Anatomy of the Treasury Market: Who Moves Yields?

Manav Chaudhary[*] Zhiyu Fu[†] Haonan Zhou[‡]

November 18, 2024

[Click here for the latest version]

Abstract

We develop a quantity-based framework to study the drivers of U.S. Treasury yields. Our method allows for flexible identification of price and factor sensitivities for heterogeneous investors in a demand-supply model using granular idiosyncratic shocks. Overall, a 1 percent demand increase for U.S. Treasury notes and bonds results in a 1 percent increase in prices, equivalent to a 10 basis points decline for the ten-year yield. We uncover substantial heterogeneity across investors and regimes in sectors' sensitivity to Treasury yield changes and aggregate factors. Using the estimated model, we decompose changes in Treasury yields over the past two decades, and document three main findings: (i) Contrary to the conventional wisdom, foreign investors contribute little to Treasury price appreciation during flight-to-safety episodes; (ii) U.S. banks and foreign investors have become more price insensitive following the global financial crisis, while the Federal Reserve has stepped up its role as a state-contingent liquidity provider; (iii) while major foreign Treasury holders are the biggest contributor to Treasury yield compression before the financial crisis, the influence of foreign demand on yields has substantially weakened since 2010.

[*]Department of Economics and Booth School of Business, University of Chicago

[†]Washington University in St. Louis, Olin School of Business

[‡]HKU Business School

We thank Ralph Koijen, Stijn van Nieuwerburgh, and Motohiro Yogo for helpful comments and seminar participants at CUHK Shenzhen, HKU and HKUST for useful interactions. All errors are ours.

1 Introduction

The U.S. Treasury market is the cornerstone of the global financial system. The overall system's stability hinges on the market's capacity to accommodate investor demand smoothly, and its yields pin down the cost of capital of firms and households. Unsurprisingly, many macro-financial phenomena of interest—such as the yield-lowering effects of foreign savings, the consequences of quantitative easing and tightening, and dynamics during flight-to-safety episodes—center around the joint dynamics of price and *quantities* in the Treasury market. The key challenge in studying such phenomena is the inherent simultaneity between investor holdings and Treasury yields. For instance, during a flight-to-safety event, yields are influenced not only by how sensitive each investor is to the underlying shock, but also by how other investors respond and absorb the resulting price pressures.

A model of Treasury demand and supply offers the lens for overcoming such difficulty. It allows us to determine who responds directly to changes in macro factors, who accommodates those demand shifts, and ultimately, to quantify the contributions of each investor to the overall change in yields. However, estimating such a system is easier said than done. Firstly, understanding macro-financial phenomena requires estimating the demand and supply for Treasuries as an asset class relative to other assets. Estimating such aggregate elasticities requires finding aggregate-level instruments for Treasury yields. Secondly, different institutions operate under different incentives and constraints, leading to varying sensitivities to Treasury yields and economic factors. Hence, a flexible methodology is required to capture these time-varying investor-leve heterogeneities.

To address these difficulties, we develop a quantity-based framework and use it to study the drivers of U.S. treasury yields over the past two decades. Through the lens of an equilibrium asset demand and supply system, we quantify the heterogeneous sensitivities of different investors to yields as well as macro factors, and trace the movement in Treasury yields in each period back to different investors. This approach offers new insights into the Treasury market. We provide the first direct estimate of the aggregate price elasticity of the Treasury market: a 1% increase in the price of the market portfolio of Treasury bonds lead to a 1% decrease in aggregate demand from all sectors. We find that there is considerable heteroegeneity in investors' price elasticities, and there have been significant changes in these elasticities overtime. Specifically, we document three new facts: (i) contrary to the conventional wisdom, foreign investors contribute little to Treasury price appreciation during flight-to-safety episodes; (ii) U.S. banks and foreign investors becomes more price insensitive after the global financial crisis, while the Federal Reserve has stepped up its role as a state-contingent liquidity provider; (iii) while major foreign Treasury holders are the biggest contributor to Treasury yield compression before the financial crisis, the influence of foreign demand on Treasury yields has substantially weakened since 2010.

We estimate the model using data on quarterly sector-level transactions and holdings of Treasuries from the Financial Accounts, combined with more granular Treasury International Capital (TIC) data for foreign Treasury holders and the Call Reports for banks. Using granular idiosyncratic shocks to other sectors as the instrument for prices in the spirit of Gabaix and Koijen (2024), we identify sector-level price and factor sensitivities. We estimate the macro multiplier for the Treasury market to be 1.0.In terms of yield this means a 1% increase in Treasury demand would result in a 10 basis point decline in the yield of a 10-year Treasury note.[1] Furthermore, our estimated system reveals substantial heterogeneity in the price elasticities between different sectors. The household sector, which includes hedge funds and personal family office investors, is relatively sensitive to changes in yields, whereas ETFs and pension funds are in general price inelastic.

Our Treasury demand and supply system relates investor's flows to Treasury price changes, macro factors, and investor-specific idiosyncratic shocks. The key identifying variation for price elasticities is from our assumption that idiosyncratic shocks to different investors are mutually independent. As a result, other investors' demand and supply shocks constitute relevant and exogenous residual supply or demand curve shifts from the standpoint of the investor of interest. The identifying assumption yields a set of cross-equation restrictions that can be handled using the generalized method of moments (GMM). Exploiting the knowledge of the model structure, we propose a systematic approach to weighting moment conditions based on investors contribution to prices and informativeness of their flows. The resulting estimator has many desirable properties. Firstly, it allows for

[1] Our estimate is smaller than the equity market macro multiplier of 5 (Gabaix and Koijen 2022), potentially reflecting greater availability of substitutability between fixed-income asset classes (Chaudhary, Fu, and Li 2023).

flexible specifications to handle heterogeneous and time-varying price elasticities. Secondly, it is easy to implement, and offers transparent and intuitive economic interpretation. It can be interpreted as a system of jointly estimated instrument-variable (IV) regressions, or, alternatively, OLS regressions corrected for simultaneity bias. Finally, despite its simplicity, our estimator based on this weighting scheme achieves asymptotic efficiency.

On top of price and factor sensitivity estimation, our framework yields a simple pricing equation that enables us to inspect various drivers of Treasury yields through decomposition. Using our estimates, we find that 58% of the observed yield changes for the past twenty years can be attributed to idiosyncratic demand shocks, while the remaining 42% are accounted for by observed macro factors and unobserved Treasury demand common across investors. Going deeper into sector-specific drivers, we document three key findings. Firstly, contrary to conventional wisdom, foreign investors are not the main drivers of Treasury yield compression during flight-to-safety episodes. Rather, domestic investors contribute the most to Treasury appreciation when global risk sentiment worsens. Secondly, while U.S. banks have become much less price elastic since 2009, the Federal Reserve has become an important liquidity supplier after the financial crisis. Finally, foreign investors' influence on Treasury yields has waned in the 2010s, marked by a significant decline in their market share. Consequently, a foreign Treasury demand shock of the same dollar amount in recent years would generate a much lower price impact compare of that before 2010.

Standard explanations for why Treasury bonds appreciate during heightened periods of risk typically assume that foreigners strongly demand the safety provided by U.S. Treasuries. Contrary to standard explanations for the "flight-to-safety" phenomenon, we show that after accounting for the endogenous price response, there is no significant evidence that foreign investors increase demand for U.S. Treasuries when the VIX rises. The main sources of countercyclical demand for Treasuries is due to the household sector, which includes hedge funds and family offices, and the Federal Reserve after 2009. This finding echoes recent literature documenting the pressure on the Treasury market from a broad-based cash crunch during the COVID-19 crisis (He, Nagel, and Song 2022; Barone et al. 2022). In addition, foreign investors' price elasticity of demand is 55% smaller after 2009 than before, indicating that these investors themselves are less likely to serve as an important price stabilizer during global downturns when other investors liquidate longer-term Treasury securities for cash.

Our results also shed light on how reduced risk-bearing capacity of U.S. banks affect the sensitivity of the Treasury market to demand and supply shocks. Following the financial crisis and the implementation of various liquidity- and leverage-based requirements, we find that U.S. banks' demand for Treasury bonds and notes becomes less price elastic by 61%, indicating a reduced capacity to accommodate demand and supply imbalances. From a macro perspective, our finding on U.S. banks' demand shift complements the empirical literature exploring the impact of shifting financial regulatory landscape on bank activities (Du, Tepper, and Verdelhan 2018; Favara, Infante, and Rezende 2024, among others). Meanwhile, we show that the Federal Reserve has increasingly acted as a state-contingent liquidity provider. Its responsiveness to VIX surges explains 55% of the overall Treasury price sensitivity to global risk sentiment fluctuations after 2008. Our finding adds support to recent theories that highlight the insurance channel of central bank asset purchases (Haddad, Moreira, and Muir 2024).

Also, we find that while foreign investors were significant drivers of Treasury yields before the financial crisis, they now play a much smaller role. We find that demand from foreign holders of the U.S. Treasury, such as China, Japan, and European countries, led to an average of 74 basis point reduction in Treasury yields per year from 2003 to 2007. In contrast, these countries have made a small negative contribution to yields since 2010. During the initial phase of the COVID-19 pandemic, foreign investors demanded fewer U.S. Treasuries, flipping the direction of influence and pushing up yields by 34 basis points. Hence, our findings suggest that if the largest foreign holders of Treasuries were to sell all their holdings, the price impact would likely be small.

Overall, our paper suggests that a framework that flexibly incorporating sectoral heterogeneity in price and factor elasticities can go a long way in explaining Treasury market dynamics. The flexibility and efficiency of our estimation procedure also appeals to understanding the demand and supply forces in a wider range of markets, for which the estimated macro market multiplier of the U.S. Treasury market – one of the largest and most liquid – could serve as a useful benchmark.

1.1 Related Literature

This paper contributes to multiple strands of literature. First and foremost, we contribute to the literature studying the heterogeneity of market players and how their shocks affect Treasury market dynamics. There is extensive literature for each type of investors, hence it is not possible for us to do full justice here. The literature studies different constraints and incentives faced by the Federal Reserve (Gagnon et al. 2011a; Haddad, Moreira, and Muir 2024; Krishnamurthy and Vissing-Jorgensen 2011), foreign investors (Jiang, Krishnamurthy, and Lustig 2021; Kekre and Lenel 2024; Tabova and Warnock 2022; He, Nagel, and Song 2022), banks (Jiang, Matvos, et al. 2024), mutual funds (Selgrad 2023; Huang et al. 2020), insurance companies (Chaudhary 2024; Koijen and Yogo 2023), and broker-dealers (Du, Hébert, and Li 2022; Vayanos and Vila 2021; Greenwood and Vayanos 2014), and their role in the Treasury market. Our primary contribution is to study the interactions of all Treasury market participants and their impact on equilibrium yields. In particular, we contribute to the literature on foreign special demand for Treasuries by providing novel evidence that shows foreign investors' contribution to flight-to-safety price pressure is relatively small and has further weakened in the recent decade. We also speak to the literature on quantitative easing by quantifying the impact of Federal Reserve's state-contingent liquidity provision. Also our finding that banks are less willing to provide liquidity to the Treasury market since the financial crisis speaks to the broader literature documenting the changing nature of bank intermediation and its broader market implications since 2008.

More closely related to our approach is the burgeoning literature on using demand-based frameworks to study asset prices.[2] We provide the first direct macro elasticity estimate for the U.S. Treasury market covering all market participants.[3] The idea that the demand shocks move Treasury prices is not new, and has long been acknowledged in the literature (Greenwood and Vayanos 2010; Krishnamurthy

[2] The recent development in the demand-based framework, pioneered by Koijen and Yogo (2019), has been widely adopted and applied to many asset classes, including equity (Gabaix and Koijen 2022), corporate bonds (Chaudhary, Fu, and Li 2023), foreign exchange rates (Camanho, Hau, and Rey 2022; Jiang, Richmond, and Zhang 2022; An and Huber 2024) and generated numerous new insights that is unavailable to frameworks that only focuses on prices.

[3] The literature on estimating macro multipliers of the equity market has been growing rapidly. See, for example, Da et al. (2018), van der Beck (2022), Hartzmark and Solomon (2022), Haddad, Huebner, and Loualiche (2024), and Li, Pearson, and Zhang (2024), most of which point to a smaller aggregate elasticity compared to our estimate on the U.S. Treasury market.

and Vissing-Jorgensen 2012; Hanson and Stein 2015). The success of the quantitative easing policies further adds support to this view. Nevertheless, it has been challenging to quantify the elasticity and the multiplier, as identification has been proven difficult in a highly liquid market such as U.S. Treasury. With recent advances in the model framework and identification techniques, an expanding set of papers have applied the demand system approach to understanding the demand for government debt (Fang, Hardy, and Lewis 2022; Cavaleri 2023; Eren, Schrimpf, and Xia 2023; Jansen, Li, and Schmid 2024). Another set of papers jointly study the demand for government debt along with other asset classes (Jiang, Richmond, and Zhang 2024; Koijen and Yogo 2024). Our paper builds on these works and further develops a more flexible asset demand system. By leveraging the new model framework and identification method, we are able to directly estimate investor-specific and time-varying price elasticities and macro factor loadings for every investor.[4] Our estimated model can fully recover the time path of the Treasury yields, allowing us to decompose and quantify changes in Treasury yields in each quarter into the sector- and factor-specific drivers.

More generally, our approach extends the granular instrumental variable literature (Gabaix and Koijen 2024) to a setup with full cross-sectional heterogeneity in price and factor elasticities. The GIV method has been broadly applied in the macro-finance literature to study various topics (see, for example, Kundu and Vats (2021), Chodorow-Reich et al. (2024), Adrian et al. (2022), and Camanho, Hau, and Rey (2022)). Several extensions to the original methods have also been proposed in the literature to allow for more general specifications: Qian (2024) proposes the heterogeneity-robust GIV (RGIV), which relies on the continuously updating GMM estimator to search for the optimal weighting. Baumeister and Hamilton (2023) propose a maximum likelihood (MLE) approach to the GIV. Compared to these more statistically driven approaches, our method leverages the economics of the model to propose a weighting scheme that achieves optimal efficiency of the estimator. Chodorow-Reich et al. (2024) studies empirical identification of shock propagation through a network. Our method builds on the insights of Chodorow-Reich et al. (2024), but further develop the optimal estimator in the case where the

[4] The idea that granular demand shocks from other sectors form residual supply curve shifts that help identify price elasticity of demand in the bond market is also explored in Lou (2012), Koijen, Koulischer, et al. (2017), Jansen (2021), and Zhou (2023), among others. We provide a systematic approach to extract this type of shocks.

network effect operates through heterogeneous responses to endogenous factors.

1.2 Roadmap

The paper is organized in five sections. Section 2 introduces the linear asset demand and supply system and our identification approach. Section 3 reports our estimate of the Treasury macro market multiplier and the associated sector-specific price elasticities. Section 4 discusses in depth the results from decomposing Treasury yield changes using the estimated system. Section 5 concludes.

2 Methodology

In this section, we propose a model framework to study the Treasury market and develop an identification strategy in the spirit of the granular instrument variable (GIV, Gabaix and Koijen 2024). The framework is sufficiently flexible to accommodate investor heterogeneity and can be applied to any asset markets with investor- or sector-level holding data.

2.1 Model Framework

Notations Throughout the text, we use bold symbols to denote vectors or matrices, and regular symbols to denote the scalars. For example, $\mathbf{q}_t$ is a vector with $q_{i,t}$ being its i-th entry. For the ease of notation, we also introduce the subscript S to denote size-weighted aggregation, such that $q_{S,t} \equiv \mathbf{S}'\mathbf{q}_t = \sum_i S_i q_{i,t}$. Similar aggregation notations are also defined for other weighting matrices throughout the text.

For the ease of exposition, we start with a simple case with constant price elasticity and factor loadings across time. In this case, asset demand of entity i is specified as follows:

$$q_{i,t} = -\zeta_i p_t + \boldsymbol{\lambda}_i \boldsymbol{\eta}_t + u_{i,t}.$$

To be concrete, we can consider $q_{i,t}$ to be the percent change in holdings of Treasury bonds by investor i at time t, p_t be the percent change of the aggregate Treasury price, ζ_i be the elasticity of investor i, $\boldsymbol{\lambda}_i$ be the loading on common factors $\boldsymbol{\eta}_t$, such as monetary policy shocks or the uncertainty index, which are assumed to be observable for the simple model. This assumption will be relaxed later. Finally, $u_{i,t}$ denotes the idiosyncratic shocks associated with investor i. For now we assume all the random variables have zero means for simplicity.

Market clears in each period. The net flow in the total market, including supply, would sum to zero. As $q_{i,t}$ is used to denote the entity-*i* flow as a percentage of the total holdings, we scale it by their respective size S_i to arrive at the market clearing condition:

$$\sum_i S_i q_{i,t} = 0.$$

The price adjusts to clear the market:

$$p_t = \frac{1}{\zeta_S} \left(\boldsymbol{\lambda}_S \boldsymbol{\eta_t} + u_{S,t} \right). \tag{2.1}$$

According to Equation (2.1), the Treasury price (or more accurately, the price change) is determined by aggregate demand shocks, scaled by inverse elasticities. The aggregate demand shocks can come from two sources: the common components $\boldsymbol{\lambda}_S \boldsymbol{\eta}_t$, and aggregates idiosyncratic shocks $u_{S,t}$. The inverse of the aggregate elasticity, $\frac{1}{\zeta_S}$, captures the price impact of one unit of demand shocks on the price. We also refer to the inverse elasticity, $\frac{1}{\zeta_S}$ as the *market multiplier*.

In the more general model to be discussed in Section 2.4, we allow for more flexible specifications: the elasticities ζ_i can be entity-specific or shared across investor groups; they can also be regime-specific or parameterized as functions of macro factors. The model can also accommodate time-varying sizes, or no size weighting at all—as we will discuss later, our methodology does not rely on the fat-tail distribution in sizes. Below we illustrate the methodology using the simple model introduced above, and the intuition from the simple model carries over to the full model.

In the data, we observe the time series of the quantities held $q_{i,t}$, price p_t and common factors $\boldsymbol{\eta}_t$. Given this model framework and the parameter estimates, we will be able to back out the idiosyncratic shock $u_{i,t}$ and use them to decompose the price movement to different investors. We now discuss model estimation.

2.2 Identification

2.2.1 Identification of Factor Loadings $\boldsymbol{\lambda}_i$: the "Missing Intercept" Problem

For many research questions, researchers are not interested in the elasticities *per se*, but more in the loadings on the common factors by different entities λ_i. For example, a researcher may want to understand the heterogeneous responses of asset demand to monetary policy shocks. In those cases, a crude estimate for λ_i would be the regression coefficients of quantities held on monetary policy shocks, possibly controlling for other observed factors but excluding the asset price. However, without an estimate of the elasticity ζ_i, this simple approach will only recover the difference in loadings between the investor and the aggregate market, but not the true level of the responsiveness.

To see the issue in more detail, consider the case with a single factor, say monetary policy shocks. In this case, the direct, OLS estimate of asset demand loading on monetary shocks, λ_i^q, is given by:

$$\lambda_i^q \equiv \frac{\mathbb{E}\left[q_{i,t}\eta_t\right]}{\mathbb{E}\left[\eta_t^2\right]} = \frac{\mathbb{E}\left[\left(-\zeta_i p_t + \lambda_i \eta_t\right)\eta_t\right]}{\mathbb{E}\left[\eta_t^2\right]} = \lambda_i - \frac{\zeta_i}{\zeta_S}\lambda_S. \tag{2.2}$$

This equation makes it clear that the estimated coefficient will be downward biased with the bias given by the market average loading multiplied by entity *i*'s price elasticity relative to the market. Hence, observing asset sales from an investor after monetary tightening ($\lambda_i^q < 0$) does not necessarily imply that their demand shifts downwards after surprise monetary policy tightening. If all price elasticities have the same sign, we can only infer their demand is lower than than the market average ($\lambda_i < \lambda_S$). Without knowing the price elasticities and the aggregate loading, however, we cannot determine the sign of the true demand loading λ_i. This issue reflects the "missing intercept" problem commonly faced in macroeconomics: when making inference using micro data, the general equilibrium effects are differenced out in the cross-section. Here the equilibrium asset price is the intercept that is omitted from the analysis.

Therefore, to correctly identify the factor loadings λ_i, we need consistent estimates for the price elasticity ζ_i. To proceed, we follow the following three steps:

1. We regress p_t and $q_{i,t}$ on $\boldsymbol{\eta}_t$ to obtain the residual p_t^{ε} and $q_{i,t}^{\varepsilon}$, and we also denote the coefficients as $\hat{\boldsymbol{\lambda}}_i^q$ and $\hat{\boldsymbol{\lambda}}_i^p$;

2. We estimate elasticities $\boldsymbol{\zeta}$ using residuals p_t^{ε} and $q_{i,t}^{\varepsilon}$ as discussed in the next section

3. We then construct the estimators for $\boldsymbol{\lambda}_i$ as:

$$\hat{\lambda}_i = \hat{\lambda}_i^q + \hat{\zeta}_i \hat{\lambda}^p.$$

In Appendix A.2, we discuss the asymptotic properties of this three-step procedure. Despite the multi-step feature of this estimation method, no adjustment is needed for the asymptotic variance for $\hat{\zeta}$. We also derive the asymptotic variance for loading estimators $\hat{\lambda}_i$ in the Appendix.

2.2.2 Identification of Price Elasticities

As discussed at the end of the previous subsection, we first regress $q_{i,t}$ and $p_{i,t}$ on common factors $\boldsymbol{\eta}_t$ and work with the residuals orthogonal to the common factors. Therefore, in this subsection, we proceed as if there is no common factor for the ease of exposition.

The key challenge in identifying elasticities is that price p_t is endogenous to demand shocks $u_{i,t}$: the idiosyncratic shocks to a sector moves the price through the market clearing. Instruments for prices are needed for identification. However, for many highly liquid markets such as US treasuries, external instruments for prices that can be used for all type of investors with satisfactory statistical power may not exist. Inspired by the insight of granular instrument variables, that the shocks specific to each investor $u_{i,t}$ are often highly idiosyncratic and orthogonal to each other (Gabaix and Koijen 2024), we use our demand and supply system to systematically extract instruments for each investor group. Formally, we make the following identifying assumption:

Assumption 1. *$u_{i,t}$ is independent from $u_{j,t}$ for any $i \neq j$.*

Economically, this assumption states that investor *i*'s trading, conditional on prices and macro factors, is unrelated to the asset demand shifts of other investors. In the real world, due to the large heterogeneity across different types of investors in terms of beliefs and institutional frictions, investors' trading motives can be highly idiosyncratic with little connection with others, and hence "idiosyncratic".

For example, foreign countries' selling Treasuries may be driven solely by their domestic concerns and unrelated to U.S. pension funds' demand. [5]

A potential threat to identification is that after controlling for observed common factors, there are still residual covariance across the idiosyncratic shocks due to unobserved factor structures. Such concerns can be addressed using methods such as principal component analysis (PCA) to extract further unobserved common factors to test the robustness of the assumption. We discuss handling of the unobserved factors at the end of this section.

Econometrically, Assumption 1 implies moment conditions by orthogonality:

$$\mathbb{E}\left[u_{i,t}u_{j,t}\right] \equiv \mathbb{E}\left[\left(q_{i,t}+\zeta_i p_t\right)\left(q_{j,t}+\zeta_j p_t\right)\right] = 0. \tag{2.3}$$

In the data, we do not directly observe the idiosyncratic shocks $u_{i,t}$, but for a given candidate estimator, $\mathbf{z}$, we can form the sample moment condition:

$$\hat{\mathbb{E}}\left[\hat{u}_{i,t}\left(z_i\right)\hat{u}_{j,t}\left(z_i\right)\right] \equiv \hat{\mathbb{E}}\left[\left(q_{i,t}+z_i p_t\right)\left(q_{j,t}+z_j p_t\right)\right] = 0.$$

Denote the number of investors as N, (2.3) covers the whole off-diagonal matrix of $u_{i,t}$ and implies a total of $N \times (N-1)/2$ moment conditions to identify at most N parameters if we allow for full heterogeneity in elasticities. The elasticities are over-identified: we in fact have too many rather than too few moment conditions. With a sufficiently long panel, we can estimate the system using the generalized method of moments (GMM) with two-step or iterative schemes to obtain the optimal weighting matrix. Nevertheless, in practice, we typically have many more moment conditions $N \times (N-1)/2$ than time periods T, and it is a known problem that the standard methods to weigh the moment conditions behave poorly in this scenario (Newey and Windmeijer 2009; Han and Phillips 2006).

Guided by the economics behind the model framework, we derive the optimal weighting for moment conditions in (2.3).[6] We refer to the corresponding estimator

[5] This assumption also assumes one investor's demand cannot be a function of other investors' contemporaneous shocks. In this way, we rule out the possibility that some investors possess and can act quickly on their information advantages. If one is able to identify such investors in the data, our approach allows to exclude these investors in the construction of the moment conditions for optimal GIV. We show later that while broker-dealers could act on clients' idiosyncratic shocks by learning from the trading patterns, excluding broker-dealers (or any other sector for this purpose) from the estimation do not materially affect our quantitative estimates.

[6] In other words, we identify a *one-step* GMM estimator that directly achieves the efficiency

as the *optimal GIV estimator*. In the simple framework discussed in this section, it turns out that the optimal GIV estimator prescribes a very simple and intuitive weighting scheme:

Definition 1. The optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ solves the following sample moment conditions ($\hat{\mathbb{E}}$ denotes the sample mean across the time dimension):

$$\hat{\mathbb{E}}\left[\hat{u}_{i,t}\left(\hat{\zeta}_i\right)\underbrace{\sum_{j\neq i} S_{j,t}u_{j,t}\left(\hat{\zeta}_j\right)}_{\hat{u}_{S(-i),t}(\hat{\boldsymbol{\zeta}})}\right] \equiv \hat{\mathbb{E}}\left[\left(q_{i,t}+\hat{\zeta}_i p_t\right)\sum_{j\neq i} S_j\left(q_{j,t}+\hat{\zeta}_j p_t\right)\right] = 0. \tag{2.4}$$

That is, to estimate the elasticity for the entity i, we weigh other entities' idiosyncratic shocks using their sizes—their respective contributions to price changes.

This estimator is optimal in the sense that it is asymptotic efficienct, formally stated in the proposition below:

Theorem 1 (Asymptotic efficiency)**.** *Given the moment conditions (2.3), Under regularity conditions,*[7] *the optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ is consistent and asymptotically normal:*

$$\sqrt{T}\left(\hat{\zeta}-\zeta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}^{\zeta}\right),$$

for $T \to \infty$. Moreover, its asymptotic variance achieves the semi-parametric efficiency bound (Chamberlain 1987), given as

$$\mathbf{V}^{\zeta} = \zeta_S^2 \times Inv\left(\begin{bmatrix} \frac{1}{\sigma_1^2}\sum_{i\neq 1} S_i^2\sigma_i^2 & S_1S_2 & S_1S_3 & \cdots & S_1S_N \\ S_1S_2 & \frac{1}{\sigma_2^2}\sum_{i\neq 2} S_i^2\sigma_i^2 & S_2S_3 & \cdots & S_2S_N \\ S_1S_3 & S_2S_3 & \frac{1}{\sigma_3^2}\sum_{i\neq 3} S_i^2\sigma_i^2 & \cdots & S_3S_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_1S_N & S_2S_N & S_3S_N & \cdots & \frac{1}{\sigma_N^2}\sum_{i\neq N} S_N^2\sigma_N^2 \end{bmatrix}\right), \tag{2.5}$$

where $\zeta_S \equiv \sum_i S_i\zeta_i$ is the aggregate elasticity and $\sigma_i^2 \equiv var(u_{i,t})^2$.

Proof. See Appendix A. □

bound.

[7] See Appendix A for the exact conditions required.

Remark. The expression (2.5) illustrates the sources of the statistical power of the optimal GIV. First, the standard error of each individual elasticity ζ_i is scaled by ζ_S – the aggregate elasticity –but not their own. In the extreme case where the market is perfectly elastic, granular idiosyncratic shocks are not able to move market price at all, making the instruments irrelevant and the standard errors blow up; second, the statistical power depends on the ratio of other sectors' size-weighted volatility, $\sum_{j\neq i} S_j^2 \sigma_j^2$ to its own residual volatility σ_i^2. The larger is the volatility in the rest of the market, the larger is the variations in the price that is exogenous to entity i that we can exploit. In the extreme case where $\sum_{j\neq i} S_j^2 \sigma_j^2 \to 0$, there is no exogenous residual supply curve shifts facing investor i, whose price elasticity is not identified.

Below we offer two sets of intuitions that shed light on the mechanics of the optimal GIV estimator and show why the simple size-weighting scheme is optimal.

Instrument-variable interpretation This estimator has a natural instrumental variable interpretation: For each entity i, the moment condition can be rewritten as:

$$\hat{\zeta}_i = -\frac{\hat{\mathbb{E}}\left[q_{i,t} \hat{u}_{S(-i),t}\left(\hat{\boldsymbol{\zeta}}\right)\right]}{\hat{\mathbb{E}}\left[p_t \hat{u}_{S(-i),t}\left(\hat{\boldsymbol{\zeta}}\right)\right]}.$$

That is, taking the elasticity for other entities as given, this moment condition uses other entities' idiosyncratic shocks as the instrument for the price when estimating entity i's elasticity. Maximizing the power in IV regressions requires maximum instrument relevance, i.e., the covariance between price and the instrument. The market clearing condition of our model implies that price is an affine function of the size-weighted aggregation of idiosyncratic shocks, so that size weighting achieves maximal relevance.

Debiased-OLS interpretation If the entity is infinitesimal so that its idiosyncratic shocks do not move the price, we can simply regress $q_{i,t}$ on p_t to estimate its elasticity using OLS, and the Gauss–Markov theorem would guarantee that OLS yields the efficient estimator. However, when entities are granular, OLS is biased, as investors' trading moves the price against them. An ideal estimator would exploit the efficiency in OLS while correcting the biases. This is exactly what the optimal

GIV does.. Lemma 1 formally shows that the optimal GIV estimator corrects for the bias arising from OLS estimation.

Lemma 1 (Debiased-OLS moment conditions)**.** *When the whole market is observed so that the market clearing condition $\sum_i S_i q_{i,t} = 0$ holds in sample for every period, the optimal GIV estimator is equivalent to the estimator that solves:*

$$\hat{\zeta}_i = -\frac{\hat{\mathbb{E}}\left[q_i p_t\right]}{\hat{\mathbb{E}}\left[p_t^2\right] - \frac{1}{\hat{\zeta}_S} S_i \hat{\sigma}_i^2} \tag{2.6}$$

where $\hat{\sigma}_i^2 \equiv \hat{\mathbb{E}}\left[\hat{u}_i\left(\hat{\zeta}_i\right)^2\right]$ is the sample variance of the idiosyncratic shocks.

Proof. Using the market clearing condition, we have:

$$\sum_{j \neq i} S_{j,t}\left(q_{j,t} + \hat{\zeta}_j p_t\right) = \hat{\zeta}_S p_t - S_i \hat{u}_i\left(\hat{\zeta}_i\right).$$

Plug it into the sample moment condition (2.4) and rearrange, we have the sample moment conditions (2.6). □

Lemma 1 states that the optimal estimator can be equivalently implemented by regressing quantities on prices and adjusting for the bias—their own impact on prices $\frac{1}{\hat{\zeta}_S} S_i \hat{\sigma}_i^2$. In addition to the OLS interpretation, Lemma 1 also offers an alternative, and sometimes more stable, algorithm for estimating the model.[8]

Unobserved common factors Failing to control for common factors will lead to a violation of Assumption 1. In particular, some common factors may not be observed directly by econometrician. To avoid common factors contaminating the identification, one idea is to use principal component analysis (PCA) to extract unobserved common factors, and use the residuals to identify the elasticities.

One potential complication is that, as pointed out by Qian (2024), with unobserved common factors and fully flexible elasticity specifications, ζ is not uniquely identified. The key intuition is that the price itself is essentially another (endogenous) factor, and alternative values of the price elasticity, combined with appropriate values of loadings on the unobserved common factors could also fit the data.

[8] A Julia package for estimating optimal GIV in its alpha stage is available upon request.

We offer two solutions to this issue, one nonparametric and the other parametric. First, suppose we observe a sector of similar entities. If within the sector the elasticity is homogeneous, then we can demean entities within this sector period by period to remove the influence of p_t. Then we can use PCA to extract the unobserved common factors from the demeaned data. This is the nonparametric approach. Second, we can alternatively parameterize the factor loadings as a linear function of characteristics of each entity, i.e.,

$$\boldsymbol{\lambda}_{i,t}\boldsymbol{\eta}_t = \boldsymbol{X}'_{i,t}\dot{\boldsymbol{\lambda}}\boldsymbol{\eta}_t,$$

in which case the factors can be controlled for using characteristics with time-specific coefficients.

Measurement Errors The quantity data can be noisy in real world applications. The quantity observed by the econometrician may be contaminated by the additional measurement errors:

$$\tilde{q}_{i,t} = q_{i,t} + \varepsilon_{i,t},$$

where $\varepsilon_{i,t}$ only enters the observed $\tilde{q}_{i,t}$ but not the real $q_{i,t}$ and hence not the pricing equation (2.1).

If the measurement errors are *classic*, i.e., it is independent with prices, common factors and each other, then the optimal GIV estimator is still consistent, as the moment conditions in (2.3) still holds with additional orthogonal terms $\varepsilon_{i,t}$. The size-weighting scheme, however, will no longer be optimal except for the special case where $var\,(\varepsilon_{i,t}) \propto var(u_{i,t})$.

Nevertheless, the existence of a residual sector may break the consistency of the estimator under classic measurement errors. A residual sector is a sector that is not directly observed in the data, but backed out from the market clearing condition. In real world applications, we typically do not observe the direct report of the holdings and flows from every market participant. For example, the household sector in the Flow of Funds data is a residual sector imputed using market clearing.

After imposing market clearing, measurement errors in the observed data will negatively enter the quantity for the residual sector r:

$$\tilde{q}_{r,t} \equiv -\frac{1}{S_r}\sum_{i\neq r} S_i\tilde{q}_{i,t} = q_{r,t} - \frac{1}{S_r}\sum_{i\neq r} S_i\varepsilon_{i,t}.$$

In this case, moment conditions between the residual sector r and other sectors with measurement errors may no longer hold, as the measurement errors lead to a negative covariance:

$$\mathbb{E}\left[\left(u_{r,t} - \frac{1}{S_r}\sum_{i\neq r} S_i\varepsilon_{i,t}\right)(u_{i,t}+\varepsilon_{i,t})\right] = -\frac{S_i}{S_r}\sigma^2_{i,\varepsilon}.$$

If one is not interested in the aggregate market elasticity, one can simply exclude the residual sector form the estimation. If there exist investors free from measurement errors $(\sigma_{i,\varepsilon} = 0)$, then the elasticity for the residual sector can also be consistently estimated by only including sectors without measurement errors.

2.3 Comparison with the Standard GIV

The size-weighting scheme proposed above may look reminiscent to readers who are familiar with the original GIV method proposed in the seminal paper of Gabaix and Koijen (2024). In this section we discuss the relationship of our approach to the method introduced in Gabaix and Koijen (2024), hereafter referred to as the standard GIV. This section can be skipped without compromising the understanding of our results in the following sections.

On top of the assumptions above, the standard GIV further imposes homogeneity in elasticities across entities.[9] To closely compare two estimators, we also consider the case with homogeneous elasticity so that $\zeta_i \equiv \zeta$, and we also consider the case where the size vector sums to 1, $\sum_i S_i = 1$.

With homogeneity, the optimal GIV estimator pools the sample moment conditions (2.4) across i weighted by the inverse of residual variance $(\sigma_i^2)^{-1}$ (the details are discussed in the next section 2.4), so that the optimal GIV estimator ζ^{OG} corre-

[9] Gabaix and Koijen (2024) also offer a method to handle heterogeneity in the online appendix, which is further developed by Chodorow-Reich et al. (2024). Here the standard GIV refers to the case with homogeneity introduced in the main text of Gabaix and Koijen (2024).

sponds to the following moment conditions in population:

$$\sum_i \frac{(\sigma_i^2)^{-1}}{\Sigma_i(\sigma_i^2)^{-1}} \mathbb{E}\left[\left(q_{i,t} + \zeta^{OGIV} p_t\right) \sum_{j\neq i} S_j \left(q_{j,t} + \zeta^{OGIV} p_t\right)\right] = 0$$

Denote the $\mathbf{\Sigma}_u = diag\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2\right)$ as the covariance matrix of the residual variance, and $\mathbf{E}_i \equiv \frac{(\sigma_i^2)^{-1}}{tr(\mathbf{\Sigma}_u^{-1})}$ as the weights using the inverse of the residual volatility, we can consolidate the corresponding moment conditions above in the matrix form:

$$\mathbf{E}'\mathbb{E}\left[\left(\mathbf{q}_t + \boldsymbol{\iota}\zeta^{OGIV} p_t\right)\left(\mathbf{q}_t + \boldsymbol{\iota}\zeta^{OGIV} p_t\right)'\right]\mathbf{S} = \mathbf{E}'\mathbf{\Sigma}_u\mathbf{S} = \frac{1}{tr(\mathbf{\Sigma}_u^{-1})}\boldsymbol{\iota}. \quad \text{(Optimal-GIV)}$$

Notice that this is a quadratic equation in ζ^{OGIV} and has to be solved nonlinearly.

The key insight in the standard GIV is that, with homogeneous elasticities, cross-sectional demeaning can eliminate the endogenous term ζp_t. Mathematically, for any vector that sums to 1, $\mathbf{X}'\boldsymbol{\iota} = 1$, we also have $\mathbf{E}'\mathbf{\Sigma}_u\mathbf{X} = \frac{1}{tr(\mathbf{\Sigma}_u^{-1})}\boldsymbol{\iota}$, so by taking the difference of these two equations, we have:

$$\mathbf{E}'\mathbb{E}\left[\left(\mathbf{q}_t + \boldsymbol{\iota}\zeta^{SGIV} p_t\right)\left(\mathbf{q}_t + \boldsymbol{\iota}\zeta^{SGIV} p_t\right)'\right](\mathbf{S} - \mathbf{X}) = 0, \quad \text{(Standard-GIV)}$$

where $\mathbf{S} - \mathbf{X}$ essentially demeans $\mathbf{q}_t$ in the cross section. Move weighting vectors into the expectation operator, this weighting scheme can be further expressed as in terms of weighted average of q:

$$\mathbb{E}\left[\left(q_{E,t} + \zeta^{SGIV} p_t\right)\left(q_{S,t} - q_{X,t}\right)\right] = 0.$$

Importantly, the term ζp_t drops out from the second term in the expectation due to differencing. As it is linear in ζ, it now be implemented with an IV regression of precision-weighted quantity $(q_{E,t})$ on p_t, using $q_{S,t} - q_{X,t}$ as the instrument. Gabaix and Koijen (2024) further show that within this class of the linear estimators, setting $\mathbf{X} = \mathbf{E}$ achieves the highest statistical power, and hence the recommended instrument is the size-minus-precision weighted q.

At first glance, the mathematical difference between the standard GIV and the optimal GIV in the simplest case is merely whether to demean $q_{i,t}$ in the cross section. This subtle difference nevertheless has an important implication on the statistical power in a sample where the distribution of the contribution to price changes across entities does not exhibit a large fat tail.

To see the power comparison, consider a simpler case with homoskedasticity, i.e., $\sigma_i \equiv \sigma$. In this case, the asymptotic variance of the optimal GIV estimator has a simple expression: [10]

$$\sqrt{T}\left(\hat{\zeta}^{OGIV} - \zeta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}_{\zeta}^{OGIV} = \frac{\zeta^2}{(N-2)\sum_i S_i^2 + 1}\right),$$

while the standard GIV converges to a different limiting distribution: [11]

$$\sqrt{T}\left(\hat{\zeta}^{SGIV} - \zeta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}_{\zeta}^{SGIV} = \frac{\zeta^2}{N\sum_i S_i^2 - 1}\right).$$

We can show that $\mathbf{V}_{\zeta}^{OGIV} > \mathbf{V}_{\zeta}^{SGIV}$. More importantly, the crucial difference is that the standard GIV relies on the fat-tail distribution in S_i for statistical power, while the optimal GIV does not. This is because when sizes are equally distributed ($S_i = \frac{1}{N}$), the size weighting and precision weighting (which is equal weighting in the case of homoskedasticity) coincides, so the standard GIV's instrument has zero variation after demeaning. This is not the case in the optimal GIV, whose statistical power still grows at the rate of $\sqrt{T}$.

This power comparison is not only theoretical but is also relevant practically. While fat-tail distribution in sizes is prevalent in real-world data, the residual volatility typically shrinks as an entity becomes larger. As shown in the expressions in (A.6), what matters for the statistical power is not only the size distribution but the the distribution of each entity's contribution to price volatility, i.e., the size-weighted volatility, $S_i^2\sigma_i^2$. To see that, consider the case where the size has a fat tail distribution but the residual volatility is inversely proportional to the size, $\sigma_i \propto \frac{1}{S_i}$. In this case, the size vector again coincide with the precision weighting, and there is no variation in the size-minus-precision instruments. Figure B.1 in the Appendix compares the power of the standard GIV and the optimal GIV estimator in simulations. With parameters calibrated to the Treasury market in the real world data, we find that the optimal GIV is twice as powerful as the standard GIV.

[10]See the Appendix A for the derivation.

[11]See Gabaix and Koijen (2024) for the derivation.

2.4 General Framework

After illustrating the main idea using the simple model framework above, we introduce the more flexible model. The general model is specified as follows:

$$\left.\begin{aligned} q_{i,t} &= -p_t \times \mathbf{C}_{i,t}'\boldsymbol{\zeta} + \mathbf{X}_{i,t}'\boldsymbol{\beta} + u_{i,t}, \\ 0 &= \textstyle\sum_i S_{i,t} q_{i,t} \end{aligned}\right\} \Longrightarrow p_t = \frac{1}{\mathbf{C}_{S,t}'\boldsymbol{\zeta}}\left[\mathbf{X}_{S,t}'\boldsymbol{\beta} + u_{S,t}\right],$$

which allows for the following flexibility:

- The elasticity is parameterized by the matrix $\mathbf{C}_{i,t}$, which can be entity-specific and time-varying;
- The control variable $\mathbf{X}_{i,t}$ can also also be entity&time-specific;
- The size vector $S_{i,t}$ is also allowed to be time-varying; we also do not require it to sum to 1.

The simple model can be recovered from the general specification by setting $\mathbf{C}_{i,t}$ to be the entity fixed effects, and $\mathbf{X}_{\mathbf{i},\mathbf{t}}$ to be the interaction of entity fixed effects with the aggregate factors $\boldsymbol{\eta}_t$.

The moment conditions used to identify the elasticity vector $\boldsymbol{\zeta}$ remain the same as before, only it is now conditional on all other exogenous variables $(\boldsymbol{\eta}_t, \mathrm{C}_t, \mathrm{S}_t)$:

$$\mathbb{E}\left[u_{i,t}u_{j,t} \mid \mathbf{X}_t, \mathbf{C}_t, \mathbf{S}_t\right] = 0. \tag{2.7}$$

We again follow the three-step procedures outlined in 2.2.1 to strip away the influence of control variables $\mathbf{X}_{i,t}$, and then form the optimal GIV estimator for $\boldsymbol{\zeta}$ using the residuals $q_{i,t}^{\varepsilon}$ and p_t^{ε}. We provide the optimal GIV estimator and its asymptotic properties below, and refer readers to Appendix A for the derivation.

Definition 2. Let $\boldsymbol{\psi}_t(\mathbf{z})$ be a vector of length $\frac{N(N-1)}{2}$ with subscript ij $(i \neq j)$ denote the row corresponding the entity pair (i,j) :

$$\psi_{ij,t}(\mathbf{z}) \equiv \hat{u}_{it}(\mathbf{z})\,\hat{u}_{jt}(\mathbf{z}) \equiv \left(q_{i,t}^{\varepsilon} + p_t^{\varepsilon}\mathbf{C}_{i,t}'\mathbf{z}\right)\left(q_{j,t}^{\varepsilon} + p_t^{\varepsilon}\mathbf{C}_{j,t}'\mathbf{z}\right).$$

Denote $\mathbf{V}^{\psi} \equiv \mathbb{E}\left[\boldsymbol{\psi}_t(\boldsymbol{\zeta})\,\boldsymbol{\psi}_t(\boldsymbol{\zeta})'\right] = Diag\left(\sigma_1^2\sigma_2^2, \sigma_1^2\sigma_3^2, \ldots, \sigma_{N-1}^2\sigma_N^2\right)$. Denote $\hat{\mathbf{W}}_t^*(\mathbf{z})$ as weighting matrix of the size $\frac{N(N-1)}{2} \times N_\zeta$ such that

$$\hat{W}^*_{ij,k,t}(\mathbf{z}) = \frac{1}{\mathbf{C}'_{S,t}\mathbf{z}} \left(\frac{S_{jt}C_{it,k}}{\hat{\sigma}_i^2(\mathbf{z})} + \frac{S_{it}C_{jt,k}}{\hat{\sigma}_j^2(\mathbf{z})} \right),$$

where $\hat{\sigma}_i^2(\mathbf{z}) \equiv \hat{\mathbb{E}}\left[\hat{u}_{it}(\mathbf{z})^2\right]$. The optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ for the general model solves the following sample moment conditions:

$$\hat{\mathbb{E}}\left[\hat{W}^*_t\left(\hat{\boldsymbol{\zeta}}\right)' \psi_t\left(\hat{\boldsymbol{\zeta}}\right)\right] = \mathbf{0}.$$

Similar to the simple model, we show the optimal GIV estimator in this case is also asymptotically efficient:

Theorem 2 (Asymptotic efficiency in the general model)**.** *Given the moment conditions (2.7), Under regularity conditions*[12]*, the optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ is consistent and asymptotically normal:*

$$\sqrt{T}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}^\zeta\right),$$

for $T \to \infty$, where

$$\mathbf{V}^\zeta = \left(\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{V}^\psi\mathbf{W}^*\right]\right)^{-1}.$$

Moreover, $\mathbf{V}^\zeta$ achieves the semi-parametric efficiency bound.

Proof. See Appendix A. □

3 Estimation

3.1 Model Specification and Data

Using the estimation method introduced in Section 2, we estimate a demand system for the Treasury market, specified as follows:

$$\left.\begin{array}{ll} f_{i,t} & = -\zeta_{i,r(t)}p_t + \boldsymbol{\lambda}_{i,r(t)}\boldsymbol{\eta}_t + \bar{f}_i + u_{i,t} \\ \sum_i S_i f_{i,t} & = 0 \end{array}\right\} \implies p_t = \frac{1}{\zeta_{S,r(t)}}\left[\boldsymbol{\lambda}_{S,r(t)}\boldsymbol{\eta}_t + \bar{f}_S + u_{S,t}\right]. \tag{3.1}$$

[12]See Appendix for the exact conditions

Below we explain the meaning of each terms and the relevant data sources. For a detailed description of data construction, see Appendix C.1.

Flows and positions We focus on medium-to-long term U.S. Treasury notes and bonds (Treasury bonds hereafter for simplicity). To capture the holdings and transactions of Treasury bonds, we rely on the Financial Accounts of the United States (also referred to as Flow of Funds), compiled by and retrieved from the Federal Reserve Board. The Financial Accounts report the level of holdings and transactions of Treasury securities at the sectoral level, such as mutual funds, banks, etc. We exclude holdings and flows of Treasury bills whenever possible.[13] The transactions are designed to only capture the exchange of the assets, and hence the "flows", but not the changes due to asset revaluation. The market clearing condition are respected in the flow data such that total dollar flows in one quarter sums to zero across all sectors (including net issuance):

$$\sum_i F_{i,t} = 0. \tag{3.2}$$

One special sector in the Financial Accounts is the household sector. As there is no direct measurement of the total assets held by household sector in the U.S., Financial Accounts data compute household sector holding from the market clearing condition (3.2) above by taking the residuals. As a result, the household sector is a catch-all sector for all investors that are not included and observed in other sectors. These investors include hedge funds, endowment funds and family offices. While we refer to them as the household sector following the labeling of the Financial Accounts, the sector should be interpreted more broadly.[14]

For two sectors in the Financial Accounts, we further break them down to a more granular level: For U.S.-chartered depository institutions (banks), we use individual bank holding of Treasury securities derived from Call Reports.[15] For

[13] In the Financial Accounts, not all sectors report separately Treasury bills and and notes&bonds. In those cases, we use the total flows and positions in all Treasury securities including bills. See Appendix C.1 for our treatment in more detail.

[14] Following the discussion on measurement errors in Section 2, to avoid spurious correlations, we only use moment conditions between the household sector (the residual sector) and sectors with accurate reporting, such as the Fed, supply, banks, etc. In Appendix Table C.3 we also report estimation results without using the household sector, and the estimates for other sectors are largely consistent with the baseline specification, suggesting the measurement error is not a major concern.

[15] We follow a similar procedure, used by Jansen, Li, and Schmid (2024) and documented in Ap-

foreign investors (RoW), we use Treasury International Capital (TIC) to further break it down to the country level.

We estimate the model at the quarterly frequency. Our baseline sample spans from 2003Q4 to 2023Q4. In Appendix Table C.3, we also report the estimates using only the Financial Accounts to extend the sample back to 1970Q1. Though with a large time-span, the elasticity estimates are highly stable.

Figure 1b provides a bird's-eye view of the holdings (in market values whenever possible) and dollar flows by different groups of investors reported in the Financial Accounts in this century. The total Treasury notes and bonds outstanding increased from less than $\$5$ trillion to more than $\$$ 20 trillion at the end of 2023, or from around 30% percent of GDP to 70%.[16] Given the drastic shift in the debt to GDP ratio over the sample period, 1% increase in debt-to-GDP in 2023 has very different implication than that in 2008. Hence in this paper, we opt to use the total outstanding market value to normalize flows, as discussed below.

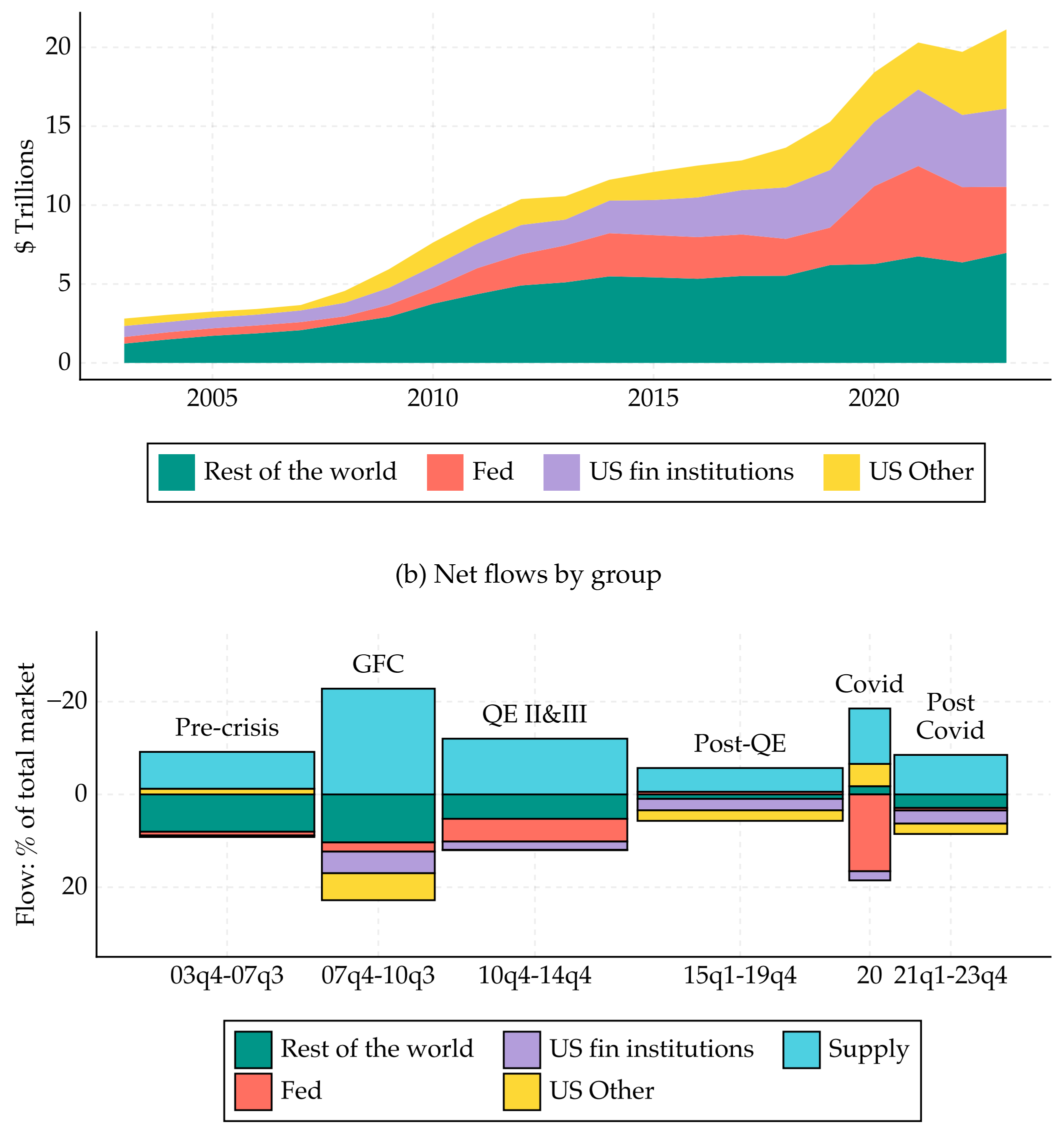
pendix C.1, to impute Treasury holdings from Treasury and agency holdings of individual banks.

[16]Total Federal debt-to-GDP ratio is around 120% at the end of 2023, comprised of Treasury notes and bonds, bills and non-marketable Treasury securities.

Figure 1: **Holdings and Flows: U.S. Treasury Notes & Bonds**

(a) Total holdings by group

(b) Net flows by group



Note: Figure 1 plots the total holding of U.S. Treasury notes and bonds by investor groups (panel (a)) and groups the flows (net transactions as percentage of the total market size) by investor groups and episodes. We aggregate investor sectors reported in Financial Account (Z.1) to four investor categories and one issuer sector ("supply"), the exact mapping available in Appendix C.1. In Panel (b), the sign of the flows are inverted, so that a higher position corresponds to a upward pressure on Treasury yields.

The flow measure $f_{i,t}$ is defined as the net dollar flows $F_{i,t}$ by investor i, normalized by their market share S_i times the total market values of Treasury bonds outstanding in the last quarter A_{t-1}^{total}:

$$f_{i,t} \equiv \frac{F_{i,t}}{S_i A_{t-1}^{total}}.$$

We use the constant market share S_i in the baseline specifications, defined as the average market share of each entity throughout the sample period:[17]

$$S_i = \frac{|\sum_t A_{i,t}|}{\sum_t A_t^{total}}.$$

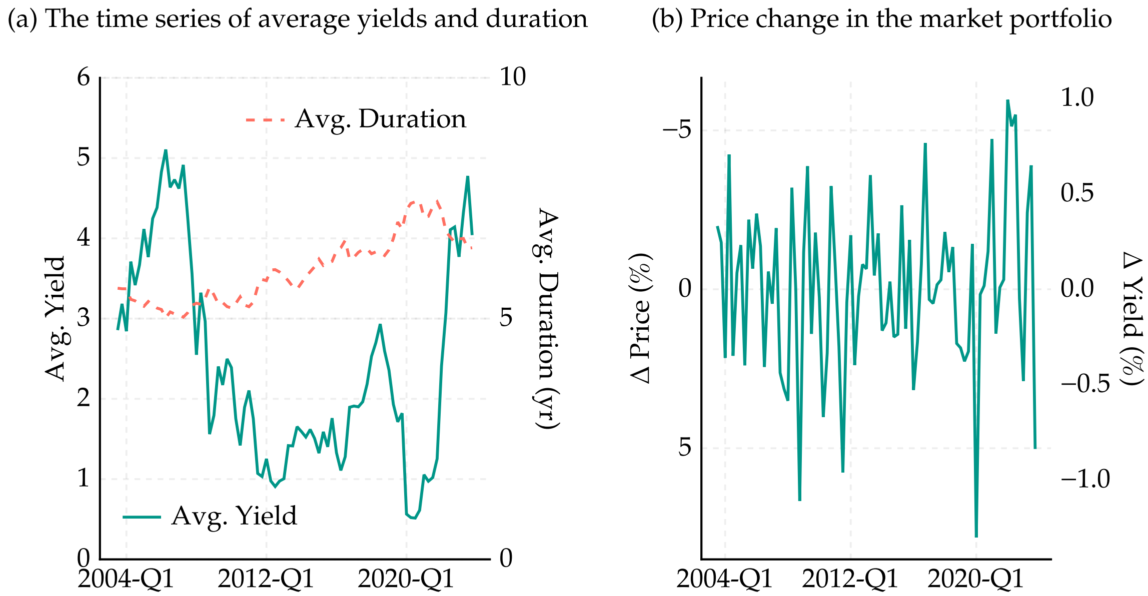
Supply enters the system as negative demand. We choose the sign convention such that S_i are always positive, and $f_{i,t}$ follows the sign of the dollar flows $F_{i,t}$, with positive numbers indicating net buying and negative numbers indicating net selling. For example, in 2023Q4, Department of the Treasury issued $\$876.5$ billion of Treasury bonds, and the total outstanding bonds by the end of 2023Q3 is $\$20370.8$ billion, so the supply flow in Q4 is $f_{supply,2023Q4} = \frac{-\$876.5}{\$20370.8} = -4.3\%$. In this way, the supply and demand can be treated uniformly in this framework: a positive ζ_{supply} means when the Treasury price is higher, supply tends to increase.

Price and Elasticity For price p_t, we use the quarter-on-quarter percent change in the price of the market portfolio of Treasury notes and bonds with maturity longer than 1. The data is downloaded from CRSP US Treasury Database. Figure 2a plots the path of average yields (the left axis) and the average duration (the right axis) of the market portfolio. The average duration over the sample period is around 6 years. Figure 2b plots the quarterly changes in the price of the market portfolio. We flip the sign of the y-axis so that an increase in the yield corresponds to a higher position in the figure. We also mark on the right axis the the corresponding changes in average yields. As coupon and principal payments are fixed, the change in prices is closely tied to the return to the market portfolio. The time series

[17] The choice of using constant size weighting is to address the following concerns: Some sectors, such as households and security brokers and dealers, have very volatile net positions, and have close-to-zero net positions for several periods. However, as their gross long and short positions are much larger, using their real-time net positions as their sizes will not properly capture their influence in the market. Using constant size weighting avoids such issues. With constant size weighting, flows can be interpreted as the change in demand as a percentage of the total market outstanding, scaled by a constant S_i.

exhibit stationarity and close to zero autocorrelation.

Figure 2: The path of average yields on U.S. Treasury notes and bonds



Note: Figure 2 plots the path of average yield, duration and price change in the market portfolio of U.S. Treasury bonds and notes from 2004 to 2023. All three objects are aggregated from CRSP from the security level weighted by amount outstanding. In Panel (b), the left y-axis corresponds to price change while the right y-axis corresponds to yield change, defined as price change scaled by average duration.

We allow different investors to have different price elasticities. In principle, our method allows for full heterogeneity in elasticities. For estimation efficiency and interpretability, we bin smaller investors into groups by their characteristics and estimate an homogeneous elasticity for each group. For example, all banks are assumed to have the same elasticity, and all types of pension funds are binned into one group as well. The detailed mapping from sectors to groups can be found in Appendix Table C.1.

As the U.S. financial market has undergone a regime shift after the global financial crisis, many sectors may behave differently before and after the crisis. For example, the Fed is only an active investor in the long-term Treasury market after 2009, and the existing literature has documented that the new regulations since the global financial crisis hamper the intermediation capacity of the banking sector (Du, Tepper, and Verdelhan 2018; Stulz, Taboada, and van Dijk 2022, among others). We also allow for these sectors to have different elasticities across regimes, denoted as the subscript $r(t)$. We assume that $r(t)$ takes two values, and set 2009Q1 as the start of the second regime.

Common factors $\boldsymbol{\eta}_t$ We control for an extensive set of common factors that help explain Treasury price fluctuations. For macro variables, we include innovations to VIX – the canonical indicator of global risk sentiment – to gauge the intensity of "flight to safety". We include quarterly changes in the effective Fed funds rate to control for the short-end movement in the interest rate and examine the pass-through of the short rate to longer-term interest rates. We also consider innovations to the USD broad index, and gap between quarterly CPI inflation and the inflation target (0.05 percent per quarter, or 2 percent per annum). As common factors are assumed to be exogenous to idiosyncratic shocks and prices in the Treasury market, we avoid controlling for factors directly related to the Treasury market such as the credit spread. Adding more macro factors does not meaningfully improve the explanatory power of flows and prices, and risks overfitting.

Two policy sectors, the Fed and the Department of Treasury (supply), pre-announce planned purchases and sales for policy transparency and forward guidance. For this reason, we also include variables predicting their flows in the common factors to capture policy anticipation and potential front-running behaviors of other sectors. For the Fed, we control for the purchase scheduled a quarter ahead.[18] For the Treasury, we include net issuance in the previous quarter to control for the predictive components in supply. After controlling for those factors, the coefficients estimated for the Fed and the supply measure the flow sensitivity due to deviation from the scheduled purchases and sales.

One potential threat to our identification strategy is that there exists a hidden factor structure among the idiosyncratic shocks $u_{i,t}$, so that residual covariance remains even after controlling for the macro factors. To ensure the robustness of our estimates, we use a version of principal component analysis (PCA) robust to outliers to extract common factors from the flows by foreign countries and banks, for which we have micro data on country- and bank-level Treasury holdings. We pick 3 common factors in the baseline as further extraction only minimally increase the explanatory power, and in Appendix Table C.3 we report the specification with

[18]Appendix C.2 discusses the construction of the scheduled purchase series.

more factors and the results are almost identical.

Unobserved idiosyncratic flows The price elasticities are identified using the moment condition that $u_{i,t}$ is orthogonal to each other (see Section 2):

$$\mathbb{E}\left[u_{i,t}u_{j,t} \mid \boldsymbol{\eta}_t\right] = 0 \;\; \forall i \neq j.$$

3.2 Elasticity Estimation

We start with the price elasticities. Table 1 reports the elasticity estimates of the baseline model (averaged across time for sectors with regime shifts). In Appendix Table C.5 we report the full table of the model estimates including factor loadings.

Aggregate Price Elasticity The estimate for aggregate market elasticity is 1.01, suggesting that a 1% demand shock lead to around $\frac{1}{1.01} = 99bps$ increase in the aggregate price of the market portfolio, equivalent to a $20bps$ decrease in yields for a five-year bond or a $10bps$ decrease for a ten-year bond.

To put the numbers into perspective, the U.S. aggregate equity market multiplier is around $5 \sim 8$ (Gabaix and Koijen 2022), and that of the corporate bond market is around 3 at the level of portfolios of bonds with the same credit rating (Chaudhary, Fu, and Li 2023). Our estimates is consistent with the prior that the Treasury market is a very liquid market.[19]

To further shed light into the estimation of the aggregate elasticity, notice that from the price equation we can write the aggregate elasticity as the ratio between the volatility of aggregate demand shocks, $f^D_{S,t} \equiv \mathbf{S}'\left(\boldsymbol{\lambda}_{r(t)}\boldsymbol{\eta}_t + \bar{\mathbf{f}} + \mathbf{u}_t\right)$, and the volatility of the price:

$$p_t = \frac{1}{\zeta_S}\left[\underbrace{\boldsymbol{\lambda}_{S,r(t)}\boldsymbol{\eta}_t + \bar{f}_S + u_{S,t}}_{f^D_{S,t}}\right] \implies \zeta_S = \frac{std(f^D_{S,t})}{\sigma_p}. \tag{3.3}$$

This equation offers a very intuitive understanding of the aggregate elasticity. The

[19]The literature often finds that "micro" price elasticity is larger than the market-wide "macro" elasticity. Koijen, Koulischer, et al. (2017) find a micro elasticity of 3 for the Euro Area government bond market, adding to the plausibility of our macro estimate.

market is elastic if there are volatile demand shocks while the price are stable, and vice versa. While we observe the volatility of the price, the demand shocks are not directly observable. Instead, we only observe the flows, which is equal to the sum of demand shocks and the response to equilibrium price changes. Nevertheless, we can use the volatility of flows to gauge a reasonable magnitude of the price elasticities.[20] The size-weighted average of the flow volatility, $\sqrt{\sum_i S_i^2 var(f_{i,t})}$, is around 3.62% at the quarter frequency, while the price volatility is around $\sigma_p = 2.6\%$, making the ratio around 1.38, the same order of magnitude as our estimate. To formally estimate ζ_S using this formula, one needs to identify the covariance of the underlying demand shocks for each investor. Our method is designed to address this identification need.

Heterogeneous elasticities across sectors Underlying the aggregate elasticity estimate are different sectors with vastly heterogeneous elasticities. Table 1 reports the elasticity of each sector, ordered by their average contribution to the aggregate elasticity, $\zeta\ share \equiv \frac{\zeta_i \times S_i}{\zeta_S}$.

Figure 3 provides an illustration of our sector-specific estimates against the OLS approach of estimating price elasticities. For the household sector and the mutual fund sector, we plot their "demand curves". We residualize sector inflow into Treasuries against common factors and plot it on the x-axis. The y-axis represents two notions of Treasury price changes. Using green circle markers, we plot the observed price changes, also residualized against the common factors. OLS estimation would thus imply that that the household sector is completely inelastic, while mutual funds have a slight downward-sloping demand curve. Simultaneity, however, renders OLS estimators biased in this context: When sectors are granular and demand is downward-sloping, a higher asset price reduces demand, but the reduced demand will counteract the price.

In orange, we plot the flows against the predicted price changes from idiosyncratic shocks from other sectors that we extract from our estimation. As discussed in the IV interpretation of in Section 2, the slopes of the orange lines correspond

[20] Roughly speaking, there are two sources of errors when approximating aggregate demand shock volatility $std(f_{S,t}^D)$ using the size-weighted flow volatility: First, the flow volatility contains the variance of flows in response to price changes, and hence it may overstate the volatility of the demand shocks and therefore overestimate the elasticity. Second, it does not capture the cross-sectional covariance in demand shocks. When demand shocks are positively correlated, it understates the aggregate demand volatility and implied elasticity.

Table 1: **Average price elasticities**

Sector	$S(\%)$	ζ	ζ **Share (%)**
Aggregate		1.01	100.0
		(0.77, 1.25)	
Households	5.74	9.8	55.74
		(5.07, 14.54)	
RoW	44.45	0.42	18.41
		(0.24, 0.61)	
Fed	22.08	0.44	9.84
		(0.11, 0.78)	
Mutual Funds	6.75	0.61	4.11
		(0.15, 1.08)	
Banks	5.26	0.67	3.41
		(0.42, 0.91)	
Insurance	2.59	0.99	2.54
		(0.66, 1.32)	
Dealers	0.81	2.04	1.63
		(-6.61, 10.69)	
Other	11.78	-0.07	-0.85
		(-0.3, 0.15)	
ETF	1.18	-0.49	-0.58
		(-0.85, -0.13)	
Pension	5.42	-0.1	-0.56
		(-0.27, 0.06)	
Supply	100.0	0.06	6.3
		(-0.01, 0.14)	

Note: Table 1 reports the aggregate price elasticity, sector-specific price elasticity of demand ζ for the 10 investor categories in the sample, and the price elasticity of supply. The sample period is 2003Q4–2023Q4. The price elasticities are identified using the optimal GIV estimator developed in Section 2. Time-series averages are taken for sectors that assume to have different elasticities before and after 2009Q1 (the sectors include Federal Reserve, Rest of World and U.S. Banks). S corresponds to the size weight used in the estimation, defined as the average market share of each entity throughout the sample period. ζ share denotes the (size-weighted) fraction of aggregated elasticity accounted for by each sector. 95% confidence intervals are reported, with the standard errors given by Theorem 2.

exactly to the (reciprocal of) the elasticity coefficients reported in Table 1.[21] Using

[21]To further ensure the exclusion restriction is respected, we can exclude all moment conditions involving the target sectors (households/mutual funds) when extracting idiosyncratic shocks for other sectors. In this way, those extracted idiosyncratic shocks are completely unrelated to the target sectors except through the price. This is corresponding to the leave-one-out estimator reported

the instrumented price changes reveals that the household sector is in fact highly elastic. Mutual funds are also slightly more elastic than what the simple correlation suggests.

The debiased-OLS interpretation of our optimal GIV estimator (see Lemma 1) helps interpret these differences. Though the mutual fund sector and the household sector are comparable in terms of sizes, the household sector has much more volatile idiosyncratic shocks than the mutual fund sector, a pattern that is also visible from the volatility of the raw flows. The optimal GIV estimator recognizes that and make larger correction for the household sector than for mutual funds.

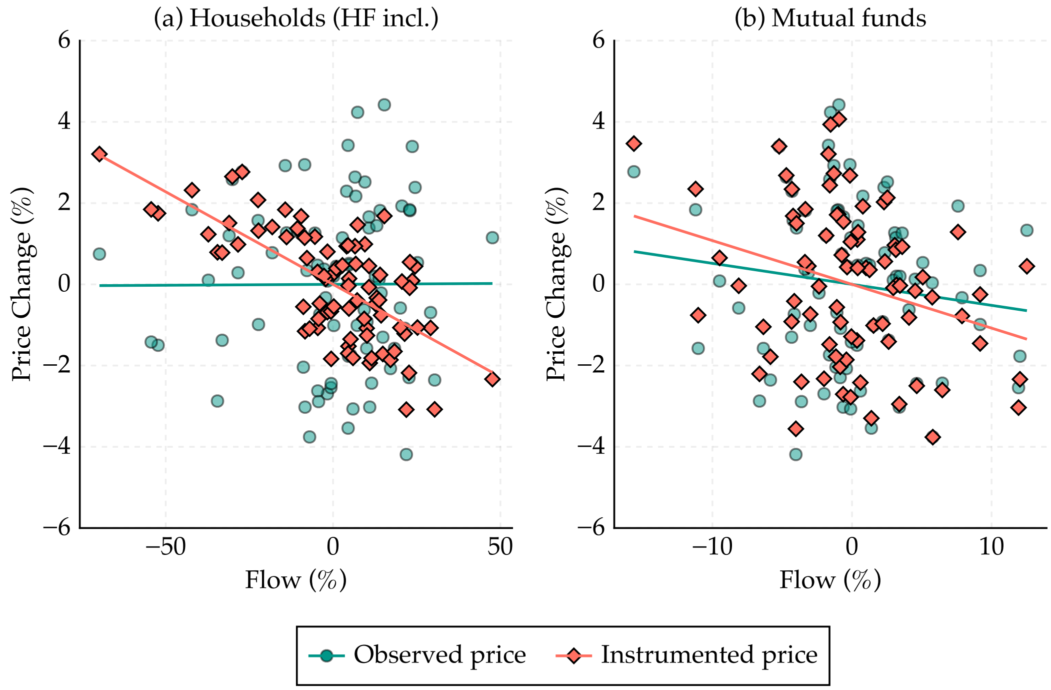
Overall, after adjusting for biases, the household sector is the most price elastic, accounting for 55% of the total liquidity provision in the market. As discussed in the data section, the household sector is a catch-all sector that includes other type of investors not covered by the Financial Accounts, such as hedge funds, family offices, separately managed account for high net-worth individuals, endowment funds by not-for-profit organizations, etc. These investors with high sophistication, less regulatory constraints and sufficient capital intuitively can act as the most nimble liquidity providers of the Treasury market.

Moving on to other sectors, we estimate investors from the rest of the world are on average less elastic than the domestic sectors, but due to its large market share, the foreign sector is the second largest contributor to the aggregate elasticity. The Fed is estimated to be price-elastic—as we discuss in details in Section 4.2, we control for the Fed's scheduled purchase so the elasticity estimates are from the surprises, primarily the decision on when to initiate or exit its asset purchase programs.

Security brokers and dealers are not contributing much to the aggregate elasticity (only 1.63%), a finding that is seemingly at odds with the large literature on the the crucial role of primary dealers in the Treasury market and the heavy policy emphasis. To reconcile this difference, we note that our estimates are based on quarterly data, while the discussion on primary dealers are typically centered around a much higher frequency, including intraday trades. Even though gross trading flows via broker-dealers every day is huge, they typically do not hold large inventories for an extended period, and hence both their net positions at quarter ends and net flows are very small. On average their net position at quarter ends are

in Appendix Table C.4.

Figure 3: **Optimal GIV vs. OLS: Two examples**

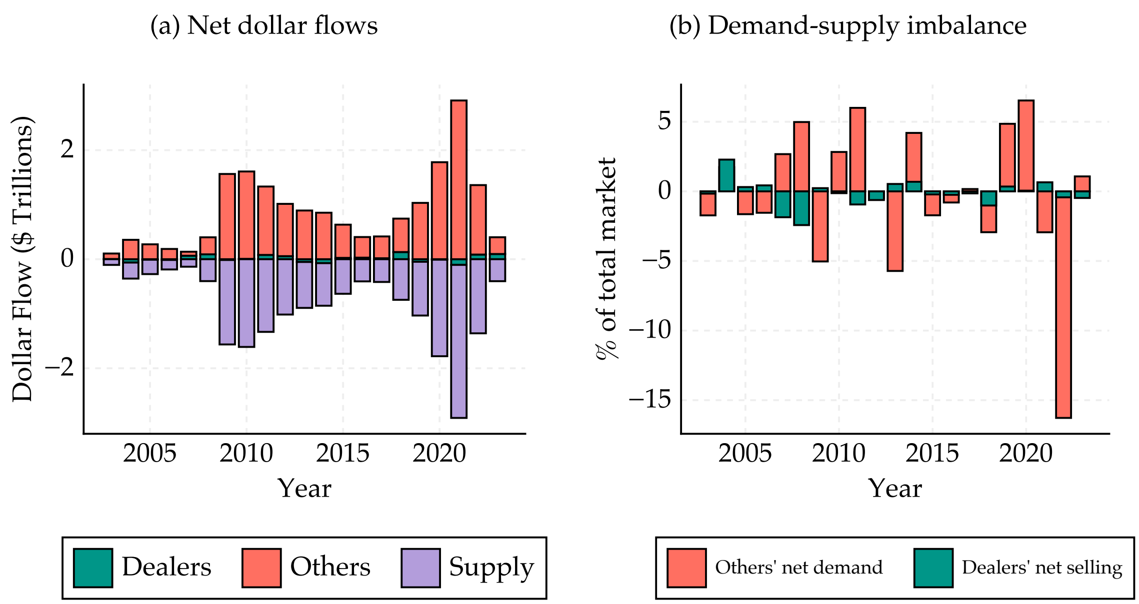


Note: Figure 3 compares the estimated price elasticities of demand using our identification approach to the OLS estimates, using the household sector and the mutual fund sector as case studies. In both panels, the green dots relate the sector inflow to U.S. Treasury notes and bonds to observed quarterly changes in the Treasury price, both residualized against the common factors included in our baseline specification. For the red dots, the y-axis plots the predicted price changes using size-weighted idiosyncratic shocks (excluding the sector of interest) estimated from the GMM procedure.

less than 1% of total market values. Panel (a) in Figure 4 compares their net dollar flows with the aggregate flows from the supply and other investors and confirms that the net flows from the broker-dealers are tiny. Given their relatively small net flows, they are unlikely to provide significant liquidity on a quarterly basis.

In Panel (b) of Figure C.1 we plot the aggregate demand-supply imbalance from other sectors, backed out from our model, and the dealers' flows. If two bars point to the same direction, then the dealers provide liquidity to the rest of the market. However, if anything, broker-dealers only cover a very small percent of the total demand imbalance, and often go in the opposite direction: For example, in 2008, when there was a net buying pressure from the rest of the economy, broker-dealers also purchased a significant share of Treasury bonds.

Figure 4: **The role of security broker-dealers**



Note: Figure 4 examines the role of security broker-dealers in affecting U.S. Treasury yields at low frequency. In Panel (a), we compare net annual flow into the Treasury market due to broker-dealers (in green) with those of other holders and the net issuance. In Panel (b), we plot a measure of demand-supply imbalance and the broker-dealers' degree of accommodation. For all sectors except the broker-dealers, the red bars aggregate the demand/supply curve shifts due to common factors and idiosyncratic shocks, $q_{it} + \zeta_{ir(t)} p_t$ to the annual frequency. The green bar, on the other hand, plots the annual total selling due to the broker-dealers. Bars pointing to the same direction indicates that dealers are accommodating the demand-supply imbalance of other sectors, by either selling when others increase demand, or buying when others sell.

Exchange-traded funds (ETFs) and pension funds are among the least elastic sectors. These findings can be explained by the passive and inactive investment behaviors of these sectors. ETFs are mostly passive, whose mandate forbids them to actively time the market. Pension funds are also known as buy-and-hold investors in the Treasury market.[22] Finally, the supply has a weak positive elasticity that is indistinguishable from zero. This is consistent with the treasury's "regular and predictable" debt management stance, that it does not seek to time the market in issuance. Every quarter, the Treasury announces its anticipated offering amounts of notes and bonds in the next quarter, and deviations from these

[22]Our finding is also consistent with Koijen, Koulischer, et al. (2017), who find that pension funds have the lowest price elasticity in the European government bond market.

announcement are small in normal periods (Cassidy and Mirani 2024).

Robustness of elasticity estimates We perform extensive robustness checks to show our elasticity estimates are highly stable across various specifications. One may be concerned that as all elasticities are estimated jointly, estimation errors for one investor can propagate into the whole system. In Table C.4 in Appendix, we report the estimation results while excluding sectors in turns. No single sector drives our results, as we recover very similar sector-level elasticity estimates across all specifications. In Table C.3, we further put the model to test under multiple other specifications: 1) we include 5 unobserved common factors recovered from PCA; 2) we use only the Financial Account data to extend the sample back to 1970; 3) we estimate the results including Treasury bills as well. The results across all these specifications are largely comparable. The elasticities including Treasury bills are estimated to be larger, with the aggregate elasticity being 2.0. This is consistent with the prior that bills are much more liquid.

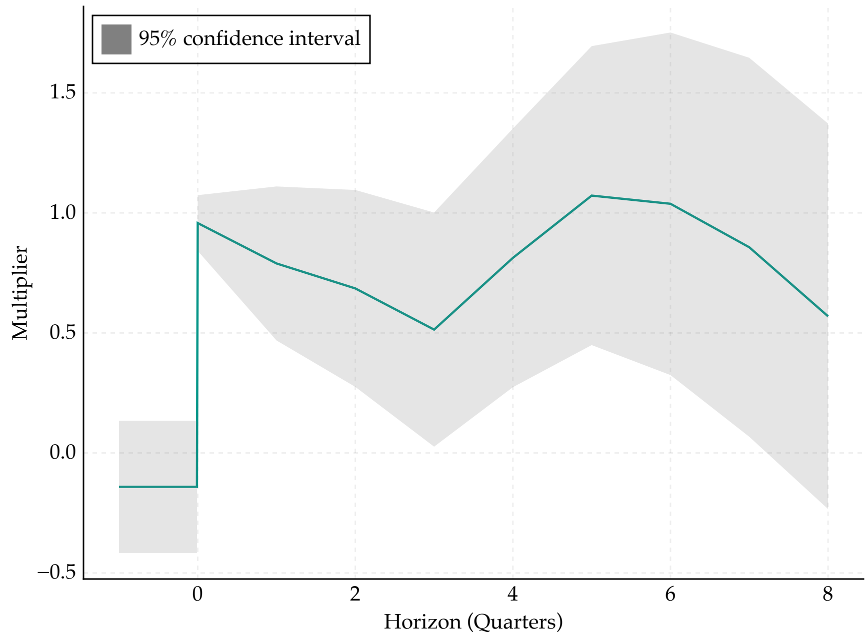
The price impact at longer horizons Like Gabaix and Koijen (2022) for the equity market, we find a persistent impact of demand and supply shocks in the Treasury market. To study how demand shocks propagate across time, we construct the aggregate idiosyncratic shocks estimated from the model $\hat{u}_{S,t} = \sum_i S_i \hat{u}_{i,t}$, and estimate the following local projections (Jordà 2005):

$$\Delta p_t^{t+h} = \alpha_h + M_h \hat{u}_{S,t} + \boldsymbol{\lambda}_h^p \boldsymbol{\eta}_t + e_{t+h}, \tag{3.4}$$

for $h = 0, 1, \ldots 8$ quarters, where Δp_t^{t+h} is the cumulative price changes from quarter t to $t+h$. We also estimate the local projection using the price change in the last quarter, Δp_{t-1} as a placebo test.

Figure 5 presents the cumulative price impact M_h of a one-percent demand shock, and the shades report the 95% confidence interval. The price change in the last quarter $(h = -1)$ does not predict the demand shock in this quarter, lending assurance that the demand shock is not priced in prior to the shock. On impact, the price goes up by 1%, consistent with our macro multiplier estimate. The price impact persists and remains significant for 8 quarters. Limited by statistical power, we are unable to reject the price impact to be zero after 8 quarters.

Figure 5: **Dynamic Price Impact of Demand and Supply Shocks**



Note: Figure 5 plots the impulse response function of Treasury price to demand and supply shocks. To get the impulse response function, we use the extracted size-weighted idiosyncratic demand and supply shocks $\hat{u}_{S,t}$ to estimate a local projection (Jordà 2005) following (3.4). Homoskedastic standard errors are used to construct the 95% confidence interval.

3.3 Price Decomposition

To understand the time-varying importance of sectors and factors in driving Treasury yield fluctuations, we use the equilibrium pricing equation (see (3.1)) to conduct a decomposition exercise. We decompose the Treasury yields in two ways—on a sector-by-sector or a factor-by-factor basis, keeping the market multiplier fixed. In the former case, a sector i's contribution to Treasury price changes are given by the sum of the fixed effect, the response to common factors, and the idiosyncratic shocks, scaled by the market multiplier:

$$\hat{p}_{i,t} = \hat{\zeta}_{S,r(t)}^{-1}(S_i\hat{\overline{f}}_i + S_i\hat{\boldsymbol{\lambda}}_{i,r(t)}\boldsymbol{\eta}_t + S_i\hat{u}_{i,t}).$$

Hats denote estimated values.[23] For the latter case, the k-th factor's contribution

[23]The idiosyncratic shock for each sector, $\hat{u}_{it}$, is backed out from the estimated linear system.

is given by the size-weighted sum of all sectors' sensitivity multiplied by the sector itself, and scaled by the aggregate elasticity:

$$\hat{p}_{\eta(k),t} = \hat{\zeta}_{S,r(t)}^{-1} \hat{\boldsymbol{\lambda}}_{S,r(t)} \eta_t(k),$$

Finally, the idiosyncratic shocks' total contribution equals $\hat{\zeta}_{S,r(t)}^{-1} \hat{u}_{S,t}$. Our estimates indicate that idiosyncratic shocks can explain 58% of the total variation of Treasury prices during our sample period.[24]

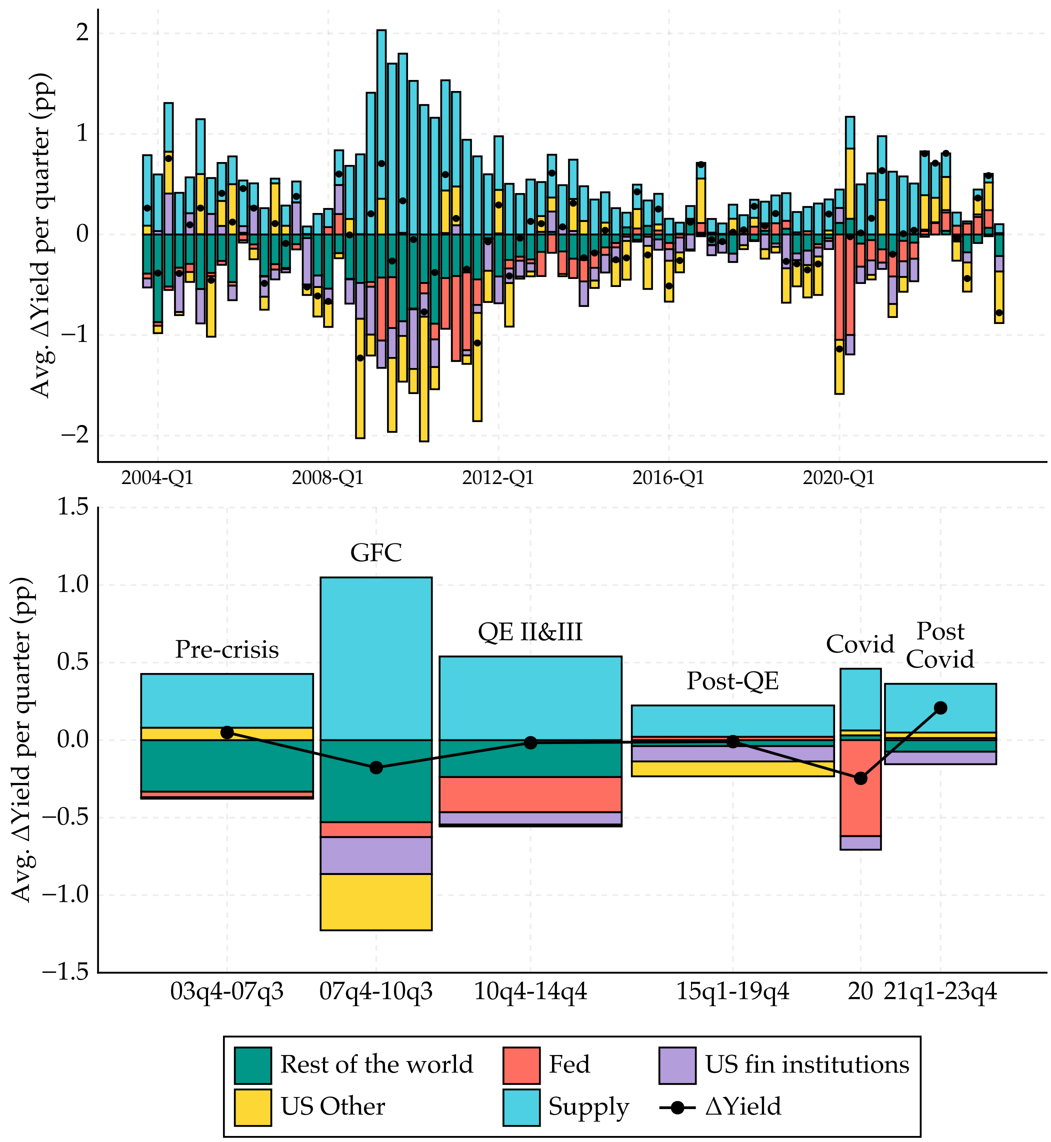
To better highlight the economics underlying the decomposition results, we group the sectors into several broad groups, singling out the influences of major U.S. financial institutions (including mutual funds, ETFs, insurance companies, pension funds and banks) as a whole, foreign investors, the Federal Reserve, and the issuer sector. We also create a group named "US Other" to include participants in the Treasury market difficult to categorize, in particular including the household sector. This group also includes non-financial sectors (non-financial companies, state and local governments), and financial firms with tiny holdings, such as bank holding companies, credit unions, GSEs, and ABS issuers.[25] Similarly, we group factors into several categories, highlighting the contribution of four factors: VIX changes, inflation, shifts in the U.S. dollar's strength, as well as the Federal Funds rate. Unobserved common demand factors, predictable Fed purchases and net issuance in the previous quarter are combined in a group labelled "common demand", to reflect the idea that these factors are either known to all market participants or capture hidden factor structures in investor demand across sectors. We scale price changes by weighted aver age duration of Treasury bonds and notes over the sample period (roughly 6 years) to arrive at yield changes in percentage point terms.

Figure 6 plots our decomposition by sector, from which we read off general patterns of time-varying, heterogeneous influences across sectors. In the upper panel, we show the yield decomposition quarter by quarter, and in the lower panel we organize them into 6 episodes for better interpretability by taking the quarterly

[24] The decomposition exercise can be understood in the spirit of Jiang, Richmond, and Zhang (2024). Each quarter, starting from zero price changes, we sequentially turn on the influence of each sector or factor and compute counterfactual market-clearing price changes. Our linear system comes with the benefit that we can read off the counterfactual prices analytically without resort to numerical calculations.

[25] Table C.1 report the detailed sector categorization.

Figure 6: **Contribution to Treasury yield fluctuations by investor and issuer groups**



Note: Figure 6 reports our main yield decomposition exercise by attributing Treasury yield movement during our sample period to each broadly-defined investor and issuer groups. The top panel reports the decomposition quarter by quarter, and the bottom panel groups the quarterly numbers to 6 distinct episodes. We aggregate investor sectors reported in Financial Account (Z.1) to four investor categories and one issuer sector ("supply"), the exact mapping available in Appendix C.1. Group-specific contribution is backed out from the pricing equation (3.1), assuming that the market multiplier is fixed at the estimated value $\hat{\zeta}_S^{-1}$. We report average yield contribution per quarter in percentage points and the width of the bars in Panel (b) reflects the length of the corresponding episode.

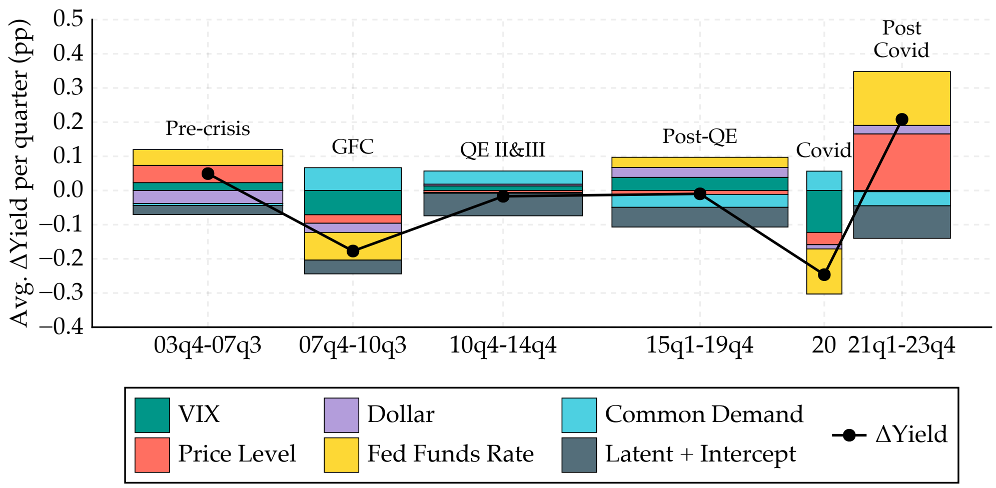
averages. Among all investor sectors, foreign investors push down Treasury yields by the most from 2003 to 2010, by an average of 33 basis points per quarter. Their contribution considerably narrows after 2010, a finding to be discussed extensively in Section 4.3. The Federal Reserve has started to participate in the Treasury market extensively since 2009. Their demand offsets 42% of the upward pressure on yield due to an expanded supply during the QE period. Meanwhile, major U.S. financial institutions, such as mutual funds and pension funds, play a relatively minor role in driving Treasury yields.

Two crisis episodes, the Global Financial Crisis (GFC) and the initial phase of the COVID-19 pandemic in 2020, offer a stark contrast. During the GFC and its aftermath (07Q4–10Q3), a variety of sectors pushed down Treasury yields. Foreign investors contribute 53 basis points per quarter, followed by U.S. household sector (36 basis points, labelled "US Other" in Figure 6) and major U.S. financial institutions (24 basis points). On the other hand, the Federal Reserve accounts for the majority of the downward yield pressure during 2020. On a quarterly basis, the Fed's purchase reduces yield by 62 basis points. The other sectors, used to be active during the GFC, account for only 2.5 basis point decline in yields, and among them, foreign investors and U.S. households marginally pushed up the yields.

In Figure 7, we plot the contribution of aggregate factors and idiosyncratic demand and supply shocks, also in percentage point terms in yearly averages. Consistent with VIX being a barometer of global risk sentiment, the negative impact of VIX on yields is more pronounced during crisis episodes. In particular, VIX surges lead to a 48 basis point decline in Treasury yields during 2020, and 84 basis points cumulatively during the financial crisis episodes (07Q4–10Q3). In Section 4.1, we further investigate in detail which sectors are fleeing to Treasury when the global risk sentiment is high.

Conventional monetary policy also exhibits substantial time variation in its influence on longer-term yields. Pre-crisis, consistent with the interest rate "conundrum" identified by Greenspan (2005), we estimate that despite the Fed's rapid rate hike of more than 4 percentage points over two years, the cumulative pass-through of the short rate on the yields on Treasury bonds and notes is 78 basis points. During the two crisis episodes, the Fed's interest rate cut lowers yields by 96 and 52 basis points respectively on a cumulative basis. The recent tightening cycle that began in early 2022 raises Treasury yields by a total of 1.8 percentage

Figure 7: **Macro and idiosyncratic drivers of Treasury yields**



Note: Figure 7 reports our yield decomposition exercise attributing yield movements to common, macro factors, idiosyncratic shocks, and group-specific fixed effects. Factor-specific contribution is backed out from the pricing equation (3.1), assuming that the market multiplier is fixed at the estimated value $\hat{\zeta}_S^{-1}$. We report average yield contribution per quarter in percentage points and the width of the bars in Panel (b) reflects the length of the corresponding episode. The bars named "common demand" nest the influence of policy factors (announced Fed purchase and lagged net issuance), and the impact of three principal components extracted from granular Treasury holding data of U.S. banks and foreign investors. The width of the bars reflects the length of the corresponding episode.

points.[26]

Hanson and Stein (2015) argue that the effect of monetary policy on long-term rates are at odds with models based on sticky prices, and propose an alternative channel through demand pressures induced by reach-for-yield motives. Our model framework allows us to investigate whose demand pressure passes through the short rate to the medium-to-long term yields. Table C.5 in the Appendix re-

[26]The decomposition exercise involving macro factors is interpreted as in terms of the deviations of the aggregate factors from a neutral benchmark level. Using different benchmarks would quantitatively affect the portion of contribution assigned to the macro factors versus the intercept (which are included together with the sector-specific idiosyncratic component in this exercise). We use economics to guide our selection of the benchmark. For most factors we use sample means as the benchmark, except for the Fed funds rate and the CPI. For the former, we use first differences and use zero as the natural benchmark. For the latter, we use the inflation target—2 percent per annum—as the benchmark. Hence, the intercept is interpreted as the average price pressure from all sectors under the scenario of a constant Fed funds rate, 2% CPI inflation, and average values in other macro factors.

ports the full coefficients of factor loadings by sectors. Almost all sectors negatively load on the Fed funds rate, suggesting that when the short rate decreases and term premia increase, these sectors' demand for Treasury notes and bonds increase. Notably, for a 1% Fed funds rate increase, banks' demand for Treasury notes and bonds decrease significantly by 3.63% ($p < 0.05$). This is also consistent with the literature on the deposit channel of the monetary policy (Drechsler, Savov, and Schnabl 2017): deposits in the banking system are highly sensitive to the short-term interest rate. As the Fed Funds rate rises, banks widen the spreads they charge on the deposits and deposits flow out of the banking system. Correspondingly, they reduce holdings in Treasuries to meet thedeposit outflow.

Consistent with the asset-liability management practices of insurance companies and pension funds (ICPFs), we find their demand for long-term bonds increases when the Fed funds rate rises. ICPF liabilities have a particularly long duration, which exposes them to interest rate risks. To hedge this risk, ICPFs purchase long-maturity bonds. However, available bonds typically have a shorter duration than their liabilities. Hence, ICPFs rely on dynamic trading strategies, which require them to take a leveraged position in bonds to synthetically match the duration of liabilities (Domanski, Shin, and Sushko 2017; Li 2024). When Fed funds rise, holding long-term yields fixed, this causes the value of assets to fall more in value than liabilities, requiring ICPFs to purchase more long-term bonds to ensure they are still duration-matched, resulting in a positive loading of the Fed funds rate.

The force of inflation has been dormant in the Treasury market in most of the sample periods we study. Nevertheless, after the COVID-19 pandemic, fiscal and monetary stimulus around the globe and supply-chain pressures pushed the inflation to a level that has not been seen since the Great Inflation in 1980s. Our estimates indicate that inflation alone pushed up the yields by 1.6 percentage points in the three years since 2021. In terms of the sectoral contribution, the household sector is by far the most sensitive sector to inflation. 1 percentage point higher inflation leads to the household sector reducing Treasury holdings by 16%. This finding corroborates empirical evidence in Nagel and Yan (2022), who show that households pay attention to news on rising inflation and respond by flowing out of nominal Treasury-related investments.[27]

[27]The Financial Accounts data do not distinguish between nominal Treasury securities and

4 Understanding Important Drivers of Treasury Yields

We use our estimated price elasticities, factor loadings, and decomposition to shed light on a number of macro-financial phenomena related to the U.S. Treasury market. We show that foreign investors are not the main contributor to Treasury yield compression when global risk sentiment worsens. While foreign investors exert considerable downward pressure on Treasury yields before 2010, both their market share and their role in driving Treasury yields weakens afterwards. Important private-sector holders of U.S. Treasuries, such as foreign investors and U.S. banks, become more inelastic after the financial crisis. In place of those sectors, the Federal Reserve started to play the role of a state-contingent liquidity provider that supports the market especially during risk-off episodes such as the COVID-19 pandemic.

4.1 Foreign Investors and "Flight to Safety"

A salient feature of U.S. Treasuries in this century is their tendency to appreciate during downturns. This countercyclicality makes Treasuries particularly attractive in an investor's portfolio as they act as a hedge against the market (Campbell, Pflueger, and Viceira 2019). In particular, a significant proportion of this hedging property arises from the "convenience yield," a component generally attributed to the special demand for U.S. Treasuries (Acharya and Laarits 2024).

Recent literature makes significant progress in resolving a number of important puzzles in international finance by resorting to the specialness of U.S. debt during global downturns. State-of-the-art models posit a countercyclical demand for U.S. Treasuries, as the safety and liquidity of U.S. government debt become particularly appealing when global risk sentiment worsens (Jiang, Krishnamurthy, and Lustig 2021; Kekre and Lenel 2024, among others). Typically in these analyses, foreign investors are considered to be the key source of the countercyclical demand. At the onset of the COVID-19 pandemic in March 2020, the sell-off of Treasuries by foreign investors triggered discussions in the academic literature as

inflation-indexed Treasury bonds (TIPS), limiting our analysis on inflation hedging behaviors largely at a high level. However, the TIPS market represents less than 10% of all marketable debt issued by the Treasury as of 2024 and was even smaller in the earlier sample. Hence, we mainly attribute household flows as related to nominal bonds and notes.

well as among policymakers on whether the special status of US Treasuries has waned (He, Nagel, and Song 2022; Weiss 2022; Vissing-Jorgensen 2021). However, perhaps surprisingly, the empirical relationship between foreign Treasury demand and global risk is in fact not well-established in the literature.

Through the lens of our estimated model, we are able to assess to what extent foreign investors drive "flight to safety". We examine the estimated flow loadings on quarterly changes in the VIX index for foreign investors – in the data, the negative comovement between Treasury yields and the VIX index is strong, making the VIX a natural barometer of investors' propensity to seek safe investments in risky times.

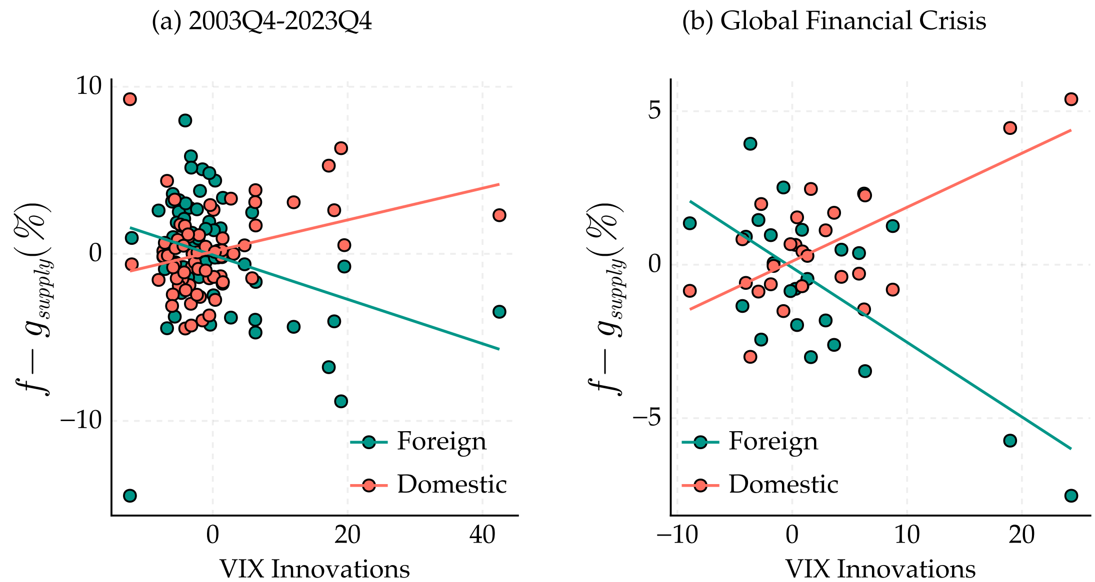
Table 2: **Sensitivities to global risk sentiments for selected sectors**

$Sector$	$S(\%)$	$\epsilon^{VIX}_{(std.)}$	$\epsilon^{VIX}_{(std.)}$ **Share (%)**
Aggregate (09-)		0.87	100.0
		(0.17, 1.57)	
Households	5.74	15.46	101.93
		(6.14, 24.79)	
Fed (09-)	22.08	2.18	55.22
		(1.38, 2.97)	
Rest of World	44.45	-0.33	-16.63
		(-0.98, 0.33)	
Supply	100.0	-0.12	-13.65
		(-0.33, 0.09)	
Mutual Funds	6.75	-2.06	-15.98
		(-3.59, -0.53)	

Note: Table 2 reports the estimated responsiveness to a rise in the VIX index for a seleted set of sectors in the data. The sample period is 2003Q4–2023Q4. The loadings are identified using the optimal GIV estimator developed in Section 2. S corresponds to the size weight used in the estimation, defined as the average market share of each entity throughout the sample period. $\varepsilon^{VIX}_{(std.)}$, the regressor of interest, is obtained by running a AR(1) time-series regression on the VIX index and taking the residual. $\varepsilon^{VIX}_{(std.)}$ share denotes the (size-weighted) fraction of aggregated VIX loading accounted for by each sector. 95% confidence intervals are reported, with the standard errors given by Theorem 2.

Table 2 reports the size-weighted aggregate VIX loading as well as sector-specific loadings for a selected set of investors. The loadings for the entire set of investor groups are reported in Table C.5 in the Appendix. Overall, one standard deviation increase in the VIX index leads to a 0.87 percent inflow into the Treasury market. The household sector displays the strongest "flight-to-safety" behavior, increasing demand for Treasuries by 15.5 percent. For foreign investors, we find no evidence that their sensitivity to VIX is significantly different from zero. If anything, the point estimate is negative, suggesting that foreign investors on average demand less U.S. Treasuries during risk-off episodes.

Figure 8: **Foreign flows (net of issuance) and global risk indicators**



Note: Figure 8 illustrates the importance of accounting for equilibrium interactions among investors and issuers to properly estimate the sensitivity of Treasury demand to shift in global risk sentiments, proxied by VIX innovations. In both panels, we plot foreign (green) and U.S. domestic (red) inflow to Treasury notes and bonds f (% of total holdings in the previous quarter), net of total issuance g (also expressed as % of total market value of outstanding issuance the previous quarter). Panel (a) plots all observations throughout the sample period. Panel (b) focuses on the Global Financial Crisis episode at the monthly frequency (2007/12-2009/12).

Our finding of a negative loading of foreign Treasury demand on VIX contrasts with the standard approach of modeling foreign investors actively redirecting asset demand towards U.S. safe assets during downturns. In fact, the clues for this result can be found using a more reduced-form approach. In Figure 8, we plot foreign inflows *net of Treasury issuance* (normalized by market size) over the entire sample period (panel (a)) against domestic flows, and find a sharp difference: On average when VIX is high, foreign investors tend to buy less, or sell more, Treasuries compared to domestic investors.. This pattern is also not a recent development since the pandemic, but is present at the monthly frequency during the Global Financial Crisis. We note that simple correlation or direct regression coefficient between flow and VIX does not capture an investor's true responsiveness of Treasury demand to rising global risk sentiment but only the relative tendency (see more discussion in Section 2). Only through an equilibrium model are we able to disentangle the endogenous demand response from the observed equilibrium holdings. Hence, this finding underscores the importance of accounting for endogenous demand and supply interaction when discussing phenomena related to international capital flows.

4.2 The Transformation of Liquidity Provision

Through the lens of our estimates, we discuss the liquidity provider role of Treasury market participants from two perspectives. First, investors' price elasticities of demand reflect their capacity to stand on the other side of the market when other investors' negative demand shocks and surprise new issuance of U.S. Treasury put downward pressure on Treasury prices. Second, liquidity provision could be state contingent. When aggregate conditions worsen and triggers selling pressure, sectors whose demand increases with adverse macro shocks would be in a better position to supply liquidity to the market and absorb the resulting imbalances.

The flexibility of our framework allows us to inspect the regime shifts in price elasticities of demand of important participants in the market. We report these elasticities for foreign investors and U.S. banks in Table 3. For foreigners and U.S. banks, we observe a 56% and 61% respective decline in price elasticities in the post-crisis era relative to the 2003–2008 period.

The decline in price elasticity for U.S. banks captures several important aspects of banking industry development post-crisis that tighten their intermediation capacity (see Du, Tepper, and Verdelhan 2018, for more discussion on the implication for other markets). More specifically, non-risk-weighted regulatory constraints, such as supplementary leverage ratio requirements, make it more costly for banks to expand balance sheets for Treasury-related intermediation activities (Favara, Infante, and Rezende 2024). Tighter risk-based regulations could strengthen the incentive for banks to substitute loans for liquid Treasury holdings (Stulz, Taboada, and van Dijk 2022). In addition, during the post-COVID monetary tightening cycle, the share of securities in banks' portfolio designated as held to maturity substantially increases, as banks exploit accounting rules to insulate book capital from the rising interest rate (Granja et al. 2024). All these forces could act to reduce banks' sensitivity to price movements.

On the other hand, the Federal Reserve has acted as an state-contingent liquidity provider post-crisis. Table 4 reports the estimates for the Fed since 2009. In these estimates, we control for anticipated purchases by the Fed as a macro factor, and all the estimates reflect the deviation of the Fed's purchases from the anticipated path.[28] The Fed is price elastic: they tend to initiate QE operations when Treasury market faces downward demand pressures. This is consistent with the objective and framework of the QE operation (Gagnon et al. 2011b), which also seeks to stabilize the Treasury market. Indeed, the language "the smooth functioning of the Treasury market" appears frequently in the FOMC minutes.[29] The Fed's QE policies also actively respond to the market distress, captured by the VIX index. We estimate that the swift response of the Fed to intervene in the Treasury market at the height of the COVID-19 crisis led to a 2 percentage point decline in the yield of Treasury bonds and notes, offsetting the upward pressure to yields mostly driven by supply expansion.[30] Finally, we also see a strong coordination of the coordination between the direct purchases and traditional monetary policies. When the Fed raises the Fed funds rate, or when inflation rises, it also simultaneously reduces its purchases.

[28]The Fed's purchases are typically announced in the FOMC meetings and pre-scheduled. In Appendix C.2, we discuss in detail the construction of the anticipated purchases, and show the deviations occured mostly in the initial and concluding phase of each QE operation in which the Fed has ample policy space to respond to contemporaneous macroeconomic and financial factors including yields.

[29]For example, the announcement of Treasury securities operations in 2020 by the New York Fed specifically stated "The Desk stands ready to adjust the size and composition of its purchase operations as appropriate to support the smooth functioning of the Treasury market". In another announcement around the same time, they explicitly mentioned that "the Desk will conduct purchase...subject to reasonable prices."

[30]One way that the state-contingency purchases further reduce Treasury yields is through the insurance channel proposed by Haddad, Moreira, and Muir (2024): supporting bond prices in bad states provides additional safety and lowers bond risk premia. Haddad, Moreira, and Muir (2024) finds that the insurance channel can account for 75% of QE's effect on yields.

Table 3: **Regime shifts for price elasticities of U.S. banks and foreigner, before and after 2009**

$Sector$	ζ	ζ **Share (%)**
Rest of World (03-08)	0.72	29.88
	(0.45, 1.0)	
Rest of World (09-23)	0.32	14.4
	(0.17, 0.47)	
U.S. Banks (03-08)	1.21	5.94
	(0.85, 1.58)	
U.S. Banks (09-23)	0.47	2.52
	(0.27, 0.67)	

Note: Table 3 reports the estimated price elasticities of demand for U.S. banks and foreign investors before and after the assumed regime shift date of 2009Q1. The model is estimated using data from 2003Q4 to 2023Q4. The price elasticities of demand are identified using the optimal GIV estimator developed in Section 2. 95% confidence intervals are reported, with the standard errors given by Theorem 2.

Table 4: **Treasury Demand of the Federal Reserve**

$Period$	$S(\%)$	ζ	$\epsilon^{VIX}_{(std.)}$	Δ **FFR**	$Inf.$
09-23	22.08	0.59	2.18	-5.43	-0.81
		(0.32, 0.87)	(1.38, 2.97)	(-7.42, -3.45)	(-1.96, 0.34)

Note: Table 4 reports a select set of estimated parameters related to the Federal Reserve's demand curve for U.S. Treasury notes and bonds. ζ refers to the price elasticity of demand. $\varepsilon^{VIX}_{(std.)}$ refers to the demand loading on a positive VIX shock, obtained by taking the residual from an AR(1) time-series regression of the VIX index. ΔFFR refers to quarterly changes in the Federal Funds rate, and $Inf.$ denotes the deviation of CPI inflation from the 2% per annum inflation target. 95% confidence intervals are reported, with the standard errors given by Theorem 2.

4.3 Foreign Demand and Treasury Yields: An Evolving Relationship

In the first decade of the 2000s, nearly 3 trillion USD was channelled from a diverse set of foreign savers to U.S. Treasury bonds and notes. High-growth emerging markets with financial market imperfections, such as China and commodity countries, build foreign reserves mostly comprised of U.S. Treasuries. The persistent current account surplus ran by developed markets such as continental Europe and Japan served as another important source of global savings.[31] Foreign investors' potential impact on long-term U.S. interest rates received substantial academic and policy debate prior to the financial crisis, going beyond understanding the borrowing cost of the U.S. government. Greenspan (2005) observe that despite consecutive Fed interest rate hikes in 2004, long-term interest rates respond little. This "interest rate conundrum" is most evident in 2006, when the effective Fed Funds rate exceeds 10-year Treasury yield by 80 basis points. Relatedly, Bernanke (2005) raised concern on financial stability risk from a low long-term interest rate. Our framework allows us to precisely attribute the contribution to Treasury yield movements by foreign investors to more granular sources and to understand their time-varying importance, and to shed light on the potential short-run interest rate impact of Treasury selloff by foreign investors.[32]

Figure 9 reports the result of our yield decomposition into components driven by major foreign investor countries and regions. Pre-crisis, China, Europe and Japan jointly account for 61% of foreign investors' yield impact, lowering Treasury yield by 0.74 percentage points per year. Of these major holders, China's foreign reserve accumulation is the most important driver of Treasury yield compression in the pre-crisis period. Over the course of four years, China's demand for U.S. Treasury brought down yields by more than 1.7 percentage points on a cumulative basis. Other major investors such as Japan and Europe contributed less. China remained the biggest mover of Treasury yields during and immediately after the financial crisis, accounting for 38% of the total impact of foreign demand. During these two episodes, Japan and Europe also extend their influence over Treasury yield. European investors' contribution amounts to 9 basis points per quarter, in a period coinciding with Euro Area's sovereign debt market turmoil and the binding

[31]For additional discussion on this topic, see Caballero (2006) and Mendoza, Quadrini, and Ríos-Rull (2009).

[32]Jiang, Richmond, and Zhang (2024) use a global demand system to examine the role of global savings in accounting for the evolution of U.S. net foreign asset positions. While we share similar decomposition methodology, our focus is on the Treasury price directly.

of the zero lower bound for U.S. nominal interest rates.[33]

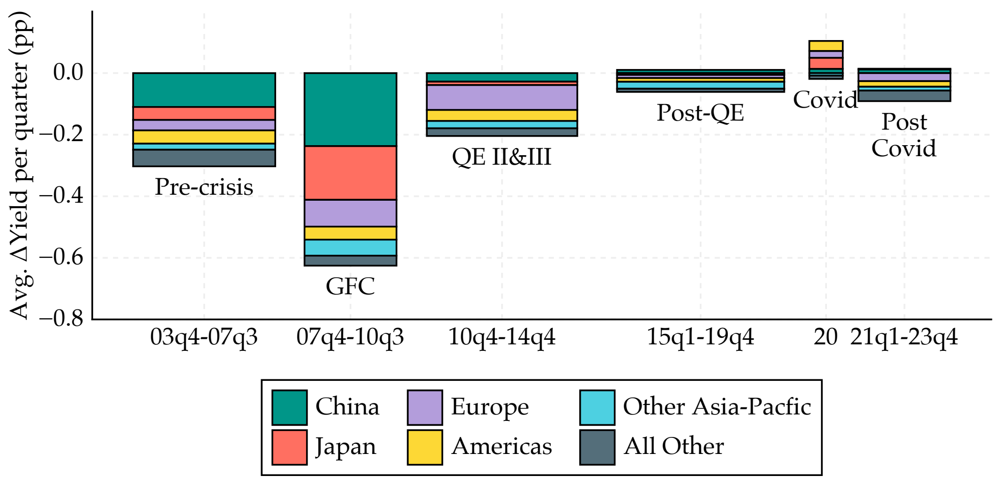
In stark contrast to the pre-crisis episodes, foreign investors' role in driving Treasury yields has substantially weakened since 2011. On a yearly basis, foreign demand pushes down Treasury yields by an average of 40 basis points, mostly driven by European investors in the aftermath of the debt crisis. On the other hand, during the initial phase of the COVID-19 pandemic, foreign investors' selling increased Treasury yields by 34 basis points. Du, Im, and Schreger (2018) document that the negative gap between long-term Treasury yield and currency-hedged G10 government yield narrowed after the financial crisis. Our estimates provide a quantity-based perspective to complement this finding, by showing that the compression of U.S. Treasury yield by foreign investors is concentrated in the first decade of the 2000s.

The weakened role of foreign investors in driving Treasury yields is consistent with previous studies documenting a smaller price impact of foreign official flows in and out of U.S. Treasuries after the financial crisis (Beltran et al. 2013). This finding also suggests that the degree of financial market disruption arising from heightening geopolitical tension could be limited. As the total supply of Treasury bonds and notes substantially expands after the financial crisis, major foreign investors account for a smaller size of the total Treasury market. Figure C.5, panel (b) in the Appendix, for example, shows that China's market share has declined from the 2010 peak of 17.5 percent to 4 percent at the end of 2023. This is driven by a faster pace of Treasury issuance while a constant position held by China before 2020, and the steady outflow of China from Treasuries since 2020. Our estimated market multiplier of 0.99 would imply that a complete, surprise sell-off of China's U.S. Treasury bonds and notes within a quarter by the end of 2023 would have only pushed up Treasury yields by 50 basis points, assuming China holds the market portfolio. To put the size of the price impact into perspective, we notice that the standard deviation of quarterly Treasury yield changes from 2003Q4 to 2023Q4 is 45 basis points. As a result, the negative price impact induced by an outflow the

[33] The concern over exchange rates could be an additional source of Treasury demand for China and Japan, both of whom actively participate in currency intervention. For both countries, the rapid Fed Funds rate cut to zero during the Global Financial Crisis exerted currency appreciation pressure. For China after 2014, the slow unwinding of Quantitative Easing led to depreciation pressure. China responded in 2015 by sharply selling its foreign reserves. Nevertheless, consistent with the idea that the sales constitutes a smaller relative demand shock compared to pre-crisis situation, the price impact is small.

size of China's entire Treasury notes and bonds holdings sits well in the normal range of quarterly yield fluctuations.

Figure 9: **Important foreign holders' contributions to yields, percentage points per year**



Note: Figure 9 decomposes the contribution of foreign investors to Treasury yield fluctuations further to the level of large countries and regions. We report average yield contribution per quarter in percentage points. The width of the bars reflects the length of the corresponding episode.

5 Conclusion

This paper introduces a novel empirical framework to understanding how demand and supply forces drive asset price fluctuations. Methodologically, our estimation approach helps flexibly and efficiently identify heterogeneous price and factor elasticities and precisely quantify the contribution of various market participants to asset price movements. Our optimal GIV estimator based on granularity can be easily applied to other markets, and will be particularly powerful in the case of concentrated markets where demand shocks play a larger role in driving market prices.

When applied to the U.S. Treasury market, our method offers novel answers to a series of important questions. We provide the first direct macro elasticity estimate of the market. We demonstrate significant structural shifts in liquidity provision marked by the global financial crisis, featuring the retreat of foreign investors and U.S. banks and the rising importance of the Federal Reserve. We also show that the "flight-to-Treasury" behavior of foreign investors during global downturns – taken as a fact in standard narratives – is not supported in the data. While our model is intended to provide a high-level characterization of the market dynamics, our estimate serves as a useful benchmark to guide future research to delve deeper into the underlying mechanisms.

References

Acharya, Viral V and Toomas Laarits (2024). "When Do Treasuries Earn the Convenience Yield? — A Hedging Perspective".

Adrian, Tobias et al. (2022). "The Term Structure of Growth-at-Risk". In: *American Economic Journal: Macroeconomics* 14.3, pp. 283–323. URL: https://www.aeaweb.org/articles?id=10.1257/mac.20180428.

An, Yu and Amy W Huber (2024). "Intermediary Elasticity and Limited Risk-Bearing Capacity". In.

Barone, Jordan et al. (Mar. 2022). *The Global Dash for Cash: Why Sovereign Bond Market Functioning Varied across Jurisdictions in March 2020*. Staff Reports 1010. Federal Reserve Bank of New York. URL: https://ideas.repec.org/p/fip/fednsr/93852.html.

Baumeister, Christiane and James D. Hamilton (2023). *A Full-Information Approach to Granular Instrumental Variables*. Working paper, UCSD. URL: https://apps.unive.it/server/eventi/78661/BH8.pdf.

Beltran, Daniel O. et al. (Feb. 1, 2013). "Foreign Holdings of U.S. Treasuries and U.S. Treasury Yields". In: *Journal of International Money and Finance* 32, pp. 1120–1143. ISSN: 0261-5606. DOI: 10.1016/j.jimonfin.2012.09.005.

Bernanke, Ben S. (2005). "The Global Saving Glut and the U.S. Current Account Deficit".

Bertaut, Carol C. and Ruth Judson (2022). *Estimating U.S. Cross-Border Securities Flows: Ten Years of the TIC SLT*. Board of Governors of the Federal Reserve System (U.S.) URL: https://ideas.repec.org/p/fip/fedgfn/2022-02-18-2.html.

Bertaut, Carol C. and Ruth A. Judson (2014). *Estimating U.S. Cross-Border Securities Positions: New Data and New Methods*. 1113. Board of Governors of the Federal Reserve System (U.S.) URL: https://ideas.repec.org/p/fip/fedgif/1113.html.

Bertaut, Carol C. and Ralph W. Tryon (2007). *Monthly Estimates of U.S. Cross-Border Securities Positions*. 910. Board of Governors of the Federal Reserve System (U.S.) URL: https://ideas.repec.org/p/fip/fedgif/910.html.

Caballero, Ricardo J. (Dec. 2006). *On the Macroeconomics of Asset Shortages*. NBER Working Papers 12753. National Bureau of Economic Research, Inc. URL: https://ideas.repec.org/p/nbr/nberwo/12753.html.

Camanho, Nelson, Harald Hau, and Hélène Rey (2022). *Global Portfolio Rebalancing and Exchange Rates*. Working Paper 24320. National Bureau of Economic Research. DOI: 10.3386/w24320.

Campbell, John Y., Carolin Pflueger, and Luis M. Viceira (Dec. 18, 2019). "Macroeconomic Drivers of Bond and Equity Risks". In: *Journal of Political Economy* 128.8, pp. 3148–3185. ISSN: 0022-3808. DOI: 10.1086/707766.

Cassidy, William and Shreye Mirani (2024). "The Debt Ceiling's Disruptive Impact: Evidence from Many Markets". In: URL: https://williamcassidy.com/files/CassidyMiraniApril112024.pdf.

Cavaleri, Filippo (June 2023). "The Convenience Yield and the Demand for U.S. Treasury Securities". In.

Chamberlain, Gary (Mar. 1, 1987). "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions". In: *Journal of Econometrics* 34.3, pp. 305–334. ISSN: 0304-4076. DOI: 10.1016/0304-4076(87)90015-7.

Chaudhary, Manav (Oct. 13, 2024). *Regulator Beliefs*. DOI: 10.2139/ssrn.4986079. Social Science Research Network: 4986079. Pre-published.

Chaudhary, Manav, Zhiyu Fu, and Jian Li (Sept. 2023). "Corporate Bond Multipliers: Substitutes Matter".

Chodorow-Reich, Gabriel et al. (July 5, 2024). *Propagation of Shocks in Networks: Identification and Applications*. DOI: 10.2139/ssrn.4885984. Pre-published.

Da, Zhi et al. (Feb. 2018). "Destabilizing Financial Advice: Evidence from Pension Fund Reallocations". In: *The Review of Financial Studies* 31.10, pp. 3720–3755. ISSN: 0893-9454. DOI: 10.1093/rfs/hhy011. eprint: https://academic.oup.com/rfs/article-pdf/31/10/3720/25693037/hhy011.pdf.

Domanski, Dietrich, Hyun Song Shin, and Vladyslav Sushko (Apr. 2017). "The Hunt for Duration: Not Waving but Drowning?" In: *IMF Economic Review* 65.1, pp. 113–153. ISSN: 2041-4161, 2041-417X. DOI: 10.1057/s41308-016-0026-9.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl (Nov. 1, 2017). "The Deposits Channel of Monetary Policy". In: *The Quarterly Journal of Economics* 132.4, pp. 1819–1876. ISSN: 0033-5533. DOI: 10.1093/qje/qjx019.

Du, Wenxin, Benjamin Hébert, and Wenhao Li (2022). "Intermediary Balance Sheets and the Treasury Yield Curve". In.

Du, Wenxin, Joanne Im, and Jesse Schreger (2018). "The U.S. Treasury Premium". In: *Journal of International Economics* 112, pp. 167–181. ISSN: 0022-1996. DOI: 10.1016/j.jinteco.2018.01.001.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan (2018). "Deviations from Covered Interest Rate Parity". In: *The Journal of Finance* 73.3, pp. 915–957. ISSN: 1540-6261. DOI: 10.1111/jofi.12620. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12620.

Eren, Egemen, Andreas Schrimpf, and Fan Dora Xia (2023). "The Demand for Government Debt". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.4466154.

Fang, Xiang, Bryan Hardy, and Karen K. Lewis (May 2022). *Who Holds Sovereign Debt and Why It Matters*. DOI: 10.3386/w30087. National Bureau of Economic Research: 30087. Pre-published.

Favara, Giovanni, Sebastian Infante, and Marcelo Rezende (2024). "Leverage Regulations and Treasury Market Participation: Evidence from Credit Line Drawdowns".

Gabaix, Xavier and Ralph S. J. Koijen (May 12, 2022). *In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis*. DOI: 10.2139/ssrn.3686935. Pre-published.

— (July 2024). "Granular Instrumental Variables". In: *Journal of Political Economy* 132.7, pp. 2274–2303. ISSN: 0022-3808. DOI: 10.1086/728743.

Gagnon, Joseph et al. (May 2011a). "Large-Scale Asset Purchases by the Federal Reserve: Did They Work?" In: *FRBNY Economic Policy Review*, p. 43.

— (Mar. 2011b). "The Financial Market Effects of the Federal Reserve's Large-Scale Asset Purchases". In: *International Journal of Central Banking* 7.1, pp. 3–43. URL: https://ideas.repec.org/a/ijc/ijcjou/y2011q1a1.html.

Granja, João et al. (Mar. 2024). *Book Value Risk Management of Banks: Limited Hedging, HTM Accounting, and Rising Interest Rates*. DOI: 10.3386/w32293. National Bureau of Economic Research: 32293. Pre-published.

Greenspan, Alan (Feb. 2005). "Testimony".

Greenwood, Robin and Dimitri Vayanos (May 1, 2010). "Price Pressure in the Government Bond Market". In: *American Economic Review* 100.2, pp. 585–590. ISSN: 0002-8282. DOI: 10.1257/aer.100.2.585.

— (Mar. 1, 2014). "Bond Supply and Excess Bond Returns". In: *The Review of Financial Studies* 27.3, pp. 663–713. ISSN: 0893-9454. DOI: 10.1093/rfs/hht133.

Haddad, Valentin, Paul Huebner, and Erik Loualiche (2024). "How Competitive Is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing".

Haddad, Valentin, Alan Moreira, and Tyler Muir (2024). "Asset Purchase Rules: How QE Transformed the Bond Market".

Han, Chirok and Peter C. B. Phillips (2006). "GMM with Many Moment Conditions". In: *Econometrica* 74.1, pp. 147–192. ISSN: 1468-0262. DOI: 10.1111/j.1468-0262.2006.00652.x.

Hanson, Samuel G. and Jeremy C. Stein (Mar. 2015). "Monetary Policy and Long-Term Real Rates". In: *Journal of Financial Economics* 115.3, pp. 429–448. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2014.11.001.

Hartzmark, Samuel M and David H Solomon (Nov. 2022). *Predictable Price Pressure*. Working Paper 30688. National Bureau of Economic Research. DOI: 10.3386/w30688.

He, Zhiguo, Stefan Nagel, and Zhaogang Song (Jan. 1, 2022). "Treasury Inconvenience Yields during the COVID-19 Crisis". In: *Journal of Financial Economics* 143.1, pp. 57–79. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2021.05.044.

Huang, Shiyang et al. (2020). "Does Liquidity Management Induce Fragility in Treasury Prices: Evidence from Bond Mutual Funds". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3689674.

Jansen, Kristy A. E. (Aug. 8, 2021). *Long-Term Investors, Demand Shifts, and Yields*. DOI: 10.2139/ssrn.3901466. Pre-published.

Jansen, Kristy A. E., Wenhao Li, and Lukas Schmid (2024). *Granular Treasury Demand with Arbitrageurs*.

Jiang, Erica Xuewei, Gregor Matvos, et al. (2024). "Monetary Tightening and US Bank Fragility in 2023: Mark-to-market Losses and Uninsured Depositor Runs?" In: *Journal of Financial Economics* 159, p. 103899. URL: https://www.sciencedirect.com/science/article/pii/S0304405X24001223?casa_token=5ZMrs79EHAcAAAAA:4vtqP5VumpjyLnkw66qcAkis0cKo94XyDFrhqAbxmUHnjxwnRmzs1S-32PQwTn67WQLOeJ7NrA.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno Lustig (2021). "Foreign Safe Asset Demand and the Dollar Exchange Rate". In: *The Journal of Finance* 76.3, pp. 1049–1089. ISSN: 1540-6261. DOI: 10.1111/jofi.13003. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13003.

Jiang, Zhengyang, Robert Richmond, and Tony Zhang (Sept. 5, 2022). *Understanding the Strength of the Dollar*. DOI: 10.2139/ssrn.4210797. Pre-published.

— (2024). "A Portfolio Approach to Global Imbalances". In: *The Journal of Finance* 79.3, pp. 2025–2076. DOI: 10.1111/jofi.13333. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13333.

Jordà, Òscar (Mar. 2005). "Estimation and Inference of Impulse Responses by Local Projections". In: *American Economic Review* 95.1, pp. 161–182. ISSN: 0002-8282. DOI: 10.1257/0002828053828518.

Kekre, Rohan and Moritz Lenel (June 2024). "The Flight to Safety and International Risk Sharing". In: *American Economic Review* 114.6, pp. 1650–1691. ISSN: 1944-7981. DOI: 10.1257/aer.20211319.

Koijen, Ralph S. J., François Koulischer, et al. (May 1, 2017). "Euro-Area Quantitative Easing and Portfolio Rebalancing". In: *American Economic Review* 107.5, pp. 621–627. ISSN: 0002-8282. DOI: 10.1257/aer.p20171037.

Koijen, Ralph S. J. and Motohiro Yogo (2019). "A Demand System Approach to Asset Pricing". In: *Journal of Political Economy* 127.4, pp. 1475–1515. DOI: 10.1086/701683. eprint: https://doi.org/10.1086/701683.

— (Mar. 2023). "Understanding the Ownership Structure of Corporate Bonds". In: *American Economic Review: Insights* 5.1, pp. 73–92. ISSN: 2640-205X. DOI: 10.1257/aeri.20210550.

— (Nov. 4, 2024). *Exchange Rates and Asset Prices in a Global Demand System*. DOI: 10.2139/ssrn.3383677. Social Science Research Network: 3383677. Pre-published.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (Oct. 2011). *The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy*. Working Paper 17555. National Bureau of Economic Research. DOI: 10.3386/w17555.

— (2012). "The Aggregate Demand for Treasury Debt". In: *Journal of Political Economy* 120.2, pp. 233–267. ISSN: 0022-3808. DOI: 10.1086/666526. JSTOR: 10.1086/666526.

Kundu, Shohini and Nishant Vats (Dec. 1, 2021). *Banking Networks and Economic Growth: From Idiosyncratic Shocks to Aggregate Fluctuations*. DOI: 10.2139/ssrn.3975942. Social Science Research Network: 3975942. Pre-published.

Li, Jennifer, Neil D Pearson, and Qi Zhang (2024). "Impact of Demand Shocks on the Stock Market: Evidence from Chinese Ipos".

Li, Ziang (2024). "Long Rates, Life Insurers, and Credit Spreads". In: *Life Insurers, and Credit Spreads (May 25, 2024)*. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4465485.

Lou, Dong (Dec. 2012). "A Flow-Based Explanation for Return Predictability". In: *The Review of Financial Studies* 25.12, pp. 3457–3489. ISSN: 0893-9454. DOI: 10.1093/rfs/hhs103. eprint: https://academic.oup.com/rfs/article-pdf/25/12/3457/24436896/hhs103.pdf.

Mendoza, Enrique G., Vincenzo Quadrini, and José-Víctor Ríos-Rull (June 2009). "Financial Integration, Financial Development, and Global Imbalances". In: *Journal of Political Economy* 117.3, pp. 371–416. ISSN: 1537-534X. DOI: 10.1086/599706.

Nagel, Stefan and Zhen Yan (2022). *Inflation Hedging on Main Street? Evidence from Retail TIPS Fund Flows*. National Bureau of Economic Research. URL: https://www.nber.org/papers/w30692.

Newey, Whitney K. (1993). "16 Efficient Estimation of Models with Conditional Moment Restrictions". In: *Handbook of Statistics*. Vol. 11. Elsevier, pp. 419–454. ISBN: 978-0-444-89577-6. DOI: 10.1016/S0169-7161(05)80051-3.

Newey, Whitney K. and Daniel McFadden (Jan. 1, 1994). "Chapter 36 Large Sample Estimation and Hypothesis Testing". In: *Handbook of Econometrics*. Vol. 4. Elsevier, pp. 2111–2245. DOI: 10.1016/S1573-4412(05)80005-4.

Newey, Whitney K. and Frank Windmeijer (2009). "Generalized Method of Moments With Many Weak Moment Conditions". In: *Econometrica* 77.3, pp. 687–719. ISSN: 1468-0262. DOI: 10.3982/ECTA6224. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA6224.

Qian, Eric (June 6, 2024). *Heterogeneity-Robust Granular Instruments*. DOI: 10.48550/arXiv.2304.01273. arXiv: 2304.01273 [econ]. Pre-published.

Selgrad, Julia (2023). "Testing the Portfolio Rebalancing Channel of Quantitative Easing". In.

Stulz, René M, Alvaro G Taboada, and Mathijs A van Dijk (Aug. 2022). *Why Are Banks' Holdings of Liquid Assets so High?* Working Paper 30340. National Bureau of Economic Research. DOI: 10.3386/w30340.

Tabova, Alexandra M. and Francis E. Warnock (Dec. 2022). *Preferred Habitats and Timing in the World's Safe Asset*. DOI: 10.3386/w30722. National Bureau of Economic Research: 30722. Pre-published.

Van der Beck, Philippe (2022). "On the Estimation of Demand-Based Asset Pricing Models". URL: https://doi.org/10.2139%2Fssrn.4111329.

Vayanos, Dimitri and Jean-Luc Vila (2021). "A Preferred-Habitat Model of the Term Structure of Interest Rates". In: *Econometrica* 89.1, pp. 77–112. ISSN: 0012-9682. DOI: 10.3982/ECTA17440.

Vissing-Jorgensen, Annette (Nov. 1, 2021). "The Treasury Market in Spring 2020 and the Response of the Federal Reserve". In: *Journal of Monetary Economics* 124, pp. 19–47. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2021.10.007.

Weiss, Colin R. (Jan. 28, 2022). "Foreign Demand for U.S. Treasury Securities during the Pandemic". In: URL: https://www.federalreserve.gov/econres/notes/feds-notes/foreign-demand-for-us-treasury-securities-during-the-pandemic-20220128.html.

Zhou, Haonan (2023). "The Fickle and the Stable: Global Financial Cycle Transmission via Heterogeneous Investors".

A Proofs

We prove the properties of the optimal GIV estimator for the general framework introduced in (2.4), restated here:

$$\left.\begin{array}{ll} q_{i,t} & = -p_t \times \mathbf{C}'_{i,t}\boldsymbol{\zeta} + \mathbf{X}'_{i,t}\boldsymbol{\beta} + u_{i,t}, \\ 0 & = \sum_i S_{i,t} q_{i,t} \end{array}\right\} \implies p_t = \frac{1}{\mathbf{C}'_{S,t}\boldsymbol{\zeta}} \left[\mathbf{X}'_{S,t}\boldsymbol{\beta} + u_{S,t}\right], \qquad \text{(A.1)}$$

where $\mathbf{C}_{i,t}$ and $\mathbf{X}_{i,t}$ are matrices of exogenous variables that can be individual-specific and time-varying, and $\boldsymbol{\zeta}$ and $\boldsymbol{\beta}$ are vectors of constant coefficients to be estimated. Recall that we use the S subscript to denote size weighted summation, i.e., $\mathbf{X}_{S,t} = \sum_i S_{i,t}\mathbf{X}_{i,t}$. We denote the length of $\boldsymbol{\zeta}$ as N_ζ and the length of $\boldsymbol{\beta}$ as N_β. For convenience, I also define $\zeta_{S,t} = \mathbf{C}'_{S,t}\boldsymbol{\zeta}$ as the aggregate elasticity, and $M_t = \frac{1}{\zeta_{S,t}}$ the "multiplier", namely, the change in p_t in response to a one-unit change in quantity.

We formally state the assumptions needed for the optimal GIV estimator:

Assumption A.1 (Full model)**.**

1. The model is specified as in (A.1), with $\mathbf{S}_t > 0$ and $\mathbf{C}'_{S,t}\boldsymbol{\zeta} \neq 0$ with probability 1.

2. For each entity i, $u_{i,t}$ is i.i.d. with $\mathbb{E}\left[u_{i,t} \mid \mathrm{C}_t, \mathrm{S}_t, \mathbf{X}_t\right] = 0$ and $\mathbb{E}\left[u_{i,t}^2 \mid \mathrm{C}_t, \mathrm{S}_t, \mathbf{X}_t\right] = \sigma_i^2 > 0$.

3. For any $i \neq j$, $u_{i,t}$ is independent to $u_{j,t}$ conditional on $(\mathrm{C}_t, \mathrm{S}_t, \mathbf{X}_t)$.

4. $\mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}'_{i,t}\right]$ are non-singular.

5. The parameter space for $(\boldsymbol{\zeta}, \boldsymbol{\beta})$ is compact and the true parameter is in the interior of the parameter space.

Notice that we do *not* require the size vector S_i sums to 1. The only assumption we impose on the sizes are positive. This is not a restriction as one can always simply swap the sign of $q_{i,t}$ and $S_{i,t}$ to satisfy the assumption. For example, the supply can be specified as a sector with size 1 and a higher supply enters as a negative $q_{i,t}$, so that their elasticity has the same sign as other sectors.

We first proceed without the exogenous variables $\mathbf{X}_{i,t}$. As we show in A.2, all results describe here go through with common factors.

A.1 Optimal GIV Estimator

For ease of reading, we reintroduce the definition of optimal GIV estimator here.

Let $\boldsymbol{\psi}_t(\mathbf{z})$ be a vector of length $\frac{N(N-1)}{2}$ with subscript ij $(i \neq j)$ denote the row corresponding the entity pair (i, j) :

$$\psi_{ij,t}(\mathbf{z}) \equiv \hat{u}_{it}(\mathbf{z}) \hat{u}_{jt}(\mathbf{z}) \equiv \left(q_{i,t} + p_t \mathbf{C}'_{i,t}\mathbf{z}\right)\left(q_{j,t} + p_t \mathbf{C}'_{j,t}\mathbf{z}\right).$$

We have the conditional moment conditions:

$$\begin{aligned}\mathbb{E}\left[\psi_{ij,t}(\boldsymbol{\zeta}) \mid \mathrm{C}_t, \mathrm{S}_t\right] &= \mathbb{E}\left[\left(q_{i,t} + p_t \mathbf{C}'_{i,t}\boldsymbol{\zeta}\right)\left(q_{j,t} + p_t \mathbf{C}'_{j,t}\boldsymbol{\zeta}\right) \mid \mathbf{C}_t, \mathbf{S}_t\right] \\ &= \mathbb{E}\left[u_{i,t}u_{j,t} \mid \mathbf{C}_t, \mathbf{S}_t\right] = 0.\end{aligned}$$

Denote $\mathbf{V}^{\psi} \equiv \mathbb{E}\left[\boldsymbol{\psi}_t(\boldsymbol{\zeta})\boldsymbol{\psi}_t(\boldsymbol{\zeta})'\right] = Diag\left(\sigma_1^2\sigma_2^2, \sigma_1^2\sigma_3^2, \ldots, \sigma_{N-1}^2\sigma_N^2\right)$. Also denote $\hat{\mathbf{W}}_t^*(z)$ as a weighting matrix of the size $\frac{N(N-1)}{2} \times N_\zeta$, whose entry corresponding to the row ψ_{ij} and the column ζ_k is given as:

$$\hat{W}^*_{ij,k,t}(\mathbf{z}) = \frac{1}{\mathbf{C}'_{S,t}\mathbf{z}}\left(\frac{S_{jt}C_{it,k}}{\sigma_i^2} + \frac{S_{it}C_{jt,k}}{\sigma_j^2}\right) \tag{A.2}$$

Definition. The optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ solves:

$$\mathbf{b}_T^{W^*}\left(\hat{\boldsymbol{\zeta}}\right) \equiv \frac{1}{T}\sum_t W_t^*\left(\hat{\boldsymbol{\zeta}}\right)'\boldsymbol{\psi}_t\left(\hat{\boldsymbol{\zeta}}\right) = 0.$$

To understand how this is derived, we define the Jacobian matrix of $\boldsymbol{\psi}_t(\mathbf{z})$ evaluated at the true parameter, $\mathbf{D}_t = \frac{\partial \boldsymbol{\psi}_t(\mathbf{z})}{\partial z}\mid_{z=\zeta}$. Its entry is given as:

$$D_{ij,t} = u_{jt}p_tC_{it,k} + u_{it}p_tC_{jt,k},$$

and then the optimal weighting matrix is given as:

$$\mathbf{W}_t^* = \left(\mathbf{V}^{\psi}\right)^{-1}\mathbb{E}\left[\mathbf{D}_t \mid \mathrm{C}_t, \mathrm{S}_t\right],$$

where we use the knowledge that $\mathbb{E}\left[u_{jt}p_t \mid \mathbf{C}_t, \mathbf{S}_t\right] = \frac{1}{\mathbf{C}'_{S,t}\boldsymbol{\zeta}}S_{jt}\sigma_j^2$.

With a complicated functional form, the weighting matrix has a very intuitive interpretation. Consider the simple case described in the main text, where each investor has its own time-invariant elasticity and size, namely, $S_{jt} = S_j$ and $C_{it,k} = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$. In this case, the optimal GIV estimator solves:

$$\frac{1}{T}\sum_t \hat{u}_{it}\left(\hat{\boldsymbol{\zeta}}\right)\underbrace{\sum_{j\neq i} S_j \hat{u}_{jt}\left(\hat{\boldsymbol{\zeta}}\right)}_{\hat{u}_{S(-i),t}\left(\hat{\boldsymbol{\zeta}}\right)} = 0 \quad \forall i.$$

That is, the optimal GIV estimator simply weigh different sectors using their sizes. If we further assume a homogeneous elasticity across all entities, the optimal GIV estimator further weigh entities sharing the same elasticity using their respective volatility, in the same spirit as the generalized least square (GLS) estimator.

$$\frac{1}{T}\sum_t \sum_i \frac{1}{\sigma_i^2}\hat{u}_{it}\left(\hat{\boldsymbol{\zeta}}\right)\hat{u}_{S(-i),t}\left(\hat{\boldsymbol{\zeta}}\right) = 0.$$

We make one additional assumption on the weighting matrices to ensure the estimator is well-behaved.

Assumption A.2. $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{W}_t^*\right]$ *is non-singular.* $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)\right] = 0$ *if and only if* $\mathbf{z} = \boldsymbol{\zeta}$.

In many cases, the second part of the assumption is actually redundant and directly follows from the first part. For example, when size weights and $\mathbf{C}_{i,t}$ are constant across time, or $\mathbf{C}_{i,t}$ are non-negative with probability 1 (such as dummy variables), non-singular conditions in $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{W}_t^*\right]$ automatically implies the unique identification. The additional assumption is only needed when we have nonrestrictive loading factors $\mathbf{C}_{i,t}$. To see this, consider another candidate $\tilde{\boldsymbol{\zeta}} \equiv \boldsymbol{\zeta} + \Delta\boldsymbol{\zeta}$. We have:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{W}_t^{*\prime}\boldsymbol{\psi}_t\left(\tilde{\boldsymbol{\zeta}}\right)\right] &= \mathbf{0} \\
&= \mathbb{E}\left[M_t\sum_{i\neq j}\left(\frac{S_{jt}\mathbf{C}_{it}}{\sigma_i^2} + \frac{S_{it}\mathbf{C}_{jt}}{\sigma_j^2}\right)\left(u_{i,t} + p_t\mathbf{C}_{i,t}'\Delta\boldsymbol{\zeta}\right)\left(u_{j,t} + p_t\mathbf{C}_{j,t}'\Delta\boldsymbol{\zeta}\right)\right] \\
&= \mathbb{E}\left[M_t\sum_{i\neq j}\left(\frac{S_{jt}\mathbf{C}_{it}}{\sigma_i^2} + \frac{S_{it}\mathbf{C}_{jt}}{\sigma_j^2}\right)\left(M_t\left(S_{jt}\sigma_j^2\mathbf{C}_{i,t}' + S_{it}\sigma_i^2\mathbf{C}_{j,t}'\right)\Delta\boldsymbol{\zeta} + \sigma_p^2\Delta\boldsymbol{\zeta}'\mathbf{C}_{i,t}\mathbf{C}_{j,t}'\Delta\boldsymbol{\zeta}\right)\right] \\
&= \mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^{*\prime}\right]\Delta\boldsymbol{\zeta} + \mathbb{E}\left[\sigma_p^2 M_t\sum_{i\neq j}\left(\frac{S_{jt}\mathbf{C}_{it}}{\sigma_i^2} + \frac{S_{it}\mathbf{C}_{jt}}{\sigma_j^2}\right)\left(\Delta\boldsymbol{\zeta}'\mathbf{C}_{i,t}\mathbf{C}_{j,t}'\Delta\boldsymbol{\zeta}\right)\right].
\end{aligned}$$

With non-negative $\mathbf{C}_{it}$ and non-singular $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{W}_t^*\right]$, the only solution to the equation above is $\Delta\boldsymbol{\zeta} = \mathbf{0}$. However, with unrestricted $\mathbf{C}_{i,t}$ potentially multiple roots can arise from this equation. One needs to verify the uniqueness of solution when complicated loading factors are used.

With this definition, we prove Theorem 2, restated below.

Theorem (Asymptotic efficiency in the general model)**.** *Given the moment conditions (2.7), under Assumption A.1 and Assumption A.2, the optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ is consistent and asymptotically normal:*

$$\sqrt{T}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}^{\zeta}\right),$$

for $T \to \infty$, where

$$\mathbf{V}^{\zeta} = \left(\mathbb{E}\left[\boldsymbol{W}_t^{*\prime}\mathbf{V}^{\psi}\boldsymbol{W}_t^*\right]\right)^{-1}.$$

Moreover, $\mathbf{V}^{\zeta}$ achieves the semi-parametric efficiency bound.

Proof. The consistency of the estimator directly follows the standard GMM argument.

To prove optimality, we first derive the asymptotic variance for an estimator $\hat{\zeta}^W$ with a generic weighting matrix that solves:

$$\mathbf{b}_T^W\left(\hat{\boldsymbol{\zeta}}^W\right) \equiv \frac{1}{T}\sum_t W_t\left(\hat{\boldsymbol{\zeta}}^W\right)'\boldsymbol{\psi}_t\left(\hat{\boldsymbol{\zeta}}^W\right) = 0. \tag{A.3}$$

The weighting scheme W_t can be a function of the given estimates and observed variables $(\mathrm{C}_t, \mathrm{S}_t)$. Denote $\mathbf{D}_t^W = \frac{\partial \mathbf{W}_t(\mathbf{z})'\boldsymbol{\psi}_t(\mathbf{z})}{\partial \mathbf{z}} \mid_{z=\zeta}$, we have

$$\begin{aligned}\mathbf{V}_{\zeta}^W &= \mathbb{E}\left[\mathbf{D}_t^{W\prime}\right]^{-1}\mathbb{E}\left[\mathbf{W}_t'\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)'\mathbf{W}_t\right]\mathbb{E}\left[\mathbf{D}_t^W\right]^{-1} &\text{(A.4)}\\ &= \mathbb{E}\left[\mathbf{D}_t'\mathbf{W}_t\right]^{-1}\mathbb{E}\left[\mathbf{W}_t'\mathbf{V}^{\psi}\mathbf{W}_t\right]\mathbb{E}\left[\mathbf{D}_t\mathbf{W}_t'\right]^{-1} &\text{(A.5)}\end{aligned}$$

where the second equation uses $\mathbb{E}\left[\mathbf{D}_t^W\right] = \mathbb{E}\left[\frac{\partial \mathbf{W}_t(\mathbf{z})'}{\partial \mathbf{z}}\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right) + \mathbf{W}_t'\frac{\partial \boldsymbol{\psi}_t(\mathbf{z})}{\partial z}\right] \mid_{z=\zeta}= \mathbb{E}\left[\mathbf{W}_t'\mathbf{D}_t\right]$. Plug in $\mathbf{W}^* = \left(\mathbf{V}^{\psi}\right)^{-1}\mathbb{E}\left[\mathbf{D}_t \mid \mathbf{C}_t, \mathbf{S}_t\right]$, we have: we have

$$\mathbf{V}_{\zeta}^{W^*} = \left(\mathbb{E}\left[\mathbf{D}_t'\left(\mathbf{V}^{\psi}\right)^{-1}\mathbf{D}_t\right]\right)^{-1} = \left(\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{V}^{\psi}\mathbf{W}^*\right]\right)^{-1}. \tag{A.6}$$

One can verify that (A.6) is also the semi-parametric efficiency bound shown by

Chamberlain (1987). To see it more straightforwardly, we can show that for any $\mathbf{W}_t \neq \mathbf{W}_t^*$, $\Delta \equiv \mathbf{V}_\zeta^W - \mathbf{V}_\zeta^{W^*}$ is positive definite. The proof for the optimality generally follows the strategy of optimal instrument in conditional models, surveyed in Newey (1993).

$$\boldsymbol{\Delta} \equiv \mathbb{E}\left[\mathbf{D}_t' \left(\mathbf{V}^\psi\right)^{-1} \mathbf{D}_t\right] - \mathbb{E}\left[\mathbf{D}_t'\mathbf{W}_t\right] \mathbb{E}\left[\mathbf{W}_t'\mathbf{V}^\psi\mathbf{W}_t\right]^{-1} \mathbb{E}\left[\mathbf{W}_t'\mathbf{D}_t\right].$$

Given $\mathbf{V}^\psi$ is a positive definite diagonal matrix, it is convenient to factor $\mathbf{V}^\psi$ into $\mathbf{D}_t$ and $\mathbf{W_t}$ by defining:

$$\begin{aligned}\tilde{\mathbf{D}}_t &= \left(\mathbf{V}^\psi\right)^{-\frac{1}{2}} \mathbf{D}_t \\ \tilde{\mathbf{W}}_t &= \left(\mathbf{V}^\psi\right)^{\frac{1}{2}} \mathbf{W}_t.\end{aligned}$$

Define $\mathbf{G}_t = \begin{bmatrix} \tilde{\mathbf{D}}_t' \\ \tilde{\mathbf{W}}_t' \end{bmatrix}$, and $\mathbf{H}' = \begin{bmatrix} \mathbf{I} & -\mathbb{E}\left[\tilde{\mathbf{D}}_t'\tilde{\mathbf{W}}_t\right] \mathbb{E}\left[\tilde{\mathbf{W}}_t'\tilde{\mathbf{W}}_t\right] \end{bmatrix}$, we can show:

$$\boldsymbol{\Delta} = \mathbf{H}'\mathbb{E}\left[\mathbf{G}_t\mathbf{G}_t'\right]\mathbf{H}.$$

Clearly, $\gneqq$ is positive semi-definite. $\boldsymbol{\Delta} = 0$ when $\tilde{\mathbf{W}}_t = \mathbb{E}\left[\tilde{\mathbf{D}}_t \mid \mathbf{C}_t, \mathbf{S}_t\right]$, or $\mathbf{W}_t = \left(\mathbf{V}^\psi\right)^{-1} \mathbb{E}\left[\mathbf{D}_t \mid \mathbf{C}_t, \mathbf{S}_t\right]$. □

A.2 Asymptotic Properties when Common Factors are Included

With common factors, we proceed with the following steps.

0. (If needed: Use PCA to extract unobserved common factors from granular data)
1. Regress $q_{i,t}$ and $p_t\mathbf{C}_{i,t}'$ on $\mathbf{X}_{i,t}$ to estimate $\hat{\boldsymbol{\beta}}^q$ and $\hat{\boldsymbol{\beta}}^{pC}$.
2. Take the residual from the first step, $q_{i,t}^\varepsilon \equiv q_{i,t} - \mathbf{X}_{i,t}'\hat{\boldsymbol{\beta}}^q$ and $\mathbf{pC}_{i,t}^{\varepsilon}{}' = p_t\mathbf{C}_{it}' - \mathbf{X}_{i,t}'\hat{\boldsymbol{\beta}}^{pC}$, and form the moment conditions:

$$\psi_{ijt}^\varepsilon(\mathbf{z}) \equiv \hat{u}_{it}^\varepsilon(\mathbf{z})\,\hat{u}_{jt}^\varepsilon(\mathbf{z}) \equiv \left(q_{i,t}^\varepsilon + \mathbf{pC}_{i,t}^{\varepsilon}{}'\mathbf{z}\right)\left(q_{j,t}^\varepsilon + p\mathbf{C}_{j,t}^{\varepsilon}{}'\mathbf{z}\right).$$

Proceed with the optimal weighting matrix defined in (A.2) to estimate $\hat{\zeta}$.

3. Form the estimator for $\boldsymbol{\beta}$ as:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{q} + \hat{\boldsymbol{\beta}}^{pC}\hat{\boldsymbol{\zeta}}.$$

The following proposition shows the asymptotic variance of $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\beta}}$.

Proposition A.1. *The asymptotic variance formula for $\hat{\boldsymbol{\zeta}}$ given by (A.6) still holds with common factors. $\hat{\boldsymbol{\beta}}$ is a consistent estimator for $\boldsymbol{\beta}$, and its asymptotic variance for $\hat{\boldsymbol{\beta}}$ is given as:*

$$Var(\hat{\boldsymbol{\beta}}) = \frac{1}{T}\left[Var^{OLS}(\hat{\boldsymbol{\beta}}) + \boldsymbol{\beta}^{pC}Var(\hat{\boldsymbol{\zeta}})\boldsymbol{\beta}^{pC\prime}\right] \tag{A.7}$$

$$Var^{OLS}\left(\hat{\boldsymbol{\beta}}\right) = \left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\right)^{-1}\left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\sigma_i^2\right)\left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\right)^{-1} \tag{A.8}$$

Proof. The proof follows Newey and McFadden (1994). The two-step estimator can be formulated as a joint GMM estimator that solves:

$$\frac{1}{T}\sum_t \mathbf{g}_t\left(\mathbf{b}, \mathbf{z}\right) \equiv \frac{1}{T}\sum_t \left[\begin{array}{c} \mathbf{g}_t^{OLS}\left(\mathbf{b}\right) \\ \mathbf{W}_t^{*\prime}\boldsymbol{\psi}_t(\mathbf{b}, \mathbf{z}) \end{array}\right] = 0$$

where $\mathbf{g}^{OLS}\left(\mathbf{b}\right)$ is the moment condition for the first step of to identify $(\boldsymbol{\beta}^{q}, \boldsymbol{\beta}^{pC})$, and the second block is optimal GIV moment conditions, taken OLS coefficients as given. As the coefficients are just-identified, the solutions to the GMM estimation are identical to the two-step estimator, and hence we can utilize properties of the GMM estimator to study the asymptotic behaviors. Use $\boldsymbol{\theta}$ to denote the vector of $(\boldsymbol{\beta}, \boldsymbol{\zeta})$ and $\hat{\boldsymbol{\theta}}$ the corresponding estimator, we have:

$$\begin{aligned}
\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{D}^{g\prime}\right)^{-1}\left(\mathbf{V}^{g}\right)\left(\mathbf{D}^{g}\right)^{-1}\right) \\
\mathbf{D}^{g} &= \mathbb{E}\left[\frac{\partial \mathbf{g}_t\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right] \\
\mathbf{V}^{g} &= \mathbb{E}\left[\mathbf{g}_t\left(\boldsymbol{\theta}\right)\mathbf{g}_t\left(\boldsymbol{\theta}\right)'\right].
\end{aligned}$$

A crucial feature of these matrices is that they are block-diagonal:

$$\mathbf{D}^{g}=\left[\begin{array}{cc}\frac{\partial \mathbf{g}_{t}^{OLS}(\mathbf{b})}{\partial \mathbf{b}}\mid_{\mathbf{b}=\boldsymbol{\beta}} & \\ & \mathbb{E}\left[\mathbf{W}_{t}^{*\prime}\mathbf{D}_{t}\right]\end{array}\right]$$

$$\mathbf{V}^{g}=\left[\begin{array}{cc}\mathbf{V}^{OLS} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}^{*\prime}\mathbf{V}^{\psi}\mathbf{W}^{*}\end{array}\right].$$

To see $\mathbf{D}^{g}$ is diagonal, notice that the estimate for $\boldsymbol{\zeta}$ does not enter the moment conditions for the OLS step by the two-step nature, and

$$\mathbb{E}\left[\frac{\partial\psi_{ij,t}\left(\boldsymbol{b},\boldsymbol{\zeta}\right)}{\partial\mathbf{b}}\mid_{\mathbf{b}=\boldsymbol{\beta}}\right]=\mathbb{E}\left[\frac{\partial\left(u_{it}-\mathbf{X}_{it}'\left(\mathbf{b}-\boldsymbol{\beta}\right)\right)\left(u_{jt}-\mathbf{X}_{jt}'\left(\mathbf{b}-\boldsymbol{\beta}\right)\right)}{\partial\mathbf{b}}\mid_{\mathbf{b}=\boldsymbol{\beta}}\right]=\mathbf{0}.$$

$\mathbf{V}^{g}$ being diagonal is a result that the error term in g_{t}^{OLS} is linear in u_{it}, while the error terms of $\boldsymbol{\psi}_{t}$ are quadratic $u_{it}u_{jt}$, so any linear combination between them have a covariance of zero.

Due to the block-diagonal structure in $\mathbf{D}^{g}$ and $\mathbf{V}^{g}$, the asymptotic variance of $\hat{\boldsymbol{\zeta}}$ is not affected by the first step estimation.

To further derive the asymptotic variance of $\hat{\boldsymbol{\beta}}$, we use the Delta method:

$$Var(\hat{\boldsymbol{\beta}})=\left[\begin{array}{cc}\mathbf{I} & \boldsymbol{\zeta}'\end{array}\right]Var(\left[\begin{array}{cc}\hat{\boldsymbol{\beta}}^{q} & \hat{\boldsymbol{\beta}}^{pC}\end{array}\right])\left[\begin{array}{cc}\mathbf{I} & \boldsymbol{\zeta}'\end{array}\right]'+\boldsymbol{\beta}^{pC}Var(\hat{\boldsymbol{\zeta}})\boldsymbol{\beta}^{pC\prime}$$

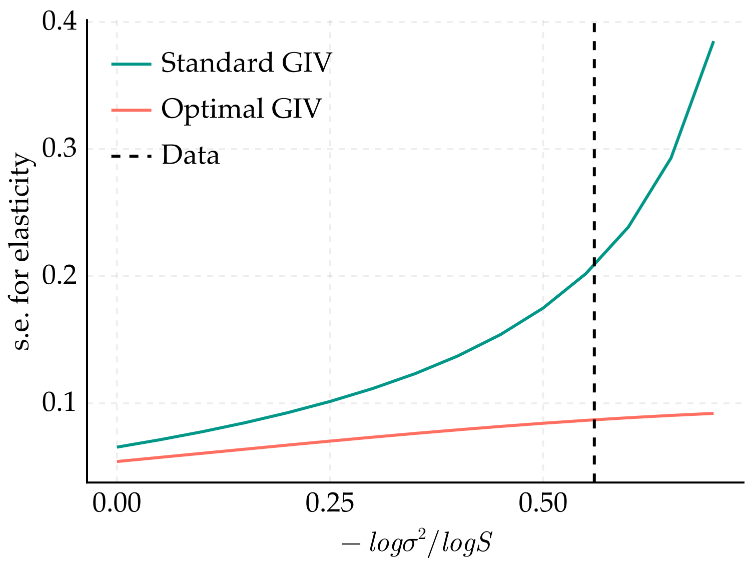
where the first part is equivalent to asymptotic variance of the OLS estimator of regressing $q_{i,t}+p_{t}\mathbf{C}_{i,t}'\boldsymbol{\zeta}$ on $\mathbf{X}_{i,t}$, which is given in (A.8). □

B Simulation

B.1 Statistical Power Comparison of GIVs

Figure B.1 compares the statistical power of the standard GIV in Gabaix and Koijen (2024) vs. the optimal GIV in simulation. We simulate the model mirroring the Treasury market in the flow of fund data. We assume there are 12 sectors with a homogeneous elasticity of 1, and the supply is inelastic. The size distribution is calibrated with the concentration in the Financial Accounts data, and we vary the rate at which volatility decreases as the size of a sector increases. Other model parameters are reported in the note accompanyiing Figure B.1. As the volatility decreases, the precision of standard GIV decays rapidly, while only mildly for the optimal GIV. When the volatility-size ratio matches that in the Financial Accounts data, the optimal GIV is twice as powerful as the standard GIV.

Figure B.1: **Statistical power comparison: standard GIV vs. optimal GIV**



Note: Figure B.1 compares the statistical power of the standard GIV in Gabaix and Koijen (2024) vs. the optimal GIV in simulation. Simulation parameters: number of sectors $N = 12$, $T = 80$ quarters, no common factors, the excess HHI in the size distribution, $h \equiv \sqrt{\sum_i S_i^2 - \frac{1}{N}} = 0.5$, elasticity $\zeta = 1.0$, and the level of the shock volatility is calibrated so that the $\sigma_p^2 = 2.5\%$.

C Empirical Appendix

C.1 More on the Data

Financial Accounts data The main dataset we use for Treasury bonds and notes holdings comes from Table FU.210 and L.210 of Federal Reserve's Z.1 release.[34] We use seasonally unadjusted transactions and holdings of Treasury notes and bonds (labelled as "other Treasury securities") whenever possible. Not all sectors in the Z.1 release separately report holdings of bills, notes and bonds. Those that

[34] https://www.federalreserve.gov/apps/fof/FOFTables.aspx

do not report this breakdown are typically small holders and are lumped into the "other" category for estimation. There are, however, a number of large sectors with no information on notes and bonds holdings directly available, including the household sector (residual), pension funds and ETFs. In these cases, we assume that their holdings are entirely in Treasury notes and bonds.

Country-level foreign holdings and bank-level holdings We use information from the Treasury International Capital system (TIC) to compile country-level holdings of long-term (bonds and notes) and short-term (bills) Treasury securities by foreign investors. For short-term holdings, we source information included in the banking liabilities survey that starts from February 2003. For notes and bonds, we resort to the benchmark-consistent estimates put together by Bertaut and Tryon (2007), Bertaut and Judson (2014), and Bertaut and Judson (2022). Gaps in the series are present from 2011Q3 to 2011Q4, when the TIC survey methodology is in transition. We use the SLT1D table to fill in the blanks.

For individual U.S. bank holdings of Treasury notes and bonds, we use data from FFIEC 031/041/051 filings, also known as the Call Reports. We download the raw data from WRDS and take several steps to clean the dataset. First, as WRDS data may contain FFIEC 002 filers (foreign branches), we identify and drop all foreign branches in the data using the relationship files from the NIC information system. Then, we largely follow the methodology of Financial Accounts to extract information on Treasury holdings. Quarterly transaction figures are computed by summing up three variables: RCON0211, RCON1286 and RCON3531 and taking the first difference. The market value of holdings are the sum of three variables: RCON0213, RCON1287 and RCON3531.[35] To get the breakdown between bills and other securities, we use an imputation method similar to Jansen, Li, and Schmid (2024). With information on the residual maturity breakdown of Treasury and agency securities (RCONA549 to RCONA554, excluding trading assets,) we compute total holdings of Treasury and agency securities and impute

[35]Only large banks (FFIEC 031 filers) report information on Treasury securities held as trading assets (RCON3531) and only fair values are reported. For Treasuries held as trading assets after 2018Q1, RCON3531 is very sparsely populated. We use RCFD3531 net of foreign office holdings as the substitute. Unlike the financial account, who assumes that U.S. Treasuries comprise 3% of the total trading book of foreign offices, we use historical realized values of RCFD3531 and RCON3531 to make the adjustment. Eventually, only two banks' RCFD3531 numbers need to be adjusted. For all other banks, we direct use RCFD3531 to impute missing RCON3531 numbers.

notes and bonds holdings assuming that the share between bills and notes and bonds are stable across the residual maturity bucket, and the data in the residual maturity buckets above one year can well approximate notes and bonds holdings.

A challenge of working with country- and bank-level holdings is that our estimation algorithm requires a balanced panel. A country or bank must consistently hold non-zero amount of Treasury securities throughout our sample period to be included individually in the data. Otherwise, they will be swept into an entity labelled "all other countries" (TIC) or "all other banks" (call report), which aggregate Treasury notes and bonds holdings of all country or bank entities that do not always hold Treasuries during the sample period. By 2023, individual bank holding and country holding cover 80% and 90% of the total holdings reported in the Z.1 release, respectively.[36]

Discrepancy The fact that not all investor sectors report Treasury notes and bonds holdings and the lack of direct estimates of notes and bonds holdings and transactions from the call report data would introduce valuation and coverage discrepancies (see Figure C.3). We sweep the discrepancy component into the household sector (the residual sector in the Z.1 methodology) in our baseline estimation.

Observed common factors Our baseline estimation makes use of the following observed macro factors: the CBOE SP500 implied volatility index (VIX, FRED ticker VIXCLS), the effective Federal Funds rate (FRED ticker FEDFUNDS), CPI (FRED ticker CPIAUCSL), and the nominal broad dollar index (splicing goods trade-weighted dollar index and goods-and-service trade-weighted dollar index, FRED tickers DTWEXBGS and DTWEXB).

Sector classification and aggregation Table C.1 reports our aggregation of Financial Account sectors to categories for GMM estimation and for yield decomposition. For italicized sectors (U.S.-chartered banks and foreign investors), we use call

[36]More specifically, we require countries to hold non-zero Treasury bonds and notes throughout from 2003Q1 to be included individually, and banks to have a non-zero position starting from 2001Q4 to be included. As countries and banks with consistently positive holdings tend to be large, our country- and bank-level holding is highly representative of the entire market. For flows, the correlation between individual bank aggregates and the Financial Accounts aggregates is 92%, and for the TIC data, the correlation is 93%.

report and TIC data to break down sector aggregates into individual bank/country holdings.

C.2 Predictability of Policy Sectors

C.2.1 The Federal Reserve

Historically, the New York Fed's Open Market Trading Desk (the Desk) used outright purchases or sales of Treasury securities as a tool to manage the supply of bank reserves to maintain conditions consistent with the Federal funds target rate set by the FOMC. Hence, small Treasury market flows were present in the NY Fed's System Open Market Account (SOMA) before 2008. The nature of these flows changed drastically during the Global Financial Crisis. Since 2009 to 2024, the Fed conducted 4 rounds of large-scale asset purchases including Treasuries and GSE debt, commonly known as Quantitative Easings (QEs), which significantly enlarged the SOMA portfolio in long-term Treasury securities.

Once activated, the scale of the QE operations can be predicted. At the start of each round of QEs, The Federal Open Market Committee (FOMC) typically announces the total amount and the pace of the purchases to provide forward guidance to market participants. Closely following the instructions of the FOMC, the Desk at the New York Fed issues detailed schedule for future purchases typically at a biweekly frequency. As the operations are prescheduled, they generally do not respond to the contemporaneous market development.[37]

To control for the scheduled Fed purchases, we construct a sequence of anticipated purchases by the Fed in each quarter based on the FOMC announcement of the Fed in the last quarter. For example, in March 18, 2009, the FOMC announced it would purchase up to $300 billion of longer-term Treasury securities over the next six months. The scheduled purchase by the Fed in Q2-3 would be $150 billion each, until further policy changes by the FOMC.

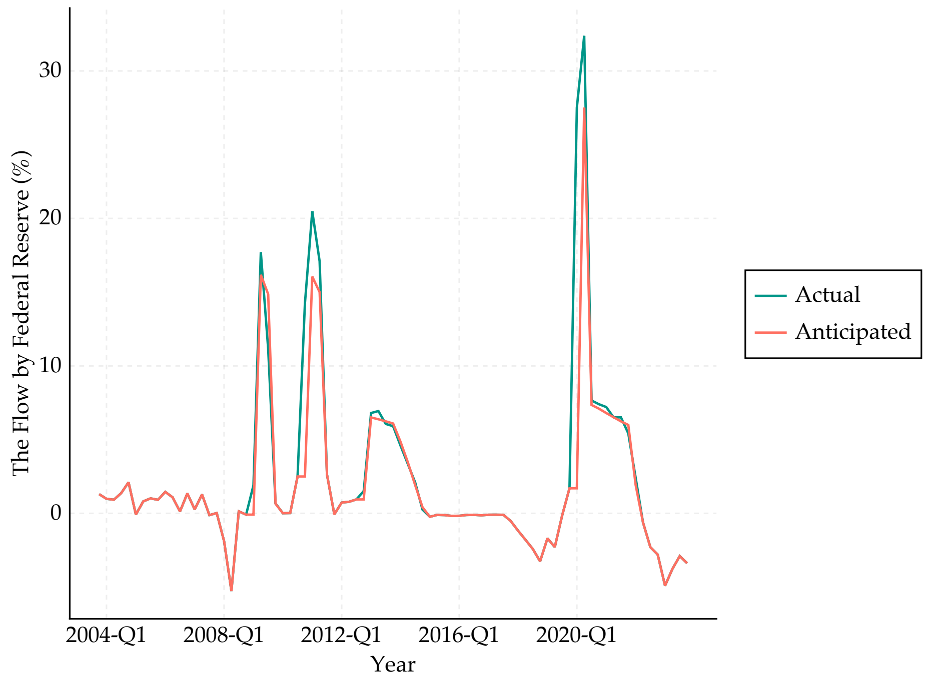
Table C.2 reports the anticipated purchases versus the actual purchases of the Fed during active QE periods. When the QE is not pre-scheduled in that quarter

[37]One exception is the initial stage of the QE in response to the COVID pandemic. Due to the extremity of the macroeconomic and financial conditions, the FOMC announced that it "will continue to purchase Treasury securities and agency mortgage-backed securities in the amounts needed to support smooth market functioning and effective transmission of monetary policy to broader financial conditions", without specifying the duration and the total amount.

(typically at the start of the QEs), we use lagged purchases by the Fed from the previous quarter as the expected purchase.[38] As investors may use more information than the FOMC announcements alone, we lean towards conservative when setting the anticipated purchases. For example, during the exit of the QE3, even though the FOMC announcement of reducing purchases are not pre-scheduled in earlier quarters, as it follows a predictable patterns and well-anticipated by investors, we consider these changes are "scheduled". Outside of the QE periods, we treat all realized, contemporaneous purchases as anticipated, as they are either due to rollovers of maturing Treasuries or reserve management purposes. As the scale of these outright purchases are small, alternative treatments have little effects on our results. Figure C.1 plots the path of the Fed's flows together with the one-quarter-ahead anticipated path sequence. As shown in the plot, for most quarters, the realized purchases closely track the anticipated purchases. The major deviations occur on the entry and exit of the QE programs. Our estimates for the Fed's sensitivity to prices and to macro factors mainly reflect the policy consideration during episodes in which Fed deviates from the anticipated schedule.

[38] Alternatively, we can also estimate an AR process of the Fed purchase outside of QE and use that as predictor. The results are not sensitive to the treatment as the deviation of the QE purchase from the regular outright operations are an order of magnitude larger.

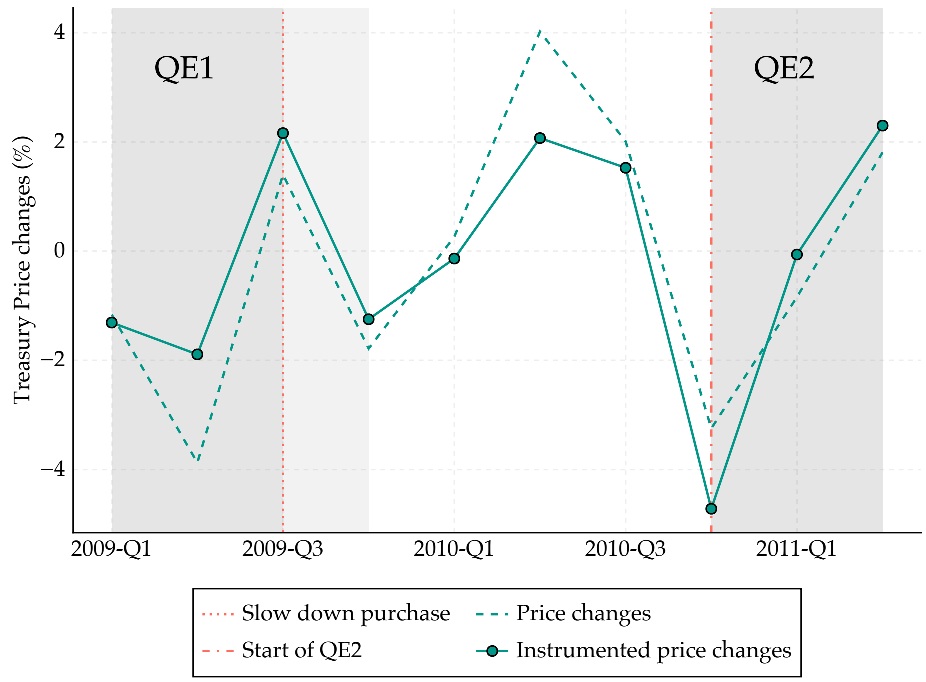
Figure C.1: **Case Study: QE Entry and Exit between 2009-2010**



Note: Figure C.1 plots the actual purchases (normalized by total Fed holdings) of U.S. Treasury notes and bonds against the constructed one-quarter-ahead anticipated Fed purchase measure in green. The anticipated path is generated by combining the actual Fed purchase path before 2009 with the pre-announced purchase schedule by the New York Fed starting from the beginning of the QE operation.

As a further illustration, in Figure C.2 we focus on the end of the QE1 in 2009Q3 and the start of QE2 in 2010Q4. The original schedule for the first round of Treasury purchases ends at 2009Q3. In that quarter, the FOMC meetings decided to change the schedule and slow down the purchase and extend to Q4 in order to "promote a smooth transition in markets as these purchases of Treasury securities are completed". This results in a negative deviation from the anticipated path. In the figure, we plot the path of the price changes of the Treasury notes and bonds (the dash line), and the price change predicted using other sectors' demand shocks (the solid line). As shown in the price path, the quarter in which they decided to slow down the purchase is the one with the largest Treasury price appreciation in the entire QE1 episode. Similarly, the start of the QE2 also coincides with the quarter that saw the largest Treasury price depreciation.

Figure C.2: **Case Study: QE Entry and Exit between 2009-2010**



Note: Figure C.2 traces Treasury price changes from 2009 to 2011, covering the end of QE1 and the start of QE2. The path of actual Treasury price changes (dashed line) is plotted against the instrumented price changes, predicted using the size-weighted granular shocks extracted from the GMM estimation.

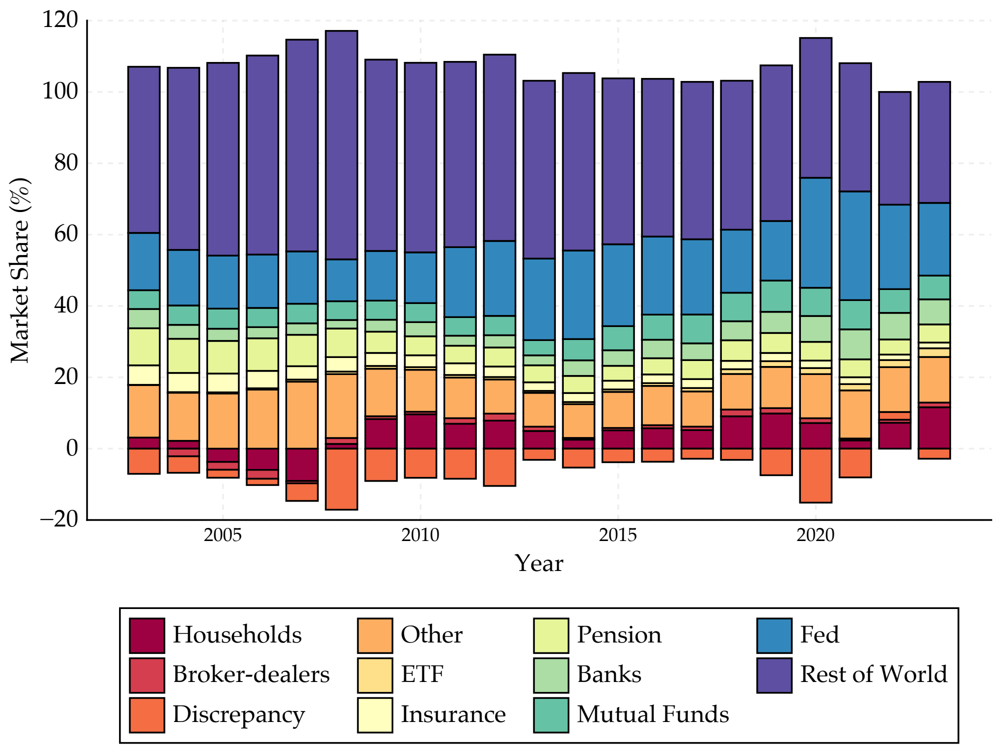
C.3 Robustness of Elasticity Estimation

Table C.3 demonstrate the robustness of our price elasticity estimation to alternative samples, controls, or dependent variables. Table C.4 shows that our finding remains robust if we remove any sector from the construction of moment conditions and reestimate the model.

C.4 Additional Tables and Figures

Table C.5 reports the full set of estimated parameters from the baseline specification.

Figure C.3: **Investor composition of U.S. Treasury notes and bonds**



Note: Figure C.3 plots the investor base of U.S. Treasury notes and bonds. We organized the holders of the U.S. Treasury identified in the Financial Account data into 10 sectors based on the classification reported in Section C.1 in the Appendix. Our procedure for calculating Treasury notes and bonds holdings from raw Financial Account data introduces a "discrepancy" component due to valuation differences and imputed holdings for sectors that do not separately report bill holdings. The market share plotted refers to each sector's market value of holdings as % of the total market value of the Treasury notes and bonds outstanding.

Table C.1: **Mapping from Sectors to Groups**

Sector	**Category (estimation)**	**Category (decomposition)**
Banks in U.S.-affiliated areas	Banks	US Financial Institution
Foreign banking offices in the U.S.	Banks	US Financial Institution
U.S.-chartered depository institutions	Banks	US Financial Institution
Exchange-traded funds	ETF	US Financial Institution
Monetary authority	Fed	Fed
Households and nonprofit organizations	Households	US Other
Property-casualty insurance companies	Insurance	US Financial Institution
Life insurance companies	Insurance	US Financial Institution
Money market funds	Other	US Financial Institution
Closed-end funds	Mutual Funds	US Financial Institution
Mutual funds	Mutual Funds	US Financial Institution
Nonfinancial corporate business	Other	US Other
Nonfinancial noncorporate business	Other	US Other
State and local governments	Other	US Other
Government-sponsored enterprises	Other	US Other
Credit unions	Other	US Other
Other financial business	Other	US Other
Issuers of asset-backed securities	Other	US Other
Holding companies	Other	US Other
State and local government employee defined benefit retirement funds	Pension	US Financial Institution
Federal government defined contribution retirement funds	Pension	US Financial Institution
Federal government defined benefit retirement funds	Pension	US Financial Institution
Private defined contribution pension funds	Pension	US Financial Institution
Private defined benefit pension funds	Pension	US Financial Institution
Rest of the world	Rest of World	Rest of World
Security brokers and dealers	Security brokers and dealers	US Financial Institution
Supply	Supply	Supply

Table C.2: The scheduled and actual purchases during QEs

	Anticipated ($ billions)	**Actual ($ billions)**	**Remarks**
2009-Q1		16	QE1 for Treasuries started (link)
2009-Q2	150	164	
2009-Q3	150	113	The FOMC decided to slow down the purchase (link)
2009-Q4	7	7	
2010-Q4		210	QE2 started (link)
2011-Q1	250	319	
2011-Q2	245	279	
2013-Q1	135	141	The FOMC announced in 2012-Q2 to QE3 (link)
2013-Q2	135	147	
2013-Q3	135	131	
2013-Q4	135	131	
2014-Q1	110	104	
2014-Q2	80	77	
2014-Q3	45	49	
2014-Q4	10	6	
2020-Q1		863	COVID-19 (link)
2020-Q2	863	1034	Total purchases not scheduled, use lagged purchases
2020-Q3	240	250	
2020-Q4	240	250	
2021-Q1	240	254	
2021-Q2	240	240	
2021-Q3	240	250	
2021-Q4	240	217	
2022-Q1	80	100	

Table C.3: **Price elasticity by sector: Alternative Specifications**

Sector	Baseline	5 factors	1970-2023	Including Bills
Aggregate	1.04	1.09	1.3	2.01
	(0.8, 1.27)	(0.83, 1.35)	(1.11, 1.5)	(1.25, 2.77)
Other	-0.27	-0.2	0.19	0.48
	(-0.5, -0.04)	(-0.42, 0.03)	(0.02, 0.36)	(0.17, 0.78)
Households	10.45	11.59	10.5	4.42
	(5.63, 15.26)	(6.16, 17.01)	(7.29, 13.7)	(0.98, 7.86)
Pension	0.11	0.11	-0.23	0.12
	(-0.05, 0.27)	(-0.05, 0.28)	(-0.39, -0.07)	(-0.11, 0.35)
Insurance	0.62	0.58	-0.67	0.15
	(0.32, 0.93)	(0.27, 0.89)	(-0.93, -0.41)	(-0.37, 0.67)
Mutual Funds	0.56	0.53	0.47	0.63
	(0.09, 1.02)	(0.05, 1.01)	(0.23, 0.72)	(-0.02, 1.27)
ETF	-0.08	-0.17	0.23	-0.39
	(-0.44, 0.27)	(-0.54, 0.19)	(0.04, 0.42)	(-0.9, 0.13)
Dealers	5.97	6.09	-1.47	-0.72
	(-2.82, 14.76)	(-3.09, 15.27)	(-8.37, 5.43)	(-11.69, 10.26)
Fed	0.39	0.45	0.03	0.15
	(0.1, 0.67)	(0.16, 0.74)	(-0.06, 0.13)	(-0.22, 0.51)
Banks	0.67	0.49	1.0	0.77
	(0.5, 0.85)	(0.31, 0.66)	(0.51, 1.49)	(0.57, 0.98)
RoW	0.43	0.41	0.39	0.6
	(0.29, 0.57)	(0.26, 0.55)	(0.27, 0.52)	(0.39, 0.81)
Supply	0.05	0.04	0.26	1.26
	(-0.02, 0.12)	(-0.03, 0.11)	(0.18, 0.33)	(0.32, 2.2)
MMF				2.45
				(0.41, 4.48)

Note: Table C.3 demonstrates the robustness of our price elasticity estimation by augmenting the model with 5 estimated principal components from granular Treasury holding data for U.S. banks and foreign investors (column 2), by extending the sample period to 1970–2023 and estimating the model only on Financial Account data (column 3), and by including Treasury bills in the estimation (column 4). 95% confidence intervals are report based on standard errors derived in Theorem 2.

Table C.4: **Price elasticity by sector: Leave-one-out Estimation**

Sector	Baseline	Other	Households	Pension	Insurance	Funds	ETF	Dealers	Fed	Banks	RoW	Supply
Aggregate	1.01	1.04	0.57	1.15	1.25	1.32	1.22	1.15	1.07	1.2	1.23	1.46
	(0.77, 1.25)	(0.78, 1.31)	(0.46, 0.68)	(0.84, 1.47)	(0.86, 1.64)	(0.83, 1.81)	(0.86, 1.57)	(0.79, 1.5)	(0.77, 1.37)	(0.84, 1.56)	(0.49, 1.96)	(0.78, 2.14)
Other	-0.07		-0.0	0.5	0.48	0.45	0.5	0.62	0.59	0.53	0.42	0.56
	(-0.3, 0.15)		(-0.18, 0.17)	(0.25, 0.74)	(0.23, 0.73)	(0.21, 0.7)	(0.25, 0.75)	(0.36, 0.87)	(0.35, 0.82)	(0.29, 0.78)	(0.18, 0.66)	(0.29, 0.82)
Households	9.8	10.43		12.32	14.63	17.08	13.54	12.61	11.97	13.85	18.08	19.6
	(5.07, 14.54)	(5.09, 15.76)		(5.91, 18.74)	(6.7, 22.57)	(7.14, 27.02)	(6.32, 20.77)	(5.48, 19.73)	(5.92, 18.02)	(6.42, 21.28)	(3.01, 33.16)	(6.23, 32.98)
Pension	-0.1	-0.02	-0.05		0.02	0.03	0.02	0.04	0.07	0.03	0.04	0.0
	(-0.27, 0.06)	(-0.19, 0.15)	(-0.18, 0.07)		(-0.16, 0.19)	(-0.14, 0.21)	(-0.15, 0.2)	(-0.14, 0.22)	(-0.1, 0.23)	(-0.15, 0.2)	(-0.12, 0.21)	(-0.18, 0.18)
Insurance	0.99	0.27	0.73	0.29		0.27	0.27	0.17	0.24	0.26	0.31	0.13
	(0.66, 1.32)	(-0.05, 0.58)	(0.49, 0.98)	(-0.03, 0.61)		(-0.05, 0.6)	(-0.06, 0.6)	(-0.16, 0.5)	(-0.07, 0.54)	(-0.06, 0.58)	(-0.0, 0.62)	(-0.21, 0.47)
Mutual Funds	0.61	0.55	0.95	0.59	0.56		0.56	0.46	0.41	0.64	0.62	0.44
	(0.15, 1.08)	(0.07, 1.03)	(0.57, 1.33)	(0.09, 1.09)	(0.06, 1.07)		(0.05, 1.07)	(-0.06, 0.97)	(-0.07, 0.88)	(0.13, 1.14)	(0.12, 1.12)	(-0.1, 0.97)
ETF	-0.49	-0.36	-0.39	-0.33	-0.35	-0.35		-0.28	-0.28	-0.33	-0.41	-0.38
	(-0.85, -0.13)	(-0.72, 0.01)	(-0.67, -0.12)	(-0.72, 0.05)	(-0.74, 0.03)	(-0.73, 0.04)		(-0.68, 0.11)	(-0.65, 0.08)	(-0.71, 0.05)	(-0.78, -0.05)	(-0.78, 0.02)
Dealers	2.04	4.88	-7.27	0.52	-0.41	-1.32	0.3		1.81	0.79	-1.11	-2.78
	(-6.61, 10.69)	(-4.32, 14.08)	(-14.04, -0.51)	(-8.72, 9.76)	(-9.73, 8.92)	(-10.64, 8.0)	(-9.09, 9.69)		(-7.08, 10.7)	(-8.49, 10.06)	(-11.2, 8.97)	(-12.75, 7.18)
Fed	0.44	0.49	0.15	0.39	0.36	0.29	0.38	0.44		0.35	0.3	0.38
	(0.11, 0.78)	(0.14, 0.84)	(-0.12, 0.41)	(0.04, 0.74)	(0.01, 0.72)	(-0.06, 0.64)	(0.02, 0.74)	(0.07, 0.81)		(0.0, 0.7)	(-0.05, 0.64)	(0.01, 0.76)
Banks	0.67	0.72	0.36	0.71	0.7	0.68	0.7	0.86	0.66		0.74	0.72
	(0.42, 0.91)	(0.47, 0.97)	(0.16, 0.55)	(0.45, 0.96)	(0.44, 0.96)	(0.42, 0.95)	(0.44, 0.96)	(0.6, 1.13)	(0.4, 0.91)		(0.49, 0.98)	(0.45, 0.99)
RoW	0.42	0.36	0.72	0.4	0.39	0.39	0.39	0.34	0.37	0.4		0.32
	(0.24, 0.61)	(0.16, 0.55)	(0.57, 0.86)	(0.2, 0.6)	(0.19, 0.59)	(0.19, 0.6)	(0.19, 0.6)	(0.13, 0.55)	(0.17, 0.56)	(0.2, 0.6)		(0.11, 0.54)
Supply	0.06	0.06	0.18	0.04	0.03	0.02	0.04	0.02	0.07	0.03	-0.0	
	(-0.01, 0.14)	(-0.01, 0.14)	(0.11, 0.26)	(-0.04, 0.11)	(-0.05, 0.11)	(-0.06, 0.09)	(-0.04, 0.11)	(-0.05, 0.1)	(-0.01, 0.14)	(-0.04, 0.11)	(-0.08, 0.08)	

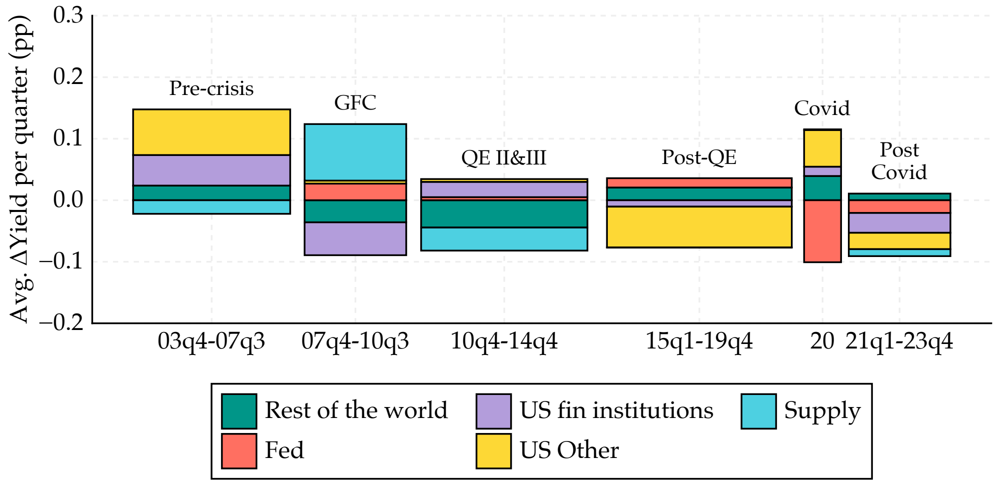
Note: Table C.4 reports the estimate sector-specific and aggregate price elasticities in our "leave-one-out" checks. For each sector, we estimate the model by removing the sector from the construction of the moment conditions report the estimated elasticities of other sectors. 95% confidence intervals are report based on standard errors derived in Theorem 2.

Table C.5: **Baseline Estimates: Full Table**

Sector	$S(\%)$	ζ	$\epsilon^{VIX}_{(std.)}$	**Δ FFR**	$Inf.$	$\epsilon^{USD}_{(std.)}$	**Lagged** f_{supply}	**Fed schedule**
Aggregate		1.01	0.75	-2.14	-1.31	-0.65	0.13	-0.34
		(0.77, 1.25)	(-0.05, 1.54)	(-3.53, -0.74)	(-2.14, -0.47)	(-1.38, 0.07)	(-0.54, 0.81)	(-1.09, 0.41)
Supply	100.0	0.06	-0.12	0.03	0.06	0.03	1.88	-0.17
		(-0.01, 0.14)	(-0.33, 0.09)	(-0.37, 0.42)	(-0.18, 0.29)	(-0.17, 0.23)	(1.71, 2.06)	(-0.35, 0.01)
RoW	44.45	0.42	-0.33	-0.56	-0.29	-0.33	-2.27	-0.84
		(0.24, 0.61)	(-0.98, 0.33)	(-1.81, 0.69)	(-1.01, 0.44)	(-0.98, 0.31)	(-2.83, -1.7)	(-1.42, -0.26)
Fed	22.08	0.44	1.61	-4.02	-0.6	-0.55	0.52	5.09
		(0.11, 0.78)	(0.07, 3.15)	(-6.34, -1.71)	(-2.02, 0.82)	(-1.8, 0.69)	(-0.82, 1.86)	(3.41, 6.76)
Other	11.78	-0.07	-1.78	-2.02	1.53	1.59	-0.55	0.87
		(-0.3, 0.15)	(-2.86, -0.7)	(-4.05, 0.01)	(0.34, 2.73)	(0.54, 2.63)	(-1.48, 0.37)	(-0.07, 1.82)
Mutual Funds	6.75	0.61	-2.06	-1.24	-0.14	0.5	-2.19	-0.77
		(0.15, 1.08)	(-3.59, -0.53)	(-4.13, 1.64)	(-1.85, 1.58)	(-0.97, 1.97)	(-3.47, -0.9)	(-2.09, 0.55)
Households	5.74	9.8	15.46	-11.51	-16.76	-9.05	-9.62	-17.01
		(5.07, 14.54)	(6.14, 24.79)	(-29.32, 6.29)	(-27.64, -5.87)	(-17.96, -0.14)	(-17.02, -2.22)	(-24.69, -9.32)
Pension	5.42	-0.1	-0.49	2.72	-0.29	-0.76	-0.98	-0.65
		(-0.27, 0.06)	(-1.49, 0.5)	(0.84, 4.6)	(-1.39, 0.81)	(-1.72, 0.2)	(-1.84, -0.13)	(-1.53, 0.22)
Banks	5.26	0.67	-0.54	-3.63	0.58	1.25	-1.0	2.3
		(0.42, 0.91)	(-2.06, 0.98)	(-6.51, -0.75)	(-1.09, 2.25)	(-0.25, 2.75)	(-2.33, 0.32)	(0.94, 3.66)
Insurance	2.59	0.99	0.19	1.73	-0.98	-0.84	-2.63	-0.71
		(0.66, 1.32)	(-0.86, 1.23)	(-0.24, 3.71)	(-2.16, 0.19)	(-1.84, 0.17)	(-3.51, -1.76)	(-1.61, 0.19)
ETF	1.18	-0.49	1.58	3.26	1.86	0.88	0.05	-0.36
		(-0.85, -0.13)	(0.44, 2.73)	(1.1, 5.43)	(0.58, 3.15)	(-0.22, 1.98)	(-0.91, 1.01)	(-1.34, 0.63)
Dealers	0.81	2.04	18.17	-10.23	-40.19	-15.6	10.3	-7.31
		(-6.61, 10.69)	(-7.05, 43.38)	(-57.99, 37.53)	(-68.68, -11.71)	(-39.89, 8.68)	(-10.76, 31.36)	(-28.97, 14.36)

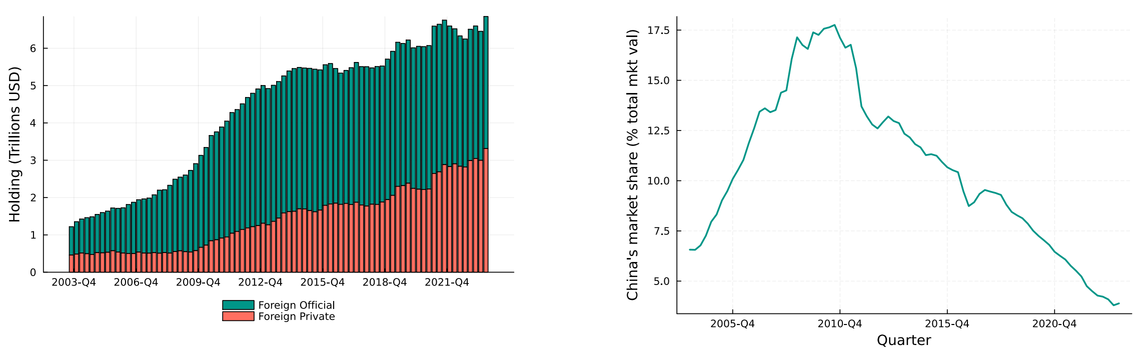
Note: Table C.5 reports the full set of estimated parameters and their associated 95% confidence intervals. ζ denotes price elasticities. S corresponds to the size weight used in the estimation, defined as the average market share of each entity throughout the sample period. Time-series averages are taken for sectors that assume to have different price elasticities before and after 2009Q1 (the sectors include Federal Reserve, Rest of World and U.S. Banks). $\varepsilon^{VIX}_{(std.)}$ and $\varepsilon^{USD}_{(std.)}$ refer to the demand loading on a positive VIX shock and USD index shock respective, obtained by taking the residual from AR(1) time-series regressions. ΔFFR refers to quarterly changes in the Federal Funds rate, and $Inf.$ denotes the deviation of CPI inflation from the 2% per annum inflation target. 95% confidence intervals are reported, with the standard errors given by Theorem 2.

Figure C.4: **Sectoral decomposition of latent demand and supply shocks**



Note: Figure C.4 further breaks down the contribution of latent demand and supply shocks (plotted in Figure 7) to the investor group level. We report average yield contribution per quarter in percentage points. The width of the bars reflects the length of the corresponding episode.

Figure C.5: **Foreign holdings of Treasury notes and bonds: Official holdings and China**



(a) Official vs. Private investors (level)

(b) China's share (% total market value)

Note: Figure C.5 provides additional context for understanding the influence of foreign demand on U.S. Treasury yields. Panel (a) plots the evolution of private and official holding of U.S. Treasury notes and bonds. Panel (b) specializes to China and plot its market share evolution over the sample period.