z_t^\iota(m)$ denotes the total market value of Treasury holdings by sector ι in maturity bucket m at time t (in billions) and $S_t(m)$ the total market value supply of Treasuries in maturity bucket m at time t (in billions). The accounting identity in equation (1) implies that the sum across 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 on marginal holdings. Aggregate regressions highlight the Fed and MMFs as the largest absorbers of U.S. Treasuries. However, a maturity breakdown reveals significant differences. In short maturities, MMFs dominate, with hedge funds (HF ROW + HF US) surpassing the Fed. In intermediate maturities, foreign officials are primary, exceeding even the Fed, but they are negligible in long maturities.

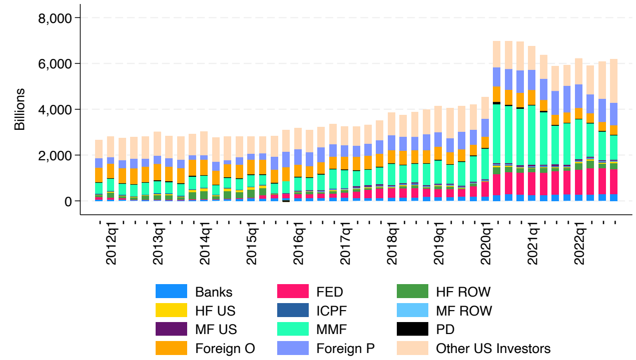
Panel (b) illustrates average holdings. We observe substantial differences between marginal and average holdings across sectors, particularly notable for hedge funds and primary dealers (PD). Specifically, HF ROW, HF US, and PD exhibit significantly higher prominence in marginal

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

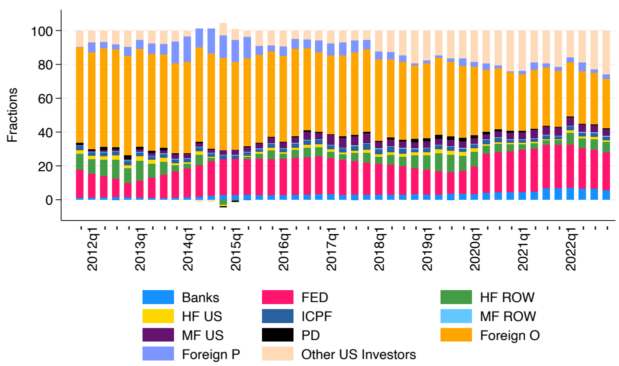
Left panels display the fraction of total U.S. Treasury outstanding (TAO) held by each investor type by maturity buckets. Right panels plot the corresponding market values (billions). Sectors are U.S. banks (Banks), Federal Reserve (FED), hedge funds outside the U.S. (HF ROW), U.S. hedge funds (HF US), 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 the quarterly sample period is 2011Q4-2022Q4.



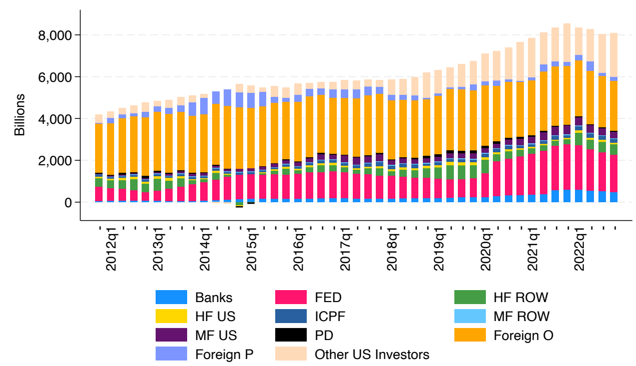
(a) $\tau < 1Y$: % Holdings as of TAO



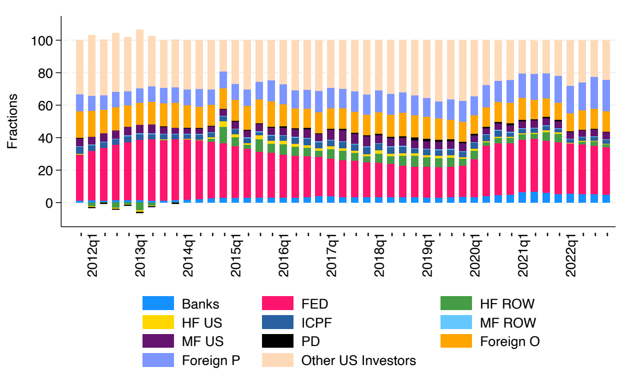
(b) $\tau < 1Y$: Total Holdings (billions)



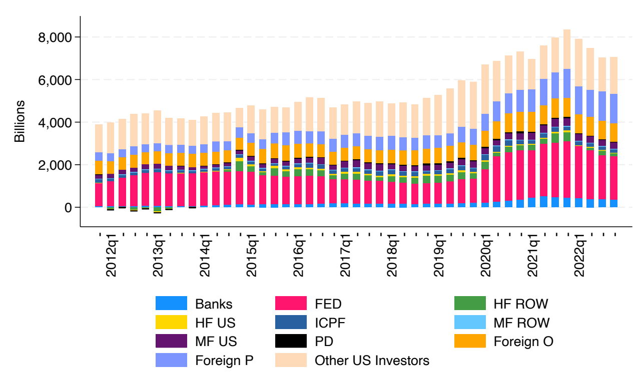
(c) $1 \leq \tau < 5Y$: % Holdings of TAO



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



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



(f) $\tau \geq 5Y$: Total Holdings (billions) holdings compared to their average holdings across all maturity buckets.

Panel (c) quantifies this gap by calculating ratios of marginal to average holdings by sector and maturity. The final row reports the average of these absolute ratios, indicating overall sector trading activity relative to their average holdings at the maturity-bucket level. HF ROW, HF US, and PD show high trading activity across maturities relative to their average holdings compared to other sectors, a characteristic indicative of arbitrage. This structural distinction is obscured at the aggregate level, as shown by the much smaller aggregate ratio in the first row of Panel (c). The next section further discusses arbitrageurs' unique features.

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. Panel (c) reports the ratio between Panel (a) and Panel (b), with an additional column "Avg. Abs. Ratio" that calculates the average of absolute ratios across the three maturity buckets. 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*		0.7	33.2	36.1***	15.9
$\tau \geq 5Y$	1.7	38.2***	15.7*	4.0*	3.4*	0.5*	6.0**		2.8**	23.1*	-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		0.9	11.7	46.5	5.5
$\tau \geq 5Y$	3.4	28.6	3.2	0.9	3.3	0.4	4.7		0.5	29.7	13.1	12.3
Panel (c): Marginal/Average Ratio												
	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	1.29	2.02	-0.43	-0.42	0.48	0.33	0.59	5.12	2.86	0.34	0.23	0.93
$\tau < 1Y$	0.68	1.00	2.95	2.58	0.38	0.00	1.00	2.26	5.67	0.29	0.52	0.12
$1Y \leq \tau < 5Y$	2.33	1.83	-4.41	-6.14	-1.88	0.67	2.79		0.78	2.84	0.78	2.89
$\tau \geq 5Y$	0.50	1.34	4.91	4.44	1.03	1.25	1.28		5.60	0.78	-0.13	0.52
Avg. Abs. Ratio	1.17	1.39	**4.09**	**4.39**	1.09	0.64	1.69	2.26	**4.02**	1.30	0.48	1.18

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 natural classification of sectors into two groups: "granular-demand investors" and "arbitrageurs". It furthermore nests each group as a special case under different parameter settings. 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, whom 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 rate 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 w.r.t. Z_t^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 risk of holding 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,

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

where each sector ι may have different beliefs about how aggregate states affect expected returns. The direct dependence on yields may reflect heuristic expectations — such as yield curve extrapolation (Fama and Bliss 1987), reaching for yield (Hanson and Stein 2015), or confusion between short-term and long-term interest-rate expectations (Shue et al. 2024).

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

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

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:

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

where we define the coefficient matrices $\psi^\iota = (\psi^\iota(1), \cdots, \psi^\iota(N))'$ and $\phi^\iota = (\phi^\iota(1), \cdots, \phi^\iota(N))'$. The outside portfolio covariance term $\mathbb{V}^\iota(R_{t+1}, \tilde{R}_{t+1}^\iota)\tilde{Z}_t^\iota$ reflects the risk interaction between Treasuries and other holdings, such as corporate bonds or foreign securities, which may vary by institution. We assume it can be decomposed into a linear function of 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).

Equation (6) shows that investor demand can, in principle, be derived from an optimization problem with beliefs, risk preferences, and non-pecuniary portfolio considerations. However, the precise structure of these inputs likely varies widely across investor types and reflects a combination of institutional constraints, regulation, and behavioral forces. In practice, factors such as capital requirements, liquidity mandates, or yield-seeking behavior differ across sectors and are difficult to observe or model directly. Attempting to specify and estimate a fully structural model for each investor type would require strong assumptions and face severe data limitations. Instead, we follow the philosophy of Koijen and Yogo (2019) and estimate sector-level demand functions flexibly from data. More formally, we lump the noise term with the inverse of the matrix in (6) as a normally distributed vector u_t^ι and express the solution in (7) as a demand function:

$$Z_t^\iota = \theta_0^\iota + B^\iota y_t - \theta^\iota \beta_t + u_t^\iota. \qquad (7)$$

We note that the model allows for cross-elasticities in that $Z_t^\iota(\tau)$ may depend on $y_t(\tau')$ for $\tau' \neq \tau$, i.e., the τ-th row of B^ι may have a non-zero element at position τ'. According to (6), this reflects a combination of asset return expectations depending on yields, and covariance across the term structure. Intuitively, the presence of cross-maturity elasticities allows granular investors to rebalance their portfolios toward higher-yielding maturities. Indeed, institutions such as insurance companies, mutual funds, and banks are known to exhibit "yield-seeking" behavior, actively reallocating across fixed-income instruments in pursuit of higher returns (Becker and Ivashina 2015; Hanson and Stein 2015; Choi and Kronlund 2018). However, we emphasize that alternative microfoundations for cross-elasticities are possible, and our main results do not depend on the specific interpretation of these demand functions.

While cross-maturity substitution by granular investors helps reduce yield differentials, this behavior is *fundamentally distinct from that of arbitrageurs* as modeled in our framework. In our framework, arbitrageurs are rational, forward-looking agents who price Treasuries based on expectations of macroeconomic states, interest rate paths, and future demand-supply imbalances. Crucially, they understand that returns are endogenously determined in equilibrium—driven by economic fundamentals and demand shocks—rather than being inferred directly from observed yields. In other words, arbitrageur expectation is the rational expectation $\mathbb{E}[R^{\tau}_{t+1} - r_t]$, which in equilibrium is driven by β_t, r_t, and supply and demand imbalances in the market. They serve as market makers, accommodate order flows across maturities, and intermediate between sectors, enforcing no-arbitrage conditions in prices subject to risk-bearing capacity. In contrast, granular investors respond directly to yields, as captured by the reduced-form demand function in equation (7), which reflects institutional frictions (such as mandates or capital regulation) or heuristic beliefs (such as reaching for yield). Because of this, the demand system in equation (7) is not suitable for arbitrageurs. Instead, we model arbitrageurs explicitly—solving their portfolio problem within the equilibrium framework in Section 4 to ensure consistency with market clearing and rational pricing.

The expression for Treasury holdings in (6) also clarifies how the model accommodates the dependence of the optimal Treasury portfolio 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 credit market shocks to influence 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, reducing 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 this provides a flexible way to capture the Fed's policy-driven behavior.

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 denote a maturity bucket as $m \in \{1,2,3\}$. Second, we add bond characteristics in the demand as control variables, although those bond characteristics will not be directly modeled in Section 4. 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 and $y_t(-m)$ denotes the weighted-average yield of the other maturity buckets. The vector $\mathbf{x}_t(m)$ is a vector of value-weighted bond characteristics for maturity bucket m: coupon, maturity bucket fixed effects, and bid-ask spread. The vector $\mathbf{Macro}_t$ denotes 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. We provide summary statistics for this set of variables in Table A3 and the correlation table in Table A4.

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, yet their below-one-year Treasury portfolio weight remains at 100%, failing to 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.[5]

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 excluding other yield from the estimation leads to a downward bias in the coefficient on own yield. 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^ι absorbs both the positive and negative effects, leading to a coefficient that is biased towards zero.[6]

In our specification, we assume that the macro variables are contemporaneously exogenous to investors, as, for instance, in Fang et al. (2025). That is, investor (latent) demand does not contemporaneously affect macro variables. We deem this assumption plausible. For instance, it is unlikely that idiosyncratic demand for Treasuries by banks or MMFs—such as that driven by changes in regulation—contemporaneously affects the GDP gap or inflation. In addition, we also assume that bond characteristics such as the pre-determined coupon rate are exogenous to latent demand. We will maintain these assumptions throughout our analysis.

We can estimate the demand system specified in equation (8) by GMM if it satisfies the

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

[6] Table A11 shows that the coefficient on own yield is attenuated closer to zero when not accounting for other yield.

moment condition:

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

The concern with this moment condition is that the error term may not be orthogonal to yields. For instance, if certain sectors have a large demand for Treasuries that is 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 propose an instrument designed to isolate fundamental demand for Treasuries, rather than relying on an instrument based on investment mandates in the spirit of Koijen and Yogo (2019). The reason is that, within the Treasury market, most institutional investors are not subject to investment mandates, except for MMFs, which have a clear mandate to operate only in the short-maturity bucket. Even seemingly long-term investors, such as pension funds, invest across the entire yield curve.[7] At the same time, instruments based on supply shocks, such as military spending shocks or the dependency ratio of the U.S. population, affect the entire Treasury market, making it difficult to argue that they influence the middle- and long-term maturity buckets differently, thereby complicating identification.

Therefore, we build on insights in Koijen and Yogo (2024) and the instrument used in Fang et al. (2025) and use the following three step procedure.[8] 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

[7] Consistent with the absence of investment mandates within Treasury markets, Jansen (2024) shows that, for safe European government bond markets, only 53% of the government bonds currently held in the portfolios of euro area investors were also in their portfolios in the previous quarter, compared to over 90% for equity markets (Koijen and Yogo 2019).

[8] Fang et al. (2025) impose market clearing at the country level by regressing investor demand and supply in a given country on the macroeconomic variables of that country. Rather than taking a cross-country approach, we adapt their methodology across the maturity spectrum by regressing holdings in each maturity bucket m on bond characteristics and macroeconomic variables in maturity bucket m.

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 $y_t(m)$. We apply the same logic to the value-weighted yield of the other buckets, for which the instrument is given by the value-weighted pseudo yield for the other maturity buckets: $\tilde{y}_t(-m)$.

Intuitively, Treasury demand and supply may vary systematically for reasons unrelated to idiosyncratic latent demand shocks. For instance, foreign officials reduce demand for short-term and medium-term Treasuries when inflation is high, because high inflation may signal a depreciation of the dollar. At the same time, the supply of short-term and medium-term Treasuries do not materially respond to inflation (see Table A5). As such, high inflation periods coincide with high implied pseudo yields. Hence, these systematic demand and supply variations are distinct from idiosyncratic latent fluctuations. Moreover, this relationship is purely coming from a supply and demand channel of inflation, because if demand and supply would respond in the same way to inflation, there would be no effect on pseudo yields. Our instrumental variable, derived from pseudo market clearing as in equation (10), exploits precisely these structural influences on yields. Empirically, this instrument is relevant. The first stage estimates of the demand system are summarized in Table A2. The corresponding Kleibergen-Paap statistic to test for weak instruments is 10.87, above the threshold of 10 for rejecting weak instruments (Stock and Yogo 2005).[9]

Crucially, the market-clearing condition in (10) generates nonlinear variation between yields and bond and macro factors, providing identifying power beyond the linear demand functions we estimate. A potential concern for identification is if investor demand itself is highly nonlinear in bond characteristics and macro.[10] To address this, we explicitly test and confirm that linear demand functions effectively explain Treasury holdings in Appendix B. We also show that higher-order nonlinear terms contribute negligible incremental R^2 to the linear demand model. In the same Appendix, we also present a stylized example to further illustrate and clarify our identification strategy.[11]

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, which are exogenous to latent demand shocks. This instrument satisfies the exclusion restriction under the identifying assumption that bond characteristics and macro variables are exogenous to investor latent demand, and linear demand functions capturing investor demand well. With the instrument, we can weaken moment

[9]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.

[10]In this instance, the coefficient on the pseudo yield would suffer from the classical omitted variable bias, because the pseudo yields and bond characteristics and macro are correlated following Equation (10).

[11]We thank Quentin Vandeweyer for discussing our paper and suggesting these examples.

condition (9) to:

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

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 U.S.**	**MMF**	**Other U.S.**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	58.200**	3.063	6.223*	127.827***	447.038**	130.233	-25.369	41.508
	[25.322]	[11.419]	[3.530]	[45.149]	[208.949]	[192.890]	[115.799]	[90.816]
$y_t(-m)$	-65.344**	0.161	-3.003	-142.372***	-629.311**	26.996	-99.204	-50.946
	[27.171]	[13.451]	[3.839]	[50.954]	[302.554]	[249.335]	[153.585]	[120.127]
Coupon Rate	-141.855***	4.358	-4.122	-127.601**	45.213	214.426	-486.904**	-323.739*
	[34.459]	[18.419]	[4.646]	[58.149]	[525.891]	[312.368]	[192.511]	[177.937]
Bid-Ask Spread	8.335	18.694***	3.157***	13.856	139.339	111.915	-103.880**	-66.715
	[7.811]	[4.531]	[1.184]	[15.654]	[131.097]	[77.105]	[46.464]	[56.264]
$\mathbb{1}\{1Y \leq \tau < 5\}$	58.064***	148.961***	13.210***	192.897***		-437.662***	2919.754***	-349.907***
	[15.201]	[4.442]	[2.088]	[25.558]		[123.159]	[92.670]	[84.446]
$\mathbb{1}\{\tau \geq 5\}$	-57.264	184.687***	10.891	53.689		554.181	134.991	28.672
	[45.645]	[20.892]	[6.661]	[86.051]		[401.451]	[225.911]	[178.260]
Credit Spread	13.576	-12.425	0.636	-39.986	-520.162***	280.422	97.032	-28.663
	[19.844]	[13.516]	[2.437]	[38.589]	[190.052]	[185.584]	[90.809]	[131.521]
Debt/GDP	661.200***	-3.418	42.325***	-3.926	5542.729***	2203.066**	-1797.427***	644.428
	[78.657]	[48.563]	[10.455]	[130.244]	[1000.035]	[920.759]	[576.206]	[540.549]
GDP Gap	10.912***	-4.579**	1.441***	12.164**	-75.801***	-9.468	-11.119	8.335
	[3.706]	[1.875]	[0.459]	[5.066]	[21.669]	[29.787]	[17.074]	[17.798]
Core Inflation	16.855**	-0.573	-2.191***	-2.760	62.817	-13.248	-75.916*	2.569
	[7.067]	[3.284]	[0.829]	[10.485]	[78.392]	[51.024]	[41.717]	[34.658]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap Statistic (*first stage*)	10.87	10.87	10.87	10.87	4.8	10.87	10.87	10.87

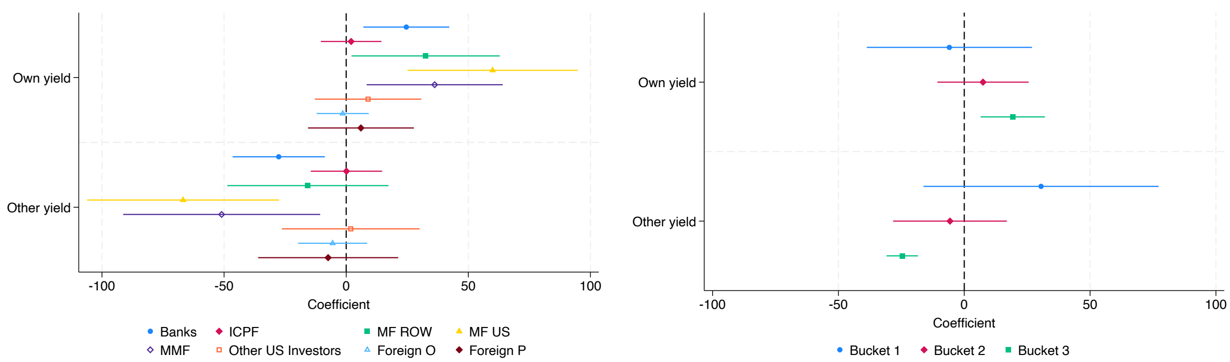
Table 3 shows the results using the IV methodology outlined in the previous section.[12] We find that all investors have downward sloping demand curves, except for the foreign official sector, whose coefficient is insignificant and small in economic magnitude. That is, granular-demand

[12] The results of the OLS estimates are in Appendix A6.

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 rise. Generally, we find that other elasticity is slightly higher than own elasticity, but the order of magnitude 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 have opposite signs.

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

Panel (a) plots the coefficients on own and other yields 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

Since Table 3 reports holdings in market values, it does not allow for direct comparison of price elasticities across investor types. Therefore, 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 the most price elastic, followed by banks.[13] ICPFs and foreign official investors are the least price elastic. Although our estimated own elasticities appear large for certain sectors, these sectors represent a small share of the total Treasury market (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

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

reduces the equilibrium market elasticity because investors tend to substitute across maturities rather than absorb net quantities. 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.

The heterogeneous cross elasticities across sectors revealed by Table 3 and Figure 3 are consistent with various mechanisms in the literature.[14] Banks show strong negative cross elasticity, reallocating toward higher-yield maturities in line with reaching-for-yield behavior (Hanson and Stein 2015), and a strong substitution across Treasuries due to liquidity regulation rules (e.g., all Treasuries are high quality liquid assets). In contrast, ICPFs display price-insensitive demand, reflecting preferred-habitat behavior (Vayanos and Vila 2021). Mutual funds, particularly U.S. funds, respond elastically to relative yields. This behavior reflects both active management and potentially behavioral factors, such as confusion between short and long rates (Shue et al. 2024), and return-chasing retail flows in bond mutual funds (Hanson et al. 2021).

MMFs, while restricted to short maturities by law, exhibit negative cross elasticity via investor flows: as longer-term yields rise, money flows out of MMFs, reducing their T-bill demand. These flow dynamics likely reflect extrapolative beliefs (Barberis et al. 2015). On the other hand, "Other U.S. investors"—primarily households that directly hold Treasuries and corporations—show muted cross-substitution. Unlike investors holding Treasuries through mutual funds or MMFs, households directly holding U.S. Treasuries benefit from tax advantages that accrue gradually over time, and they do not continuously observe market-value fluctuations, reducing their propensity for frequent portfolio adjustments to prices. Moreover, households directly holding Treasuries are wealthier individuals[15] that have better financial literacy and usage of professional advice, thus less subject to behavioral biases (Campbell 2006). Finally, foreign official investors exhibit strongly segmented demand, favoring short-to-intermediate maturities due to safety, liquidity needs, and reserve management guidelines. Their behavior is price-inelastic and consistent with the global savings glut narrative (Bernanke 2005). In contrast, foreign private investors are more yield-sensitive and may substitute across maturities or borders to reach for yield, akin to their domestic counterparts.

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,[16] while foreign

[14] For more detailed explanations of each sector, refer to Appendix C.1.

[15] According to this Survey of Consumer Finance report, "Ownership of any type of bond is concentrated among the highest tiers of the income and wealth distributions."

[16] 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.

official investors reduce their demand at that time. Furthermore, ICPFs have a large demand for long-term Treasuries, while foreign officials have a strong 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.

We conduct several robustness checks to further validate our findings and results are shown in Appendix C.2. First, to address concerns regarding potential endogeneity of bond characteristics and macroeconomic variables—thus ensuring our instrumental variable meets the exclusion restriction—Table A8 reports results from a specification where pseudo yields are inferred solely from coupon, maturity, GDP gap, and Debt/GDP, explicitly excluding bid-ask spread, credit spread, and core inflation. Second, Table A9 shows that the coefficients on own and other yields remain qualitatively similar even when we exclude macroeconomic controls from our baseline specification. Third, to evaluate potential substitution effects among Treasuries, MBS, and swaps, we augment our macroeconomic factors by including proxies such as the MBS spread (the 10-year MBS rate minus the 10-year Treasury yield) and the swap spread (the 10-year swap rate minus the 10-year Treasury yield). Table A10 demonstrates that our main results remain both qualitatively and quantitatively robust to these additional substitution channels. Finally, in Figure A3 we show that the coefficients do not materially change when we conduct a leave-one-quarter-out procedure in the demand estimations, so our results are not driven by any particular period in the sample. Overall, the robustness of our findings across these alternative specifications underscores the reliability of our instrumental variable approach.

3.4. Demand Functions of the Fed

For the Fed, we estimate its 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 presents the results. Notably, in the long-term maturity bucket, the Fed behaves similarly to granular-demand investors by increasing its long-term Treasury holdings when long-

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)
$y_t(m)$	-27.527	113.988	387.636**
	[92.968]	[168.508]	[155.940]
$y_t(-m)$	142.188	-86.527	-494.410***
	[132.311]	[210.090]	[76.189]
Coupon Rate	-23.035	-2499.609***	209.001
	[185.377]	[245.202]	[260.905]
Bid-Ask Spread	200.459***	77.362	-179.949***
	[57.755]	[73.179]	[65.603]
Credit Spread	34.025	139.717	-216.203
	[76.791]	[139.273]	[134.794]
Debt/GDP	3704.212***	295.297	4494.148***
	[334.351]	[646.766]	[1068.331]
GDP Gap	-6.754	-22.694	-49.643**
	[7.081]	[14.714]	[21.650]
Core Inflation	42.220	-68.298	158.069***
	[36.540]	[44.978]	[29.323]
Observations	45	45	45
Kleibergen-Paap Statistic (*first stage*)	4.8	18.22	23.52

term yields are elevated. Controlling for macroeconomic conditions and bond-specific features, this pattern suggests that the Fed expands its holdings in response to yield movements that are not driven by macro or bond fundamentals. This behavior aligns with qualitative evidence from Fed communications and QE episodes. For instance, in a March 2013 speech,[17] Fed Chair Ben Bernanke referenced a “growing body of research” showing that large-scale asset purchases reduce term premia and thus lower long-term interest rates. These premia are largely shaped by financial market conditions rather than macroeconomic fundamentals. The Fed has also deployed its balance sheet to offset increases in long-term yields perceived as technical, such as during the March 2020 Treasury market turmoil. Supporting this view, Haddad et al. (2024b) incorporate a “QE rule” explicitly responsive to yield levels. Theoretically, Caballero et al. (2024) show that such yield-

[17] See this script of the speech: Bernanke Speech Link.

based interventions—termed "financial conditions targeting"—can be optimal, even if financial conditions are not direct policy objectives.

Moreover, Table 4 also indicates significant "cross elasticity" for the Fed in the long-maturity bucket. This is consistent with the practice of pairing rate hikes with reductions in long-term asset holdings for policy consistency. The Fed's own documentation explicitly frames a "dual tightening" process. The 2014 and 2017 FOMC statements[18] made clear that short-term rate increases would come first, and balance sheet runoff would follow. In October 2018, Simon Potter of the NY Fed observed that "the FOMC had increased the federal funds target rate from 0% to 2-2¼%" and "the FOMC has reduced the size of the portfolio from nearly $4.3 trillion to about $4.0 trillion." Financial markets well understood these policies. For example, a Bloomberg news article[19] highlighted that the Fed was signaling a balance sheet reduction to begin shortly after the first rate increase, quoting Fed communications that this combination would strengthen the impact on financial conditions. Like for the yield elasticities, it is important to highlight that the cross-elasticities we capture are responses to policy shocks, rather than changes in the short-term rate because of macro-economic conditions.

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, suggesting prominent fiscal accommodations by the Fed.

To assess whether our main results are affected by the zero lower bound (ZLB) or inflation expectations—two key drivers of QE policy emphasized in the macroeconomic literature—we re-estimate the Fed regressions with additional controls. Specifically, we include the 5-year inflation swap rate as a proxy for inflation expectations and the spread between the federal funds rate (FFR) and the shadow rate from Wu and Xia (2016) to account for the ZLB. The coefficient on own yield decreases slightly from 387.6 to 341.8, and the coefficient on other yield shifts from –494.4 to –454.9, with both remaining statistically significant. These results suggest that the Fed's estimated demand elasticities are robust to concerns about confounding effects from the ZLB or inflation expectations.

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.

[18] See this summary for the history of the FOMC's policy normalization: FOMC Summary Link.

[19] See this link of the article: Bloomberg Article Link.

3.5. Distinct Demand of Arbitrageurs

As discussed in Section 3.1, arbitrageurs, in their pure form, are rational investors who understand how expected returns are determined by macroeconomic states and market dynamics. Unlike granular-demand investors, arbitrageurs do not hold Treasuries based on non-pecuniary benefits or reaching-for-yield behavior; rather, they actively and flexibly accommodate demand imbalances across maturities and markets in pursuit of arbitrage profits.

Table 5. **Demand System Results - Hedge Funds and Primary Dealers**

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 foreign hedge funds, U.S. hedge funds, or primary dealers 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, GDP Gap, and Core Inflation. We orthogonalize the coupon and the bid-ask spread with respect to maturity fixed effects. The quarterly sample period is from 2011Q4-2022Q4. HAC standard errors with optimal lags are reported in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Hedge Funds ROW	**Hedge Funds U.S.**	**Primary Dealers**
	(1)	(2)	(3)
$y_t(m)$	-231.289***	-49.700**	-36.556**
	[89.504]	[21.447]	[16.179]
$y_t(-m)$	285.071**	56.038**	62.381***
	[112.327]	[27.227]	[19.131]
Coupon Rate	73.196	-7.586	32.908
	[133.673]	[32.793]	[24.431]
Bid-Ask Spread	52.436	8.230	16.651**
	[50.274]	[13.538]	[7.517]
$\mathbb{1}\{1Y \leq \tau < 5\}$	309.739***	71.398***	55.931***
	[63.442]	[14.516]	[12.269]
$\mathbb{1}\{\tau \geq 5\}$	461.975***	98.824**	84.713***
	[168.887]	[39.559]	[31.321]
Credit Spread	67.828	2.873	17.063
	[104.522]	[26.826]	[18.438]
Debt/GDP	386.182	-37.772	270.318***
	[346.447]	[90.099]	[60.072]
GDP Gap	3.085	1.993	-2.668
	[16.121]	[4.184]	[3.287]
Core Inflation	-19.557	-8.373	-15.249***
	[27.033]	[6.715]	[3.984]
Observations	135	135	135
Kleibergen-Paap Statistic (*first stage*)	10.87	10.87	10.87

To highlight the difference between arbitrageurs and granular-demand investors, Table 5 shows

Table 6. **Short positions in U.S. Treasuries**

This table reports the fraction of periods in which investors held short positions in U.S. Treasuries, categorized by sector and maturity bucket. The fraction is calculated as the number of periods in which a given sector was short in a specific maturity bucket, divided by the total number of periods in the sample. Arbitrageurs include three sectors: foreign hedge funds, domestic hedge funds, and primary dealers. For explanations of other sector abbreviations, refer to the notes of Table 2. The numbers are in percentage points and the quarterly sample period is from 2011Q4 to 2022Q4. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Banks	Fed	ICPF	MF ROW	MF US	MMF	Arbitrageurs	Other US	Foreign O	Foreign P
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\mathbb{1}\{\tau < 1Y\}$	0	0	0	0	0	0	2	0	0	0
$\mathbb{1}\{1Y \leq \tau < 5Y\}$	0	0	0	0	0	0	4	4	0	0
$\mathbb{1}\{\tau \geq 5Y\}$	0	0	0	0	0	0	16	0	0	0

demand regressions for arbitrageurs following the same specification as other investor sectors. We find that arbitrageurs' demand exhibits loadings opposite to those typically observed for granular-demand investors as in Table 3: arbitrageurs' Treasury holdings decrease with own yields and increase with yields in other maturities. This negative own-yield elasticity is precisely the opposite of the typical positive elasticity observed for granular-demand investors. In addition, we find that the R^2 values from OLS regressions are on average 35% for arbitrageurs, compared to around 90% for most non-arbitrageurs, see Table A6 and A7 of the Appendix. This confirms that outside portfolios play a disproportionately large role for arbitrageurs and we need to structurally estimate their overall risk exposure rather than focusing solely on their Treasury holdings.

Similar empirical patterns are documented by Du et al. (2023b), who show that higher term spreads are associated with lower holdings by arbitrageurs, opposite to the typical yield-seeking behavior. Moreover, Table 6 shows that only arbitrageurs (hedge funds and primary dealers) held short positions in Treasuries over our sample period.[20] Specifically, arbitrageurs went short 2%, 4%, and 16% of the time in maturity bucket $\mathbb{1}\{\tau < 1Y\}$, $\mathbb{1}\{1Y \leq \tau < 5Y\}$, and $\mathbb{1}\{\tau \geq 5Y\}$, respectively. These results underscore arbitrageurs' unique role in providing market elasticity and benefiting from yield discrepancies across markets.

Given these conceptual and empirical differences, we adopt a structural approach to modeling arbitrageurs. Rather than interpreting their positions as reflecting stable demand functions, we explicitly model arbitrageurs as risk-averse investors who solve an optimization problem that determines their Treasury holdings. We structurally estimate their risk aversion and their exposure to latent outside portfolios, enabling a realistic and quantitatively precise characterization of

[20]For our granular demand investors, we find that Other U.S. Investors had short positions in U.S. Treasuries in the middle maturity bucket. However, as this is the residual sector, we cannot rule out that part of this sector also contains arbitrageurs.

arbitrageur behavior and its impact on equilibrium outcomes. This structural approach represents a central methodological contribution of our analysis, distinguishing our analysis from conventional demand estimation.

4. An Equilibrium Model of the Treasury Market

The previous section highlighted three key findings. First, granular-demand investors and the Fed exhibit downward-sloping demand curves. Second, their demand displays significant cross-maturity substitution. Third, arbitrageurs' holdings respond to yields in the opposite direction relative to other investors, and they systematically take short positions in Treasuries, reflecting their active arbitrage activities distinct from the broader market.

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 using 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 identify the magnitude of monetary policy shocks. Third, we account for latent outside assets held by arbitrageurs, adding realism by recognizing that arbitrageurs also hold outside assets, so the price of risk is not entirely driven by their 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 directly correspond to our estimated demand functions. Model 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 $R_{t+1} \equiv R_{t+1}^{(1)} = (1-P_t^{(1)})/P_t^{(1)} = \exp(r_t) - 1 \approx r_t$.

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, $\bar{\beta}$ represents the steady-state, and Φ is a matrix determining the persistence of the process.

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 Stein and Sunderam (2018), 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 and investors (including granular-demand investors and the Fed) 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 total government debt supply, which is $\sum_\tau S_t(\tau)$, as coming from a budget equation of the government, where Treasury supply adjusts to meet the need for government financing driven by macroeconomic conditions and the interest rate. Moreover, (20) captures the maturity-specific issuance, which as discussed in Greenwood et al. (2015b) is determined by fiscal needs (large immediate deficits often financed with short-term bills) and market conditions including the prevailing rate and the term premium (which is largely captured by the macro state β_t and Fed policy rate r_t). Therefore, our modeling of government debt supply reflects fiscal dynamics and the government's strategic decisions regarding maturity-specific issuance.

We model a representative arbitrageur as a special case 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,

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

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. Accordingly, the matrix of the demand response to Treasury yields is

$$\begin{pmatrix} a & -b \\ -b & a \end{pmatrix}. \tag{28}$$

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 the 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

$$S_t^{(\tau)} = \bar{S}^{(\tau)} + \zeta(\tau)\beta_t, \tag{29}$$

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,

$$\begin{aligned} p_t^{(1)} &= -r_t, \\ p_t^{(2)} &= -\frac{1+\rho_r+\gamma\sigma_r^2 b}{1+\frac{a}{2}\gamma\sigma_r^2} r_t - \frac{\gamma\sigma_r^2(\zeta(2)+\theta(2))}{1+\frac{a}{2}\gamma\sigma_r^2}\beta_t + \frac{\gamma\sigma_r^2}{1+\frac{a}{2}\gamma\sigma_r^2} u_t(2) + \frac{\frac{1}{2}-\gamma\bar{S}^{(2)}+\gamma\theta_0(2)}{\frac{1}{\sigma_r^2}+\frac{a}{2}\gamma}, \end{aligned} \tag{30}$$

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 (30), we summarize the drivers of Treasury price variation in the following proposition.

Proposition 1 (Decomposition of Treasury Pricing)**.** *One-period Treasury yield is only driven by monetary policy rate r_t, while the long-term Treasury yield is also affected by macro shocks and latent demand shocks.*

Proposition 1 is an intuitive result that follows directly from equation (30). 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 implies that a demand shock, either transient or permanent, has a smaller price impact if it comes from shorter maturities. This smaller impact is because shorter-maturity demand shocks are better accommodated by arbitrageurs given that arbitraging short-maturity Treasuries involves lower risks. In the limit case, 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 price sensitivity of Treasuries 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" of the market. Then the long-term Treasury price becomes

$$p_t^{(2)} = -\frac{2b}{a} r_t - \frac{2}{a}(\zeta(2) + \theta(2))\beta_t + \frac{2}{a} u_t(2) + \frac{2}{a}(\theta_0(2) - \bar{S}^{(2)}). \tag{31}$$

This is a case where Treasury prices are entirely driven by supply and demand absent any 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

$$p_t^{(2)} = -(1+\rho_r)r_t + \frac{1}{2}\sigma_r^2, \tag{32}$$

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.

Next, according to the non-arbitrageur demand in (19), along with the simplified elasticity matrix in (27) and solution in (30), we obtain non-arbitrageur holdings as

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

Arbitrageur holdings are $X^{(2)} = S_t^{(2)} - Z_t^{(2)}$, which in this simplified model is

$$X_t^{(2)} = \frac{\frac{1}{\sigma_r^2}\bar{S}^{(2)} + \frac{1}{4}a - \theta_0(2)\frac{1}{\sigma_r^2}}{\frac{1}{\sigma_r^2} + \frac{a}{2}\gamma} - \frac{\frac{a}{2}(1+\rho_r) - b}{1+\frac{a}{2}\gamma\sigma_r^2} r_t + \frac{\zeta(2)+\theta(2)}{1+\frac{a}{2}\gamma\sigma_r^2}\beta_t - \frac{1}{1+\frac{a}{2}\gamma\sigma_r^2}u_t(2). \tag{34}$$

We note that if $\gamma \to \infty$, non-arbitrageur's long-term Treasury holding (33) converges to $\bar{S}^{(2)} + \zeta(2)\beta_t$, which is the total debt supply in this simplified model as in (29), and arbitrageurs' long-term Treasury holding in (34) become zero. On the other hand, when $\gamma \to 0$, arbitrageurs fully absorb demand shocks from other investors, leading to loading of -1 on u_t in (34).

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 an overreaction 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 an underreaction 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. This reduces the amount that 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: 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 (34)). This counteracts the first force and may cause the yield to be even higher than predicted by 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 a one-period and its price is fully determined by the monetary policy rate. In the full model, we expect permanent shocks to have a stronger effect, as they exert a greater influence on the arbitrageurs' pricing kernel, 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 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 (see equation (A20) in Appendix D and discussions before it). This is a typical result in most asset pricing models that assumes idiosyncratic risks are not priced. It is important to note that it does not imply that latent demand shocks have no price impact, since u_t can still directly affect prices via demand pressure.

Define the expected return on Treasuries of maturity τ as $\mu_t^{(\tau)} \equiv \mathbb{E}_t[R_{t+1}^{(\tau)}]$, where $R_{t+1}^{(\tau)} = \exp(r_{t+1}^{(\tau)}) - 1 \approx r_{t+1}^{(\tau)} + \frac{1}{2}\mathbb{V}_t[r_{t+1}^{(\tau)}]$. The approximation becomes exact when we take a continuous-time approach[21]. 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).

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

$$\mu_t^{(\tau)} - r_t = \hat{A}(\tau-1)'\underbrace{\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\Sigma\hat{A}(\hat{\tau}-1)X_t(\hat{\tau})\right) + \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t\right)}_{\lambda_{\beta,t}} + A_r(\tau-1)'\underbrace{\gamma\left(\sum_{\hat{\tau}=2}^{N}\left(\sigma_r^2 A_r(\hat{\tau}-1)X_t(\hat{\tau})\right) + \sigma_r\tilde{\sigma}_r\tilde{X}_t\right)}_{\lambda_{r,t}}, \tag{35}$$

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 (35) 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 moments 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

$$\begin{aligned} \Sigma^{1/2}\tilde{\sigma}\tilde{X}_t &= \Psi\beta_t + \Lambda r_t + \psi \\ \sigma_r\tilde{\sigma}_r\tilde{X}_t &= \Psi_r\beta_t + \Lambda_r r_t + \psi_r, \end{aligned} \tag{36}$$

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 additional parameters must be estimated jointly within the full model. Intuitively, outside-asset exposure captures any non-Treasury risk borne by arbitrageurs that may interact with their Treasury

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

pricing, including positions in interest rate swaps (Du et al. 2023b), Treasury futures (Barth et al. 2021), and other fixed-income portfolios.

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[22]

$$X_t(\tau) = \left(\bar{S}(\tau) + \zeta(\tau)'\beta_t + \zeta_r(\tau)'r_t\right) - \left(\theta_0(\tau) - \alpha(\tau)'p_t - \theta(\tau)'\beta_t + u_t(\tau)\right). \tag{37}$$

Expanding the expected return $\mu_t^{(\tau)}$ and plugging the equilibrium arbitrageur holdings X_t^τ of (37) into the pricing equation (35), 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 on each term 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 pricing of Treasury securities.

To solve and estimate the model, we denote the Treasury log price (scaled to have a face value of 1) in the data as $p_t^o(\tau)$ for maturity τ, and the model-implied Treasury log price as $p_t(\tau)$, which is determined in equilibrium by (26). Similarly, we denote the actual arbitrageur holding as $X_t^o(\tau)$ for maturity τ, and the model-implied holding as $X_t(\tau)$. Then we estimate the remaining parameters by minimizing the squared sum of percentage deviations for both quantities and prices:

$$\min_{\{\gamma, A_u, \Psi, \Psi_r, \Lambda, \Lambda_r, \psi, \psi_r\}} \mathbb{E}\left[\sum_t \sum_\tau \left(\frac{X_t(\tau) - X_t^o(\tau)}{\bar{X}^{abs}(\tau)}\right)^2 + \sum_t \sum_\tau \left(p_t(\tau) - p_t^o(\tau)\right)^2\right] \tag{38}$$

subject to the equilibrium restrictions that discipline these parameters, detailed in Appendix D.3. In (38), we divide the quantity difference by $\bar{X}^{abs}(\tau)$, which is the average of $|X_t(\tau)|$ in the data, so that both quantity deviation and price deviation are expressed in percentages (note that the difference of a log price is already a percentage difference in price). In implementing (38), we restrict attention to integer-year maturities for log prices to reduce redundancy in fitting the term structure, and we focus on medium- and long-maturity buckets for quantities.

The intuition for how the estimation problem in (38) disciplines arbitrageur risk aversion γ and outside-portfolio exposure is as follows. First, risk aversion γ serves as the bridge between quantities and prices. In the simple model, as shown in equations (30) and (34), when there are macro shocks $\beta_t(2)$, the equilibrium response of the long-term Treasury price is $-\gamma\sigma_r^2(\zeta(2) +$

[22] 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.

$\theta(2))/(1+\frac{\alpha}{2}\gamma\sigma_r^2)$, and the response of arbitrageur holdings is $(\zeta(2)+\theta(2))/(1+\frac{\alpha}{2}\gamma\sigma_r^2)$. Thus, the ratio of the price response to the holding response is $-\gamma\sigma_r^2$. A similar relationship holds for permanent demand shocks $\theta_0(2)$ and latent demand shocks u_t. Since σ_r is directly observed in the data, the relative response of long-term Treasury prices to arbitrageur holdings enables us to infer the risk aversion parameter γ. Second, as shown in Appendix D.4, outside-portfolio exposure enters the determination of term structure loadings. For example, arbitrageur's interest-rate swap holdings will affect arbitrageurs' exposure to interest-rate risks, which heterogeneously impacts the pricing of the Treasury term structure. Therefore, by matching the entire term structure, we recover the arbitrageurs' outside-portfolio exposure.

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 monetary policy rule parameters have been relatively stable since 1990 (Clarida et al. 2000; Mehra and Sawhney 2010). Our detailed estimation is in Appendix E. We find that 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 expectations channel through which monetary policy affects long-term yields, which plays a critical role in understanding how the yield curve responds to monetary policy shocks ε_{t+1}^r.

We estimate problem (38) using our main data sample from 2011Q4 to 2022Q4. In Figure A6 of Appendix E.1, we show that the model-implied yields can fit the term structure reasonably well, both across maturities and over time.

The resulting absolute risk-aversion parameter is $\gamma = 0.04$. As shown in later sections, this leads to a relatively elastic Treasury market. The novelty of our approach is that we rely on the connection between yields and 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 leverages granular data that allow us to distinguish arbitrageurs explicitly from granular-demand investors.

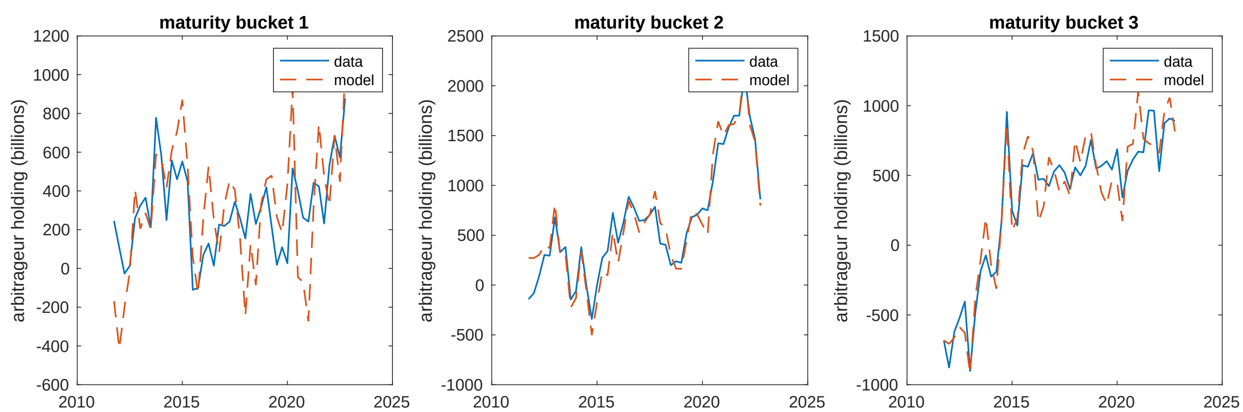
Finally, to evaluate the robustness of our estimation results, particularly to the specification of demand functions, we use a bootstrap procedure to shuffle the latent demand component and redo the instrumental variable construction, demand-function estimation, and model estimations. In Appendix D.5, we show that the standard deviation of the distribution for estimated γ is 0.014, which indicates a tight distribution around the baseline estimation of $\gamma = 0.04$.

4.5. Arbitrageur Holdings: Model v.s. Data

In Figure 4, we compare our model-implied arbitrageur Treasury holdings with actual data aggregated from hedge fund and broker-dealer holdings. We construct model-implied arbitrageur holdings as total Treasury supply minus non-arbitrageur demand according to the market clearing condition in equation (25).

Figure 4. **Arbitrageur Holdings: Model v.s. Data.**

In this figure, we plot arbitrageur holdings in the data versus in the model. The model-implied arbitrageur holding at maturity τ is $X_t(\tau) = S_t(\tau) - Z_t(\tau)$, which is the total Treasury supply minus non-arbitrageur demand.



The close alignment between the model-implied positions and observed arbitrageur holdings highlights the success of our structural equilibrium model in capturing the quantity-price dynamics of arbitrageur demand. This tight match not only confirms the validity of our estimated risk-aversion parameter but also reinforces the identification strategy underlying our empirical and structural estimation. By explicitly linking observed arbitrageur holdings to Treasury prices, our approach quantifies how arbitrageur behavior shapes Treasury market outcomes.

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. The explicit characterization of quantities and prices distinguishes our paper from the extant literature on latent factors (Ang and Piazzesi 2003; Bikbov and Chernov 2010; Joslin et al. 2014).

5.1. Decomposing Treasury Pricing

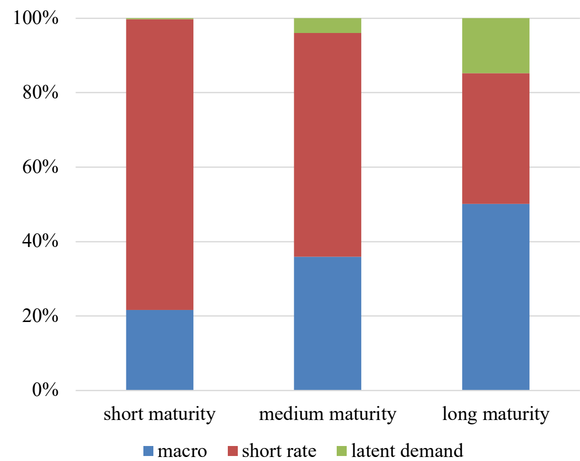
We start by taking guidance from the model to decompose Treasury prices and their variation into their underlying drivers. First, using equation (26), we decompose yields into contributions from macroeconomic states, monetary policy rate, as well as latent demand. Second, from equations (17) and (20), we obtain a decomposition of yield changes into sector-specific demand shocks and supply shocks, with $\Delta Z_t^\iota = -\theta(\tau)\Sigma^{1/2}\varepsilon_t + u_t$ as the demand shock of sector ι and $\Delta S_t = \zeta'\varepsilon_t + \zeta_r\sigma_r\varepsilon_t^r$ as supply shocks. 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 5 shows that the relative contribution of economic forces varies across the term structure, in line with the model predictions in Proposition 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 has almost 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 expands. In particular, for long-maturity Treasuries, macroeconomic variables can explain about half of the variation in yields.

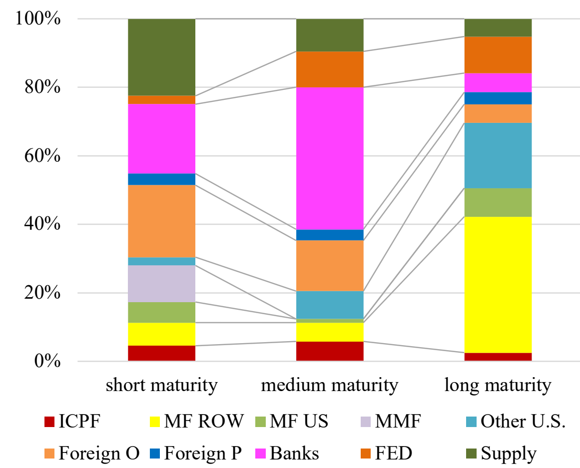
As shown in Figure 5 Panel (b), the second decomposition varies across the maturity structure of Treasuries. Although banks hold only a small share of the total market (3.4% as shown in Table 2), they play a sizable role in transmitting shocks to the Treasury market, especially in the medium-maturity bucket (1∼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

Figure 5. **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.

predominantly depends on how actively a sector responds to shocks.

5.2. Arbitrageur Risk Aversion and Treasury Market Elasticity

In our model with arbitrageurs, the response of Treasury yields to shocks is shaped not only by granular-investors' individual demand elasticities, but, critically, also by arbitrageur's 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

$$p_t = \alpha^{-1}\left((\theta_0 - \theta\beta_t + u_t) - \left(\bar{S} + \zeta\beta_t + \zeta_r r_t\right)\right). \tag{39}$$

Thus, absent arbitrageurs, the equilibrium price response to a demand shock is simply α^{-1}. In contrast, 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 7, we illustrate the equilibrium Treasury price response (in %) at each maturity bucket to a latent demand shocks that is 1% of outstanding amount 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, given a shock to short-maturity demand, the response of the long-maturity Treasury price is 20 times larger than that of the short-maturity Treasury price. Moreover, demand shocks on longer-maturity Treasuries are more powerful, as reflected by larger multipliers associated with shocks to longer maturities.

Table 7. **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 those in Panel (a).

Panel (a): With Arbitrageur		price change (%) of	
	short maturity	medium maturity	long maturity
shock on short maturity	0.001	0.009	0.020
shock on medium maturity	0.012	0.089	0.224
shock on long maturity	0.022	0.173	0.509
Panel (b): Without Arbitrageur			
shock on short maturity	0.128	0.585	3.011
shock on medium maturity	0.650	2.104	12.688
shock on long maturity	0.298	1.133	5.038
Panel (c): Price Impact Ratio (Panel (b)/Panel (a))			
shock on short maturity	111.406	68.661	147.312
shock on medium maturity	54.679	23.561	56.653
shock on long maturity	13.619	6.536	9.897

In Panel (b), we effectively remove arbitrageurs by setting $\gamma = \infty$, and examine the corresponding price multipliers, obtained by scaling (39) 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 7, which reports the ratio of the price impacts in the case without arbitrageurs and the baseline case. On average, the price impact in the case without arbitrageur is more than 20 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.

Finally, using long-run average values of total Treasury supply in each maturity bucket as weights, we convert the numbers in Panel (a) of Table 7 into a total market multiplier[23], which is 0.37. This implies that for a \$100 billion dollar demand shock to the whole Treasury market, the total Treasury valuation increases by \$37 billion dollars. On the other hand, the multiplier is 0.93 for a representative permanent demand shock. 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 that the Treasury market is in aggregate quite elastic. In Appendix D.5, we further evaluate the uncertainty of our estimation for the Treasury market elasticity using a bootstrap procedure that shuffles the latent demand component. We find that with almost certainty the Treasury market is more elastic than both the corporate bond market and the stock market.

Notably, however, once we remove arbitrageurs by setting $\gamma = \infty$, the estimated Treasury market multiplier increases to 8.74, 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 effects of demand shocks.

5.3. The Term Structure of Market Elasticity

As shown in the previous subsection, the sensitivity of Treasury prices to demand shocks varies substantially across maturities. To quantify this heterogeneity, we compute the market elasticity $\mathscr{E}(\tau)$ at maturity τ, defined as the percentage change in total Treasury valuation resulting from a 1% increase in demand for Treasuries at that maturity, relative to the total market value. We refer to $\mathscr{E}(\tau)$, which is a function of τ, as the term structure of market elasticity.

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

Figure 6. **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.

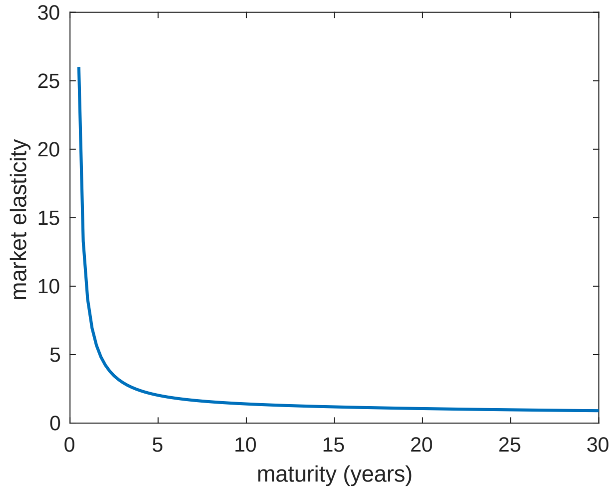


Figure 6 reveals a sharply downward-sloping term structure of market elasticity: elasticity is around 14 at the one-year maturity but falls below 1 at longer maturities. This steep decline reflects the greater risk that arbitrageurs must bear when absorbing demand shocks in longer-dated securities.

The downward slope helps reconcile seemingly contradictory findings in the literature—namely, that quantitative easing (QE) has substantial effects on long-term yields, while shifts in T-bill supply have minimal price impact. For instance, Krishnamurthy and Vissing-Jorgensen (2011) and D'Amico and King (2013) show that central bank purchases of long-term bonds significantly reduce yields by compressing term premia. In contrast, Greenwood et al. (2015c) find that even large changes in T-bill supply shift convenience yields by only a few basis points, consistent with highly elastic demand at the short end. Our results also align with recent estimates of aggregate Treasury demand elasticity of the non-T-bill segment, such as an elasticity of roughly 2 in Eren et al. (2023) and around 1 in Chaudhary et al. (2024).

6. Conventional and Unconventional Monetary Policies

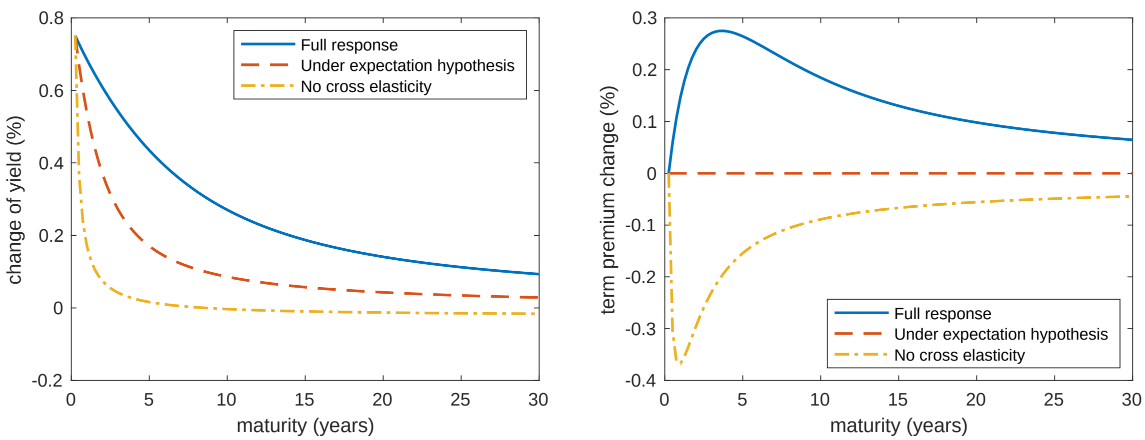
In this section, we use the model to analyze monetary policy shocks and quantitative easing.

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 7, 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 7, although the expectation component declines quickly over maturity, approaching zero at around the 15-year maturity, the full yield response in the model remains strong even at the 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).

Figure 7. **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

Although the 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. In models where the expectation hypothesis holds, the term premium should not respond. 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. This is because it lowers Treasury prices and thereby increases non-arbitrageur demand, prompting arbitrageurs to reduce their Treasury holdings, which in turn lowers the 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 leaving a larger share 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 7 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[24].

Table 8. **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

Our mechanism hinges on the portfolio adjustments and rebalancing of granular-demand investors in response to monetary policy shocks. The granular model not only allows us to examine these dynamics closely but also incorporates the government's strategic debt issuance in reaction to such shocks. Table 8 illustrates the resulting supply and demand adjustments following a monetary policy tightening. Since total debt supply is governed by the contemporaneous macro variable Debt/GDP—which remains unchanged in this experiment—the net change in Treasuries outstanding is zero. However, with tighter monetary policy raising the term premium, the government shifts issuance toward short-term bonds and reduces long-term supply. This can be viewed as minimizing

[24] 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.

issuance costs, as higher term premia make long-term borrowing less attractive.

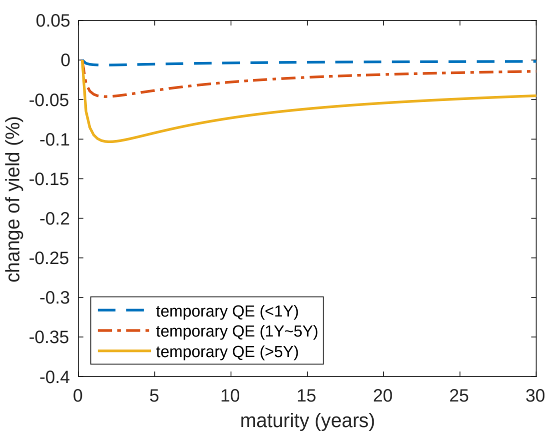
As shown in Figure 7, short-term yields rise more than long-term yields.[25] In response, nearly all investor types rebalance toward higher-yielding short-term Treasuries, reflecting strong cross-substitution effects—particularly among highly elastic money market funds (MMFs) and other U.S. investors. The Federal Reserve, consistent with monetary tightening, aggressively sells long-term Treasuries. This behavior, driven by a strong negative sensitivity to short-term rates (see Table 4), results in sales that exceed the reduction in long-term supply. Arbitrageurs absorb most of the net sales of long-term Treasuries, which contributes to elevated risk premia. Notably, the foreign official sector reduces holdings across all maturities, likely in an effort to defend their currencies amid a stronger dollar or to provide liquidity in response to domestic capital outflows.

6.2. The Impact of Quantitative Easing

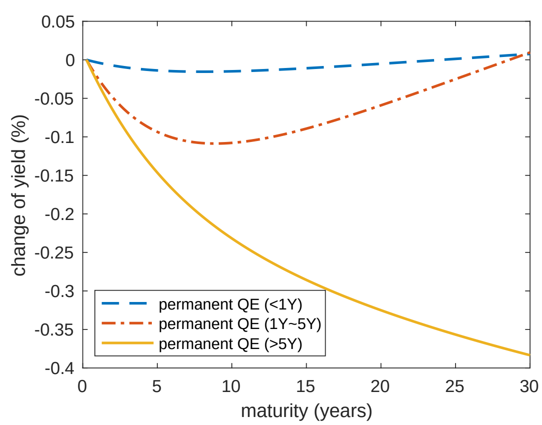
Through the lens of our model, we can interpret quantitative easing (QE) policies as changes in the Fed's demand, either temporary or permanent. We distinguish between transient QE, modeled as an increase in u_t, and permanent QE, which amounts to an increase in θ_0. Proposition 4 suggests that QE increases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

Figure 8. **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

Figure 8 shows the quantitative impact of a transient QE and a permanent QE. In both cases, we

[25]This implies a decline in the term spread during monetary policy tightening. The term spread refers to the difference between long-term and short-term yields, whereas the term premium captures the difference between actual yields and the expected path of short-term rates absent a risk premium.

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 8 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-term Treasuries. It is thus natural that the Fed usually purchases long-term Treasuries in QE programs. These patterns are significantly amplified in the case of permanent QE, as panel (b) of Figure 8 illustrates. Moreover, 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. (2023b), 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 8. Second, the expected duration of the QE purchase is between one quarter (panel (a) of Figure 8) and permanent (panel (b) of Figure 8). 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, nesting 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 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 the cross-substitution of granular-demand investors forces arbitrageurs to increase their long-term Treasury holdings and increase the price of risk. Finally, the effectiveness of QE is significantly driven by the expected persistence of the Fed's interventions.

We view our paper as a building block of a framework 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.

Barberis, N., Greenwood, R., Jin, L., and Shleifer, A. (2015). X-capm: An extrapolative capital asset pricing model. *Journal of financial economics*, 115(1):1–24.

Barth, D., Kahn, R. J., et al. (2021). Hedge funds and the treasury cash-futures disconnect. *OFR WP*, pages 21–01.

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. (2005). The global saving glut and the us current account deficit.

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*.

Caballero, R. J., Caravello, T. E., and Simsek, A. (2024). Financial conditions targeting.

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

Campbell, J. Y. (2006). Household finance. *The journal of finance*, 61(4):1553–1604.

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

Chaudhary, M., Fu, Z., and Zhou, H. (2024). Anatomy of the treasury market: Who moves yields? *Available at SSRN*.

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).

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

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.

D'Amico, S. and King, T. B. (2013). Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply. *Journal of financial economics*, 108(2):425–448.

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, C. and Xiao, K. (2024). Dissecting bond market transmission of monetary policy. *Available at SSRN*.

Fang, X., Hardy, B., and Lewis, K. K. (2025). Who holds sovereign debt and why it matters. *Review of Financial Studies*. Forthcoming.

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.

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

Greenwood, R., Hanson, S., Stein, J. C., and Sunderam, A. (2023a). A quantity-driven theory of term premia and exchange rates. *The Quarterly Journal of Economics*, 138(4):2327–2389.

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

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

Greenwood, R., Hanson, S. G., Rudolph, J. S., and Summers, L. (2015b). The optimal maturity of government debt. *The* $13 trillion question : How America manages its debt, pages 1--41$.

Greenwood, R., Hanson, S. G., and Stein, J. C. (2015c). A comparative-advantage approach to government debt maturity. *The Journal of Finance*, 70(4):1683–1722.

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

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. (2024). 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. (2025). How credit cycles across a financial crisis. *The Journal of Finance*.

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. *National Bureau of Economic Research*.

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.

Mehra, Y. P. and Sawhney, B. (2010). Inflation measure, taylor rules, and the greenspan-bernanke years. *FRB Richmond Economic Quarterly*, 96(2):123–151.

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.

Shue, K., Townsend, R., and Wang, C. (2024). Categorical thinking about interest rates.

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

Siriwardane, E., Sunderam, A., and Wallen, J. L. (2022). Segmented arbitrage. *National Bureau of Economic Research*.

Stein, J. C. and Sunderam, A. (2018). The fed, the bond market, and gradualism in monetary policy. *The Journal of Finance*, 73(3):1015–1060.

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. *Harvard University*.

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

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 2025).

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.

Additionally, our identification relies on the linear relationship between demand and the bond characteristics and macro variables. In order to test if demand is linear in the bond characteristics and the macro factors, we run regressions of total demand by granular demand investors and the

Table A1. **Identification: Linear versus Nonlinear Demand**

This table presents regression results of total demand from granular demand investors and the Fed on bond characteristics and macroeconomic variables. The specifications include both the levels of these variables and their squared terms to account for potential nonlinear effects. We also include an F-test and the corresponding p-value to test for the joint significance of the squared terms. The dependent variable is the market value of U.S. Treasuries held in aggregate by the granular demand investors and the Fed in maturity bucket m at time t. The bond characteristics include Coupon Rate and Bid-Ask Spread and the macro variables include Credit Spread, Debt/GDP, Credit Spread, GDP Gap, and Core Inflation. Column (1) and (2) show the results for $\tau < 1Y$, Column (3) and (4) for $1Y \leq \tau < 5$, and Column (5) and (6) for $\tau \geq 5$. The quarterly sample period is from 2011Q4-2022Q4. We report robust standard errors in brackets; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	$\mathbb{1}\{\tau < 1Y\}$	$\mathbb{1}\{\tau < 1Y\}$	$\mathbb{1}\{1Y \leq \tau < 5Y\}$	$\mathbb{1}\{1Y \leq \tau < 5Y\}$	$\mathbb{1}\{\tau \geq 5Y\}$	$\mathbb{1}\{\tau \geq 5Y\}$
Coupon Rate	-1274.809*	-1358.667*	-2773.208***	-3148.019***	709.191***	756.717***
	[497.488]	[539.382]	[389.843]	[506.726]	[131.352]	[176.144]
Bid-Ask Spread	566.736***	608.175**	-127.058	-193.853	-33.026	-12.400
	[142.978]	[169.776]	[78.018]	[97.093]	[56.401]	[73.605]
Credit Spread	502.511*	1941.087	-50.680	565.899	-398.073***	-1401.146
	[205.979]	[1644.713]	[187.132]	[1807.396]	[98.048]	[764.456]
Debt/GDP	12046.650***	-33819.475	2509.604***	5603.266	7544.221***	30351.381**
	[736.660]	[22664.816]	[666.473]	[16000.784]	[526.681]	[9889.370]
GDP Gap	-62.782*	71.424	49.434	97.636	-20.777	-117.506**
	[24.036]	[88.203]	[25.996]	[66.470]	[23.867]	[33.111]
Core Inflation	-74.134	-410.489	-203.696***	29.666	229.652***	208.663
	[63.865]	[395.634]	[49.210]	[294.310]	[30.379]	[190.717]
Coupon Rate^2		-3117.703		-1332.612		19.040
		[2901.439]		[1121.441]		[199.809]
Bid-Ask Spread^2		254.770		112.233		-16.937
		[300.708]		[71.065]		[78.800]
Credit Spread^2		-603.893		-255.652		474.973
		[817.309]		[873.828]		[374.298]
Debt/GDP^2		27770.846*		-2072.702		-13571.490*
		[13504.637]		[9618.073]		[6222.986]
GDP Gap^2		8.385		8.562		-11.530***
		[8.906]		[5.509]		[2.801]
Core Inflation^2		45.281		-32.753		5.282
		[52.321]		[34.769]		[25.320]
R-squared	0.931	0.943	0.841	0.857	0.952	0.966
Observations	45	45	45	45	45	45
F-statistic		1.59		1.41		3.46
p-value		.18		.24		.01

Table A2. **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)
$\tilde{y}_t(m)$	0.736***	0.446***
	[0.051]	[0.052]
$\tilde{y}_t(-m)$	0.667***	0.866***
	[0.124]	[0.079]
Coupon Rate	-0.369*	-0.862***
	[0.203]	[0.158]
Bid-Ask Spread	-0.077	-0.121*
	[0.075]	[0.067]
$\mathbb{1}\{1Y \leq \tau < 5\}$	-1.115***	-0.104
	[0.345]	[0.238]
$\mathbb{1}\{\tau \geq 5\}$	-0.112	-0.823***
	[0.271]	[0.209]
Credit Spread	0.826***	0.890***
	[0.207]	[0.173]
Debt/GDP	2.042***	0.990
	[0.781]	[0.707]
GDP Gap	-0.223***	-0.215***
	[0.051]	[0.032]
Core Inflation	0.137**	0.131***
	[0.064]	[0.050]
R-squared	0.885	0.891
Observations	135	135

Fed on these variables and add second-order terms to test for their joint significance. Table A1 shows the results. For these tests, we use robust standard errors rather than HAC standard errors to derive meaningful R^2s. The table reveals that the improvement in R^2 by adding second order terms is minimal and only increases by a maximum of 1.7% compared to about 90% already explained by the linear model. Moreover, for two of the three buckets we cannot reject the null that the second order variables are jointly insignificant. The third bucket rejects the null likely due to saturation of the model with an R^2 close to 100%. Hence, we conclude that our assumption of linear demand in bond characteristics and macro variables holds in our sample.

Finally, we check the relevance of our IV and show the first-stage results in Table A2. We find an overall strong first stage with a high R^2.

C. Additional Empirical Analysis

C.1. Sector-by-Sector Explanations of Demand Elasticities

Below, we provide more detailed explanations for our empirical findings in Table 3 at the sector level.

Banks

Banks show strong own-yield and negative cross elasticities, indicating that they shift toward maturities offering higher yields. This behavior is consistent with reaching-for-yield as theorized in Hanson and Stein (2015): when longer maturities yield more, banks expand duration to enhance returns. Banks face minimal regulatory barriers to this behavior since all Treasuries qualify as high-quality liquid assets (HQLA) under liquidity coverage rules. While they are subject to interest rate risk, many hold Treasuries in held-to-maturity accounts or manage risk with hedging. Their behavior deviates from pure preferred habitat, instead reflecting yield-seeking optimization.

Insurance Companies and Pension Funds (ICPFs)

ICPFs display minimal substitution across maturities. Both own- and cross elasticities are statistically weak, indicating price-insensitive demand. This behavior closely aligns with the preferred-habitat theory (Vayanos and Vila 2021), driven by the need to match long-duration liabilities. Regulatory capital regimes and accounting rules further reinforce this rigidity.

While reaching-for-yield has been documented in insurers' corporate bond portfolios (Becker and Ivashina 2015), their Treasury portfolios remain anchored in long maturities for liability management, rather than yield optimization. This suggests that ICPFs' Treasury demand reflects their structural mandates rather than valuation considerations. Moreover, our findings do not support the "hunt-for-duration" channel (Domanski et al. 2017), which predicts that ICPFs would increase their demand for Treasuries when yields are low because the duration of liabilities typically exceeds that of assets, implying upward-sloping demand curves for this sector. In contrast, although statistically insignificant, our estimates, if anything, suggest that ICPFs' demand for Treasuries increases with yield.

Mutual Funds (U.S. and ROW)

Mutual funds—particularly U.S. funds—show strong cross-substitution. Table 3 indicates they significantly reallocate in response to relative yield changes. This is in line with active portfolio management and reaching for yield, where fund managers adjust duration to maximize returns. Another possibility is that investors confuse short rates with long rates (Shue et al. 2024) and when the short-term rate increases, they expect the long-term rate to increase as well and therefore reduce long-term bond holdings.

While some funds face benchmark-based duration mandates (e.g., short-term or intermediate-term bond funds), sector-wide substitution likely reflects shifts across funds and return-chasing investor flows. Mutual funds behave as yield-sensitive investors, and their activity enforces relative pricing across the maturity spectrum. Foreign mutual funds exhibit similar, albeit weaker, substitution, possibly due to currency hedging frictions or greater passive investment shares.

Money Market Funds (MMFs)

MMFs are constrained by SEC Rule 2a-7 to invest in very short maturities, and they thus cannot directly substitute across the curve. However, MMFs show negative cross elasticity: their Treasury holdings fall when long-term yields rise. This reflects investor-level substitution. When longer yields become more attractive, investors withdraw funds from MMFs, shrinking MMF demand for T-bills. Hence, MMFs' demand is indirectly yield-sensitive, reflecting their AUM response to relative yields across maturities.

The AUM response to yield could reflect various behavioral reasons. For example, with extrapolative beliefs (Barberis et al. 2015), investors may interpret a rise in short-term interest rates as a signal of a continuing upward trend in rates, leading them to expect further increases. Anticipating future rate increases and associated price declines for long-term bonds, they reduce long-term bond holdings and shift toward T-bills or other cash-like instruments.

Foreign Official Sector

Foreign official institutions (e.g., central banks, reserve managers) have strong preferred habitats. These holders buy Treasuries for safety/liquidity or exchange rate management, not to maximize return, so we expect little substitution across maturities based on yield, as shown in Table 3. They favor shorter maturities (up to 5 or 10 years) to ensure liquidity.

Even if 30-year yields rise substantially, a central bank like the Bank of Japan is unlikely to start buying 30-year bonds, because that introduces too much volatility into their reserves. Foreign official demand is often described as "price inelastic" or "quantity-driven." Bernanke (2005) argued that the large foreign official purchases in the 2000s ("global savings glut") significantly depressed U.S. long-term yields – indicating that the official sector kept buying Treasuries despite yields falling, indicating that these foreign officials hold Treasuries for reasons beyond chasing yields.

Additionally, institutional guidelines for reserve managers often include duration limits and liquidity requirements[4] that favor shorter-maturity Treasuries and limit the substitution with longer-maturity Treasuries. Tabova and Warnock (2021) assemble confidential security-level holdings data and show that official foreign reserve investors are largely price-insensitive and duration-constrained compared to private investors. In their sample (2003–2019), the average duration of foreign official Treasury portfolios was just about 4 years.

Foreign Private Sector

Foreign private investors consist of a heterogeneous group of sectors – including foreign money market funds, banks, households, etc., but excluding foreign hedge funds and mutual funds. The insignificant elasticities we estimate likely reflect the aggregation of both inelastic investors, such as pension funds and wealthy foreign households that directly invest in Treasuries, and more yield-sensitive sectors, such as foreign banks.

Other U.S. Investors

As shown in Table 3, although money market funds (MMFs) and mutual funds exhibit strong behavioral cross-maturity substitution, the "Other U.S. Investors" sector shows notably muted responsiveness to yield differentials. This residual category—comprising U.S. households, non-profits, and non-financial corporations—likely reflects a distinct investor base with different behavioral dynamics. Unlike MMFs and mutual funds, whose demand aggregates the actions of

[4]Refer to this IMF document, "Guidelines for Foreign Exchange Reserve Management: Accompanying Document and Case Studies" for more details.

retail investors subject to return-chasing and extrapolation, direct holders of Treasuries tend to be wealthier households or corporations with more stable, longer-horizon objectives. On the one hand, Campbell (2006) shows that wealthier households are less subject to behavioral biases. On the other hand, corporations' Treasury demand often serves purposes such as liquidity reserves, tax optimization, or balance sheet management. Moreover, in direct Treasury holdings, the absence of performance benchmarking reduces salience and reallocation incentives. As a result, the "Other U.S." sector contributes less to yield curve rebalancing compared to MMFs and the mutual fund sector.

C.2. Additional Empirical Tables and Figures

Table A3. **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 A4. **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 A5. **Treasury Supply by Maturity Bucket Regressed on Macro Variables.**

This table shows the linear regression of Treasury supply on macro states and FFR, generating the predicted pseudo supply $\hat{S}_t(m)$ as in equation (10). The dependent variable is the face 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. Macro variables include Credit Spread, Debt/GDP, 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; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	$\mathbb{1}\{\tau < 1Y\}$	$\mathbb{1}\{1Y \leq \tau < 5Y\}$	$\mathbb{1}\{\tau \geq 5Y\}$
FFR	310.616***	60.015	82.741
	(49.905)	(54.381)	(80.185)
Credit Spread	5.181	−389.588	−435.872**
	(180.102)	(262.206)	(195.209)
GDP Gap	−160.022***	81.181**	29.774
	(36.798)	(40.011)	(36.760)
Core Inflation	−16.964	59.005	338.715***
	(59.437)	(38.779)	(36.593)
Debt/GDP	13,060.030***	5,055.045***	5,172.802***
	(868.860)	(529.398)	(331.032)
Constant	−5,214.768***	4,279.710***	2,099.422***
	(538.465)	(261.016)	(177.848)
R-squared	0.959	0.917	0.948
Observations	45	45	45

Table A6. **Demand System Results OLS - Granular Demand Investors**

This table shows the OLS estimates of our demand system specified in Equation (8) for granular demand investors. 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	85.494	-203.031***	-84.390
	[14.325]	[6.016]	[1.574]	[15.409]	[82.959]	[97.681]	[61.419]	[57.919]
$y\_t(-m)$	-57.315***	5.909	-0.143	-31.597**	105.763	-19.388	77.017	112.609
	[14.624]	[7.585]	[1.881]	[15.447]	[109.295]	[122.759]	[67.085]	[75.499]
Coupon Rate	-135.101***	10.707	-1.277	-17.017	512.046	188.228	-304.136***	-163.445
	[23.444]	[12.861]	[2.925]	[33.696]	[390.244]	[234.501]	[114.897]	[121.617]
Bid-Ask Spread	7.806	19.847***	3.413***	24.162*	19.679	127.947*	-80.549**	-54.510
	[7.638]	[4.834]	[1.143]	[13.148]	[96.878]	[66.090]	[38.206]	[47.171]
$\mathbb{1}\{1Y \leq \tau < 5\}$	58.001***	151.759***	14.013***	224.702***		-413.227***	2983.216***	-308.538***
	[12.694]	[5.180]	[1.626]	[20.429]		[118.826]	[68.496]	[89.159]
$\mathbb{1}\{\tau \geq 5\}$	-51.544**	197.389***	15.426***	231.527***		594.012***	456.849***	274.318**
	[23.479]	[12.992]	[2.969]	[26.707]		[206.674]	[114.795]	[114.354]
Credit Spread	11.640	-13.700	-0.022	-65.416**	-205.221	292.543	57.080	-66.427
	[18.426]	[12.313]	[2.458]	[32.228]	[151.589]	[191.322]	[84.304]	[124.695]
Debt/GDP	697.984***	-3.425	47.785***	200.882*	7587.435***	1767.479**	-1590.849***	998.612*
	[63.868]	[43.153]	[9.359]	[106.771]	[537.329]	[875.355]	[511.343]	[511.718]
GDP Gap	10.028***	-4.249**	1.406***	11.035***	-69.571***	3.795	-8.555	4.775
	[3.480]	[1.753]	[0.435]	[4.187]	[24.891]	[28.587]	[16.147]	[16.596]
Core Inflation	15.072**	0.410	-2.171***	-1.397	-90.427*	16.186	-63.522*	0.139
	[6.703]	[3.189]	[0.766]	[8.157]	[52.450]	[50.835]	[35.251]	[33.024]
R-squared	0.903	0.914	0.843	0.855	0.946	0.672	0.979	0.501
Observations	135	135	135	135	45	135	135	135

Table A7. **Demand System Results OLS - Arbitrageurs**

This table shows the OLS estimates of our demand system specified in Equation (8) for arbitrageurs. 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; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Hedge Funds ROW	**Hedge Funds US**	**Primary Dealers**
	(1)	(2)	(3)
$y_t(m)$	-68.567**	-19.616**	-14.513***
	[29.300]	[8.544]	[4.500]
$y_t(-m)$	65.147*	19.421*	23.678***
	[35.848]	[9.959]	[6.302]
Coupon Rate	-140.771*	-43.928**	-3.169
	[79.542]	[20.305]	[13.894]
Bid-Ask Spread	37.617	5.037	15.641***
	[40.489]	[12.009]	[5.221]
$\mathbb{1}\{1Y \leq \tau < 5\}$	257.064***	61.283***	49.626***
	[44.466]	[11.485]	[8.360]
$\mathbb{1}\{\tau \geq 5\}$	140.606**	41.245***	37.132***
	[56.932]	[15.295]	[10.293]
Credit Spread	118.723	11.295	26.136*
	[76.123]	[20.614]	[14.158]
Debt/GDP	-117.435	-109.161	154.216***
	[240.617]	[68.557]	[41.627]
GDP Gap	8.872	2.501	-0.645
	[12.232]	[3.356]	[2.553]
Core Inflation	-13.943	-8.507	-11.905***
	[20.322]	[5.542]	[2.913]
R-squared	0.311	0.270	0.461
Observations	135	135	135

Table A8. **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)
$y_t(m)$	78.196***	4.907	6.188**	117.642***	471.601**	125.251	-86.829	33.522
	[23.555]	[9.752]	[3.071]	[37.660]	[230.134]	[173.149]	[93.808]	[85.772]
$y_t(-m)$	-89.119***	-2.741	-3.451	-136.023***	-671.503**	44.714	-27.864	-47.399
	[25.907]	[11.829]	[3.451]	[43.857]	[326.263]	[227.847]	[125.541]	[115.816]
Coupon Rate	-165.562***	1.608	-4.470	-120.110**	20.422	229.717	-415.418**	-319.004*
	[32.472]	[17.705]	[4.357]	[52.757]	[536.890]	[287.511]	[161.687]	[165.068]
Bid-Ask Spread	6.150	18.571***	3.216***	15.613	145.563	111.140	-96.971**	-65.176
	[7.958]	[4.476]	[1.191]	[15.253]	[132.492]	[77.103]	[43.707]	[55.967]
$\mathbb{1}\{1Y \leq \tau < 5\}$	51.288***	148.402***	13.268***	196.885***		-437.072***	2940.741***	-346.647***
	[13.909]	[4.549]	[1.926]	[23.606]		[119.357]	[80.143]	[82.927]
$\mathbb{1}\{\tau \geq 5\}$	-95.281**	180.859***	10.735*	70.437		569.008	251.048	41.153
	[43.675]	[17.692]	[5.828]	[73.137]		[361.306]	[183.234]	[169.850]
Credit Spread	19.036	-11.749	0.746	-41.359	-538.702***	276.179	80.675	-29.389
	[21.477]	[13.788]	[2.433]	[37.797]	[202.674]	[186.150]	[84.199]	[130.339]
Debt/GDP	616.777***	-11.322	39.771***	-12.250	5426.418***	2277.501**	-1670.201***	630.213
	[80.524]	[48.394]	[10.137]	[125.189]	[973.190]	[916.997]	[557.553]	[542.993]
GDP Gap	11.171***	-4.457**	1.509***	12.832**	-76.233***	-11.172	-11.675	9.058
	[3.782]	[1.911]	[0.478]	[5.146]	[22.134]	[30.395]	[16.903]	[18.066]
Core Inflation	16.602**	-0.391	-2.049**	-0.961	71.632	-16.604	-74.638*	4.394
	[6.636]	[3.324]	[0.830]	[9.949]	[75.307]	[50.707]	[38.189]	[33.756]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap Statistic (*first stage*)	14.87	14.87	14.87	14.87	3.96	14.87	14.87	14.87

Table A9. **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)
$y_t(m)$	140.405***	0.076	7.749*	133.329***	1298.367***	275.336	-284.083**	104.933
	[36.767]	[9.785]	[4.011]	[46.887]	[413.445]	[201.078]	[118.357]	[82.124]
$y_t(-m)$	-154.952***	-4.741	-5.115	-123.073**	-2414.033***	-278.883	191.167	-129.484
	[41.832]	[12.054]	[4.103]	[47.927]	[527.114]	[240.109]	[131.240]	[97.465]
Coupon Rate	-356.228***	7.889	-9.626**	-136.202***	-5023.019***	-232.009	188.570	-501.501***
	[46.889]	[14.105]	[4.392]	[44.761]	[1641.224]	[243.233]	[157.571]	[130.670]
Bid-Ask Spread	23.685	15.935***	6.093***	22.626*	1134.988**	166.055**	-119.868**	-47.032
	[15.988]	[4.268]	[1.285]	[13.605]	[461.121]	[83.460]	[49.157]	[58.472]
$\mathbb{1}\{1Y \leq \tau < 5\}$	29.451	150.760***	12.721***	188.626***		-474.409***	3009.025***	-371.107***
	[23.605]	[4.972]	[2.476]	[28.977]		[121.559]	[94.939]	[87.103]
$\mathbb{1}\{\tau \geq 5\}$	-209.854***	186.526***	7.854	54.965		217.756	619.016***	-93.327
	[71.513]	[18.501]	[7.674]	[88.048]		[403.208]	[217.163]	[155.416]
Constant	327.600***	45.907***	8.934**	126.097***	4035.950***	1564.885***	691.906***	887.239***
	[45.890]	[12.739]	[3.681]	[42.635]	[600.965]	[235.982]	[129.916]	[117.939]
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 A10. **Demand System Results - IV extending substitutable assets**

This table shows the IV estimates of our demand system specified in equation (8), extending the set of substitutable assets by controlling for the MBS spread and swap spread. 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, Core Inflation, MBS Spread (10-year MBS rate minus 10-year US Treasury yield), and Swap Spread (10-year swap spread minus 10-year US Treasury yield). 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 U.S.**	**MMF**	**Other U.S.**	**Foreign O**	**Foreign P**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$y_t(m)$	60.441**	6.017	6.615**	131.797***	322.522**	132.659	-24.781	49.908
	[25.689]	[11.626]	[3.257]	[43.196]	[149.441]	[191.194]	[116.662]	[92.637]
$y_t(-m)$	-64.467**	0.791	-2.597	-138.630***	-438.730**	17.336	-105.172	-45.500
	[26.890]	[13.440]	[3.579]	[49.464]	[218.732]	[247.411]	[150.515]	[120.386]
Coupon Rate	-143.059***	1.461	-3.700	-124.271**	419.432	186.656	-502.682***	-322.860*
	[35.057]	[19.009]	[4.223]	[54.778]	[498.595]	[311.476]	[194.010]	[186.947]
Bid-Ask Spread	3.081	13.696***	1.308	-3.491	185.832	145.183*	-82.502*	-94.345*
	[7.972]	[4.899]	[1.069]	[15.345]	[132.859]	[83.510]	[47.099]	[56.327]
$\mathbb{1}\{1Y \leq \tau < 5\}$	56.974***	147.574***	12.996***	190.764***		-437.853***	2920.045***	-354.190***
	[14.944]	[4.974]	[1.951]	[24.000]		[120.144]	[93.700]	[82.592]
$\mathbb{1}\{\tau \geq 5\}$	-59.917	180.953***	10.542*	49.985		546.491	131.480	19.712
	[45.842]	[21.101]	[6.291]	[82.960]		[394.689]	[225.541]	[177.878]
Credit Spread	-1.842	-28.762**	-3.985**	-83.923**	-383.149***	344.292**	140.049*	-102.867
	[19.894]	[13.069]	[1.906]	[38.109]	[138.220]	[165.209]	[80.945]	[128.173]
Debt/GDP	732.156***	71.675	63.631***	198.642	5397.403***	1907.343*	-1996.437***	986.283
	[94.187]	[60.603]	[11.812]	[140.699]	[963.313]	[1096.832]	[649.803]	[659.498]
GDP Gap	11.502***	-3.611**	1.452***	12.410***	-91.584***	-4.966	-8.707	9.757
	[3.799]	[1.814]	[0.402]	[4.737]	[20.914]	[31.188]	[17.972]	[19.705]
Core Inflation	14.941**	-2.988	-2.578***	-6.597	56.330	-13.153	-75.153*	-5.046
	[7.468]	[3.557]	[0.799]	[10.417]	[69.296]	[57.444]	[43.896]	[38.841]
MBS Spread (10y)	52.896	63.852**	12.083**	118.188	-509.337	-61.435	-55.457	222.447
	[47.602]	[29.483]	[5.474]	[90.069]	[368.293]	[439.300]	[203.808]	[269.882]
Swap Spread (10y)	-29.301	-12.360	-17.804***	-161.422***	410.632	498.906	302.304	-217.934
	[39.802]	[27.900]	[3.961]	[52.601]	[328.850]	[355.315]	[206.015]	[286.442]
Observations	135	135	135	135	45	135	135	135
Kleibergen-Paap Statistic (*first stage*)	10.49	10.49	10.49	10.49	13.95	10.49	10.49	10.49

Table A11. **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)
$y_t(m)$	18.805	3.159	4.412***	41.993**	80.663**	146.508**	-85.178**	10.794
	[12.095]	[4.748]	[1.460]	[16.880]	[39.923]	[65.004]	[39.935]	[36.619]
Coupon Rate	-80.425***	4.207	-1.299	6.244	414.985	189.046	-393.641***	-275.844**
	[20.834]	[13.479]	[2.072]	[33.584]	[407.796]	[194.242]	[92.037]	[109.692]
Bid-Ask Spread	10.569	18.688***	3.260***	18.723	46.508	110.992	-100.489**	-64.973
	[9.237]	[4.593]	[1.164]	[13.893]	[95.683]	[73.939]	[42.969]	[53.514]
$\mathbb{1}\{1Y \leq \tau < 5\}$	69.689***	148.932***	13.745***	218.227***		-442.465***	2937.404***	-340.843***
	[15.713]	[4.446]	[1.763]	[22.707]		[118.149]	[75.120]	[83.973]
$\mathbb{1}\{\tau \geq 5\}$	26.035	184.483***	14.719***	235.182***		519.766***	261.455***	93.617
	[18.821]	[8.318]	[2.586]	[27.878]		[140.154]	[77.136]	[66.840]
Credit Spread	-1.703	-12.388	-0.067	-73.276**	-243.628*	286.734	73.836	-40.575
	[20.002]	[13.304]	[2.537]	[35.540]	[133.309]	[192.879]	[86.202]	[126.559]
Debt/GDP	848.132***	-3.878	50.917***	403.366***	7277.569***	2125.836**	-1513.628***	790.172
	[85.111]	[38.513]	[10.283]	[135.568]	[390.054]	[874.337]	[457.565]	[504.371]
GDP Gap	7.829**	-4.572**	1.299***	5.446	-69.348***	-8.194	-15.800	5.931
	[3.659]	[1.964]	[0.413]	[4.188]	[22.415]	[31.448]	[13.126]	[16.014]
Core Inflation	11.988	-0.561	-2.415***	-13.365	-68.659*	-11.237	-83.305*	-1.226
	[8.945]	[3.458]	[0.847]	[9.185]	[38.799]	[53.004]	[43.824]	[35.935]
Constant	-485.798***	45.933	-27.617***	-194.561**	-4070.085***	-651.540	2105.054***	203.502
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 (domestic and foreign, left y-axis) and primary dealers (right y-axis) over time.

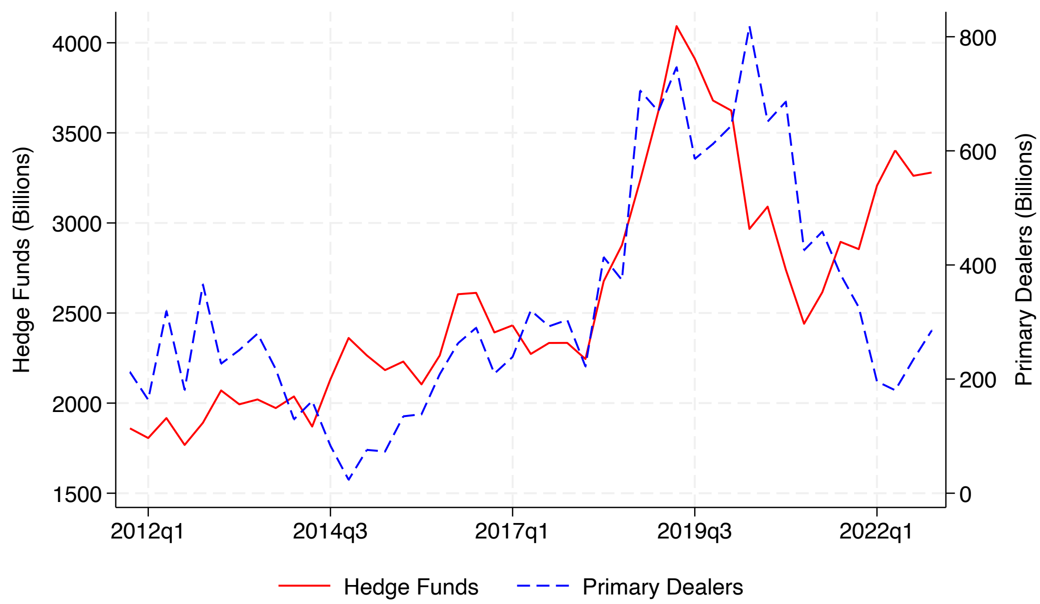


Figure A2. **Morningstar Aggregate Holdings by Domestic and Foreign Bond Funds**. This graph shows the aggregate holdings of US and foreign bond funds in USD (trillions) over time.

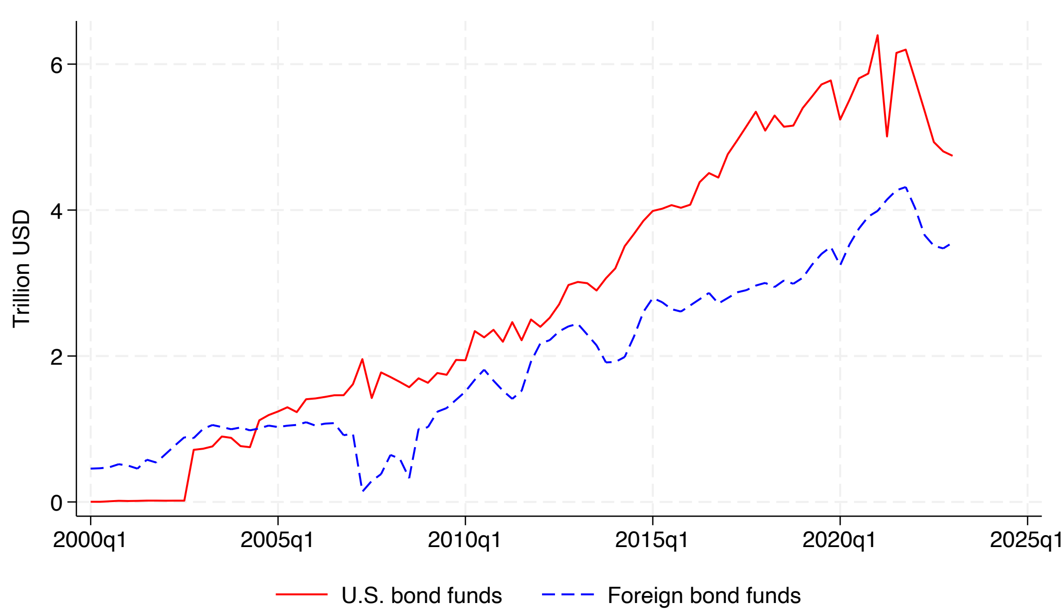
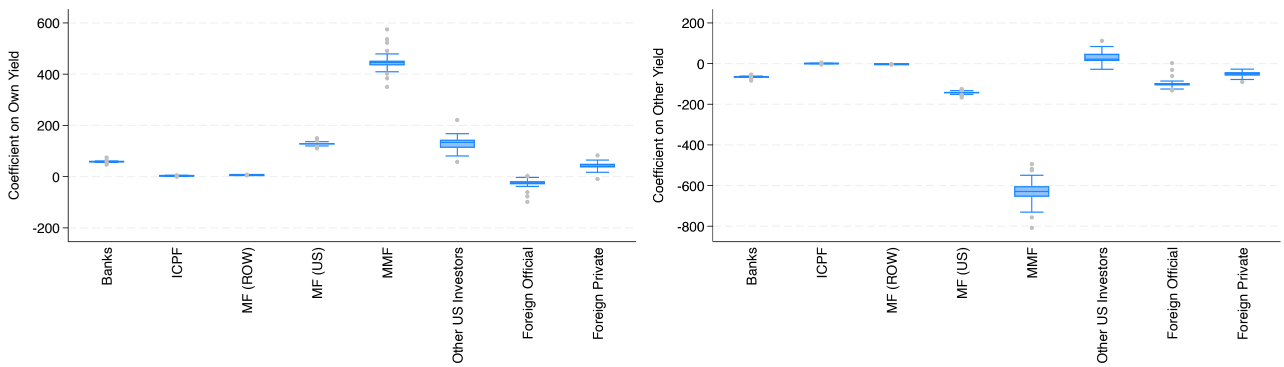


Figure A3. **Leave-One-Quarter-Out Coefficient Distributions by Sector** Panel (a) plots the distribution of leave-one-quarter-out coefficient estimates on *own yield* by investor type. Panel (b) shows the corresponding distribution for *other yield* coefficients. The analysis covers eight investor sectors: U.S. banks (Banks), 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), other U.S. investors (Other U.S. Investors), foreign official institutions (Foreign O), and foreign private investors (Foreign P). Each distribution reflects 45 estimates from rolling leave-one-quarter-out regressions over the sample period 2011Q4-2022Q4. The box plot summarizes these distribution by showing the median, interquartile range (IQR), and identifying outliers beyond 1.5 times the IQR. Among all sectors, estimations of MMFs have more variations driven by a relatively weaker first stage.



(a) Own Yield (b) Other Yield

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

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

which becomes exact when we take a continuous-time approach. Refer to Greenwood et al. (2023b) 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

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

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

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

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

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

In that case, we obtain expected next-period wealth

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

and variance of next-period wealth

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

Consequently, we can write the FOC of arbitrageurs in (35) 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

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

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),

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

With the assumption in (36), and the affine expression of p_t in (26), we rewrite (A20) as

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

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

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

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

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

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

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 (30)

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

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

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

$$A_r(1)\rho_r - A_r(2) - 1 = A_r(1)\gamma\left(\sigma_r^2 A_r(1)\alpha(2)' A_r\right)$$

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

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

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

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

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

$$A_u(2)' = \gamma\sigma_r^2\left(-\frac{a}{2}A_u(2)' - (0,-1)\right)$$
$$A_u(2)' = \frac{1}{1+\gamma\sigma_r^2\frac{a}{2}}(0,\gamma\sigma_r^2).$$

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 (30).

Proof of Proposition 1

According to equation (30), $p_t^{(1)}$ 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 (30), 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,

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

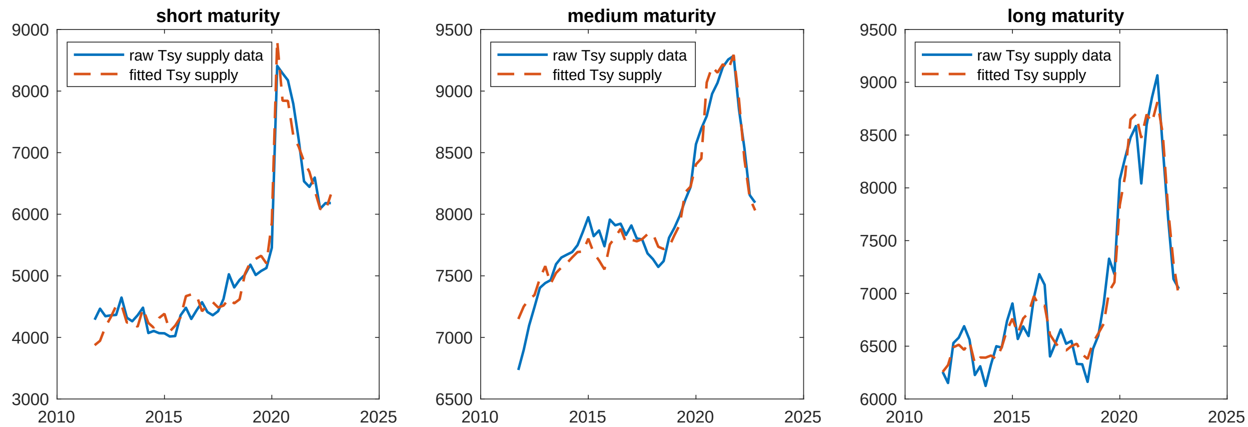


Figure A4. Treasury Supply: Data versus Model Fitting.

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 A4, 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:

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

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 (38) 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,

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

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 $p_t = A\beta_t + A_r r_t + A_u u_t + C$ according to equation (26). The latent demand term u_t is unobservable with mean 0 and uncorrelated with β_t and r_t. We note that the problem in (A31) 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:

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

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

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

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,

$$\begin{aligned}\underbrace{A(\tau)'-A(\tau-1)'\Phi}_{\text{left hand side}}&=\underbrace{A_r(\tau-1)}_{\text{dep var}}(\phi_r'\Phi-\lambda_{\beta,r})-\underbrace{\hat{A}(\tau-1)'}_{\text{dep var}}\lambda_{\beta,\beta},\\ \underbrace{-A_r(\tau)+A_r(\tau-1)\rho_r-1}_{\text{left hand side}}&=\underbrace{\hat{A}(\tau-1)'}_{\text{dep var}}\lambda_{r,\beta}+\underbrace{A_r(\tau-1)}_{\text{dep var}}\lambda_{r,r}.\end{aligned}\tag{A34}$$

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

$$\begin{aligned}\hat{A}^{\text{shift}}&=\left(0,\hat{A}(1),...,\hat{A}(N-1)\right)'\\ A_r^{\text{shift}}&=\left(0,A_r(1),...,A_r(N-1)\right)'\end{aligned}\tag{A35}$$

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

$$\begin{aligned}&\left(\hat{A}^{\text{shift}}\gamma\Sigma\sum_{\hat{\tau}=2}^{N}\hat{A}(\hat{\tau}-1)\alpha(\hat{\tau})'+A_r^{\text{shift}}\gamma\sigma_r^2\sum_{\hat{\tau}=2}^{N}A_r(\hat{\tau}-1)\alpha(\hat{\tau})'+I\right)A_u\\ =&\hat{A}^{\text{shift}}\gamma\Sigma(\hat{A}^{\text{shift}})'+A_r^{\text{shift}}\gamma\sigma_r^2(A_r^{\text{shift}})'\end{aligned}\tag{A36}$$

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,

$$\min_{\{\gamma,A_u,\lambda_{\beta,\beta},\lambda_{\beta,r},\lambda_{r,\beta},\lambda_{r,r},\psi,\psi_r\}} \mathbb{E}\left[\sum_t \sum_\tau \left(\frac{X_t(\tau)-X_t^o(\tau)}{\bar{X}^{abs}(\tau)}\right)^2 + \sum_t \sum_\tau (p_t(\tau)-p_t^o(\tau))^2\right], \tag{A37}$$

subject to iteration equations in (A32) and (A33) that give rise to $\{A, A_r, \hat{A}\}$, equation (A36) that solves for A_u, equation (A25) that solves for C, log price vector p_t as in (26), and arbitrageur Treasury holding $X_t(\tau)$ given by equation (37). Taking expectation implies that the price impact of latent demand u_t will drop out, resulting in the following equivalent optimization

$$\min_{\{\gamma,A_u,\lambda_{\beta,\beta},\lambda_{\beta,r},\lambda_{r,\beta},\lambda_{r,r},\psi,\psi_r\}} \sum_t \sum_\tau \left(\frac{X_t(\tau)-X_t^o(\tau)}{\bar{X}^{abs}(\tau)}\right)^2 + \sum_t \sum_\tau (A\beta_t + A_r r_t + C - p_t^o(\tau))^2,$$

which will be the objective function in our implementation. As discussed in Section 4.3, we restrict to integer-year maturities for log prices to reduce redundancies, so we use 1, 2, ..., 30 year maturities for yield fitting, and we focus on medium- and long-maturity buckets for quantities, which are relatively more important for disciplining the risk premium.

After we estimate problem (A37), 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

$$
\begin{aligned}
&A(\tau-1)'(I-\Phi)\bar{\beta}+A_r(\tau-1)(\bar{r}-\phi_r'\Phi\bar{\beta})+\frac{1}{2}\hat{A}(\tau-1)'\Sigma\hat{A}(\tau-1)\\
&+\frac{1}{2}(A_r(\tau-1)\sigma_r)^2+\frac{1}{2}A_u(\tau-1)'\Sigma^u A_u(\tau-1)-\hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\left(\bar{S}(\hat{\tau})-\theta_0(\hat{\tau})\right)+\psi\right)\\
&-A_r(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2A_r(\hat{\tau}-1)\left(\bar{S}(\hat{\tau})-\theta_0(\hat{\tau})\right)+\psi_r\right)\\
=&\hat{A}(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\Sigma\hat{A}(\hat{\tau}-1)\alpha(\hat{\tau})'\right)C+A_r(\tau-1)'\gamma\left(\sum_{\hat{\tau}=2}^{N}\sigma_r^2A_r(\hat{\tau}-1)\alpha(\hat{\tau})'\right)C+C(\tau)-C(\tau-1)
\end{aligned}
\tag{A38}
$$

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

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

where $\mathbf{1}_\tau$ is an N-dimensional vector that is one for element τ but zero otherwise. Then equation (A38) 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.

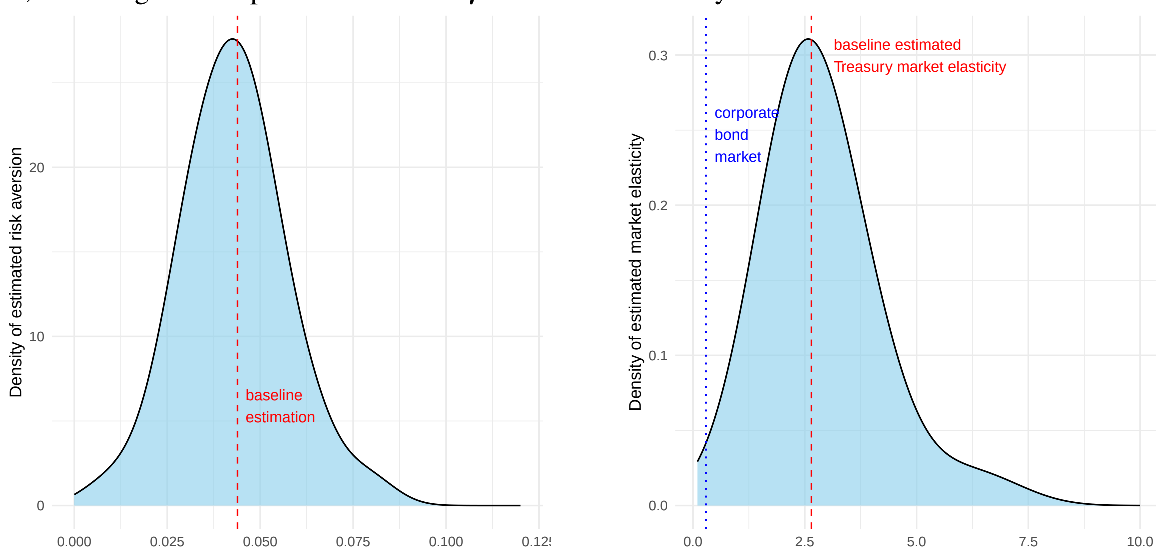
D.5. Bootstrapping Standard Errors of Estimated Model Parameters

We quantify estimation uncertainty, particularly for the arbitrageurs' risk aversion parameter γ, using a bootstrap approach. Since our estimation involves multiple steps, the bootstrap approach is the most suitable method for obtaining standard errors in the estimation.

Specifically, we generate bootstrap samples by randomly drawing (with replacement) from latent demand and generate bootstrapped demand. Using the bootstrapped demand, we reconstruct our pseudo-yield instrument and redo the IV estimation of demand functions. With these updated demand-function parameters, we re-estimate structural parameters, including arbitrageur risk aversion, price of risk parameters, and outside-asset exposure. We repeat this process 100 times, obtaining bootstrap distributions for each parameter. From these distributions, we calculate confidence intervals that transparently reflect the uncertainty and potential skewness inherent in our estimates.

Figure A5. Distribution of Estimations via Bootstrap.

This figure illustrates the distribution of bootstrap estimation of γ and market elasticity, and their average values (dashed red lines). We generate bootstrap samples by randomly drawing (with replacement) from latent demand u_t and generate the simulated Treasury yields according to the same linear mapping from latent demand to Treasury yields. For each random draw, we re-estimate structural parameters, including arbitrageur risk aversion, price of risk parameters, and outside-asset exposure. We repeat this process 100 times, obtaining bootstrap distributions for γ and market elasticity.



(a) Distribution of Estimated Risk Aversion γ

(b) Distribution of Estimated Market Elasticity

As shown in Figure A5(a), our bootstrap analysis of the arbitrageurs' risk-aversion parameter γ yields a distribution around the baseline estimation. The 90% confidence interval (cutoffs at 5% and 95% quantiles) for γ is (0.022,0.066) and one standard deviation of the γ distribution is 0.014. As a result, our model estimation confidently reveals that the risk aversion parameter is significantly larger than 0 and tight around the estimated value of 0.04.

In Figure A5(b), we illustrate the distribution of estimated market elasticity. Alongside the distribution of Treasury market elasticity, we also plot the corporate bond market elasticity, which is the inverse of the price multiplier and thus 1/3.5 according to Chaudhary et al. (2022). We find that the 90% confidence interval (cutoffs at 5% and 95% quantiles) for market elasticity is (1.57,6.15) and one standard deviation of the market elasticity distribution is 2.25. The probability of Treasury market elasticity larger than 1/3.5 is almost 1. Consequently, we conclude that with extremely high probability, the Treasury market is more elastic than the corporate-bond market.

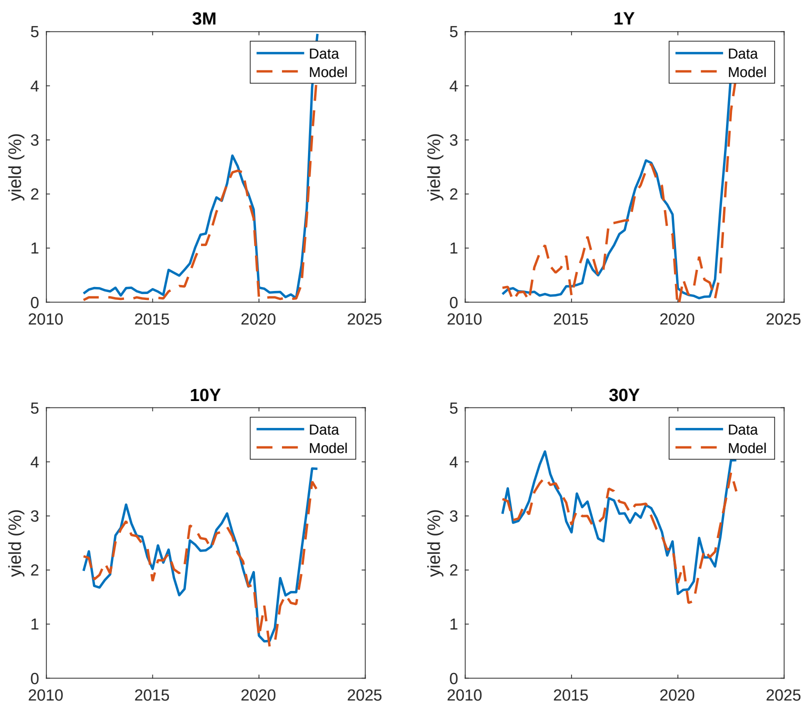
E. Additional Quantitative Results

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

In Figure A6, 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 information from fundamental economic variables and demand shocks, which is more challenging than a typical affine term structure model that includes multiple 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.

Figure A6. **Model Fit on the Dynamics of Treasury Yields.**

Model-implied yields are constructed using equation (26).

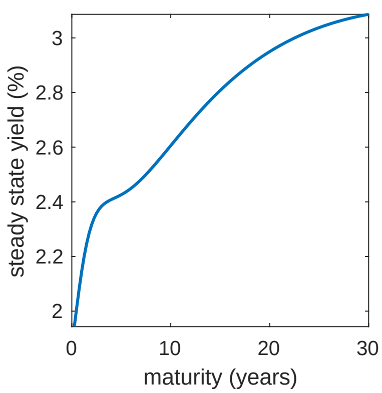


In Figure A7 panel (a), we illustrate the steady-state yield curve. This steady-state yield curve is upward-sloping and mainly reflects the average shape of the yield curve in our model estimation period.

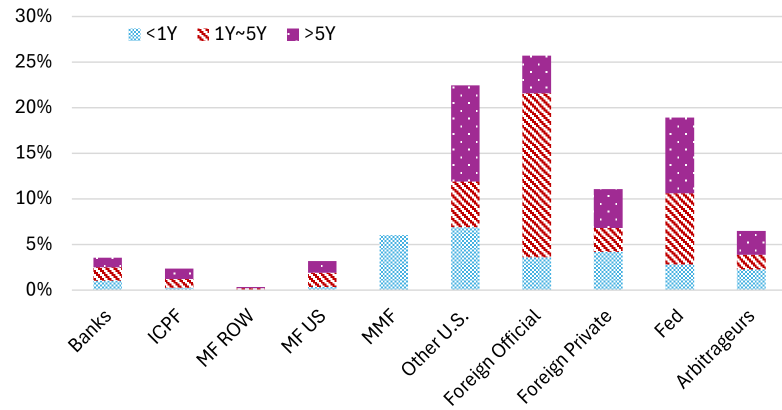
In Figure A7 panel (b), we illustrate the steady-state portfolio allocations across different sectors. The model implies that in the long run, foreign investors are still the largest holder among

Figure A7. **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

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.

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 change in total Treasury valuation as a fraction of total Treasury valuation is

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

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

$$\mathscr{M} = \frac{\sum_{\tau'} S(\tau')A_u(\tau',\cdot)u}{S^{total}}/1\% = \frac{1}{S^{total}}S'A_uS = \omega' A_uS. \tag{A39}$$

This market multiplier is closely related to the bucket-level multiplier in Table 7, where the value in row τ and column τ' represents the percentage change of price at maturity τ' in response to a 1% extra latent demand of maturity τ,

$$\mathscr{M}(\tau',\tau) = \frac{A_u(\tau',\tau) * 1\% * S(\tau)}{1\%} = A_u(\tau',\tau)S(\tau) \tag{A40}$$

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

$$\mathscr{M} = \sum_{\tau'}\sum_{\tau} \omega(\tau')\mathscr{M}(\tau',\tau) \tag{A41}$$

Using equation (A41) and the values in Table 7 Panel (a), we obtain a total market multiplier of 0.37, i.e., a 1% representative demand shock on the whole Treasury market increases total Treasury valuation by 0.37%. This can also be equivalently stated as a \$1 billion demand shock increasing Treasury valuation by \$0.37 billion. Following the same procedure, we can use Panel (b) of Table 7 to calculate the Treasury market multiplier in the case without arbitrageurs, which leads to a value of 8.74. Moreover, we can also calculate the multiplier for a subset $\mathscr{T}$ of maturities,

$$\mathscr{M}(\mathscr{T}) = \sum_{\tau'\in\mathscr{T}}\sum_{\tau\in\mathscr{T}} \frac{S(\tau')}{\sum_{l\in\mathscr{T}} S(l)}\mathscr{M}(\tau',\tau). \tag{A42}$$

For example, lumping all maturities above one year, we get a market multiplier of 0.48.

Next, we show the aggregate multiplier for permanent demand shocks. In Table A12, 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 7. Using Panel (a) of A12, we can calculate the Treasury market multiplier for permanent demand shock as 0.93, which is higher than the case of latent demand shock, because permanent demand shocks significantly change the risk premium. Moreover, lumping all maturities above one year, we get a market multiplier of 1.25.

Panel (b) of Table A12 is identical to Panel (b) of Table 7, 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.

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

Table A12. **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	0.002	0.004
shock on medium maturity	0.016	0.116	0.211
shock on long maturity	0.072	0.572	1.690
Panel (b): Without Arbitrageur			
shock on short maturity	0.128	0.585	3.011
shock on medium maturity	0.650	2.104	12.688
shock on long maturity	0.298	1.133	5.038
Panel (c): Price Impact Ratio (Panel (b)/Panel (a))			
shock on short maturity	367.051	239.447	745.121
shock on medium maturity	39.799	18.092	60.179
shock on long maturity	4.148	1.980	2.981

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,

$$\left(\frac{\sum_{\tau'} S(\tau') A_u(\tau', \tau) * (1\% * S^{total})}{S^{total}} \right) / 1\%, \tag{A43}$$

Then we define the market elasticity at maturity τ as $\mathscr{E}(\tau)$ as the inverse of the market multiplier at maturity τ in (A43),

$$\mathscr{E}(\tau) = \frac{1}{\sum_{\tau'} S(\tau') A_u(\tau', \tau)} \tag{A44}$$