ISSN 0956-8549

Anatomy of the Treasury Market: Who Moves Yields?

By

Manav Chaudhary
Julie Zhiyu Fu
Haonan Zhou

FINANCIAL MARKETS GROUP DISCUSSION PAPER NO. 959

June 2026

Anatomy of the Treasury Market: Who Moves Yields?

Manav Chaudhary* Julie Zhiyu Fu† Haonan Zhou‡

First draft: November 15, 2024

This draft: April 19, 2026

[Click here for the latest version]

Abstract

Factor regressions provide a model-agnostic way to identify what drives Treasury yields, but not which investors respond. We develop an equilibrium framework that decomposes yield-factor regressions into investor-level drivers. An optimal estimator that weights idiosyncratic shocks across investors identifies each investor's demand response to yields and economic factors. Aggregating through market clearing recovers the standard factor regression, now decomposed by investor. Treasury demand is highly inelastic, with substantial heterogeneity across investors and time. Since 2008, foreign investors' influence has waned while the Federal Reserve's has grown. During flight-to-safety episodes, domestic—not foreign—investors drive the sharp decline in yields.

*Department of Finance, London School of Economics. M.Chaudhary6@lse.ac.uk

†Washington University in St. Louis, Olin School of Business. z.fu@wustl.edu

‡HKU Business School. haonanz@hku.hk

We thank Philippe van der Beck, William Cassidy, Aditya Chaudhry, Greg Duffee, Zhengyang Jiang, Xiang Fang, Xavier Gabaix, Iryna Kaminska, Rohan Kekre, Ralph Koijen, Lira Mota, Stefan Nagel, Andreas Neuhierl, Stijn van Nieuwerburgh, Simon Oh, Quentin Vandeweyer, and Motohiro Yogo for helpful comments and seminar participants at CUHK Shenzhen, HKU, HKUST, Bank of Korea, Peking University, WashU Olin, Brown University, HBS Junior Finance Conference, University of Zurich, the Chicago Booth Treasury Markets Conference (2025), Second UIC Finance Conference, the 2025 OFR Rising Scholar Conference, 2025 Sovereign Bond Markets International Conference, 2025 NBER Summer Institute Asset Pricing, EFA 2025, and University of Virginia for useful interactions. We thank Marco (Erchuan) Zhang for excellent research assistance. All errors are ours.

1 Introduction

Treasury yields move when changes in macroeconomic and financial conditions shift investors' desire to hold Treasuries. Factor regressions offer a way to learn what drives yields, without making assumptions about the structure of the economy: regress yields on a set of factors measuring the economic environment, and read off what moves them and by how much. But by jumping straight from factors to yields, they skip over which investors are responding and how much each contributes to the equilibrium yield change. For many questions, such as the effects of quantitative easing or the role of foreign savings in compressing yields, answering not only what factors drive yields but who responds to these factors is central.

Identification is the key challenge in decomposing the aggregate factor-yield relationship into its investor-level drivers. When a factor changes, investor holdings respond through two channels that must be separately identified. The first is direct: the factor shifts demand. The second is indirect: the direct demand shift creates an imbalance in the market, which causes yields and investor holdings to adjust in order for the market to clear.

This paper develops an equilibrium framework that separately identifies these two channels. Focusing on Treasuries as an asset class, we estimate investor-level factor and price loadings—how each investor's holdings respond to macroeconomic conditions and yields. This investor-level specification arises naturally as a log-linear approximation to a broad class of portfolio choice problems. It captures heterogeneity in investor objectives, constraints, and beliefs empirically, without requiring the researcher to take a stand on its structural origins. Aggregating these responses through market clearing recovers the standard factor regression—but now with a decomposition of each factor's yield effect into its investor-level drivers. Identification comes from the framework's equilibrium structure: idiosyncratic shocks to one investor's demand serve as instruments for yields in other investors' demand equations, allowing us to separate direct factor responses from indirect price-mediated adjustments.

We use our framework to conduct three exercises. First, we estimate price and factor sensitivities at both the market and investor levels—where investors are defined as aggregate sectors in the Financial Accounts. At the sector level, we find substantial heterogeneity in how different sectors respond to yields and macroeconomic factors. At the market level, aggregate demand is quite inelastic: a 1% increase in total Treasury demand raises prices by 0.83%, equivalent to a 12.8-basis-point decline given the Treasury market's average duration of 6.5 years. Second, we decompose historical yield movements into their sector-level determinants and find a notable shift in the Treasury market's structure since the financial crisis. Before the crisis, foreign investors were the primary drivers of yield compression, but their influence has since waned. Meanwhile, foreign investors have retreated from their role as liquidity providers, whereas the Federal Reserve has stepped in, providing a liquidity backstop when yields rise too high. Third, using sectors' heterogeneous loading on the market risk factor, we examine behavior during flight-to-safety episodes. Contrary to conventional wisdom, we find that domestic private investors increase their demand for Treasuries during market turmoil, whereas foreign investors, if anything, reduce demand.

Our modeling framework focuses on Treasuries as an asset class, abstracting from substitution and arbitrage across the yield curve. The log-linear specification allows us to empirically capture heterogeneous investor-level loadings on factors and yields, without taking a theoretical stance on the underlying differences in objectives, constraints, or beliefs that give rise to them. Substitution patterns across assets are similarly difficult to model explicitly. We show that our log-linear system absorbs the complex feedback effects into the investor's loading on Treasury yield and common factors.[1]

To estimate the model, we compile quarterly sector-level transactions and Treasury holdings from the Financial Accounts, supplementing them with country-level Treasury International Capital (TIC) data for foreign holders and Call Reports for banks. We construct a dataset of Treasury transactions for representative sectors that are mutually exclusive and collectively exhaustive of Treasury market activity. Hence, once we estimate the demand for the sectors, using market clearing, we can fully decompose yield changes into their sector-level drivers.

Estimating the model requires addressing the endogeneity of the sector's holdings and Treasury yields. The standard solution is to find idiosyncratic shocks to one sector's holdings and use these as instruments for yields in another sector's demand equation. Such instruments are exogenous if idiosyncratic shocks are uncorrelated across sectors and relevant if the reallocation induced by the shock impacts yields. The challenge in our setting is that we require instruments for all sectors. Existing Treasury market

[1]Structurally, our yield loading is a total (as opposed to partial) elasticity: the response of demand to a change in Treasury yields, accounting for equilibrium adjustments in substitute asset prices. A partial elasticity, by contrast, holds other prices fixed.

instruments—such as regulatory changes and balance sheet shocks—are unsuitable as they are limited to specific sectors or short-lived episodes. Finding appropriate instruments is further complicated by the need to use shocks that meaningfully affect Treasury prices—a significant challenge given the size of the Treasury market. If shocks induce only small price movements, instruments may lack relevance, weakening estimator power and introducing bias due to weak instrument problems.

To address these challenges, we develop an optimal GMM estimator that leverages our model's structure to identify idiosyncratic sector shocks and efficiently estimate parameters. In our framework, a sector's Treasury allocation depends on three components: (i) equilibrium yields, (ii) common drivers such as macroeconomic factors and other unobserved common factors, and (iii) idiosyncratic sector-specific drivers. Given the considerable heterogeneity in sectoral constraints and objectives, these idiosyncratic drivers are a significant source of variation in holdings. Our estimation strategy extracts these demand shifters for one sector and uses them as instruments for yields in another sector's demand equation. The identifying assumption—that idiosyncratic drivers are orthogonal across sectors—admits an overidentified system, which we estimate by optimally weighting moment conditions, giving greater weight to shocks from sectors that exert a stronger influence on equilibrium yields. Our estimator is an optimal GMM implementation of Granular Instrumental Variables (GIV; Gabaix & Koijen, 2024), allowing for flexible estimation in the presence of heterogeneous investors and time-varying price elasticities.[2]

First, we quantify price and factor sensitivities at both the sector and market levels. At the sector level, we find significant heterogeneity in price elasticities across sectors. On the one hand, the household sector—which captures the holdings of hedge funds, endowment funds, and family offices—is relatively sensitive to yield changes. On the other hand, ETFs and pension funds' allocation and the government's Treasury issuance decisions are very price inelastic, responding minimally to yield changes.

Moreover, price elasticities have changed considerably since the financial crisis for key sectors, suggesting a shift in the "macro structure" of liquidity provision. Foreign investors have become less price elastic, a significant development given their large presence in the Treasury market. In contrast, the Federal Reserve increasingly acts as a state-contingent liquidity provider, purchasing Treasuries when yields and market stress (as measured by the VIX) are high. This pattern aligns with the insurance channel of central

[2]We make our optimal GIV estimator package available at https://github.com/FuZhiyu/OptimalGIV.jl (Julia) and https://github.com/FuZhiyu/optimalgiv (Python).

bank asset purchases (Haddad, Moreira, & Muir, 2024).

At the market level, the aggregate demand for Treasuries is relatively inelastic. We estimate a macro multiplier of 0.83, meaning that a 1% increase in Treasury demand (as a percentage of total outstanding) raises the overall market value of Treasuries by 0.83%. Expressed in yield terms, this corresponds to a 12.8-basis-point decline for Treasury bonds with the average duration (6.5 years).[3]

Second, applying market clearing, we link Treasury yields to sector-level allocation decisions, decomposing the aggregate yield movement into its investor-level components. This decomposition reinforces the notion that there has been a notable shift in the Treasury market structure since the financial crisis. Before 2008, foreign investors were the primary drivers of yield movements, exerting downward pressure of 120 basis points per year between 2003 and 2007—consistent with the literature's emphasis on the global savings glut during this period. However, their influence on U.S. yields has declined markedly since the crisis. Meanwhile, the Federal Reserve has emerged as a central driver of yields through its direct Treasury purchases, particularly during and following crisis episodes via quantitative easing operations. This marks a significant shift from the pre-crisis period when the Fed's influence on yields operated primarily through policy rate decisions rather than Treasury purchases.

Third, we use our model to examine investor behavior in flight-to-safety episodes. These are periods when rising market risk coincides with declining Treasury yields. A common explanation for this pattern is that foreign investors engage in flight-to-safety, increasing their demand for U.S. Treasuries as market risk rises. However, we find no significant evidence that foreign investors increase their Treasury demand when the VIX rises. In fact, we observe net foreign selling pressure of Treasuries during flight-to-safety episodes. This behavior is more consistent with a precautionary saving view that investors save in Treasuries during normal times and sell them during flight-to-safety episodes, taking advantage of the hedging property of U.S. Treasuries in particular (Acharya & Laarits, 2024; Acharya & Pedersen, 2005; Bianchi, Hatchondo, & Martinez, 2018; Brunnermeier, Merkel, & Sannikov, 2024). In contrast, the domestic sector significantly increases its Treasury demand in response to rising market risk, consistent with flight-to-safety behavior.

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

1.1 Related Literature

A burgeoning literature highlights the important role of market participants in shaping Treasury market yields, and this paper makes several key contributions to this literature. Most notably, Koijen, Koulischer, Nguyen, and Yogo (2017) employs a demand system for government bonds in the euro area to examine the portfolio rebalancing mechanism of quantitative easing; Jiang, Richmond, and Zhang (2024) estimates the portfolio choice model using cross-country bilateral portfolio holdings data to understand the exorbitant privilege the U.S. commands on its debt; Jiang, Richmond, and Zhang (2025) extends the framework to further study its effect on the strength of the U.S. dollar. Focusing on the U.S. Treasury market, Cavaleri (2023) studies different sectors' demand for convenience services, Eren, Schrimpf, and Xia (2023) estimate the sectoral elasticity for Treasuries using monetary policy surprises, and Jansen, Li, and Schmid (2024) estimates institutional investors' demand elasticities and substitution patterns across the term structure of the Treasury yield curve. Among these, Jansen et al. (2024) is closest to our work. They marry a demand-based approach with a Vayanos and Vila structural model to study how sector-specific demand shocks propagate through the term structure via risk-averse arbitrageurs who intermediate across maturities. We explore a complementary problem: unpacking the reduced-form factor regression into its investor-level drivers—identifying who responds to which macroeconomic factors and how much each sector contributes to equilibrium yield movements. This motivates our methodological choices—a relatively model-agnostic framework that generalizes factor regressions to account for investor holdings, and an optimal GIV estimator that instruments for all sectors simultaneously without relying on sector-specific external instruments.

By providing the first direct macro elasticity estimate for U.S. Treasuries, we advance the literature studying the role of demand shocks in driving overall Treasury market yields (Greenwood & Vayanos, 2010; Hanson & Stein, 2015; Krishnamurthy & Vissing-Jorgensen, 2012), while also contributing to the broader demand-system approach to asset pricing.[4] Our macro multiplier estimate of 0.83 suggests that the U.S. Treasury market absorbs demand shocks more effectively than the U.S. equity market, which has a macro multiplier of 5 (Da, Larrain, Sialm, & Tessada, 2018; Gabaix & Koijen, 2022; Haddad, Huebner, & Loualiche, 2025; Hartzmark & Solomon, 2025), and the U.S. corporate bond market, which has a portfolio-level multiplier of 3.5 (Chaudhary et al., 2023). Our Trea-

[4]Initially applied to the cross-section of stocks (Koijen & Yogo, 2019), variants have since been used to study various asset classes, including the equity market (Gabaix & Koijen, 2022), bond market (Bretscher, Schmid, Sen, & Sharma, 2020; Chaudhary et al., 2023; Fang & Xiao, 2025), and exchange rates (An & Huber, 2024; Camanho, Hau, & Rey, 2022; Jiang et al., 2025; Koijen & Yogo, 2024).

sury macro-multiplier estimates also align with the magnitudes implied by existing price impact event studies (e.g., Bernanke, Reinhart, & Sack, 2004; Gagnon, Raskin, Remache, & Sack, 2011a; Warnock & Warnock, 2009, among others).

Our findings further deepen our understanding of the role of different sectors in the Treasury market. Our decomposition demonstrates that foreign investors were the main drivers of yield compression in the pre-crisis period, consistent with the global savings glut hypothesis (Bernanke, 2005; Caballero, 2006; Jiang, Richmond, & Zhang, 2024). However, we show that their influence has since diminished considerably. Our analysis also sheds light on two competing views in the literature about foreign demand for Treasuries during flight-to-safety episodes, thereby advancing our understanding of U.S. exorbitant privilege (Campbell, Pflueger, & Viceira, 2019; He, Nagel, & Song, 2022; Kekre & Lenel, 2024). The flight-to-safety view holds that foreign investors have heightened demand for Treasuries during crises (Jiang, Krishnamurthy, & Lustig, 2021; Jiang, Krishnamurthy, & Lustig, 2024; Kekre & Lenel, 2024), while the precautionary saving view suggests that foreign investors save in Treasuries during normal times and sell them to raise liquidity during crises (Bianchi et al., 2018, among others). Our estimates provide empirical support for the precautionary saving view. Our findings also echo the literature documenting the outsized role of quantitative easing by the Federal Reserve in driving yields down since the financial crisis (Haddad et al., 2024; Krishnamurthy & Vissing-Jorgensen, 2011). We also speak to the broader literature studying demand for Treasuries by particular sectors, such as banks (Jiang, Matvos, Piskorski, & Seru, 2024), mutual funds (Huang, Jiang, Liu, & Liu, 2025; Selgrad, 2023), insurance companies (Chaudhary, 2024; Koijen & Yogo, 2023), and broker-dealers (Du, Hébert, & Li, 2023; Greenwood & Vayanos, 2014; Vayanos & Vila, 2021).

Econometrically, our approach extends the granular instrumental variable (GIV) literature (Gabaix & Koijen, 2024). GIV has been increasingly applied in macro-finance settings where finding market-level instruments is typically difficult (Adrian, Grinberg, Liang, Malik, & Yu, 2022; Camanho et al., 2022; Chodorow-Reich, Gabaix, Koijen, & Viviano, 2024; Kubitza, Sigaux, & Vandeweyer, 2024; Kundu & Vats, 2021). The GIV extension here accommodates heterogeneous price elasticities and factor loadings across investors without requiring explicit parameterization based on observable characteristics. We also propose an economically motivated optimal Generalized Method of Moments (GMM) estimator. It assigns greater weight to shocks from investors with greater influence on equilibrium yields. This ensures instrument relevance—a key challenge in a market as large as U.S. Treasuries—and yields an efficient estimator. The approach builds on other efficient GIV variants, such as Baumeister and Hamilton (2023) and Qian (2024), which characterize optimality statistically rather than economically. It has since been adopted in recent work on demand-driven price impact in equities (Chaudhry & Li, 2025) and currencies (Kim & Kim, 2025).

1.2 Roadmap

The remainder of the paper is organized as follows. Section 2 introduces an equilibrium model of the Treasury market that can be taken to the data. Section 3 outlines our estimation and identification approach. Section 4 reports our estimate of the Treasury macro multiplier. Section 5 discusses the sector- and factor-specific drivers of Treasury yield and examines their time variation. Section 6 examines the role of foreign investors in explaining the countercyclical aggregate demand for Treasuries. Section 7 concludes.

2 The Treasury Market: Data and Model Framework

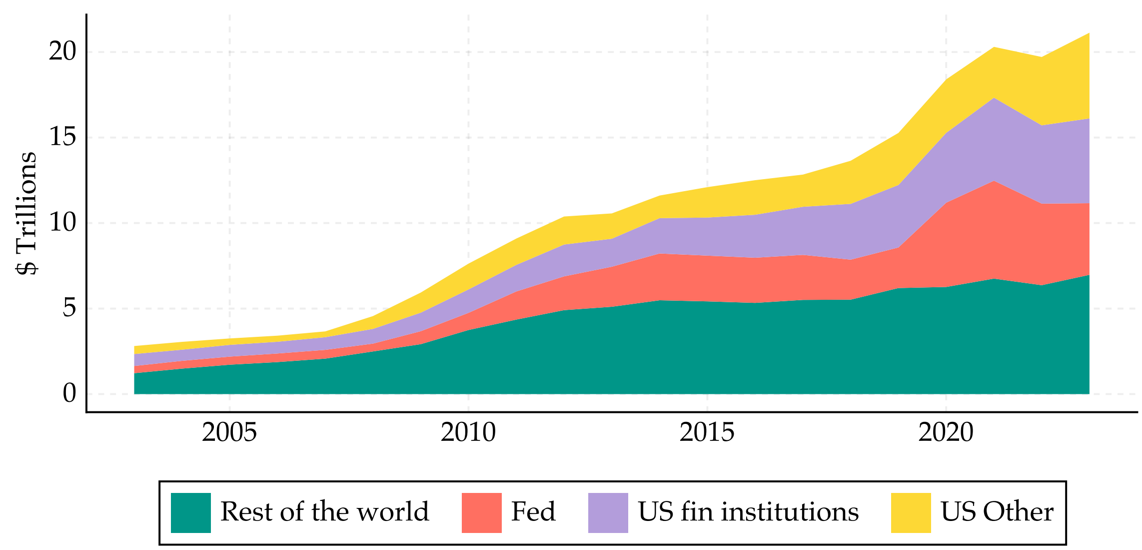
Our analysis focuses on the market portfolio of Treasury notes and bonds (Treasury securities with maturities exceeding one year, referred to as *Treasury bonds*). Yields on this portfolio are important indicators for the overall funding costs of the U.S. government and serve as benchmarks for analyzing other long-term interest rates in the economy. For this purpose, we combine sectoral holdings data, to be introduced in Section 2.1, and an empirically tractable model framework that we outline in Section 2.2. To sharply characterize how Treasury bond prices are shaped by investors' allocation and government issuance decisions on aggregate, we abstract away from modelling the entire Treasury yield curve. Our empirical framework, however, can accommodate multiple assets and maturities, and we formally discuss how accounting for substitutes affects the interpretation of our estimates later in the analysis.

2.1 Data

The primary source of our Treasury bonds holdings and issuance data is the Federal Reserve's Financial Accounts (Z.1). This dataset reports positions and transactions (flows) in Treasury notes and bonds at the investor sector level. We complement these data with more granular information for banks and the foreign sector. For banks, we use Call Reports to collect individual commercial banks' holdings of and flows in and out of U.S. Treasury securities, following the procedure of Jansen et al. (2024) to impute Treasury holdings from reported Treasury and agency securities. For the foreign sector, we draw on Treasury International Capital (TIC) data to disaggregate holdings by country.[5]

Figure 1 shows the evolution of the market value of Treasury bonds holdings from 2003Q4 to 2023Q4, our sample period. Treasury bonds are held by four broad groups: (i) U.S. financial institutions, including banks, mutual funds, insurers, and pension funds; (ii) other U.S. investors, primarily households and non-financial institutions; (iii) the Federal Reserve; and (iv) foreign investors. Over our sample period, the total supply of Treasury bonds increased from less than $5 trillion to more than $20 trillion, rising from roughly 30% of GDP to 70%.[6]

Figure 1: **Holdings by Group: U.S. Treasury Notes & Bonds**



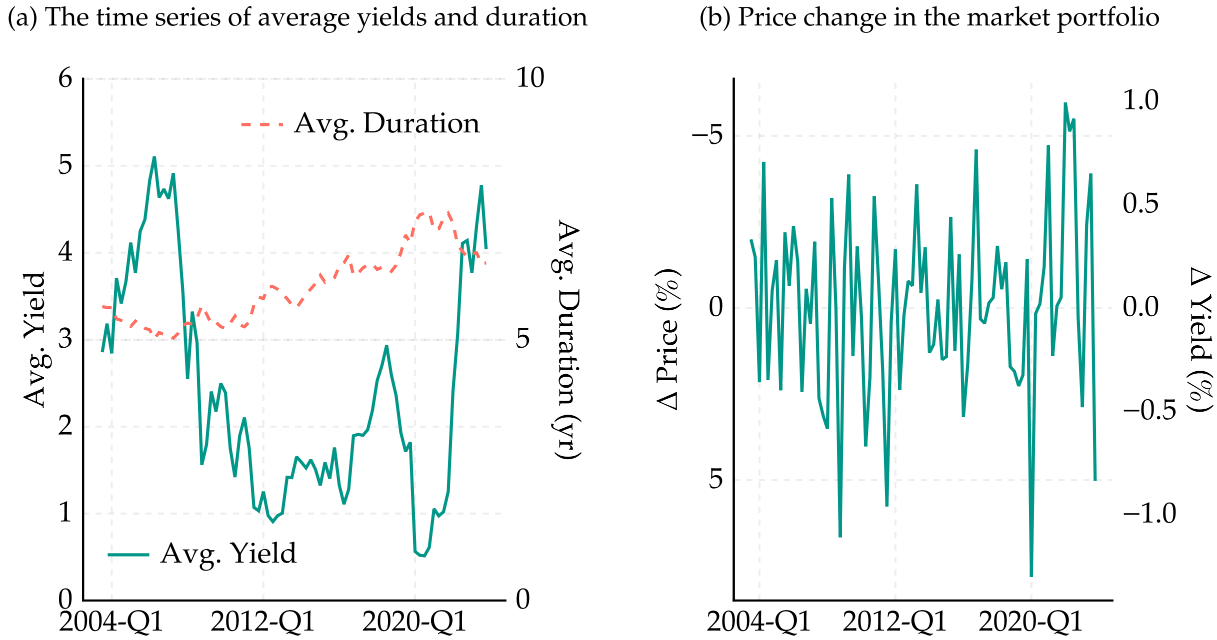
Note: Figure 1 plots the total market value of U.S. Treasury notes and bonds holdings by investor groups (panel (a)). We aggregate investor sectors reported in Financial Account (Z.1) to four investor categories. *US Financial institutions* include banks, ETFs, mutual funds, pension funds, insurance companies, and security brokers and dealers. *US Other* include other smaller domestic categories, most notably households and nonfinancial corporates. Appendix Table B.1 reports the detailed mapping.

We measure the Treasury bond yield using the CRSP US Treasury Database. Using security-level data, we calculate the average yield of a market-weighted portfolio of Treasury notes and bonds that have maturities exceeding one year. Figure 2a shows that

[5]Appendix B.1 provides more details on data construction, including the use of the entire Treasury securities holdings for sectors that do not report breakdowns, the imputation procedure for individual bank holdings, and the adjustment of reported discrepancies.

[6]Total federal debt stood at about 120% of GDP at the end of 2023, including notes and bonds, bills, and non-marketable securities.

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.

average yields during our sample period fluctuate between 0.5 and 5 percent, while the average duration of the market portfolio ranges from 5 to 7 years. Figure 2b plots the quarterly changes in the price of the market portfolio. We flip the sign of the y-axis so that a drop in the price, or equivalently, an increase in the yield, corresponds to a rise. We also mark the corresponding changes in average yields on the right axis.

2.2 An Equilibrium Model

We outline our equilibrium model of the Treasury market. The model allows us to connect the yield we observe on Treasury bonds to the data we have collected on sectors' allocation decisions and the government's issuance decisions. The framework is parsimonious yet sufficiently flexible to incorporate sector-specific differences in objectives, constraints, beliefs, and substitution patterns across assets.

Notations We index sectors by i (e.g., insurance companies, the Federal Reserve). Treasury supply is treated as one of the sectors with negative demand. For banks and foreign investors, we disaggregate further to the individual bank or country level, and refer to these banks and countries also as sectors. We use bold symbols for vectors and matrices ($\mathbf{q}_t$) and regular symbols for scalars ($q_{i,t}$). Key variables include: $Q_{i,t}$ for sector i's Treasury holdings in face value terms, $A_{i,t}$ for the holdings in market value terms, P_t for the Treasury market portfolio price, and y_t for the average yield.

Portfolio choice Investor sectors allocate to Treasuries under distinct frictions. Hedge funds target short-horizon returns subject to Value-at-Risk (VaR) limits and redemption risk; insurers buy-and-hold to match long-duration liabilities under capital rules; banks hold Treasuries for liquidity management, compliance, and duration hedging. Fully microfounding each sector is unwieldy given these interacting constraints. In order to flexibly capture these diverse drivers of Treasury demand, we leverage the fact that the first-order approximation of a large class of portfolio choice problems can be expressed as a log-linear function of prices and demand shifters:

$$\Delta q_{i,t} = \overline{\Delta q}_i - \zeta_i \Delta p_t + \nu_{i,t}. \tag{2.1}$$

Here, $\Delta q_{i,t}$ is the percent change in sector i's Treasury holdings, Δp_t is the percent change in the price of the Treasury market portfolio, ζ_i is sector i's price elasticity, $\overline{\Delta q}_i$ is a sector-specific, time-invariant intercept, and $\nu_{i,t}$ is a demand shifter. This expression follows from log-linearizing a Treasury demand or supply function around the long-run mean and differencing. The approach lets the data capture the motives behind portfolio choice via elasticities and shifters, and it locally nests a wide range of models. In Appendix A.6, we discuss in detail the structural interpretation of elasticity ζ_i and shifters $\nu_{i,t}$ in various portfolio choice models, ranging from standard mean-variance portfolio choice to demand for convenience and learning-from-price models.

Interpreting the elasticity ζ_i in a multi-asset setting. Equation (2.1) represents a single-asset demand curve. In Appendix A.7, we show that this formulation nests a richer multi-asset demand system in which Treasury holdings depend not only on their own price but also on the prices of substitute assets. The key insight is that, by imposing market clearing conditions for these substitutes, we can express their equilibrium prices as functions of Treasury price changes and demand shifts. This collapses the multi-asset system to a reduced-form single-asset demand curve with a *total elasticity* ζ_i, which embeds the feedback effects from cross-asset substitution.

To gain intuition, consider substitution between cash Treasuries and Treasury futures as in basis trades. In this case, ζ_i captures the overall responsiveness of sector i's *cash Treasury* holdings to a 1 percent change in Treasury prices induced by Treasury demand shifts, accounting for the resulting equilibrium adjustments in futures prices. Its reciprocal corresponds to the price impact of a unit demand shock to Treasuries. This contrasts with a partial own-price elasticity, which holds Treasury futures prices fixed. For many policy applications—such as quantitative easing or large-scale foreign official flows—the total elasticity is the relevant parameter, as policymakers care about overall Treasury price movements. Our approach is designed to answer such questions.[7]

In short, the single-asset representation aggregates the matrix of partial own- and cross-price elasticities into a single parameter ζ_i, reducing dimensionality and improving tractability.

Demand shifters $\nu_{i,t}$ The demand shifters $\nu_{i,t}$ have two components: common factors $\boldsymbol{\eta}_t$ that affect all sectors and idiosyncratic shocks $u_{i,t}$ specific to each sector. This gives us:

$$\Delta q_{i,t} = \overline{\Delta q}_i - \zeta_i \Delta p_t + \underbrace{\boldsymbol{\lambda}_i \boldsymbol{\eta}_t + u_{i,t}}_{\nu_{i,t}}. \tag{2.2}$$

Common factors $\boldsymbol{\eta}_t$ capture economy-wide drivers such as monetary policy changes, risk sentiment shifts, or drifts in the inflation expectation. While all sectors are exposed to these factors, they respond with different intensities captured by sector-specific loadings $\boldsymbol{\lambda}_i$. For example, insurance companies may be more sensitive to long-term interest rate changes than hedge funds. The idiosyncratic component $u_{i,t}$ captures sector-specific drivers such as regulatory changes, portfolio rebalancing needs, or liquidity constraints that are orthogonal to common factors.

Equilibrium We connect sector-level demand and supply to Treasury yields through market clearing. In equilibrium, all Treasury bonds must be held by someone—the net flow across all sectors, including Treasury supply, must sum to zero. Formally, we weight each sector's flow $\Delta q_{i,t}$ by its average market share S_i,[8] to arrive at the market clearing condition:

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

[7] Our estimates do not speak to questions regarding substitutability between Treasuries and futures and the arbitrage frictions between these assets. These questions require explicitly modeling both assets.

[8] Our empirical counterpart of S_i is the long-run average market share rather than the current period share (see Section 3). This aligns with our log-linear approximation around the steady state and avoids noise from transitory volatility in market shares.

Treasury price adjusts to clear the market. To find the equilibrium price, we substitute our demand equations into the market clearing condition. This gives us the pricing equation:

$$\Delta p_t = \frac{1}{\zeta_S}\left(\overline{\Delta q}_S + \underbrace{\boldsymbol{\lambda}_S \boldsymbol{\eta_t} + u_{S,t}}_{\nu_{S,t}}\right), \tag{2.3}$$

where the S subscript denotes size-weighted aggregation across all sectors. For example, $\zeta_S \equiv \sum_i S_i \zeta_i$ is the aggregate elasticity and $\boldsymbol{\lambda}_S \equiv \sum_i S_i \boldsymbol{\lambda}_i$ is the vector of aggregate factor loadings.

Equation (2.3) shows that Treasury price changes are determined by aggregate demand shifts, scaled by the inverse of the aggregate elasticity, $\frac{1}{\zeta_S}$. We refer to this inverse elasticity as the *macro multiplier*—it captures the percentage change in Treasury prices from a 1% change in aggregate net demand. Importantly, this multiplier reflects the total equilibrium effect, including amplification from cross-asset substitution as Treasury price movements propagate through other markets and feed back into Treasury demand. We provide formal derivations based on a multi-asset demand system in Appendix A.7.

Once we obtain estimates for sector-specific elasticities ζ_i and factor loadings $\boldsymbol{\lambda}_i$, we can decompose any Treasury price movement into contributions from individual sectors and common factors. The contribution of sector i to the price change is given by:

$$\Delta p_{i,t} = \frac{1}{\zeta_S} S_i \left(\overline{\Delta q}_i + \boldsymbol{\lambda}_i \boldsymbol{\eta}_t + u_{i,t}\right). \tag{2.4}$$

This decomposition reveals which investors drive Treasury price movements during different episodes. For example, consider a positive inflation surprise that causes Treasury prices to decline. Our framework can determine both the direct effect of the inflation shock and which specific investors (banks, foreign central banks, etc.) amplify or dampen this shock through their trading responses.

To implement this decomposition, we need to estimate the elasticities ζ_i and factor loadings $\boldsymbol{\lambda}_i$. This presents a challenge: prices and quantities are jointly determined in equilibrium, creating simultaneity bias in standard estimation approaches. In Section 3, we lay out our identification strategy.

3 Identification and Estimation

3.1 Specification Details

We estimate the following generalized version of the demand function for each sector i, and then connect it to the yield using the market clearing condition:

$$\Delta q_{i,t} = -\zeta_{i,r(t)}\Delta p_t + \boldsymbol{\lambda}_{i,r(t)}\boldsymbol{\eta}_t + \overline{\Delta q}_i + u_{i,t} \tag{3.1}$$

where $\Delta q_{i,t}$ is the percent change in holdings of Treasury bonds by sector i at time t, Δp_t is the percent change in the price of the Treasury market portfolio, $\zeta_{i,r(t)}$ is sector i's price elasticity in regime $r(t)$, $\boldsymbol{\lambda}_{i,r(t)}$ is sector i's factor loadings in regime $r(t)$, $\boldsymbol{\eta}_t$ is the common demand shifter, $\overline{\Delta q}_i$ captures sector-i fixed effect, and $u_{i,t}$ is sector i's idiosyncratic demand shifter. Below we describe each component in more detail.

Sample period Our baseline analysis covers the period from 2003Q4 to 2023Q4. We also explore data starting from 1970Q1, based solely on Financial Accounts data (without Call Reports or TIC data), and arrive at similar estimates relative to the baseline.

Sector definitions and groupings We use the sectoral definitions from the Financial Accounts. The complete list of sectors and their mapping to the aggregated sectors is provided in Appendix Table B.1. For interpretability and statistical power, we pool coefficients of similar sectors in our estimation. For example, all banks covered in the Call Reports are constrained to have the same elasticity, as are all pension fund types from the Financial Accounts. One sector worth highlighting in the Financial Accounts is the "household sector," which serves as the residual sector to ensure the sum of all sectors' flows equals zero. This sector should be broadly interpreted as a catch-all category for investors not included or observed in other sectors, such as hedge funds, endowment funds, and family offices.[9] In addition, we separately consider the Federal Reserve and the Treasury in this analysis, by estimating their price elasticities and factor loadings separately, in order to highlight the potential distinct roles played by each policy entity as buyers and sellers of Treasuries in shaping price dynamics.

[9] In particular, household holdings of Treasuries are substantially driven by hedge funds domiciled in the Cayman Islands, while TIC data heavily undercounts Cayman-held Treasury securities (Barth et al., 2025). Such funds are predominantly managed by U.S. investors (Cayman Islands Monetary Authority, 2023). Enhanced Financial Accounts and Form PF data (2013 onwards) suggests that domestic and foreign hedge fund positions can account for more than 70 percent of household sector holdings of Treasury securities in recent years.

Flows and positions The flow measure $\Delta q_{i,t}$ is defined as the net dollar flows $F_{i,t}$ by investor i, normalized by the sector's long-run average market share S_i times the total market values of Treasury bonds outstanding in the last quarter A_{t-1}^{total}:

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

More specifically, the long-run average of the market share S_i is constructed as the average market share of each entity throughout the sample period:

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

where $A_{i,t}$ denotes the market value of sector i's Treasury holdings as of time t. We scale the denominator by the long-run average share rather than the previous quarter's share for two main reasons. First, using the long-run average share aligns with our interpretation of the log-linear demand system as a first-order approximation of the portfolio choice problem around the steady state (measured by the long-run average). Second, the long-run average offers a statistical advantage as it is not affected by transitory volatility in market shares, which could otherwise lead to noisy estimates.[10]

Supply enters the system as negative demand. We choose the sign convention such that S_i are always positive, and $\Delta q_{i,t}$ 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, the Department of the Treasury issued $\$876.5$ billion of Treasury bonds, and the total outstanding bonds by the end of 2023Q3 was $\$20{,}370.8$ billion, so that the flow from the supply side in Q4 is $\Delta q_{supply,2023Q4} = \frac{-\$876.5}{1\times\$20{,}370.8} = -4.3\%$. 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.

Common factors Common factors play two crucial roles in our analysis. First, they enable us to decompose and interpret price changes in the Treasury market. Second, as we will outline later, our identification strategy relies on properly controlling for common factors to isolate idiosyncratic demand shocks.

We incorporate common factors in our model guided by both economic theory and their empirical explanatory power for yields. To this end, we include a comprehensive set

[10] We have also examined the analysis using time-varying market shares and found that the results are generally consistent with our baseline findings.

of common factors. These include several key macro-financial variables: (i) innovations to the VIX (CBOE Volatility Index), which serves as a primary indicator of global risk sentiment and measures "flight to safety" intensity; (ii) quarterly changes in the effective Fed funds rate to account for short-term interest rate movements and their transmission to longer-term rates; (iii) innovations to the USD broad index; and (iv) quarterly changes to long-term inflation expectations (Cleveland Fed).[11] We also include consensus forecasts from the Survey of Professional Forecasters (SPF) regarding expected changes in one-quarter and ten-year Treasury yields for the current quarter. These forecasts capture market expectations about yield curve changes not reflected in other variables. We do not include other prices that are also endogenously affected by idiosyncratic shocks in the Treasury market, such as credit spreads. Our robustness exercises, documented in Appendix B.4, indicate that including additional macro factors provides minimal improvement in explanatory power while increasing the risk of overfitting.

In addition to these macroeconomic factors, we need to account for the fact that policymakers' announcements influence Treasury bond demand across all sectors. Two key policy entities, the Federal Reserve (Fed) and the Department of the Treasury (supply), regularly pre-announce their planned purchases and sales to provide transparency and forward guidance to market participants. To capture the anticipatory effects and potential front-running behaviors these announcements may induce, we include variables that can predict the policy sectors' flows in our common factors. Specifically, for the Fed, we control for purchases scheduled a quarter ahead.[12] For the Treasury, we include net issuance in the previous quarter to control for the anticipated component in Treasury supply. This approach allows us to isolate and measure the flow sensitivity due to deviations from scheduled purchases and sales when estimating coefficients for these policy sectors.

To capture any remaining common factors, we also extract unobserved factors from granular data. Specifically, we estimate principal components (PCs) externally from the Survey of Professional Forecasters (SPF) panel data on 10-year yield forecasts, and internally from country-level and bank-level Treasury flows. Our baseline specification includes four externally estimated PCs from the SPF yield forecasts. In robustness checks, we expand this to include up to 10 external PCs and also incorporate internally estimated PCs from bank and foreign investor flows. The results remain essentially unchanged.

[11] More broadly, accounting for macro factors is crucial to understanding Treasury price dynamics because they convey information not captured by the cross-section of yields (Ludvigson & Ng, 2009).

[12] Appendix B.2 discusses the construction of the scheduled purchase series.

Accounting for structural breaks We leverage the flexibility of our framework to include regime-specific price elasticities and/or factor loadings for a selected set of sectors (the Fed and the foreign sector in our baseline estimation), to study the impact of significant changes in monetary policy framework and foreign investor demand since the Global Financial Crisis. The indicator $r(t)$ in (3.1) captures the regime-specific coefficients.

3.2 Identification Assumption

The main challenge in estimating the model is addressing the endogeneity arising from the interaction between Treasury prices and investors' allocation decisions. The standard approach in the literature is to search for idiosyncratic demand shifters from one sector and use these to instrument the price in another sector's demand equation. These instruments are valid if one sector's idiosyncratic shifters are orthogonal to the other sectors' idiosyncratic shifters. And they are relevant as long as the idiosyncratic shocks are large enough to have a meaningful price impact. For example, flow-induced-trading by mutual funds has been successfully used as a source of idiosyncratic demand shocks to estimate stock and corporate bond elasticities (Chaudhary et al., 2023; Lou, 2012).

The challenge with such approaches in our setting is that we require instruments for all sectors. While previous research has used regulatory changes and balance sheet shocks as sources of idiosyncratic demand shocks, these instruments are unsuitable for our analysis because they are typically limited to particular sectors or short-lived episodes. Furthermore, finding appropriate instruments is especially difficult in the Treasury market due to its enormous size—individual sector shocks often are not large enough to meaningfully affect Treasury prices, resulting in instruments not being relevant.

To address these challenges, we leverage our model's structure to extract idiosyncratic sector shocks. In our model, each sector's Treasury allocation depends on three components: (i) price (Δp_t), (ii) common demand shifters ($\boldsymbol{\eta}_t$), and (iii) idiosyncratic sector-specific shifters ($u_{i,t}$). The considerable heterogeneity in sectoral constraints and objectives makes these idiosyncratic drivers a significant source of variation in holdings. Our estimation strategy extracts these idiosyncratic sector-specific demand shifters from one sector and uses them as instruments for the yield in another sector.

Similar to standard instrumental variable approaches, our identification relies on idiosyncratic sector-specific drivers being orthogonal across sectors. For instance, selling Treasuries by a foreign country may be driven solely by their domestic concerns, and hence this can be used as an instrument for U.S. pension funds' demand. Formally, our identification assumption is based on conditional independence of idiosyncratic sector-specific shifters ($u_{i,t}$) across sectors:

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

Assumption 1 is also used in the Granular Instrumental Variables (GIV; Gabaix & Koijen, 2024) approach, which has seen increasing application in macro-finance research (Adrian et al., 2022; Camanho et al., 2022; Chodorow-Reich et al., 2024; Kubitza et al., 2024; Kundu & Vats, 2021). As such, our estimator to be outlined below can be viewed as an extension of the standard GIV estimator, but with important differences. We offer a detailed comparison in Appendix A.4.[13]

For ease of exposition, we develop the optimal GIV estimator treating Assumption 1 as exact. Strict independence of the idiosyncratic shocks requires that we capture all common variation; despite our extensive controls, some residual common factors may remain in $u_{i,t}$. In Section 3.4, we characterize the estimator's behavior under such violations, develop sharp diagnostics to quantify any resulting bias, and propose corrective strategies. The bias rises only modestly with residual common variation, and our robustness exercises indicate that such variation is unlikely to be consequential.

3.3 The Optimal GIV Estimator

Based on Assumption 1, we propose an economically motivated optimal generalized method of moments (GMM) estimator that weights investor shocks according to their market influence. This approach assigns greater importance to shocks from investors who have more significant impact on equilibrium yields, thereby enhancing the relevance of idiosyncratic demand shocks as the instrument for yields.

For ease of exposition, we proceed *as if* there are no common factors in the model. Operationally, this can be achieved by first residualizing $\Delta q_{i,t}$ and Δp_t against the common factors and taking the residuals. We discuss the handling of common factors in Section 3.4. For exposition, here we also assume that elasticities are fully heterogeneous across sectors and constant across regimes. Appendix A.1 derives the properties of our estimator in a fully general demand system with time-varying elasticities, factor loadings

[13]The key difference from the standard implementation is that our GIV extension *flexibly* accommodates heterogeneous price elasticities and factor loadings across investors without requiring explicit parameterization based on observable characteristics. In addition, our optimal GMM implementation more efficiently exploits information contained in idiosyncratic shocks.

and sector sizes. We show how to incorporate external common factors in Appendix A.3.

Our identification Assumption 1 implies the following moment restrictions,

$$\mathbb{E}\left[u_{i,t}u_{j,t}\right] = 0, \quad i, j = 1, \ldots, N, i \neq j. \tag{3.2}$$

The set of moment conditions (3.2) yields an overidentified system of equations. More specifically, they place restrictions on the entire off-diagonal covariance matrix of idiosyncratic sector-specific shifters, giving us $N \times (N-1)/2$ moment conditions to identify just N sector-specific price elasticities. We would like to optimally weight the moment conditions. In theory, we could adopt efficient versions of generalized method of moments (GMM) estimators off the shelf. However, a practical challenge is that the number of moment conditions $N \times (N-1)/2$ could far exceed the length of the time-series T in the sample. In such settings, statistical methods to find the optimal weighting matrix, such as two-step GMM, are known to perform poorly (Han & Phillips, 2006; Newey & Windmeijer, 2009).

Guided by the economics of our model, we derive an optimal weighting scheme for the moment conditions. This optimal estimator assigns greater weight to moment conditions from entities with larger market shares. The rationale is straightforward: idiosyncratic shocks of entities with larger market shares generate greater price impact, making them more relevant as instruments. We refer to this size-weighted approach as the *optimal GIV estimator*.

Formally, the optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ solves the following sample moment conditions:

Definition 1. The sample moment conditions for the optimal GIV estimator are given by

$$\hat{\mathbb{E}}\left[\hat{u}_{i,t}\left(\hat{\zeta}_i\right) \underbrace{\sum_{j\neq i} S_j \hat{u}_{j,t}\left(\hat{\zeta}_j\right)}_{\hat{u}_{S(-i),t}\left(\hat{\boldsymbol{\zeta}}\right)}\right] \equiv \hat{\mathbb{E}}\left[\left(\Delta q_{i,t} + \hat{\zeta}_i \Delta p_t\right) \sum_{j\neq i} S_j \left(\Delta q_{j,t} + \hat{\zeta}_j \Delta p_t\right)\right] = 0. \tag{3.3}$$

for all i, where $\hat{\mathbb{E}}$ denotes the sample mean across the time dimension.

In Appendix A.2 we show that the estimator is asymptotically efficient.

Theorem 1 (Asymptotic efficiency)**.** *Given the moment conditions (3.2), under regularity conditions laid out in Appendix A, the optimal GIV estimator* $\hat{\boldsymbol{\zeta}}$ *is consistent and asymptotically normal:*

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

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

$$\mathbf{V}^{\zeta} = \zeta_S^2 \times \text{Inv}\left(\begin{bmatrix} \frac{1}{\sigma_1^2}\sum_{i\neq 1} S_i^2\sigma_i^2 & S_1S_2 & S_1S_3 & \cdots & S_1S_N \\ S_1S_2 & \frac{1}{\sigma_2^2}\sum_{i\neq 2} S_i^2\sigma_i^2 & S_2S_3 & \cdots & S_2S_N \\ S_1S_3 & S_2S_3 & \frac{1}{\sigma_3^2}\sum_{i\neq 3} S_i^2\sigma_i^2 & \cdots & S_3S_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_1S_N & S_2S_N & S_3S_N & \cdots & \frac{1}{\sigma_N^2}\sum_{i\neq N} S_i^2\sigma_i^2 \end{bmatrix}\right), \tag{3.4}$$

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

The expression (3.4) illustrates the sources of statistical power in the optimal GIV estimator. First, the standard error of each sector's elasticity, ζ_i, is scaled by the aggregate elasticity, ζ_S. This reflects that demand shocks have greater power when their price impact is larger. In the extreme case where the market is perfectly elastic ($\zeta_S \to \infty$), demand shocks cannot move prices at all, rendering the instruments irrelevant and causing standard errors to approach infinity. Second, the standard errors depend on the ratio of other sectors' size-weighted shock volatility, $\sum_{j\neq i} S_j^2\sigma_j^2$, to the sector's own shock volatility, σ_i^2. Higher volatility in the rest of the market creates larger exogenous price variations that can be used to instrument sector i's demand, enhancing instrument power. Conversely, when $\sum_{j\neq i} S_j^2\sigma_j^2 \to 0$, there is no exogenous price variation available to estimate sector i's price elasticity.

Below, we provide two ways to illustrate why our optimal GIV estimator corrects for the simultaneity bias and why size weighting maximizes the power of our estimates.

Instrumental-variable interpretation Our estimator has a natural instrumental variable (IV) interpretation. For each entity i, the moment condition (3.3) can be rewritten as:

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

Note that this is nothing other than an IV estimator where we instrument price changes using size-weighted idiosyncratic shocks of entities other than i. Maximizing the power of an IV estimator requires maximum instrument relevance, i.e., maximizing the covariance between price changes and the instrument. The market clearing condition of our model implies that the price change is an affine function of the size-weighted aggregation of idiosyncratic shocks, so that size weighting achieves maximal relevance.

Bias-corrected OLS interpretation For an infinitesimal entity—whose idiosyncratic shocks have negligible impact on the price—we can estimate its elasticity by simply regressing $\Delta q_{i,t}$ on Δp_t using OLS. Furthermore, under our assumption of temporally homoskedastic idiosyncratic shocks, this OLS estimator would be of optimal efficiency. However, when entities are large enough to move the price, OLS becomes biased. An ideal estimator would exploit the efficiency in OLS while correcting the biases. Our optimal GIV achieves both. Lemma 1 formally shows that the optimal GIV estimator corrects for the bias arising from OLS estimation.

Lemma 1 (Bias-corrected OLS moment conditions)**.** *The optimal GIV estimator is equivalent to the estimator that solves,*

$$\hat{\zeta}_i = \underbrace{-\frac{\hat{\mathbb{E}}\left[\Delta q_{i,t}\Delta p_t\right]}{\hat{\mathbb{E}}\left[\Delta p_t^2\right]}}_{\hat{\zeta}_i^{OLS}} + \hat{\zeta}_S^{-1}\frac{S_i\hat{\sigma}_i^2}{\hat{\mathbb{E}}\left[\Delta p_t^2\right]} \tag{3.5}$$

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

Proof. Using the market clearing condition, we have:

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

Plugging it into the sample moment condition (3.3) and rearranging, we have the sample moment conditions (3.5). □

Lemma 1 states that the optimal estimator can be equivalently implemented by regressing quantities on prices and adjusting for the bias of the OLS estimator $\hat{\zeta}_i^{OLS}$ by the sector's own impact on prices, $\frac{\hat{\zeta}_S^{-1}S_i\hat{\sigma}_i^2}{\hat{\mathbb{E}}[\Delta p_t^2]}$. Notice that the bias disappears when the sector's own impact on prices is zero, i.e., when the sector is infinitesimal, or the market is perfectly elastic.[14] In addition to the OLS interpretation, Lemma 1 also offers an alternative, and sometimes more stable, algorithm for estimating the model.

3.4 Common Factors and the Robustness of Identification

Our identification method relies on the assumption that, after controlling for common factors, idiosyncratic shocks are orthogonal across sectors. The primary threat to identification is the presence of residual correlations in idiosyncratic shocks ($u_{i,t}$). We address this concern in a number of ways and provide a sharp diagnostic tool to assess the robustness of our estimates.

Our observable macro-financial factors, ranging from macroeconomic variables to market expectations about future yields, are comprehensive and explain a significant portion of Treasury price movements. In particular, we use Principal Components Analysis to extract additional common factors from Treasury yield forecasts of the Survey of Professional Forecasters. Collectively, these observable factors explain 58.6% of the price variance.[15]

If we fail to control for all common demand shocks beyond the observable demand shifters included in our estimation, our estimates will be biased. In Appendix A.5, we give propositions that sharply characterize the dependence of the magnitude of the bias on the residual comovement with price variation. Formally, the biased estimate of the aggregate elasticity $\hat{\zeta}_S^b$ can be approximated by:

$$\hat{\zeta}_S^b \approx \zeta_S \sqrt{1 - \mathcal{R}^2} \tag{3.6}$$

where $\mathcal{R}^2$ represents the share of the remaining price variance explained by common factors not controlled for, i.e., $\mathcal{R}^2 = \frac{\text{Var(hidden factors)}}{\text{Var}(\Delta p_t^{resid.})}$. The term $1 - \mathcal{R}^2$ captures the share of remaining price variance explained by true idiosyncratic shocks.[16] The higher the share of common comovement not controlled for, the stronger the downward bias in the estimated

[14] Using our empirical estimates, Appendix B.3 provides an illustration of the effect bias correction has on the simple OLS estimates for sector-level price elasticity estimates.

[15] In Appendix B, we demonstrate the fit of our observable controls by progressively adding macro-financial factors into the set of controls and reporting the R^2.

[16] The bias formula shares the same spirit as the price impact bound developed by van der Beck, Bretscher, and Fu (2025), which stresses the tension between heterogeneity across investors $\sqrt{1-\rho}$ and the market elasticity when inferring from the price and flow volatility. While both formulas involve R-squared, a subtle difference is that here $\mathcal{R}^2$ is the R-squared of *remaining price changes* explained by common factors, while the homogeneity ρ for the price impact bound is the R-squared of total demand shocks explained by cross-investor average.

aggregate elasticity relative to the truth.

Based on (3.6), it is likely that the impact of residual comovement on our estimates would be relatively modest. For instance, if hidden factors explain 25% of the remaining price variance ($\mathcal{R}^2 = 0.25$), our estimated elasticity would be biased downward by only about 13% ($\sqrt{1-0.25} \approx 0.87$). To generate a 50% bias in our aggregate estimates, hidden factors would need to explain 75% of the remaining price variance. Given our comprehensive set of controls and PCA-extracted factors, such a high level of unexplained common comovement seems unlikely.

To add to the credence of this intuition, we develop an alternative set of moment conditions in Appendix A.3.2 that take into account the factor structure for residuals. We discuss identifiability, and develop a method to jointly estimate "internal PCA" from holdings data along with other parameters of the model. Figure B.8 in the Appendix shows that internally extracted factors indeed provide little explanatory power for Treasury price changes beyond what observable controls provide, suggesting that the bias from residual comovement with prices would indeed be small.

For sector-specific estimates, strategic behavior among sophisticated investors also raises valid concern. For example, fast-moving investors might front-run a large but slow-moving investor's positive demand shock. Such strategic interactions would violate Assumption 1 and bias our estimates. To address this concern, we leverage the over-identifying nature of our GMM procedure to implement a leave-one-out estimation approach, by excluding one sector at a time from the set of moment conditions to test the sensitivity of our results to particular sectors. We find that our estimates are generally robust to excluding any single sector, suggesting that strategic interactions are not a significant concern in our setting.[17]

4 The Price Elasticity of the Treasury Market

In this section, we present our findings on the price elasticity of the Treasury market. We first analyze the elasticity of aggregate demand for Treasury bonds. We then examine sectoral heterogeneity in demand sensitivity to Treasury price changes. Finally, to under-

[17] In Appendix A.8, we discuss a few more conceptual issues in identification, including handling missing sectors and measurement errors. For sectors with potential measurement errors in holdings data, we can adopt the same leave-one-out approach to exclude moment conditions involving such sectors, and check the robustness of our estimates. We also shed light on how our approach is immune to the "missing intercept" problem that arises in reduced-form regressions of flows on factors without controlling for prices.

stand who provides liquidity to the Treasury market and how this has evolved over time, we investigate which sectors contribute most to the aggregate demand elasticity and how this has changed since the financial crisis.

Aggregate Price Elasticity The top row of Table 1 reports the aggregate price elasticity of the Treasury market. Our estimate for the aggregate market elasticity is 1.2. This implies a macro multiplier of 0.83 (reciprocal of elasticity). The parameter implies that a 1% demand shock results in a 0.83% increase in the aggregate market value of Treasury bonds. In terms of yields, this translates to a 12.8 basis point decrease (given the duration of outstanding Treasury bonds is around 6.5 years).

Compared with multiplier estimates for other asset classes in the literature, our macro multiplier estimate of 0.83 suggests that the U.S. Treasury market can more effectively absorb demand shocks compared to the U.S. equity market, which has a macro multiplier closer to 5 (Gabaix & Koijen, 2022), and the U.S. corporate bond market, which has a portfolio-level multiplier of 3.5 (Chaudhary et al., 2023). These differences in macro multipliers across asset classes are consistent with the principle that multipliers increase when outside assets are poor substitutes, which makes it harder for arbitrageurs to hedge risks when countering demand shocks (Chaudhary et al., 2023).

The literature has also estimated the price impact of various events on Treasury yields, including foreign FX interventions, quantitative easing, Treasury repurchase operations, and other policy interventions. These studies provide reduced-form price impact estimates around events associated with demand/supply shocks. By making a few assumptions on the implicit size of the associated quantity changes, we can map these price-impact estimates to multipliers. While imperfect, they provide a useful benchmark for our estimates. Appendix B.5 shows that our Treasury macro-multiplier estimates align with the magnitudes implied by existing price-impact studies.

Sector-specific price elasticities The aggregate elasticity estimate of 1.2 masks significant heterogeneity across different market participants. Table 1 reports the price elasticity of demand of each sector, and shows each sector's contribution to the aggregate elasticity. This contribution, labeled as "ζ share" in the table, is calculated as ζ share $\equiv \frac{\zeta_i \times S_i}{\zeta_S}$, where ζ_i is the sector's elasticity, S_i is its market share, and ζ_S is the aggregate elasticity. The contribution of a sector to the aggregate elasticity can be viewed as the extent to which a given sector helps provide liquidity to the overall Treasury market. Sectors with high contributions are either very elastic and, hence, respond strongly to price changes by ac-

Table 1: **Average price elasticities**

Sector	$S(\%)$	ζ	ζ **Share (%)**
Aggregate		1.2	100.0
		(0.9, 1.5)	
Households	5.74	10.54	50.38
		(4.68, 16.41)	
RoW	44.45	0.54	20.0
		(0.32, 0.76)	
Fed	22.08	0.61	11.2
		(0.21, 1.02)	
Banks	5.26	0.95	4.14
		(0.73, 1.16)	
Dealers	0.81	3.36	2.26
		(-6.48, 13.21)	
Insurance	2.59	0.6	1.29
		(0.24, 0.97)	
Mutual Funds	6.75	0.18	0.99
		(-0.39, 0.74)	
ETF	1.18	-0.75	-0.74
		(-1.19, -0.31)	
Other	11.78	0.05	0.53
		(-0.22, 0.32)	
Pension	5.42	-0.09	-0.41
		(-0.28, 0.09)	
Supply	100.0	0.12	10.34
		(0.05, 0.2)	

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 3. Time-series averages are taken for sectors that are assumed to have different elasticities before and after 2009Q1 (the sectors include Federal Reserve and Rest of World). S corresponds to the size weight used in the estimation, defined as the (absolute value of) 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 A.1 and Proposition A.1.

commodating demand shocks, or they may have large market shares and, due to their size, are able to absorb large demand shocks.

The household sector is the most price-elastic sector. Though relatively small, it accounts for roughly half of the total liquidity provision in the market due to its high price elasticity. As discussed in the data section, the household sector is a catch-all category that includes various investors not directly measured by the Financial Accounts, such as hedge funds, family offices, separately managed accounts for high net-worth individuals, and endowment funds of not-for-profit organizations. These sophisticated investors, operating with fewer regulatory constraints and substantial capital, can naturally serve as the most nimble liquidity providers in the Treasury market.

Examining other sectors, we find that foreign investors are less price-elastic than domestic investors. However, due to their substantial market share, foreign investors still rank as the second-largest liquidity providers after the household sector. The Federal Reserve demonstrates significant price elasticity in our estimates. As we elaborate in Section 5.1.2, our methodology controls for the Fed's scheduled purchases, meaning that our elasticity estimates capture market responses to unexpected policy decisions, such as the timing of initiating or terminating asset purchase programs. And as we will discuss, the Fed's role as a Treasury market liquidity provider did not become significant until the Global Financial Crisis.

Security brokers and dealers contribute only 2.26% to the aggregate elasticity. This modest contribution reflects the fact that we are focusing on demand and supply forces at a quarterly horizon. While broker-dealers play a crucial role in intermediating daily trading volumes, they maintain relatively small net positions over quarterly horizons, resulting in them playing a minor role in our analysis. For instance, their average quarter-end net position represents less than 1% of total market value (see Appendix Figure B.7 for more details).

Exchange-traded funds (ETFs) and pension funds are among the least elastic sectors. The inelasticity of ETFs is consistent with their passive trading strategies. The point estimate for ETFs is slightly negative, potentially reflecting return-chasing behavior by retail investors. Overall, they contribute very little to the aggregate elasticity, due to their inelasticity and small market share. Pension funds are well-known as buy-and-hold investors in the Treasury market, which aligns with our finding of their low price elasticity. Our findings are also consistent with Koijen et al. (2017), who find that pension funds have the lowest price elasticity in the European government bond market.

On the supply side, we estimate a positive elasticity close to zero. This result reflects the Treasury's commitment to a "regular and predictable" debt management approach, whereby they avoid timing the market when issuing Treasury bonds. The Treasury typically announces its planned offering amounts of notes and bonds for the upcoming quarter, with minimal deviations from these announcements during normal periods (Cassidy

& Mirani, 2024). Hence, after controlling for the preannounced supply schedules, there is little room for the supply to respond to price changes.

The price impact at longer horizons We find that demand and supply shocks have a persistent price impact lasting at least for a few quarters, similar to what Gabaix and Koijen (2022) found for the aggregate equity market. To formally examine how these demand shocks propagate over time, we employ a local projection approach following Jordà (2005),

$$\Delta p_t^{t+h} = \alpha_h + M_h \hat{u}_{S,t} + \boldsymbol{\lambda}_h^p \boldsymbol{\eta}_t + e_{t+h}, \tag{4.1}$$

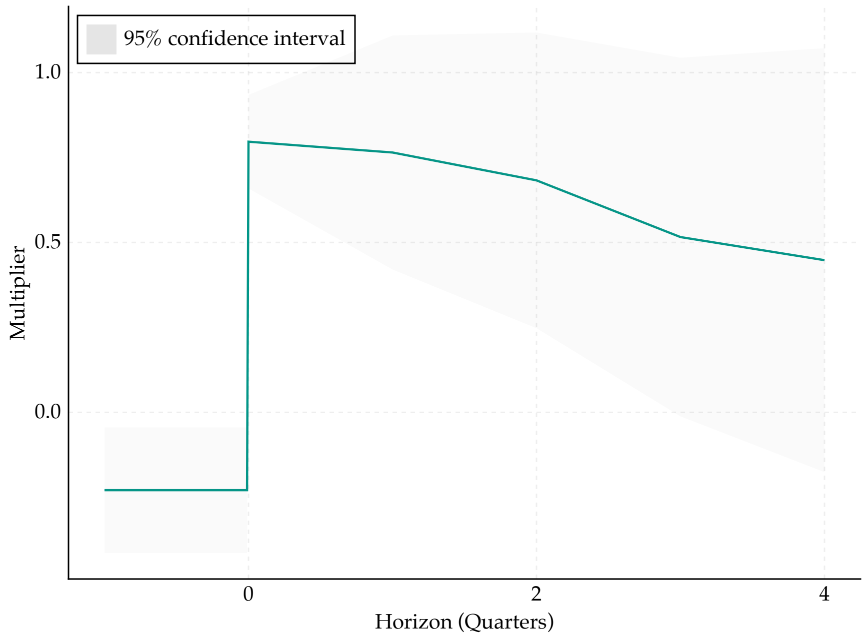
where $h = 1, \dots, 4$ quarters and Δp_t^{t+h} represents the cumulative price changes from quarter t to quarter $t+h$. As a placebo test, we also use the price change from the previous quarter, Δp_{t-1}, as the independent variable and relate it to contemporaneous shocks.

Figure 3 plots the cumulative price impact M_h of a one-percent demand shock with the 95% confidence intervals. The price change in the quarter prior to the shock ($h = -1$) does not predict the demand shock in this quarter, consistent with demand shocks not being priced in advance. On impact, the price increases by 0.83%, consistent with our macro multiplier estimate, and it slowly decays, and at one-year horizon, it drops to around 0.5%. Beyond this horizon, statistical power is limited, preventing us from determining if and when the effect dissipates.[18]

The price impact across the term structure Our procedure yields a macro multiplier for the market portfolio of Treasury notes and bonds, as well as a set of aggregate and sectoral demand and supply shocks. The nature of these shocks can be further evaluated in light of the literature on the factor structure of yield curves (Litterman & Scheinkman, 1991, among others). In Appendix Figure B.12, we estimate contemporaneous versions of the price impact regression (4.1), replacing the dependent variable with price changes implied in Treasury zero-coupon yields (Gurkaynak, Sack, & Wright, 2007) and term premia (Adrian, Crump, & Moench, 2013) across various tenors. As shown in the figure, the price impact estimates are increasing with maturities, and line up closely with a duration-based extrapolation of the price impact on the market portfolio assuming uniform impact on yields across maturities. This result suggests that our demand shocks mostly drive

[18] It is a common finding that the price impact mean reverts over the longer horizon. The market is generally more elastic than in the short run (van der Beck, 2025). Hence, our estimates of the elasticity and the dynamic price impact are mostly informative of Treasury price variations at the quarterly frequency, but do not directly speak to the long-run responses of the Treasury market.

Figure 3: **Dynamic Price Impact of Demand and Supply Shocks**



Note: Figure 3 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 (4.1). Homoskedastic standard errors are used to construct the 95% confidence interval.

the shifts in the level factor of the yield curve, a finding consistent with our objective of explaining changes in the overall level of yields.

Robustness of elasticity estimates

Our elasticity estimates are robust to a variety of alternative specifications and robustness checks.

Leave-one-out estimates Our identification strategy uses idiosyncratic shocks from other sectors to instrument for price in another sector's demand. This strategy works as long as idiosyncratic shocks across different sectors are independent of one another. As discussed in Section 3.4, a potential concern is that demand shocks may be correlated across sectors. For example, sophisticated investors may be able to predict the idiosyncratic shocks by other sectors and hence front-run them. To address this concern, we perform a leave-one-out analysis by sequentially excluding one sector at a time from our estimation procedure. While this approach reduces statistical power, if our identification assumption is valid, the estimated elasticities should be similar across the leave-one-out specifications. Table B.3 in the Appendix presents these results. The average "leave-one-out" aggregate price elasticity is 1.26, very close to our baseline estimates. Similarly, for large sectors such as foreign investors and the U.S. Treasury, their elasticity estimates remain broadly consistent across all leave-one-out specifications. Taken as a whole, these findings suggest that our identification strategy is robust to potential strategic behaviors between sectors.

Alternative specifications: more factors, extending to the 1970s, and including bills Table B.5 presents additional robustness checks, showing our elasticity estimates are relatively stable across various alternative specifications. First, we include more common factors extracted from yield curve forecasts, by controlling for 5 principal components from 1-year yield forecast and 5 principal components from 10-year yield forecast. Second, we repeat our analysis on a longer sample extended back to the 1970s. To extend the sample, we use only Financial Account data, as country-level and bank-level holdings data are not available for this period. We find the elasticity estimates are similar for this longer, albeit less granular, sample. Third, we explore how our results are impacted by expanding our definition of Treasury bonds to include Treasury bills (Treasuries with maturity less than one year). Consistent with Treasury bills being more "cash-like" and having more close substitutes, we find that with this expanded definition, our elasticity estimates are twice as high, reaching 2.3.[19]

Diagnostics using idiosyncratic shocks $\hat{u}_{i,t}$ Our estimated demand and supply shocks at both the sector level ($\hat{u}_{i,t}$) and the market level $\hat{u}_{S,t}$, can serve as diagnostic tools to evaluate the validity of our identification strategy. In Appendix B.6, we provide several supporting results. First, by using China as a case study, we show that the extracted demand shocks from China are well aligned with narrative evidence on the role of China's domestic factors in influencing its Treasury demand (Figure B.11 and its surrounding notes). In addition, we report pairwise sample correlations of extracted demand and supply shocks in a heatmap (Figure B.13) and find that the correlations are in general close to zero. For a small number of pairs with moderate correlations, we further leverage overidentification to exclude these pairs from the moment conditions and show in Table B.5 that our estimation results are robust.

[19]Table B.5 also shows that our results are unaffected by using an alternative strategy to deal with reported discrepancy, as discussed in Section B.1 in the Appendix.

Internal principal components and bias formula diagnostics Figure B.8 shows that our aggregate and sector-specific price elasticities are quantitatively unaffected by adding internally estimated principal components of flows (see discussion in Section 2 and A.3.2 in the Appendix for the estimation procedure) as additional common factors. Panel (a) plots our baseline estimates against estimates obtained from adding one to five internal principal components. Panel (b) reports the R^2 and adjusted R^2 obtained by projecting Treasury price changes onto the estimated internal principal components. The maximum incremental R^2 is 7% with 5 internal PCs, and the adjusted R^2 is very close to zero. Guided by the approximate bias formula (3.6), we conclude that the bias arising from our baseline specification is small.

4.1 Time-Varying Elasticities: Post-GFC Changes in Liquidity Provision

Leveraging the flexibility of our framework, we document a significant regime shift in Treasury market liquidity provision since the Global Financial Crisis (GFC). A sector's price elasticity of demand reflects its willingness to stand on the other side of the market when demand and supply shocks move Treasury prices. Most notably, since the GFC, we find that foreign investors have become much less price-sensitive, while the Federal Reserve has emerged as a state-contingent liquidity provider, stepping in when Treasury prices fall too much.

Table 2 reports the price elasticities of foreign investors and the Fed before and after the GFC. Foreign investors' price elasticity declined by 43% since the GFC. Meanwhile, the Fed, which was completely price inelastic prior to the financial crisis, now exhibits significant price elasticity.[20] The Fed's shift primarily reflects its interventions during crisis periods, highlighting its role as a state-contingent liquidity provider—willing to provide a backstop when Treasury prices fall too much.

The substantial decline in the price sensitivity of foreign investors is reflected in the apparent lack of accommodation by important foreign buyers such as China and Japan

[20] The Fed's elasticity before the financial crisis is estimated to be precisely zero because we assume all changes in the Fed's Treasury holdings outside of quantitative-easing regimes are anticipated and controlled as part of a common factor. These changes typically result from open market operations to achieve the FOMC's interest rate targets. Alternative approaches—treating all changes as unanticipated or using an AR(1) process to control for expected changes—all yield near-zero elasticity estimates for the Fed in the pre-crisis period. As a byproduct, the pre-crisis estimates of the Federal Reserve's Treasury demand loading on all common factors except the anticipated purchases are almost identical to zero. See Appendix B.2 for more details.

after 2010. A variety of factors could explain this transformation, including a lower precautionary demand and the post-crisis shrinkage of current account balances of China and Japan, which reduces their capacity to absorb imbalances, a more diversified reserve management strategy by official sectors towards riskier assets, as well as geopolitical tension. These factors are overwhelmingly related to foreign official holdings. We provide more evidence consistent with this interpretation, by estimating the regime-dependent price elasticities and factor loadings of foreign official and private investors using TIC data. Panel (a) of Appendix Table B.8 shows that the decline in price elasticity is primarily driven by foreign official investors (from 2.28 pre-crisis to 0.6 post-crisis). While foreign private investors were able to partially fill in by increasing their elasticity, the overall elasticity shift is dominated by the larger official sector. Panel (b) of Table B.8 shows that foreign official investors in general exhibit a more adverse response to factors that increase the duration and inflation risk of Treasury bonds relative to foreign private investors.

Table 2: **Regime shifts for price elasticities of foreign investors and the Fed, before and after 2009**

Sector	ζ	ζ **Share (%)**
Fed (03-08)	0.0	-
	-	
Fed (09-23)	0.83	15.13
	(0.48, 1.18)	
Rest of World (03-08)	0.79	29.94
	(0.49, 1.1)	
Rest of World (09-23)	0.45	16.52
	(0.26, 0.64)	

Note: Table 2 reports the estimated price elasticities of demand for foreigners and the Fed 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 3. 95% confidence intervals are reported, with the standard errors given by Theorem A.1 and Proposition A.1.

Overall, these regime shifts mark a substantial change in the market structure of liquidity provision in the Treasury market. Foreign investors, once dominant players, have significantly reduced their active participation, while the Federal Reserve has emerged as a crucial liquidity provider during periods of market stress. In Section 5, we examine this transformation further by decomposing observed yield fluctuations and quantifying the contributions of different sectors to Treasury market dynamics.

5 Yield Decomposition and Structural Changes

Having estimated price elasticities and factor sensitivities of different sectors, we now turn to quantifying their contributions to Treasury yield movements over time. We use the equilibrium pricing equation (2.3) to decompose observed Treasury yield fluctuations from two complementary perspectives. First, we analyze the contribution of each sector to yield changes, revealing which investors have been most influential in driving yields at different points in time. Second, we examine the contribution of common factors, identifying which economic forces most strongly drive Treasury yield fluctuations through investor demand responses.

5.1 Decomposition of Gross Contribution by Sector

Written in a general form, sector i's contribution to Treasury price changes is given by the sum of the fixed effect, the response to common factors, and the idiosyncratic shocks, scaled by the macro multiplier:

$$\Delta \hat{p}_{i,t} = \frac{1}{\hat{\zeta}_{S,r(t)}} S_i \left(\widehat{\overline{\Delta q}}_i + \hat{\boldsymbol{\lambda}}_{i,r(t)} \boldsymbol{\eta}_t + \hat{u}_{i,t} \right),$$

where hats denote estimated values.[21]

To clarify the economic insights from our decomposition, we consolidate sectors into five groups. We highlight: (1) major U.S. financial institutions (mutual funds, ETFs, insurance companies, pension funds, and banks), (2) foreign investors, (3) the Federal Reserve, and (4) Treasury supply from the Department of the Treasury. We also create a fifth group, "US Other," which captures remaining Treasury market participants, primarily households along with non-financial companies, state and local governments, and smaller financial firms (bank holding companies, credit unions, GSEs, and ABS issuers).[22]

Figure 4 presents our yield decomposition by sector. The upper panel shows quarterly contributions over time, while the lower panel aggregates these results into six distinct episodes, using quarterly averages, for better interpretability. Our analysis reveals two major shifts in Treasury market dynamics over the past two decades. First, foreign investors were the dominant force in pushing down Treasury yields from 2003 to 2010, contributing an average reduction of 36 basis points per quarter (and 30 basis points per

[21] The idiosyncratic shock for each sector, $\hat{u}_{it}$, is backed out from the estimated linear system.
[22] Table B.1 provides the complete sector categorization details.

quarter prior to the financial crisis). However, their influence significantly diminished after 2010, as we will explore in Section 5.1.1. Second, the Federal Reserve emerged as a crucial market participant beginning in 2010, with its demand offsetting 42% of the upward yield pressure from expanded Treasury supply during the QE II and QE III period. We examine this new role of the Fed in Section 5.1.2. Throughout this period, major U.S. financial institutions, including mutual funds and pension funds, had a comparatively minor impact on Treasury yields.

5.1.1 The Evolving Influence of Foreign Investors

In the first decade of the 2000s, foreign investors channeled nearly 3 trillion USD into U.S. Treasuries, becoming a dominant force in the market. Current account surplus countries—including China, Japan, oil exporters, and European nations—built substantial Treasury holdings as they accumulated foreign reserves. This phenomenon sparked significant academic and policy debate about the impact of foreign investors on U.S. interest rates. Greenspan (2005) highlighted the "interest rate conundrum" where long-term rates remained low despite Fed rate hikes, most notably in 2006 when the Fed Funds rate exceeded the 10-year Treasury yield by 80 basis points. Similarly, Bernanke (2005) raised concerns about financial stability risks from persistently low long-term rates driven by this "global savings glut."[23] Our framework allows us to quantify the role different countries played in driving Treasury yield fluctuations. In this way, we can more systematically evaluate the impact of foreign investors.[24]

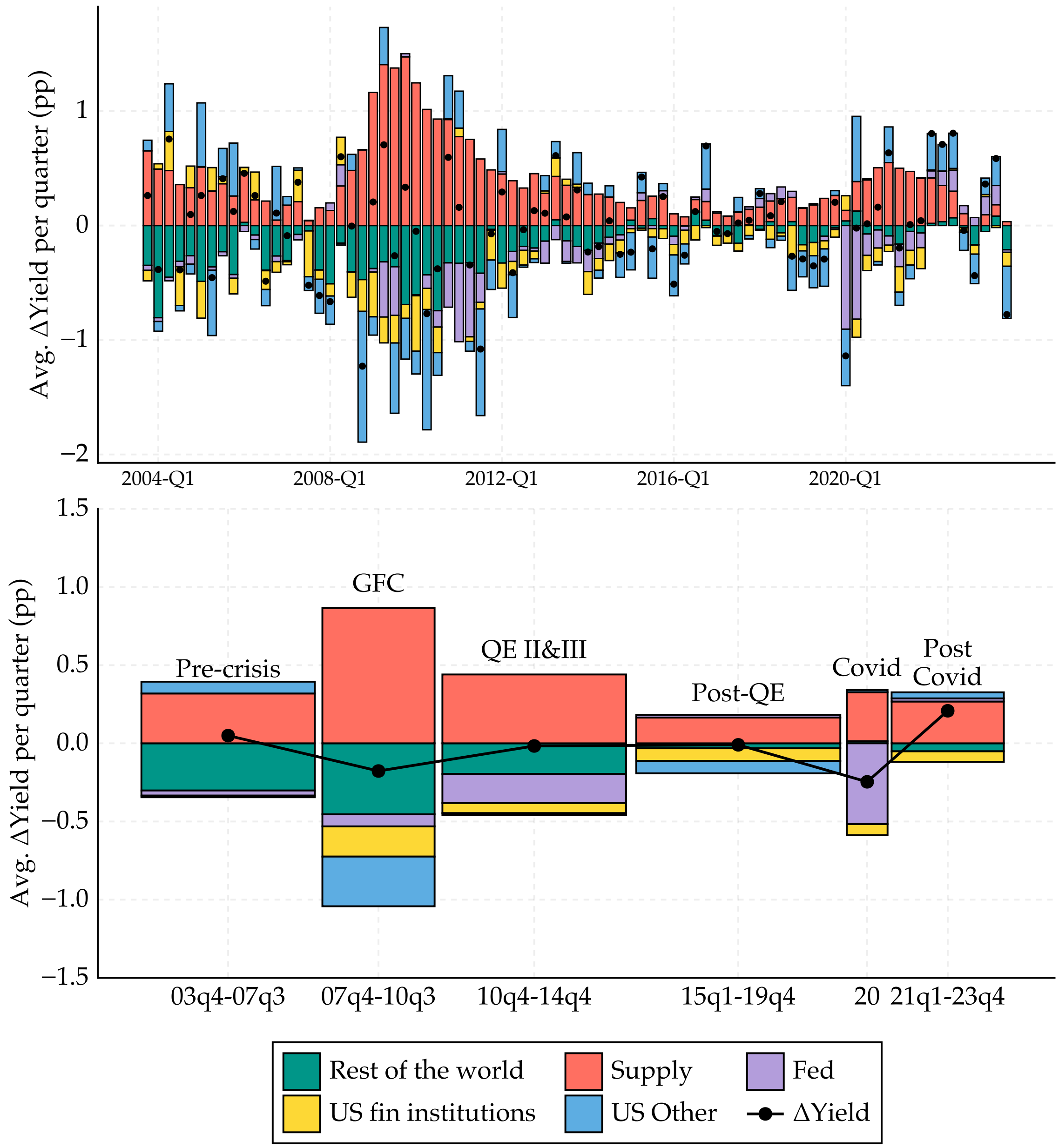
Figure 4 shows that foreign investors played a dominant role in driving Treasury yields before the Global Financial Crisis (GFC), but their influence has significantly diminished since then. Figure 5 provides a more detailed breakdown of Treasury yield changes by country and region. In the pre-crisis period, China, Europe, and Japan collectively accounted for 65% of the foreign sector's impact on yields. Consistent with the global savings glut hypothesis, these foreign investors reduced Treasury yields by 0.78 percentage points annually. China was particularly influential, depressing yields by a cumulative 1.7 percentage points over four years.

The aftermath of the financial crisis marked a significant shift in foreign investors'

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

[24]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.

Figure 4: **Contribution to Treasury yield fluctuations by investor and issuer groups**



Note: Figure 4 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 Accounts (Z.1) to four investor categories and one issuer sector ("supply"), the exact mapping available in Appendix B.1. Group-specific contribution is backed out from the pricing equation (2.3), assuming that the macro 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.

influence on the Treasury market. China's role diminished considerably, while European investors temporarily increased their presence during the Euro Area sovereign debt crisis.[25] Since then, foreign investors' downward pressure on Treasury yields eased. In fact, during the initial phase of the COVID-19 pandemic, foreign investors sold Treasuries, contributing to a 5.5 basis point increase in yields over 2020.

Our findings on the evolving role of foreign investors align with findings by Du, Im, and Schreger (2018), who document a narrowing negative gap between long-term Treasury yields and currency-hedged G10 government yields after the financial crisis. Our analysis complements their work by providing a quantity-based perspective, demonstrating that foreign investors' compression of U.S. Treasury yields was primarily concentrated in the first decade of the 2000s. The finding is also 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, Kretchmer, Marquez, & Thomas, 2013). We offer a few complementary explanations for this transformation in the previous section.

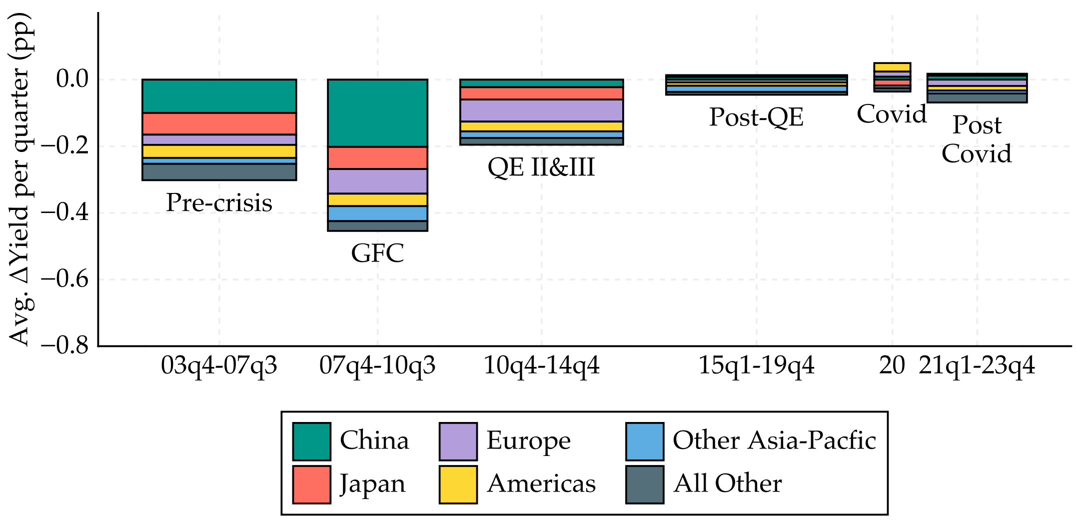
The shrinking market share and diminished price impact of foreign investors suggest that Treasury market disruptions would likely be limited if foreign investors sold Treasury holdings in response to geopolitical tensions. Since the financial crisis, the total supply of Treasuries has expanded substantially. This expansion has reduced foreign investors' share of the total Treasury market. For example, Figure B.10, panel (b) in the Appendix, shows that China's market share declined from its 2010 peak of 17.5 percent to just 4 percent by the end of 2023. This decline resulted from two factors: rapid Treasury issuance while China maintained constant holdings before 2020, and China's steady withdrawal from Treasuries since 2020.

Our framework allows us to quantitatively assess the potential market impact of foreign investor actions that have been the subject of significant policy concern. Consider a hypothetical scenario that has received considerable attention: China liquidating its entire U.S. Treasury holdings within a quarter. This scenario—reportedly proposed but rejected during the global financial crisis—has been characterized as China's economic "nuclear option" during trade disputes (Paulson, 2013; Sorkin, 2018). Our estimated macro multiplier provides a first-order approximation of the direct yield impact. Given our macro

[25]Exchange rate concerns likely motivated Treasury purchases by China and Japan, both active currency intervention participants. The rapid Fed Funds rate cut during the Global Financial Crisis created appreciation pressure, while the unwinding of QE after 2014 led to depreciation pressure for China, prompting foreign reserve sales in 2015. However, these sales had minimal price impact compared to pre-crisis demand shocks.

multiplier of 0.83, a complete, unanticipated liquidation of China's Treasury holdings as of end-2023 would increase Treasury yields by approximately 50 basis points. For context, this impact falls within the normal range of quarterly yield fluctuations, as the standard deviation of quarterly Treasury yield changes from 2003Q4 to 2023Q4 is 45 basis points. The situation would have been far more severe had this scenario occurred in 2008, when China held nearly 20% of U.S. Treasuries. At that time, such a liquidation would have triggered a 2.5 percentage point spike in yields (using a multiplier of 0.83, China's market share of 20%, and a 6.5-year duration of the market portfolio of Treasuries), a dramatic move that could significantly amplify global financial instability during the crisis.

Figure 5: **Important foreign holders' contributions to yields, percentage points per quarter**



Note: Figure 5 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.

5.1.2 The Federal Reserve as a State-Contingent Liquidity Provider

While foreign investors' influence has waned, the Federal Reserve has emerged as a critical state-contingent liquidity provider in the Treasury market since the Global Financial Crisis (GFC). A comparison of the GFC and COVID-19 episodes reveals a striking evolution in the composition of market participants driving Treasury yields during periods of stress. During the GFC and its aftermath (07Q4–10Q3), yield compression stemmed from multiple sectors: foreign investors (45 basis points per quarter), U.S. households (32 basis points, labeled "US Other" in Figure 4), and major U.S. financial institutions (19 basis points). In contrast, during the COVID-19 pandemic in 2020, the Federal Reserve dominated market intervention, single-handedly reducing yields by 51.7 basis points per quarter through its purchases. Such impact mostly reflects the Fed's active response to market distress (with a loading on VIX of 2.35), consistent with it playing the role of a state-contingent liquidity provider.[26] In contrast, other sectors on the demand side collectively contributed only a 4.2 basis point decline during the same episode, with foreign investors and U.S. households actually exerting slight upward pressure on yields.

Table 3 presents our estimates for Federal Reserve behavior since 2009. These estimates control for anticipated Fed purchases as a common factor, with all results reflecting deviations from the expected purchase path.[27]

Our analysis reveals the Fed is price elastic, initiating QE operations when Treasury markets experience downward demand pressures. This behavior aligns with the stated objectives of QE operations (Gagnon, Raskin, Remache, & Sack, 2011b), which include Treasury market stabilization. The phrase "the smooth functioning of the Treasury market" appears frequently in FOMC minutes, underscoring this priority.[28]

Finally, our estimates demonstrate the coordination between the Fed's direct Treasury purchases and conventional monetary policy. When the Federal Funds rate increases or when inflation goes up, the Fed simultaneously reduces its Treasury purchases.

5.2 Decomposition by Factors

While our sectoral decomposition provides valuable insights into which investors drive Treasury yield movements, it combines the effects of multiple demand shifters. To gain deeper understanding of the underlying economic forces, we now complement our analysis with a factor-based decomposition approach. We employ a two-step methodology: first decomposing yield changes according to the economic factors that shift demand

[26] State-contingent purchases can further reduce Treasury yields through the insurance channel proposed by Haddad et al. (2024): supporting bond prices during market stress provides additional safety and lowers bond risk premia. Haddad et al. (2024) finds this insurance channel accounts for approximately 75% of QE's effect on yields.

[27] The Fed typically announces purchases after FOMC meetings according to a pre-scheduled plan. Appendix B.2 details our methodology for constructing anticipated purchases and shows that deviations primarily occurred during the initial and concluding phases of QE operations, when the Fed had flexibility to respond to contemporaneous macroeconomic and financial conditions, including Treasury yields.

[28] For example, a 2020 New York Fed announcement 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." Another contemporaneous announcement specified that "the Desk will conduct purchase...subject to reasonable prices."

Table 3: **Treasury Demand of the Federal Reserve**

$Period$	$S(\%)$	ζ	$\epsilon^{VIX}_{(std.)}$	Δ **FFR**	$Inf.$
09-23	22.08	0.83	2.35	-6.18	-1.42
		(0.48, 1.18)	(1.4, 3.29)	(-8.46, -3.9)	(-2.34, -0.49)

Note: Table 3 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 quarterly changes in long-term inflation expectation. 95% confidence intervals are reported, with the standard errors given by Theorem A.1 and Proposition A.1.

curves (such as risk sentiment, inflation, and monetary policy), then identifying which sectors transmit these factor-driven pressures to market prices. This approach is similar to traditional reduced-form factor pricing models that regress prices on observed macroeconomic variables, as it reveals which factors most significantly drive Treasury yield fluctuations. However, our method has the additional advantage of identifying which sectors are most influential in transmitting specific economic forces to Treasury prices.

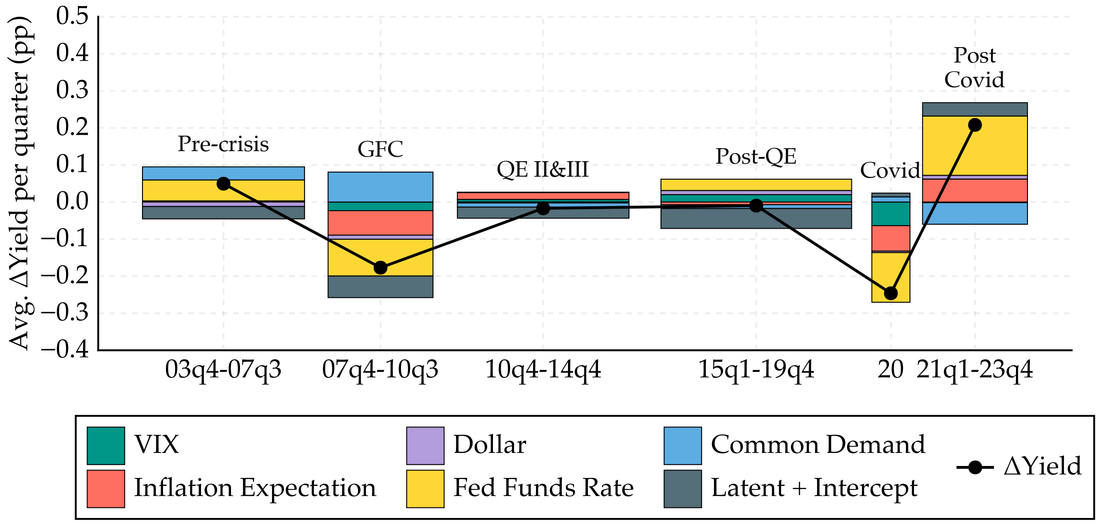
Formally, the contribution of the *k*-th factor to yield changes is given by:

$$\Delta \hat{p}_{\eta(k),t} = \hat{\zeta}^{-1}_{S,r(t)} \hat{\lambda}_{S,r(t)}(k) \eta_t(k), \tag{5.1}$$

where $\hat{\lambda}_{S,r(t)}(k) = \sum_i S_i \hat{\lambda}_{i,r(t)}(k)$ represents the size-weighted average sensitivity to factor k across all sectors. The contribution from idiosyncratic shocks equals $\hat{\zeta}^{-1}_{S,r(t)} \hat{u}_{S,t}$, which our estimates indicate accounts for around 40% of the total variation in Treasury prices during our sample period.

Figure 6 illustrates the contribution of various factors to Treasury yield movements, expressed in quarterly averages (percentage points). Accordingly, Table B.7 in the Appendix further allocates the contribution of factors to each sector by showing the full table of factor loadings. We organize these factors into distinct categories, focusing on four key drivers: VIX changes (risk sentiment), inflation expectation, U.S. dollar strength fluctuations, and Federal Funds rate adjustments. We group other common demand factors, predictable Fed purchases, and previous-quarter net issuance under the label "common demand"—representing factors either widely known to market participants or capturing common patterns of investor demand across sectors potentially unobservable. The decomposition includes sector-specific intercepts to account for average price pressure from all sectors under "neutral" market conditions.[29]

Figure 6: **Macro and idiosyncratic drivers of Treasury yields**



Note: Figure 6 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 (5.1), assuming that the macro 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.

Consistent with VIX being a barometer of global risk sentiment, its negative impact on yields is most pronounced during crisis periods. During the 2020 crisis, VIX surges led to a 25.5 basis point decline in Treasury yields in four quarters, while during the financial crisis (07Q4-10Q3), they contributed to a similar 26 basis point cumulative decline. We further investigate which sectors flee to Treasuries during periods of elevated global risk in Section 6.

Our estimation can also inform the transmission of conventional monetary policy. The estimated average loading on Fed Funds Rate changes is -2.68 (Table B.7). Given an average duration of 6.5 years and a market elasticity of 1.2, this means that the "pass-through" coefficient of Fed Funds Rate changes to Treasury yields, defined according to

[29]This decomposition interprets macro factors as deviations from neutral benchmark levels. Different benchmarks would redistribute contributions between macro factors and the intercept (which appears along with sector-specific idiosyncratic components). We select economically meaningful benchmarks: sample means for most factors and zero for Fed funds rate changes. Thus, the intercept represents average price pressure under conditions of constant Fed funds rate and average values for other macro factors.

(2.3), is 0.34: A 100 basis point rate hike on average leads to a 34 basis points increase in yields. The pass-through is stronger in the post-crisis period when the Fed's own Treasury purchases coordinate with the rate change, amplifying the effect of monetary policy. Examining specific episodes, we find that in the pre-crisis years, despite the Fed's rapid rate hike of more than 4 percentage points from 2004 to 2006, the cumulative impact of the short rate on the yields on Treasury bonds and notes is less than 100 basis points, again illustrating the interest rate "conundrum" faced by the Fed (Greenspan, 2005). During the Global Financial Crisis and the COVID-19 crisis, the Fed's interest rate cuts lowered yields by 109 and 54 basis points respectively on a cumulative basis. The recent tightening cycle that began in early 2022, on the other hand, raised Treasury yields by a total of 1.8 percentage points.

Hanson and Stein (2015) argue that monetary policy's effect on long-term rates contradicts sticky-price models, proposing instead a channel through demand pressures induced by reach-for-yield motives. Our framework allows us to identify which sectors transmit short rate changes to medium- and long-term yields. Table B.7 in the Appendix reports the full coefficients of factor loadings by sectors. Consistent with the asset-liability management practices of insurance companies and pension funds (ICPFs), we find that their demand for long-term bonds increases when the Fed funds rate rises, holding long-term yields fixed. When short-term rates rise while long-term yields remain unchanged, the yield curve twists. Both assets and liabilities fall in value, but because assets have shorter duration, they fall more than liabilities. To restore their duration hedge, ICPFs need to increase the duration of their assets, which they achieve by buying more long-term bonds, such as Treasuries.

On the other hand, banks' demand for Treasury notes and bonds negatively loads on the Fed funds rate. For a 1% Fed funds rate increase, banks' demand for Treasury notes and bonds decreases by 2.43%. This aligns with research on the deposit channel of monetary policy (Drechsler, Savov, & Schnabl, 2017): as the Fed Funds rate rises, banks widen deposit spreads, causing deposit outflows and corresponding reductions in Treasury holdings.

Following the COVID-19 pandemic, global fiscal and monetary stimulus combined with supply-chain pressures pushed inflation to levels not seen since the 1980s. Our estimates indicate that inflation expectations alone pushed yields up by 0.68 percentage points in the three years since 2021. In contrast, deflationary expectations significantly pushed down Treasury yields by a similar amount per quarter during the Global Financial

Crisis. Among sectors, households show the greatest sensitivity to inflation expectations, reducing Treasury holdings by 13% for a one-standard-deviation increase in inflation expectations. This finding supports evidence from Nagel and Yan (2022), who demonstrate that households pay attention to inflation news and respond by withdrawing from nominal Treasury-related investments.[30]

6 Treasury Investors and "Flight to Safety"

We use our estimated price elasticities, factor loadings, and decomposition to investigate the drivers of "flight to safety" in the Treasury market—the phenomenon where yields decline as global risk sentiment increases.

Two competing views exist in the literature about how investors' demand for Treasuries behaves during crises. The *flight-to-safety view* holds that foreign investors have heightened demand for Treasuries during crises, thereby driving the countercyclical pattern observed in Treasury prices (Jiang et al., 2021; Jiang, Krishnamurthy, & Lustig, 2024; Kekre & Lenel, 2024). This view underpins much of the literature on the special status of the U.S. dollar, and many models built on this assumption seek to explain the dollar's exorbitant privilege. In contrast, the *precautionary saving view*, often applied to characterizing foreign official sector's U.S. dollar reserve management, suggests that Treasuries serve as precautionary saving vehicles for foreign countries during normal times but are drawn down to raise liquidity during disruptions in international financial markets (e.g., Bianchi et al., 2018; Eisenbach & Phelan, 2025). Under this view, foreign investors are not the primary driver of Treasuries' countercyclical properties but rather benefit from the hedging property that they appreciate during crises. These two views yield different predictions about investor trading behavior and carry distinct theoretical implications for the source of U.S. dollar exorbitant privilege.

Despite the theoretical importance of these assumptions, their empirical validity has remained largely unexamined. The foreign sell-off of Treasuries at the onset of the COVID-19 pandemic in March 2020 intensified debates in academic and policy circles about whether the special status of U.S. Treasuries has waned (He et al., 2022; Vissing-Jorgensen, 2021; Weiss, 2022).

[30] The Financial Accounts data do not distinguish between nominal Treasury securities and inflation-indexed Treasury bonds (TIPS), limiting our inflation hedging analysis to a high level. However, the TIPS market represents less than 10% of all marketable Treasury debt as of 2024 and was even smaller earlier in our sample. Therefore, we primarily attribute household flows to nominal bonds and notes.

Our estimated model allows us to assess the extent to which foreign investors drive flight-to-safety dynamics. We analyze the estimated flow loadings on quarterly VIX index changes for foreign investors. Since Treasury yields and the VIX index exhibit strong negative comovement in the data, the VIX serves as an effective indicator of investors' tendency to seek safe assets during periods of heightened risk.

Table 4: **Sensitivities to global risk sentiments for selected sectors**

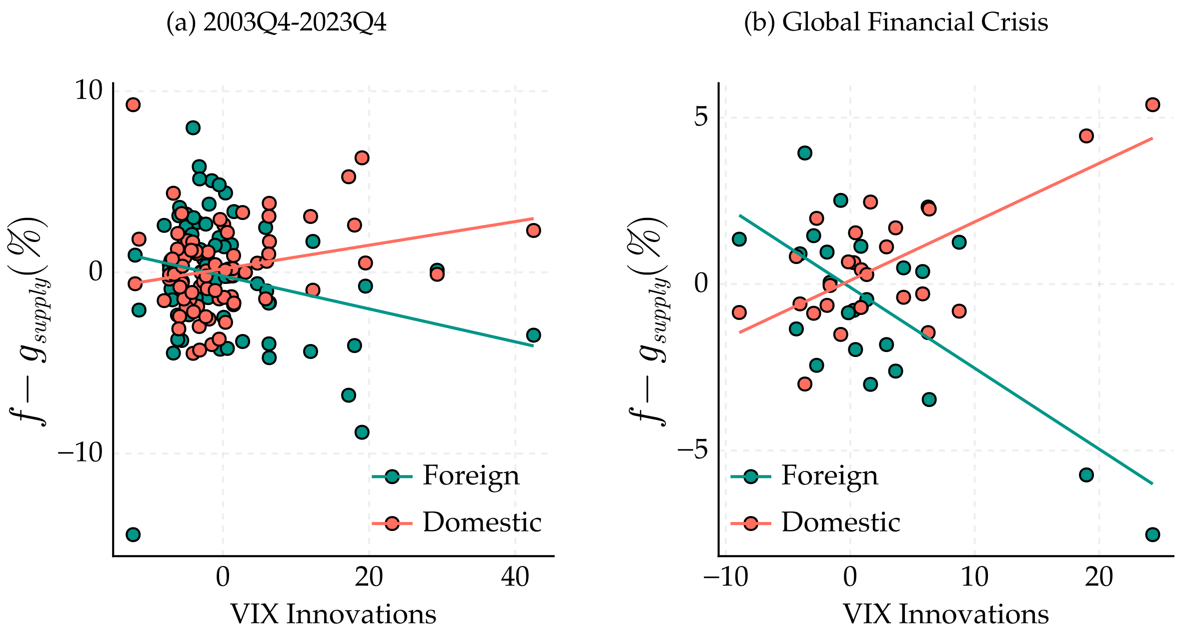
$Sector$	$S(\%)$	$\epsilon^{VIX}_{(std.)}$	$\epsilon^{VIX}_{(std.)}$ **Share (%)**
Aggregate (09-)		0.55	100.0
		(-0.2, 1.31)	
Households	5.74	11.3	117.14
		(1.48, 21.12)	
Fed (09-)	22.08	2.35	93.61
		(1.4, 3.29)	
Rest of World	44.45	-0.4	-31.83
		(-1.12, 0.33)	
Supply	100.0	-0.04	-6.32
		(-0.23, 0.16)	
Mutual Funds	6.75	-2.18	-26.55
		(-3.82, -0.53)	

Note: Table 4 reports the estimated responsiveness to a rise in the VIX index for a selected set of sectors in the data. The sample period is 2003Q4–2023Q4. The loadings are identified using the optimal GIV estimator developed in Section 3. 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 an 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 standard errors given by Theorem A.1 and Proposition A.1.

Table 4 presents the size-weighted aggregate VIX loading and sector-specific loadings for selected investors, with the complete set of investor groups reported in Table B.7 in the Appendix. A one standard deviation increase in the VIX index generates a 0.55 percent inflow into the Treasury market. Notably, households exhibit the strongest flight-to-safety behavior, increasing their Treasury demand by 11 percent. In contrast, foreign investors show no statistically significant sensitivity to VIX changes. Their point estimate is negative, suggesting they reduce their demand for U.S. Treasuries during risk-off episodes rather than increase it.[31]

[31] In Appendix B.4, we show in Table B.4 that the point estimates for foreign investors' VIX loading remain quantitatively stable when excluding any other sector from the moment conditions used to estimate the system. Table B.6 further shows that when we extend the estimation to cover Treasury bills, there is still no

Figure 7: **Foreign flows (net of issuance) and global risk indicators**



Note: In Figure 7 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).

Figure 7 illustrates this pattern by plotting foreign inflows *net of Treasury issuance* (normalized by market size) and domestic flows net of issuance against changes in the VIX index. Panel (a) shows the entire sample period, revealing a stark difference: when VIX is high, foreign investors purchase fewer Treasuries—or even sell them—compared to domestic investors. This relationship is not merely a recent phenomenon; it was also evident at monthly frequency during the Global Financial Crisis (panel (b)). Importantly, a direct regression coefficient between flows and VIX would not capture an investor's true *demand* responsiveness to rising global risk sentiment, but would reflect the relative tendencies between foreign and domestic investors, as well as the sensitivity of Treasury supply. In contrast, the coefficient in Table 4 estimates demand sensitivity to risks while accounting for the endogenous response to price movements.

Our finding that foreign Treasury demand loads negatively on VIX challenges the common assumption—based on the flight-to-safety view—that foreign investors exhibit stronger countercyclical demand for Treasuries. Instead, it aligns with the precautionary

significant evidence supporting the flight-to-safety view, despite the overall loading on VIX being positive and significant.

saving view, which posits that foreign investors reduce Treasury holdings during crises to raise liquidity. This finding raises further questions about the source of the safe haven status of the U.S. dollar, given that foreign investors do not increase their demand for U.S. Treasuries during crises. We leave this question for future research.

7 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 elasticities and factor loadings 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 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, V. V., & Laarits, T. (2024). *When Do Treasuries Earn the Convenience Yield? — A Hedging Perspective*.

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, *77*(2), 375–410. doi:10.1016/j.jfineco.2004.06.007

Adrian, T., Crump, R. K., & Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics*, *110*(1), 110–138. doi:10.1016/j.jfineco.2013.04.009

Adrian, T., Grinberg, F., Liang, N., Malik, S., & Yu, J. (2022). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics*, *14*(3), 283–323. Retrieved from https://www.aeaweb.org/articles?id=10.1257/mac.20180428

Ahmed, R., & Rebucci, A. (2024). Dollar reserves and U.S. yields: Identifying the price impact of official flows. *Journal of International Economics*, *152*, 103974. doi:10.1016/j.jinteco.2024.103974

An, Y., & Huber, A. W. (2024). Intermediary Elasticity and Limited Risk-Bearing Capacity.

Barth, D., Beltran, D., Hoops, M., Kahn, J., Liu, E., & Perozek, M. (2025). The Cross-Border Trail of the Treasury Basis Trade. *FEDS Notes*. Retrieved from https://www.federalreserve.gov/econres/notes/feds-notes/the-cross-border-trail-of-the-treasury-basis-trade-20251015.html

Baumeister, C., & Hamilton, J. D. (2023). *A full-information approach to granular instrumental variables*. Working paper, UCSD. Retrieved from https://apps.unive.it/server/eventi/78661/BH8.pdf

Beltran, D. O., Kretchmer, M., Marquez, J., & Thomas, C. P. (2013). Foreign Holdings of U.S. Treasuries and U.S. Treasury Yields. *Journal of International Money and Finance*, *32*, 1120–1143. doi:10.1016/j.jimonfin.2012.09.005

Bernanke, B., Reinhart, V., & Sack, B. (2004). Monetary policy alternatives at the zero bound: An empirical assessment. *Brookings Papers on Economic Activity*, *2004*(2), 1–100.

Bernanke, B. S. (2005). *The global saving glut and the U.S. current account deficit*. Retrieved from https://www.federalreserve.gov/boarddocs/speeches/2005/200503102/

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

Bertaut, C. C., & Judson, R. A. (2014). *Estimating U.S. Cross-Border Securities Positions: New Data and New Methods* (No. 1113). Board of Governors of the Federal Reserve System (U.S.) Retrieved from https://ideas.repec.org/p/fip/fedgif/1113.html

Bertaut, C. C., & Tryon, R. W. (2007). *Monthly estimates of U.S. cross-border securities positions* (No. 910). Board of Governors of the Federal Reserve System (U.S.) Retrieved from https://ideas.repec.org/p/fip/fedgif/910.html

Bianchi, J., Hatchondo, J. C., & Martinez, L. (2018). International Reserves and Rollover Risk. *American Economic Review*, *108*(9), 2629–2670. doi:10.1257/aer.20140443

Bretscher, L., Schmid, L., Sen, I., & Sharma, V. (2020, December 7). *Institutional Corporate Bond Demand* (SSRN Scholarly Paper No. ID 3756280). Social Science Research Network. doi:10.2139/ssrn.3756280

Brunnermeier, M. K., Merkel, S., & Sannikov, Y. (2024). Safe Assets. *Journal of Political Economy*, *132*(11), 3603–3657. doi:10.1086/730547

Caballero, R. J. (2006, December). *On the macroeconomics of asset shortages* (NBER Working Papers No. 12753). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/12753.html

Camanho, N., Hau, H., & Rey, H. (2022). *Global Portfolio Rebalancing and Exchange Rates* (Working Paper No. 24320). National Bureau of Economic Research. doi:10.3386/w24320

Campbell, J. Y., Pflueger, C., & Viceira, L. M. (2019). Macroeconomic Drivers of Bond and Equity Risks. *Journal of Political Economy*, *128*(8), 3148–3185. doi:10.1086/707766

Cassidy, W., & Mirani, S. (2024). The Debt Ceiling's Disruptive Impact: Evidence from Many Markets. Retrieved from https://williamcassidy.com/files/CassidyMiraniApril112024.pdf

Cavaleri, F. (2023). The Convenience Yield and the Demand for U.S. Treasury Securities.

Cayman Islands Monetary Authority. (2023). Investments Statistical Digest. Retrieved from https://www.cima.ky/investments-statistical-digest

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, *34*(3), 305–334. doi:10.1016/0304-4076(87)90015-7

Chaudhary, M. (2024, October 13). Regulator Beliefs. doi:10.2139/ssrn.4986079. Social Science Research Network: 4986079

Chaudhary, M., Fu, Z., & Li, J. (2023, September). *Corporate bond multipliers: Substitutes matter*.

Chaudhry, A., & Li, J. (2025, June 25). Endogenous Elasticities: Price Multipliers Are Smaller for Larger Demand Shocks. doi:10.2139/ssrn.5325371. Social Science Research Network: 5325371

Chodorow-Reich, G., Gabaix, X., Koijen, R. S. J., & Viviano, D. (2024, July 5). Propagation of Shocks in Networks: Identification and Applications. doi:10.2139/ssrn.4885984

Da, Z., Larrain, B., Sialm, C., & Tessada, J. (2018). Destabilizing financial advice: Evidence from pension fund reallocations. *The Review of Financial Studies*, *31*(10), 3720–3755. doi:10.1093/rfs/hhy011. eprint: https://academic.oup.com/rfs/article-pdf/31/10/3720/25693037/hhy011.pdf

Domanski, D., Shin, H. S., & Sushko, V. (2017). The Hunt for Duration: Not Waving but Drowning? *IMF Econ Rev*, *65*(1), 113–153. doi:10.1057/s41308-016-0026-9

Drechsler, I., Savov, A., & Schnabl, P. (2017). The Deposits Channel of Monetary Policy. *Q J Econ*, *132*(4), 1819–1876. doi:10.1093/qje/qjx019

Du, W., Hébert, B., & Li, W. (2023). Intermediary balance sheets and the treasury yield curve. *Journal of Financial Economics*, *150*(3), 103722. doi:10.1016/j.jfineco.2023.103722

Du, W., Im, J., & Schreger, J. (2018). The U.S. treasury premium. *Journal of International Economics*, *112*, 167–181. doi:10.1016/j.jinteco.2018.01.001

Eisenbach, T. M., & Phelan, G. (2025). Fragility of Safe Asset Markets. *The Review of Financial Studies*. Retrieved April 16, 2026, from https://dx.doi.org/10.1093/rfs/hhaf064

Eren, E., Schrimpf, A., & Xia, F. D. (2023). The Demand for Government Debt. *SSRN Journal*. doi:10.2139/ssrn.4466154

Fang, C., & Xiao, K. (2025, August 1). What Do $40 Trillion of Portfolio Holdings Say about Monetary Policy Transmission? doi:10.2139/ssrn.5025417. Social Science Research Network: 5025417

Fu, Z., Li, J., & Xie, Y. (2022). Convenience yield, inflation expectations, and public debt growth. *Inflation Expectations, and Public Debt Growth (February 5, 2022)*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4027631

Gabaix, X., & Koijen, R. S. J. (2022, May 12). In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis. doi:10.2139/ssrn.3686935

Gabaix, X., & Koijen, R. S. J. (2024). Granular Instrumental Variables. *Journal of Political Economy*, *132*(7), 2274–2303. doi:10.1086/728743

Gagnon, J., Raskin, M., Remache, J., & Sack, B. (2011a). Large-Scale Asset Purchases by the Federal Reserve: Did They Work? *FRBNY Economic Policy Review*, 43.

Gagnon, J., Raskin, M., Remache, J., & Sack, B. (2011b). The financial market effects of the federal reserve's large-scale asset purchases. *International Journal of Central Banking*, *7*(1), 3–43. Retrieved from https://ideas.repec.org/a/ijc/ijcjou/y2011q1a1.html

Greenspan, A. (2005, February). *Testimony: Federal Reserve Board's semiannual Monetary Policy Report to the Congress*.

Greenwood, R., Hanson, S. G., & Stein, J. C. (2015). A Comparative-Advantage Approach to Government Debt Maturity. *The Journal of Finance*, *70*(4), 1683–1722. doi:10.1111/jofi.12253

Greenwood, R., & Vayanos, D. (2010). Price Pressure in the Government Bond Market. *American Economic Review*, *100*(2), 585–590. doi:10.1257/aer.100.2.585

Greenwood, R., & Vayanos, D. (2014). Bond Supply and Excess Bond Returns. *Rev Financ Stud*, *27*(3), 663–713. doi:10.1093/rfs/hht133

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, *70*(3), 393–408. JSTOR: 1805228. Retrieved from https://www.jstor.org/stable/1805228

Gurkaynak, R. S., Sack, B., & Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, *54*(8), 2291–2304. Retrieved from https://ideas.repec.org//a/eee/moneco/v54y2007i8p2291-2304.html

Haddad, V., He, Z., Huebner, P., Kondor, P., & Loualiche, E. (2026, February 27). Causal Inference for Asset Pricing. doi:10.2139/ssrn.5187305. Social Science Research Network: 5187305

Haddad, V., Huebner, P., & Loualiche, E. (2025). How Competitive Is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing. *American Economic Review*, *115*(3), 975–1018. doi:10.1257/aer.20230505

Haddad, V., Moreira, A., & Muir, T. (2024). *Asset Purchase Rules: How QE Transformed the Bond Market*.

Han, C., & Phillips, P. C. B. (2006). GMM with Many Moment Conditions. *Econometrica*, *74*(1), 147–192. doi:10.1111/j.1468-0262.2006.00652.x

Hanson, S. G., & Stein, J. C. (2015). Monetary policy and long-term real rates. *Journal of Financial Economics*, *115*(3), 429–448. doi:10.1016/j.jfineco.2014.11.001

Hartzmark, S. M., & Solomon, D. H. (2025). Market-Wide Predictable Price Pressure. *American Economic Review*, *115*(9), 3171–3213. doi:10.1257/aer.20231725

He, Z., Nagel, S., & Song, Z. (2022). Treasury inconvenience yields during the COVID-19 crisis. *Journal of Financial Economics*, *143*(1), 57–79. doi:10.1016/j.jfineco.2021.05.044

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, *22*(3), 477–498. doi:10.1016/0022-0531(80)90056-3

Huang, S., Jiang, W., Liu, X., & Liu, X. (2025). Does Liquidity Management Induce Fragility in Treasury Prices? Evidence from Bond Mutual Funds. *Rev Financ Stud*, *38*(2), 337–380. doi:10.1093/rfs/hhae082

Jansen, K. A. E., Li, W., & Schmid, L. (2024). *Granular Treasury Demand with Arbitrageurs*.

Jiang, E. X., Matvos, G., Piskorski, T., & Seru, A. (2024). Monetary tightening and US bank fragility in 2023: Mark-to-market losses and uninsured depositor runs? *Journal of Financial Economics*, *159*, 103899. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304405X24001223

Jiang, Z., Krishnamurthy, A., & Lustig, H. (2021). Foreign Safe Asset Demand and the Dollar Exchange Rate. *The Journal of Finance*, *76*(3), 1049–1089. doi:10.1111/jofi.13003. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13003

Jiang, Z., Krishnamurthy, A., & Lustig, H. N. (2024). *Dollar Safety and the Global Financial Cycle* (SSRN Scholarly Paper No. ID 3328808). Social Science Research Network. doi:10.2139/ssrn.3328808

Jiang, Z., Richmond, R., & Zhang, T. (2024). A portfolio approach to global imbalances. *The Journal of Finance*, *79*(3), 2025–2076. doi:10.1111/jofi.13333. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13333

Jiang, Z., Richmond, R. J., & Zhang, T. (2025). Understanding the strength of the dollar. *Journal of Financial Economics*, *168*, 104052. doi:10.1016/j.jfineco.2025.104052

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, *95*(1), 161–182. doi:10.1257/0002828053828518

Kaminska, I., & Zinna, G. (2020). Official Demand for U.S. Debt: Implications for U.S. Real Rates. *Journal of Money, Credit and Banking*, *52*(2–3), 323–364. doi:10.1111/jmcb.12660

Kekre, R., & Lenel, M. (2024). The flight to safety and international risk sharing. *American Economic Review*, *114*(6), 1650–1691. doi:10.1257/aer.20211319

Kim, J., & Kim, M. (2025, May 22). Deciphering CIP Deviation: Who Moves It and Why It Matters? doi:10.2139/ssrn.5271538. Social Science Research Network: 5271538

Koijen, R. S. J., Koulischer, F., Nguyen, B., & Yogo, M. (2017). Euro-Area Quantitative Easing and Portfolio Rebalancing. *American Economic Review*, *107*(5), 621–627. doi:10.1257/aer.p20171037

Koijen, R. S. J., & Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, *127*(4), 1475–1515. doi:10.1086/701683. eprint: https://doi.org/10.1086/701683

Koijen, R. S. J., & Yogo, M. (2023). Understanding the Ownership Structure of Corporate Bonds. *American Economic Review: Insights*, *5*(1), 73–92. doi:10.1257/aeri.20210550

Koijen, R. S. J., & Yogo, M. (2024, November 4). Exchange Rates and Asset Prices in a Global Demand System. doi:10.2139/ssrn.3383677. Social Science Research Network: 3383677

Krishnamurthy, A., & Vissing-Jorgensen, A. (2011, October). *The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy* (Working Paper No. 17555). National Bureau of Economic Research. doi:10.3386/w17555

Krishnamurthy, A., & Vissing-Jorgensen, A. (2012). The Aggregate Demand for Treasury Debt. *Journal of Political Economy*, *120*(2), 233–267. doi:10.1086/666526. JSTOR: 10.1086/666526

Kubitza, C., Sigaux, J.-D., & Vandeweyer, Q. (2024). The implications of CIP deviations for international capital flows. *Fama-Miller Working Paper, Chicago Booth Research Paper*, 24–18. Retrieved from https://www.ecb.europa.eu/press/conferences/shared/pdf/20241107_money_markets/Sigaux_paper.pdf

Kundu, S., & Vats, N. (2021, December 1). Banking Networks and Economic Growth: From Idiosyncratic Shocks to Aggregate Fluctuations. doi:10.2139/ssrn.3975942. Social Science Research Network: 3975942

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315–1335. doi:10.2307/1913210. JSTOR: 1913210

Litterman, R. B., & Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *The Journal of Fixed Income*, *1*(1), 54–61. doi:10.3905/jfi.1991.692347

Lou, D. (2012). A flow-based explanation for return predictability. *The Review of Financial Studies*, *25*(12), 3457–3489. doi:10.1093/rfs/hhs103. eprint: https://academic.oup.com/rfs/article-pdf/25/12/3457/24436896/hhs103.pdf

Ludvigson, S. C., & Ng, S. (2009). Macro Factors in Bond Risk Premia. *Rev. Financ. Stud.*, *22*(12), 5027–5067. doi:10.1093/rfs/hhp081

Mendoza, E. G., Quadrini, V., & Ríos-Rull, J.-V. (2009). Financial integration, financial development, and global imbalances. *Journal of Political Economy*, *117*(3), 371–416. doi:10.1086/599706

Nagel, S. (2016). The Liquidity Premium of Near-Money Assets*. *The Quarterly Journal of Economics*, *131*(4), 1927–1971. doi:10.1093/qje/qjw028

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

Newey, W. K. (1993, January 1). 16 Efficient estimation of models with conditional moment restrictions. In *Handbook of Statistics* (Vol. 11, pp. 419–454). doi:10.1016/S0169-7161(05)80051-3

Newey, W. K., & McFadden, D. (1994, January 1). Chapter 36 Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (Vol. 4, pp. 2111–2245). doi:10.1016/S1573-4412(05)80005-4

Newey, W. K., & Windmeijer, F. (2009). Generalized Method of Moments With Many Weak Moment Conditions. *Econometrica*, *77*(3), 687–719. doi:10.3982/ECTA6224. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA6224

Paulson, H. M., Jr. (2013). *On the Brink: Inside the Race to Stop the Collapse of the Global Financial System–With Original New Material on the Five Year Anniversary of the Financial Crisis*. Hachette UK.

Qian, E. (2024, June 6). Heterogeneity-robust granular instruments. doi:10.48550/arXiv.2304.01273. arXiv: 2304.01273 [econ]

Selgrad, J. (2023). Testing the Portfolio Rebalancing Channel of Quantitative Easing.

Sorkin, A. R. (2018). The Unknowable Fallout of China's Trade War Nuclear Option. *The New York Times: Business*. Retrieved from https://www.nytimes.com/2018/10/09/business/dealbook/china-trade-war-nuclear-option.html

Swanson, E. T. (2011). Let's twist again: A high-frequency event-study analysis of operation twist and its implications for QE2. *Brookings Papers on Economic Activity*, *2011*(1), 151–188.

van der Beck, P. (2025, May 16). Short- versus Long-run Demand Elasticities in Asset Pricing. doi:10.2139/ssrn.4111329. Social Science Research Network: 4111329

van der Beck, P., Bretscher, L., & Fu, J. Z. (2025, June 13). A Bound on Price Impact and Disagreement. doi:10.2139/ssrn.5293784. Social Science Research Network: 5293784

Vayanos, D., & Vila, J.-L. (2021). A Preferred-Habitat Model of the Term Structure of Interest Rates. *ECTA*, *89*(1), 77–112. doi:10.3982/ECTA17440

Vissing-Jorgensen, A. (2021). The Treasury Market in Spring 2020 and the Response of the Federal Reserve. *Journal of Monetary Economics*, *124*, 19–47. doi:10.1016/j.jmoneco.2021.10.007

Warnock, F. E., & Warnock, V. C. (2009). International capital flows and U.S. interest rates. *Journal of International Money and Finance*, *28*(6), 903–919. doi:10.1016/j.jimonfin.2009.03.002

Weiss, C. R. (2022). Foreign Demand for U.S. Treasury Securities during the Pandemic. Retrieved from https://www.federalreserve.gov/econres/notes/feds-notes/foreign-demand-for-us-treasury-securities-during-the-pandemic-20220128.html

Wolcott, E. L. (2020). Impact of Foreign Official Purchases of US Treasuries on the Yield Curve. *AEA Papers and Proceedings*, *110*, 535–540. doi:10.1257/pandp.20201124

A Methodological Appendix

A.1 General Framework

Here we introduce the more flexible model. Notation: for entities $i = 1, \dots, N$ and periods $t = 1, \dots, T$, let $q_{i,t} \in \mathbb{R}$ denote the quantity of entity i at time t, $p_t \in \mathbb{R}$ the price at time t, and $S_{i,t} > 0$ the (exogenous) size of entity i at time t. Let $\mathbf{C}_{i,t} \in \mathbb{R}^{N_\zeta}$ parameterize elasticities and $\mathbf{X}_{i,t} \in \mathbb{R}^{N_\beta}$ collect exogenous controls. Let $\boldsymbol{\zeta} \in \mathbb{R}^{N_\zeta}$ denote the elasticity vector and $\boldsymbol{\beta} \in \mathbb{R}^{N_\beta}$ denote the loading vector. We use the subscript S to denote size-weighted sums (e.g., $\mathbf{C}_{S,t} := \sum_i S_{i,t}\, \mathbf{C}_{i,t}$). The general model is specified as follows:

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

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 be entity&time-specific;
- The size vector $S_{i,t}$ is also allowed to be time-varying; we do not require it to sum to 1. The only assumption we impose on sizes is positivity, $S_{i,t} > 0$. 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 its elasticity has the same sign as other sectors.

The simple model can be recovered from the general specification by setting $\mathbf{C}_{i,t}$ to be the entity fixed effects, and $\mathbf{X}_{i,t}$ 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 $(\mathbf{X}_t, \mathbf{C}_t, \mathbf{S}_t)$:

$$\mathbb{E}\left[u_{i,t} u_{j,t} \mid \mathbf{X}_t, \mathbf{C}_t, \mathbf{S}_t\right] = 0 \text{ for all } i \neq j. \tag{A.2}$$

We discuss the handling of control variables $\mathbf{X}_{i,t}$ in detail in Appendix A.3. At the high level, when $\mathbf{X}_{i,t}$ are observable, we partial out the effect of controls by (i) regressing $q_{i,t}$ on $\mathbf{X}_{i,t}$ to obtain $q^{\varepsilon}_{i,t}$ and (ii) regressing p_t on $\mathbf{X}_{S,t} := \sum_i S_{i,t}\mathbf{X}_{i,t}$ to obtain p^{ε}_t. We then form the optimal GIV estimator for $\boldsymbol{\zeta}$ using the residuals $q^{\varepsilon}_{i,t}$ and p^{ε}_t. We provide the definition of the optimal GIV estimator and its asymptotic properties below, and provide the proof in Appendix section A.2.

Definition A.1. 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) :

$$\psi_{ij,t}(\mathbf{z}) \equiv \hat{u}_{i,t}(\mathbf{z})\,\hat{u}_{j,t}(\mathbf{z}) \equiv \left(q^{\varepsilon}_{i,t} + p^{\varepsilon}_t \mathbf{C}'_{i,t}\mathbf{z}\right)\left(q^{\varepsilon}_{j,t} + p^{\varepsilon}_t \mathbf{C}'_{j,t}\mathbf{z}\right).$$

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

$$\hat{\mathbf{W}}^*_{ij,k,t}(\mathbf{z}) = \frac{1}{\mathbf{C}'_{S,t}\mathbf{z}}\left(\frac{S_{j,t}C_{it,k}}{\hat{\sigma}_i^2(\mathbf{z})} + \frac{S_{i,t}C_{jt,k}}{\hat{\sigma}_j^2(\mathbf{z})}\right), \tag{A.3}$$

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:

$$\hat{\mathbb{E}}\left[\hat{\mathbf{W}}^*_t\left(\hat{\boldsymbol{\zeta}}\right)'\boldsymbol{\psi}_t\left(\hat{\boldsymbol{\zeta}}\right)\right] = \mathbf{0}. \tag{A.4}$$

While (A.3) looks complicated, it has a very intuitive interpretation.

Example 1. 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 weights different sectors using their sizes—entities with larger influence are weighted more in the estimation.

Example 2. If we further assume homogeneous elasticity across all entities, the optimal GIV estimator further weights entities sharing the same elasticity using their own residual volatilities:

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

This is akin to the generalized least squares (GLS) estimator: entities with noisier observations are weighted less in estimation.

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

Theorem A.1 (Asymptotic efficiency in the general model)**.** *Given the moment conditions (A.2), under regularity conditions*[32]*, the optimal GIV estimator $\hat{\boldsymbol{\zeta}}$ is consistent and asymptotically normal:*

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

for $T\to\infty$, where

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

Moreover, $\mathbf{V}_\zeta^{W*}$ achieves the semiparametric efficiency bound.

Proof. See Appendix A.2. □

A.2 Proof of Theorem A.1

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 $\mathbb{E}\left[u_{i,t}^2\mid \mathrm{C}_t,\mathrm{S}_t,\mathbf{X}_t\right]=\sigma_i^2>0$.[33]
3. For any $i\neq j$, $u_{i,t}$ is independent of $u_{j,t}$ 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.

[32] See Appendix A.2 for the exact conditions.

[33] This i.i.d. condition is used for efficiency. Consistency of $\hat{\boldsymbol{\zeta}}$ only requires $\mathbb{E}[u_{i,t}\mid\mathbf{C}_t,\mathbf{S}_t,\mathbf{X}_t]=0$ and finite second moments. If $u_{i,t}$ exhibits serial correlation, the optimal (efficient) weighting replaces σ_i^2 by the long-run variance $\Omega_i:=\sum_{\ell=-\infty}^{\infty}\gamma_i(\ell)$ with $\gamma_i(\ell)=\mathbb{E}[u_{i,t}u_{i,t-\ell}]$ (i.e., HAC/Newey–West). The resulting optimal weight retains the same form with σ_i^2 replaced by Ω_i.

In this subsection, we proceed as if there were no exogenous variables $\mathbf{X}_{i,t}$. As we show in A.3, all results described here go through with observed common factors.

We proceed in the following steps. We first derive the asymptotic properties for a broad class of GMM estimators with generic weighting matrices. Then we derive the semiparametric efficiency bound for the class of GMM estimators. The first two steps follow the standard GMM literature. Finally, we show that the optimal GIV estimator in Equation (A.4) achieves the semiparametric efficiency bound.

A.2.1 Asymptotic Properties of GMM Estimators

For ease of reading, we restate the moment conditions of the model here.

Let $\mathbf{z}$ be a generic parameter vector. The full moment conditions of the model can be summarized with a vector $\boldsymbol{\psi}_t(\mathbf{z})$ of length $\frac{N(N-1)}{2}$ stacking pairs with indices $i < j$ (we take $i < j$ to avoid duplicating symmetric cross-moments); the subscript ij corresponds to the entity pair (i, j):

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

We have the following moment conditions that under the true parameter $\boldsymbol{\zeta}$, the conditional expectation of $\boldsymbol{\psi}$ is zero:

$$\begin{aligned}\mathbb{E}\left[\psi_{ij,t}(\boldsymbol{\zeta}) \mid \mathbf{C}_t, \mathbf{S}_t\right] &= \mathbb{E}\left[\left(q_{i,t} + p_t \mathbf{C}'_{i,t}\boldsymbol{\zeta}\right)\left(q_{j,t} + p_t \mathbf{C}'_{j,t}\boldsymbol{\zeta}\right) \mid \mathbf{C}_t, \mathbf{S}_t\right] \\ &= \mathbb{E}\left[u_{i,t} u_{j,t} \mid \mathbf{C}_t, \mathbf{S}_t\right] = \mathbf{0}. \end{aligned} \tag{A.5}$$

Notice that this is an overidentified system, as we have $\frac{N(N-1)}{2}$ moment conditions while N_ζ parameters to be estimated.

Consider the class of GMM estimators $\hat{\boldsymbol{\zeta}}^W$ indexed by the weighting matrix $\mathbf{W}_t(\mathbf{z})$. The weighting matrix is measurable with respect to the filtration generated by $\mathbf{C}_t$ and $\mathbf{S}_t$. The estimator $\hat{\boldsymbol{\zeta}}^W$ solves:[34]

[34]This formulation nests the standard $\hat{\boldsymbol{\zeta}}^A = \arg\min_{\boldsymbol{z}} \sum_t \boldsymbol{\psi}_t(\boldsymbol{z})'\mathbf{A}\boldsymbol{\psi}_t(\boldsymbol{z})$, which is equivalent to a time-invariant weighting matrix for the moment conditions after taking the first order condition. As the moment conditions are conditional expectations, the efficient weighting matrix is typically time-varying. See Newey (1993).

$$\mathbf{b}_T^{\mathbf{W}}\left(\hat{\boldsymbol{\zeta}}^W\right) \equiv \frac{1}{T}\sum_t W_t\left(\hat{\boldsymbol{\zeta}}^W\right)' \boldsymbol{\psi}_t\left(\hat{\boldsymbol{\zeta}}^W\right) = 0. \tag{A.6}$$

As a shorthand, we also define $\mathbf{W}_t \equiv \mathbf{W}_t(\boldsymbol{\zeta})$ as the weighting scheme evaluated at the true parameter.

To ensure the estimator is well-behaved, we make one additional assumption on the weighting matrices.

Assumption A.2. $\mathbb{E}\left[\mathbf{W}_t{}'\mathbf{W}_t\right]$ *is non-singular.* $\mathbb{E}\left[\mathbf{W}_t{}'\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)\right] = 0$ *if and only if* $\mathbf{z} = \boldsymbol{\zeta}$.[35]

The lemma below characterizes the asymptotic properties of the GMM estimator $\hat{\boldsymbol{\zeta}}^W$.

Lemma A.1 (Asymptotic Properties of GMM Estimator.)**.** *Under Assumption A.2, the GMM estimator* $\hat{\boldsymbol{\zeta}}^W$ *is consistent and asymptotically normal, and its asymptotic variance is given by:*

$$\mathbf{V}_\zeta^W = \mathbb{E}\left[\mathbf{D}_t'\mathbf{W}_t\right]^{-1}\mathbb{E}\left[\mathbf{W}_t'\mathbf{V}^{\psi}\mathbf{W}_t\right]\mathbb{E}\left[\mathbf{W}_t'\mathbf{D}_t\right]^{-1} \tag{A.7}$$

where $\mathbf{J}_t = \frac{\partial \boldsymbol{\psi}_t(\mathbf{z})}{\partial \mathbf{z}}\mid_{\mathbf{z}=\boldsymbol{\zeta}}$ *is the (raw) Jacobian matrix of* $\boldsymbol{\psi}_t\left(\mathbf{z}\right)$ *evaluated at the true parameter, we define* $\mathbf{D}_t \equiv \mathbb{E}[\mathbf{J}_t \mid \mathbf{C}_t, \mathbf{S}_t]$*, and* $\mathbf{V}^{\psi} \equiv \mathbb{E}\left[\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)'\right] = \operatorname{diag}\left(\sigma_1^2\sigma_2^2, \sigma_1^2\sigma_3^2, \ldots, \sigma_{N-1}^2\sigma_N^2\right)$ *is the covariance matrix of* $\boldsymbol{\psi}_t\left(\boldsymbol{\zeta}\right)$*.*

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

[35] For the optimal weighting matrix $\mathbf{W}_t^*$, the second part of this assumption follows from the first part in many cases. 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 imply unique identification. The additional assumption is only needed when we have nonrestrictive loading factors $\mathbf{C}_{i,t}$. To see this, consider another candidate $\tilde{\boldsymbol{\zeta}} \equiv \boldsymbol{\zeta} + \Delta\boldsymbol{\zeta}$. Write $M_t := 1/(\mathbf{C}_{S,t}'\boldsymbol{\zeta})$ and let $\sigma_p^2 := \operatorname{Var}(p_t \mid \mathbf{C}_t, \mathbf{S}_t)$. We have:

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

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.

derive the asymptotic variance, following the standard GMM formula, we have:

$$\mathbf{V}_{\zeta}^{W} = \mathbb{E}\left[\mathbf{D}_{W,t}'\right]^{-1}\mathbb{E}\left[\mathbf{W}_t'\boldsymbol{\psi}_t(\boldsymbol{\zeta})\,\boldsymbol{\psi}_t(\boldsymbol{\zeta})'\,\mathbf{W}_t\right]\mathbb{E}\left[\mathbf{D}_{W,t}\right]^{-1}$$

where $\mathbf{D}_{W,t} = \frac{\partial(\mathbf{W}_t(\mathbf{z})'\boldsymbol{\psi}_t(\mathbf{z}))}{\partial \mathbf{z}} \mid_{z=\zeta}$. Evaluating $\mathbf{D}_{W,t}$ at $\mathbf{z} = \boldsymbol{\zeta}$, we have:

$$\mathbb{E}\left[\mathbf{D}_{W,t}\right] = \mathbb{E}\left[\frac{\partial \mathbf{W}_t(\mathbf{z})'}{\partial \mathbf{z}}\boldsymbol{\psi}_t(\boldsymbol{\zeta}) + \mathbf{W}_t'\frac{\partial \boldsymbol{\psi}_t(\mathbf{z})}{\partial \mathbf{z}}\right]\Bigg|_{\mathbf{z}=\boldsymbol{\zeta}} = \mathbb{E}\left[\mathbf{W}_t'\mathbf{D}_t\right]$$

Using the law of iterated expectations, we have: $\mathbb{E}\left[\mathbf{W}_t'\boldsymbol{\psi}_t(\boldsymbol{\zeta})\,\boldsymbol{\psi}_t(\boldsymbol{\zeta})'\,\mathbf{W}_t\right] = \mathbb{E}\left[\mathbf{W}_t'\mathbf{V}^{\psi}\mathbf{W}_t\right]$. Plugging them in, we obtain Equation (A.7). □

A.2.2 Semiparametric Efficiency Bound

The semi-parametric efficiency bound for the class of GMM estimators is given by the following lemma:

Lemma A.2 (Semiparametric efficiency bound)**.** *Under Assumptions A.1-A.2, the asymptotic variance of the GMM estimator* $\hat{\boldsymbol{\zeta}}^{W}$ *is minimized at the optimal weighting matrix* $\mathbf{W}_t^*$ *given by:*

$$\mathbf{W}_t^* = \left(\mathbf{V}^{\psi}\right)^{-1}\mathbf{D}_t, \tag{A.8}$$

and the lower bound of the asymptotic variance is given by:

$$\mathbf{V}_{\zeta}^{W*} = \mathbb{E}\left[\mathbf{D}_t'\left(\mathbf{V}^{\psi}\right)^{-1}\mathbf{D}_t\right]^{-1} = \mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^*\right]^{-1}. \tag{A.9}$$

Proof. The proof for the optimality generally follows the strategy of optimal instrument in conditional models, surveyed in Newey (1993).

To show Equation (A.9), note that Since $\mathbf{W}_t^* = (\mathbf{V}^{\psi})^{-1}\mathbf{D}_t$ and $\mathbf{V}^{\psi}$ is nonrandom (diagonal in $\sigma_i^2\sigma_j^2$), we have the pointwise identity

$$\mathbf{D}_t'(\mathbf{V}^{\psi})^{-1}\mathbf{D}_t = \mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^*.$$

Taking expectations on both sides yields Equation (A.9).

To show $\mathbf{V}_{\zeta}^{W*}$ is the lower bound of the asymptotic variance, we need to show that for any $\mathbf{W}_t$, $\mathbf{V}_{\zeta}^{W} - \mathbf{V}_{\zeta}^{W^*}$ is positive semi-definite. Given the covariance matrices are symmetric positive definite, it is equivalent to show $\boldsymbol{\Delta} \equiv \left(\mathbf{V}_\zeta^{W^*}\right)^{-1} - \left(\mathbf{V}_\zeta^{W}\right)^{-1}$ is positive semi-definite:

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

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:

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

Define $\mathbf{G}_t = \begin{bmatrix} \tilde{\mathbf{D}}_t' \\ \tilde{\mathbf{W}}_t' \end{bmatrix}$, and $\mathbf{H}' = \begin{bmatrix} \mathbf{I} & -\mathbb{E}\left[\tilde{\mathbf{D}}_t'\tilde{\mathbf{W}}_t\right] \mathbb{E}\left[\tilde{\mathbf{W}}_t'\tilde{\mathbf{W}}_t\right]^{-1} \end{bmatrix}$, we can express $\boldsymbol{\Delta}$ as:

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

Clearly, $\boldsymbol{\Delta}$ is positive semi-definite. Moreover, $\boldsymbol{\Delta} = 0$ when $\tilde{\mathbf{W}}_t = \tilde{\mathbf{D}}_t$, or $\mathbf{W}_t = \left(\mathbf{V}^\psi\right)^{-1} \mathbf{D}_t$.
□

A.2.3 Deriving the optimal GIV estimator

The remaining step is to show Equation (A.8) gives the expression used in the optimal GIV estimator in Equation (A.3).

To show that, let $\mathbf{J}_t \equiv \frac{\partial \boldsymbol{\psi}_t(\mathbf{z})}{\partial \mathbf{z}} |_{\mathbf{z}=\boldsymbol{\zeta}}$ denote the raw Jacobian evaluated at the true parameter. Its entry is given as:

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

Note that from the model structure, we have:

$$p_t = \frac{1}{\mathbf{C}_{S,t}'\boldsymbol{\zeta}}\left[\mathbf{X}_{S,t}'\boldsymbol{\beta} + u_{S,t}\right]$$

Hence we have:

$$\mathbb{E}\left[u_{j,t}p_t \mid \mathbf{C}_t, \mathbf{S}_t\right] = \frac{1}{\mathbf{C}_{S,t}'\boldsymbol{\zeta}} S_{j,t}\sigma_j^2.$$

And therefore, Since $\mathbf{D}_t = \mathbb{E}[\mathbf{J}_t \mid \mathbf{C}_t, \mathbf{S}_t]$, we have

$$D_{ij,k,t} = \mathbb{E}\left[J_{ij,k,t} \mid \mathbf{C}_t, \mathbf{S}_t\right] = \frac{1}{\mathbf{C}_{S,t}'\boldsymbol{\zeta}}\left[S_{j,t}\sigma_j^2 C_{it,k} + S_{i,t}\sigma_i^2 C_{jt,k}\right].$$

Recall that $\mathbf{V}^{\psi} \equiv \mathbb{E}\left[\boldsymbol{\psi}_t(\boldsymbol{\zeta})\,\boldsymbol{\psi}_t(\boldsymbol{\zeta})'\right] = \operatorname{diag}\left(\sigma_1^2\sigma_2^2, \sigma_1^2\sigma_3^2, \ldots, \sigma_{N-1}^2\sigma_N^2\right)$. Plug into Equation (A.8), we have:

$$\mathbf{W}^*_{ij,k,t} = \left(V_{ij}^{\psi}\right)^{-1} D_{ij,k,t} = \frac{1}{\mathbf{C}_{S,t}'\boldsymbol{\zeta}}\left(\frac{S_{j,t}C_{it,k}}{\sigma_i^2} + \frac{S_{i,t}C_{jt,k}}{\sigma_j^2}\right), \tag{A.10}$$

which is exactly the weighting matrix used in the optimal GIV estimator in Equation (A.3).

Proof of Theorem 1. Finally, the proof to the Theorem 1 in the main text is a special case of the proof above, where $S_{jt} = S_j$ and $C_{it,k} = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$. The only remaining step is to derive explicitly the asymptotic variance in the special case.

In this case, the optimal weighting matrix is given as:

$$W^*_{ij,k,t} = \frac{1}{\zeta_S}\left(\frac{S_j C_{it,k}}{\sigma_i^2} + \frac{S_i C_{jt,k}}{\sigma_j^2}\right) = \begin{cases} \dfrac{S_j}{\zeta_S\,\sigma_i^2}, & k = i, \\ \dfrac{S_i}{\zeta_S\,\sigma_j^2}, & k = j, \\ 0, & k \neq i, j. \end{cases}$$

Plugging this into $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^*\right]$ with the expression for $\mathbf{V}^{\psi}$, we obtain the closed-form entries for the $N \times N$ matrix $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^*\right]$. Define $\Sigma \equiv \sum_m S_m^2\sigma_m^2$. Then for any $k, \ell \in \{1, \ldots, N\}$,

$$\left[\mathbb{E}[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^*]\right]_{k\ell} = \frac{1}{\zeta_S^2}\begin{cases} \dfrac{\Sigma}{\sigma_k^2} - S_k^2, & k = \ell, \\ S_k S_\ell, & k \neq \ell. \end{cases} \tag{A.11}$$

Hence, the asymptotic variance in this case is given as:

$$\mathbf{V}_{\zeta}^{W*} = \zeta_S^2 \times \text{Inv}\left(\begin{bmatrix} \frac{1}{\sigma_1^2}\sum_{i\neq 1} S_i^2\sigma_i^2 & S_1S_2 & S_1S_3 & \cdots & S_1S_N \\ S_1S_2 & \frac{1}{\sigma_2^2}\sum_{i\neq 2} S_i^2\sigma_i^2 & S_2S_3 & \cdots & S_2S_N \\ S_1S_3 & S_2S_3 & \frac{1}{\sigma_3^2}\sum_{i\neq 3} S_i^2\sigma_i^2 & \cdots & S_3S_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_1S_N & S_2S_N & S_3S_N & \cdots & \frac{1}{\sigma_N^2}\sum_{i\neq N} S_i^2\sigma_i^2 \end{bmatrix}\right). \tag{A.12}$$

A.3 Common Factors and Identification of Factor Loadings $\boldsymbol{\lambda}_i$

We first consider the case with external common factors. They can be directly observable, such as macrofinancial variables, or unobservable principal components extracted from granular data *outside of the model*. Then we discuss the method to extract the unobserved common factors *within the model*.

A.3.1 External Common Factors

With external common factors, we proceed with the following steps. We have an additional set of moment conditions to identify the coefficients on $\mathbf{X}_{i,t}$:

$$\mathbb{E}\left[\mathbf{X}_{i,t}u_{i,t}\right] = \mathbb{E}\left[\mathbf{X}_{i,t}(q_{i,t} + p_t\mathbf{C}'_{i,t}\boldsymbol{\zeta} - \mathbf{X}'_{i,t}\boldsymbol{\beta})\right] = \mathbf{0}. \tag{A.13}$$

Hence, in principle, $\boldsymbol{\zeta}$ and $\boldsymbol{\beta}$ are estimated jointly in a GMM system with both sets of moment conditions. Algorithmically, for each guess of $\boldsymbol{\zeta}$, we can estimate $\boldsymbol{\beta}$ by regressing $q_{i,t} + p_t\mathbf{C}'_{i,t}\boldsymbol{\zeta}$ on $\mathbf{X}_{i,t}$; then we use the residuals to form the moment conditions in (A.2).

We can further simplify the algorithm by utilizing the linearity of the OLS estimator. Notice that regressing $q_{i,t} + p_t\mathbf{C}'_{i,t}\boldsymbol{\zeta}$ on $\mathbf{X}_{i,t}$ is equivalent to regressing $q_{i,t}$ on $\mathbf{X}_{i,t}$ and $p_t\mathbf{C}'_{i,t}$ on $\mathbf{X}_{i,t}$, and combine the residuals. Hence, we can follow the three-step algorithm below:

0. (If needed: Use PCA to extract unobserved common factors from external datasets, and include them in $\mathbf{X}_{i,t}$)
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, $q^{\varepsilon}_{i,t} \equiv q_{i,t} - \mathbf{X}'_{i,t}\hat{\boldsymbol{\beta}}^q$ and $\mathbf{pC}^{\varepsilon}_{i,t}{}' = p_t\mathbf{C}'_{it} - \mathbf{X}'_{i,t}\hat{\boldsymbol{\beta}}^{pC}$,

and form the moment conditions:

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

Proceed with the optimal weighting matrix defined in (A.3) to estimate $\hat{\zeta}$.[36]

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.9) 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:*

$$Var(\hat{\boldsymbol{\beta}}) = \frac{1}{T}\left[Var^{OLS}(\hat{\boldsymbol{\beta}}) + \boldsymbol{\beta}^{pC}Var(\hat{\boldsymbol{\zeta}})\boldsymbol{\beta}^{pC\prime}\right] \tag{A.14}$$

$$Var^{OLS}\left(\hat{\boldsymbol{\beta}}\right) = \left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\right)^{-1}\left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\sigma_i^2\right)\left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\right)^{-1} \tag{A.15}$$

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

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

where $\mathrm{g}^{OLS}(\mathbf{b})$ 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.[37] 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

[36] Note that we still use the original $\mathbf{C}_{i,t}$ in constructing the weighting without residualizing it, as $\frac{\partial \psi_{ij,t}}{\partial \zeta_k} \equiv D_{ij,k,t} = u_{jt}p_tC_{it,k} + u_{it}p_tC_{jt,k}$ regardless whether we have the common factors $X_{i,t}$ on the right-hand side.

[37] Here, we take the shortcut by formulating the GMM system in terms of $\boldsymbol{\beta}^q$ and $\boldsymbol{\beta}^{pC}$ rather than $\boldsymbol{\beta}$ directly as in (A.13). The derivation is slightly more complicated as the (A.13) involves both $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ in the moment conditions, but the results are equivalent.

estimator, we have:

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

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

$$\mathbf{D}^{g}=\begin{bmatrix}\frac{\partial\mathbf{g}_t^{OLS}(\mathbf{b})}{\partial\mathbf{b}}\mid_{\mathbf{b}=\boldsymbol{\beta}} & \\ & \mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{D}_t\right]\end{bmatrix}$$
$$\mathbf{V}^{g}=\begin{bmatrix}\mathbf{V}^{OLS} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}^{*\prime}\mathbf{V}^{\psi}\mathbf{W}^{*}\end{bmatrix}.$$

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 $u_{it}u_{jt}$ for pairs $i\neq j$. By the cross-sectional independence assumption (Assumption A.1 part 3), any linear combination between them has 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:

$$Var(\hat{\boldsymbol{\beta}})=\begin{bmatrix}\mathbf{I} & \boldsymbol{\zeta}\end{bmatrix}'Var(\begin{bmatrix}\hat{\boldsymbol{\beta}}^{q} & \hat{\boldsymbol{\beta}}^{pC}\end{bmatrix})\begin{bmatrix}\mathbf{I} & \boldsymbol{\zeta}\end{bmatrix}'+\boldsymbol{\beta}^{pC}Var(\hat{\boldsymbol{\zeta}})\boldsymbol{\beta}^{pC\prime}$$

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.15). □

A.3.2 Internal Common Factor Estimation

In this section we study identification when common demand shocks are unobservable, and to be estimated together with the other model parameters.

Factor structure for residuals. Suppose the residual $u_{i,t}$ contains common factors $\boldsymbol{\eta}_t$ and idiosyncratic shocks, with $\boldsymbol{\epsilon}_t := (\epsilon_{1,t}, \dots, \epsilon_{N,t})'$:

$$u_{i,t} := \boldsymbol{\lambda}_i' \boldsymbol{\eta}_t + \epsilon_{i,t}$$

where

$$\mathbb{E}[\boldsymbol{\eta}_t] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \mathbf{I}_K,\ \mathbb{E}[\boldsymbol{\epsilon}_t] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'] = \boldsymbol{\Sigma}_\epsilon = \mathrm{diag}(\sigma_1^2, \dots, \sigma_N^2),\ \mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\epsilon}_t'] = \mathbf{0}. \quad \text{(A.16)}$$

We normalize the factors to have unit variance. We also denote $\boldsymbol{\Lambda}$ as the $N \times K$ matrix formed by stacking $\boldsymbol{\lambda}_i'$ row-wise.

Let $\hat{u}_{i,t}(\boldsymbol{\zeta}) = q_{i,t} + p_t\, \mathbf{C}_{i,t}' \boldsymbol{\zeta}$ denote the residual for a candidate elasticity vector $\boldsymbol{\zeta}$. In this case the moment condition for pair (i,j) has an additional term due to the unobserved common factors:

$$\psi_{ij,t}(\boldsymbol{\zeta}; \boldsymbol{\Lambda}) \equiv \hat{u}_{i,t}(\boldsymbol{\zeta})\, \hat{u}_{j,t}(\boldsymbol{\zeta}) - \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_j, \qquad i \neq j. \quad \text{(A.17)}$$

Stacking over (i,j) yields $\frac{N(N-1)}{2}$ moments. We can use the GMM estimator for $\boldsymbol{\zeta}$ as in Appendix A.1, replacing the zero off-diagonal restriction with (A.17) to capture the additional unobserved common variation.

Weighting. With latent factors, $\mathrm{Cov}\big(\psi_{ij,t}, \psi_{kl,t}\big)$ are generally non-zero because of common factors $\boldsymbol{\eta}_t$. Hence, the covariance matrix of the moments V^ψ is no longer diagonal and hence the existing weighting is no longer optimal. In principle, V^ψ can be represented using low-rank matrices involving $\boldsymbol{\Lambda}$, and hence we can still derive the optimal weighting matrix using these representations. In practice, when unobserved factor-induced comovement is relatively small, the diagonal weighting from Appendix A.1 remains close to optimal and efficient for computation, so we adopt it in our baseline implementation. Note that due to off-diagonals being non-zero, the formula for standard errors no longer applies; to address this, we compute bootstrap standard errors.

Unknown loadings: profiled internal PCA. We treat $\boldsymbol{\Lambda}$ as a nuisance parameter and profile it out i.e. solve for it in the inner loop. For each trial $\boldsymbol{\zeta}$, compute residuals $\hat{u}_{i,t}(\boldsymbol{\zeta})$

and the sample covariance $\widehat{\boldsymbol{\Sigma}}_{uu}(\boldsymbol{\zeta}) = T^{-1}\sum_t \hat{\mathbf{u}}_t(\boldsymbol{\zeta})\,\hat{\mathbf{u}}_t(\boldsymbol{\zeta})'$. We estimate the factor loadings $\widehat{\boldsymbol{\Lambda}}(\boldsymbol{\zeta})$ using heteroskedasticity-robust PCA methods (e.g., HeteroPCA or deflated HeteroPCA).

The off-diagonal moment restrictions become

$$\widehat{\psi}_{ij}(\boldsymbol{\zeta}) = \hat{u}_{i,t}(\boldsymbol{\zeta})\,\hat{u}_{j,t}(\boldsymbol{\zeta}) - \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\zeta})'\,\hat{\boldsymbol{\lambda}}_j(\boldsymbol{\zeta}), \qquad i \neq j, \tag{A.18}$$

A.3.3 Identifiability

When we estimate $\boldsymbol{\Lambda}$ internally, uniqueness of the solution is no longer guaranteed and depends on the exact specification. As a rule of thumb, we show that identification fails when $N_\zeta + K > N$. Below we provide precise conditions that clarify when the moment condition has a non-unique solution.

Moment representation. Using matrix notation, we define the observable second moments as:

$$\boldsymbol{\Sigma}_{qq} := \mathbb{E}[\mathbf{q}_t\mathbf{q}_t'] \in \mathbb{R}^{N\times N}, \qquad \boldsymbol{\Sigma}_{qp} := \mathbb{E}[\mathbf{q}_t p_t] \in \mathbb{R}^{N\times 1}, \qquad \sigma_p^2 := \mathbb{E}[p_t^2] \in \mathbb{R},$$

For tractability, in this section we consider the elasticity loading matrix $\mathbf{C}$ to be time-invariant. The moment conditions are then given in terms of observables and parameters:

$$\boldsymbol{\Sigma}_{qq} \;+\; \boldsymbol{\Sigma}_{qp}\,(\mathbf{C}\boldsymbol{\zeta})' \;+\; (\mathbf{C}\boldsymbol{\zeta})\,\boldsymbol{\Sigma}_{qp}' \;+\; \sigma_p^2\,(\mathbf{C}\boldsymbol{\zeta})(\mathbf{C}\boldsymbol{\zeta})' = \boldsymbol{\Lambda}\,\boldsymbol{\Lambda}' \;+\; \boldsymbol{\Sigma}_\epsilon. \tag{A.19}$$

We denote $(\boldsymbol{\zeta}_\star, \boldsymbol{\Lambda}_\star, \boldsymbol{\Sigma}_\star)$ as the true parameters. Identification means $(\boldsymbol{\zeta}_\star, \boldsymbol{\Lambda}_\star, \boldsymbol{\Sigma}_\star)$ is the unique solution to the moment condition (A.19). Equivalently, let $(\tilde{\boldsymbol{\zeta}}, \tilde{\boldsymbol{\Lambda}}, \tilde{\boldsymbol{\Sigma}})$ denote deviations from the truth.Identification requires that the only solution to the perturbed system (A.20) is $\tilde{\boldsymbol{\zeta}} = \mathbf{0}$, $\tilde{\boldsymbol{\Lambda}} = \mathbf{0}$, and $\tilde{\boldsymbol{\Sigma}} = \mathbf{0}$.

$$(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})\,\boldsymbol{\Sigma}_{qp}' + \boldsymbol{\Sigma}_{qp}\,(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' + \sigma_p^2\Big((\mathbf{C}\,\tilde{\boldsymbol{\zeta}})(\mathbf{C}\,\boldsymbol{\zeta}_\star)' + (\mathbf{C}\,\boldsymbol{\zeta}_\star)(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' + (\mathbf{C}\,\tilde{\boldsymbol{\zeta}})(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})'\Big) = \boldsymbol{\Lambda}_\star\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Lambda}}\,\boldsymbol{\Lambda}_\star' + \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Sigma}}. \tag{A.20}$$

To characterize identification more clearly, we study local identifiability around the true parameters by focusing on the first-order approximation to (A.20). Dropping the quadratic terms and refactoring, the first-order approximation is

$$(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})\left(\boldsymbol{\Sigma}_{qp} + \sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_\star\right)' + \left(\boldsymbol{\Sigma}_{qp} + \sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_\star\right)(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' - \tilde{\boldsymbol{\Sigma}} = \;\boldsymbol{\Lambda}_\star\,\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Lambda}}\,\boldsymbol{\Lambda}_\star'. \tag{A.21}$$

Proposition A.2. *The parameters are first-order identified if and only if the only solution to the following equation is* $\tilde{\boldsymbol{\zeta}} = \mathbf{0}$*:*

$$\mathbf{Q}^{\Lambda}\,\mathbf{C}\,\tilde{\boldsymbol{\zeta}}\,\big(\mathbf{S}'\boldsymbol{\Sigma}_{\star}\big)\mathbf{Q}^{\Lambda} \;+\; \mathbf{Q}^{\Lambda}\,\big(\boldsymbol{\Sigma}_{\star}\,\mathbf{S}\big)\,\tilde{\boldsymbol{\zeta}}'\mathbf{C}'\,\mathbf{Q}^{\Lambda} = \zeta_S^{\star}\,\mathbf{Q}^{\Lambda}\,\tilde{\boldsymbol{\Sigma}}\,\mathbf{Q}^{\Lambda}, \tag{A.22}$$

where $\mathbf{R}^{\Lambda} := \boldsymbol{\Lambda}_{\star}(\boldsymbol{\Lambda}_{\star}'\boldsymbol{\Lambda}_{\star})^{-1}\boldsymbol{\Lambda}_{\star}'$*, and* $\mathbf{Q}^{\Lambda} := \mathbf{I} - \mathbf{R}^{\Lambda}$ *as the null space projector with respect to* $\boldsymbol{\Lambda}_{\star}$*, so that* $\mathbf{R}^{\Lambda}\boldsymbol{\Lambda}_{\star} = \boldsymbol{\Lambda}_{\star}$ *and* $\mathbf{Q}^{\Lambda}\boldsymbol{\Lambda}_{\star} = \mathbf{0}$*.*

Proof. Necessity: Expressing $\boldsymbol{\Sigma}_{qp}$ and σ_p^2 in terms of the parameters, we have

$$\boldsymbol{\Sigma}_{qp} + \sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_{\star} = \frac{\boldsymbol{\Lambda}_{\star}\boldsymbol{\Lambda}_{S,\star}' + \boldsymbol{\Sigma}_{\star}\,\mathbf{S}}{\zeta_S^{\star}}, \quad \text{where } \boldsymbol{\Lambda}_{S,\star} := \mathbf{S}'\boldsymbol{\Lambda}_{\star},\ \zeta_S^{\star} := \mathbf{S}'\mathbf{C}\,\boldsymbol{\zeta}_{\star}.$$

Multiplying (A.21) by $\mathbf{Q}^{\Lambda}$ on both sides and simplifying using $\mathbf{Q}^{\Lambda}\boldsymbol{\Lambda}_{\star} = \mathbf{0}$ yields (A.22).

Sufficiency: Suppose (A.22) holds for a nonzero $\tilde{\boldsymbol{\zeta}}$ and a diagonal $\tilde{\boldsymbol{\Sigma}}$. Define the left-hand side of (A.21) as

$$\mathbf{L} := (\mathbf{C}\,\tilde{\boldsymbol{\zeta}})\,\big(\boldsymbol{\Sigma}_{qp} + \sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_{\star}\big)' \;+\; \big(\boldsymbol{\Sigma}_{qp} + \sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_{\star}\big)\,(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' \;-\; \tilde{\boldsymbol{\Sigma}}.$$

By (A.22) and $\mathbf{Q}^{\Lambda}\boldsymbol{\Lambda}_{\star} = \mathbf{0}$, we have $\mathbf{Q}^{\Lambda}\,\mathbf{L}\,\mathbf{Q}^{\Lambda} = \mathbf{0}$. Now set

$$\tilde{\boldsymbol{\Lambda}} := \mathbf{L}\,\boldsymbol{\Lambda}_{\star}\,(\boldsymbol{\Lambda}_{\star}'\boldsymbol{\Lambda}_{\star})^{-1} \;-\; \tfrac{1}{2}\,\mathbf{R}^{\Lambda}\,\mathbf{L}\,\boldsymbol{\Lambda}_{\star}\,(\boldsymbol{\Lambda}_{\star}'\boldsymbol{\Lambda}_{\star})^{-1}.$$

Using $\mathbf{L} = \mathbf{L}'$ and $(\mathbf{R}^{\Lambda})' = \mathbf{R}^{\Lambda}$, we compute

$$\boldsymbol{\Lambda}_{\star}\,\tilde{\boldsymbol{\Lambda}}' = \mathbf{R}^{\Lambda}\,\mathbf{L} - \tfrac{1}{2}\,\mathbf{R}^{\Lambda}\,\mathbf{L}\,\mathbf{R}^{\Lambda}, \qquad \tilde{\boldsymbol{\Lambda}}\,\boldsymbol{\Lambda}_{\star}' = \mathbf{L}\,\mathbf{R}^{\Lambda} - \tfrac{1}{2}\,\mathbf{R}^{\Lambda}\,\mathbf{L}\,\mathbf{R}^{\Lambda}.$$

Adding yields

$$\boldsymbol{\Lambda}_{\star}\,\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Lambda}}\,\boldsymbol{\Lambda}_{\star}' = (\mathbf{R}^{\Lambda}\,\mathbf{L} + \mathbf{L}\,\mathbf{R}^{\Lambda}) - \mathbf{R}^{\Lambda}\,\mathbf{L}\,\mathbf{R}^{\Lambda}.$$

Using the decomposition $\mathbf{L} = (\mathbf{R}^{\Lambda} + \mathbf{Q}^{\Lambda})\,\mathbf{L}\,(\mathbf{R}^{\Lambda} + \mathbf{Q}^{\Lambda})$, the right-hand side equals $\mathbf{L} - \mathbf{Q}^{\Lambda}\,\mathbf{L}\,\mathbf{Q}^{\Lambda} = \mathbf{L}$, because $\mathbf{Q}^{\Lambda}\,\mathbf{L}\,\mathbf{Q}^{\Lambda} = \mathbf{0}$. Therefore, (A.21) holds, proving lack of first-order identification. □

The general characterization of identification depends on the exact $\mathbf{C}$. However, corollaries below offer two sufficient conditions when the first-order identification fails.

Corollary 1. *The parameters are not first-order identified when* $N_{\zeta} + K > N$*.*

Proof. When $N_\zeta + K > N$, we have $\mathcal{S} := \mathrm{col}(\mathbf{C}) \cap \mathrm{col}(\mathbf{\Lambda}_\star) \neq \{\mathbf{0}\}$. Pick any non-zero $\mathbf{v} \in \mathcal{S}$ and choose $\tilde{\boldsymbol{\zeta}}$ such that $\mathbf{C}\,\tilde{\boldsymbol{\zeta}} = \mathbf{v}$. Since $\mathbf{v} \in \mathrm{col}(\mathbf{\Lambda}_\star)$, we have $\mathbf{Q}^\Lambda \mathbf{C}\,\tilde{\boldsymbol{\zeta}} = \mathbf{Q}^\Lambda \mathbf{v} = \mathbf{0}$. With $\tilde{\mathbf{\Sigma}} = \mathbf{0}$, Equation (A.22) holds. □

The corollary states that the number of elasticity coefficients plus the number of factors is greater than the number of entities. Therefore, with fully heterogeneous elasticities we cannot identify factors internally. Empirically, identification is restored by imposing structure—e.g., grouping entities to share elasticities or restricting K—so that $\mathrm{col}(\mathbf{C})$ is sufficiently "transverse" to $\mathrm{col}(\mathbf{\Lambda}_\star)$ and the dimension count is favorable.

Corollary 2. *The parameters are not first-order identified when* $\mathbf{\Sigma}_\star \mathbf{S} \in \mathrm{col}(\mathbf{\Lambda}_\star)$.

Proof. When $\mathbf{\Sigma}_\star \mathbf{S} \in \mathrm{col}(\mathbf{\Lambda}_\star)$, we have $\mathbf{Q}^\Lambda \mathbf{\Sigma}_\star \mathbf{S} = \mathbf{0}$, hence Equation (A.22) holds for any $\tilde{\boldsymbol{\zeta}}$. □

This is akin to the granularity condition required by Gabaix and Koijen (2024) in the original GIV setup—the condition is satisfied when the size–volatility distribution is uniform, so $\mathbf{\Sigma}_\star \mathbf{S} \propto \mathbf{1}$ and there exists a common factor with homogeneous loading. Intuitively, to disentangle price elasticities from unobserved common factors, we rely on investors' influence on the price $S_i\,\sigma_i^2$—larger and more volatile investors have larger price impacts. However, when their influence is perfectly aligned with the loadings on the common factors, we can no longer distinguish elasticity from factor loadings.

A.4 Comparison with the Standard GIV

The size-weighting scheme in the optimal GIV may look reminiscent to readers who are familiar with the original GIV method proposed in the seminal paper of Gabaix and Koijen (2024). Indeed, these two estimators are closely linked, but with subtle differences. 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.

On top of the assumptions above, the standard GIV further imposes homogeneity in elasticities across entities.[38] 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, $\sum_i S_i = 1$.

[38] 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).

With homogeneity, the optimal GIV estimator pools the sample moment conditions (3.3) across i weighted by the inverse of residual variance $(\sigma_i^2)^{-1}$ (see Section A.1), so that the optimal GIV estimator ζ^{OGIV} corresponds 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:

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

Denote $\text{Off}(\mathbf{A})$ as the matrix that sets the diagonal elements of $\mathbf{A}$ to zero, we can also express the moment conditions as:

$$\mathbf{E}' \,\text{Off}\left(\mathbb{E}\left[\left(\mathbf{q}_t + \boldsymbol{\iota}\zeta^{OGIV} p_t\right)\left(\mathbf{q}_t + \boldsymbol{\iota}\zeta^{OGIV} p_t\right)'\right]\right)\mathbf{S} = 0.$$

The estimating equation is quadratic in ζ because the moment condition multiplies two terms that each depend linearly on ζ, so we solve for ζ^{OGIV} numerically.

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 $\mathbf{E}'\mathbf{\Sigma}_u\mathbf{X} = \frac{1}{tr(\mathbf{\Sigma}_u^{-1})}$, so by taking the difference of these two equations, we have:

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

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 can 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: [39]

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

while the standard GIV converges to a different limiting distribution: [40]

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

We can show that $\mathbf{V}_{\zeta}^{OGIV} < \mathbf{V}_{\zeta}^{SGIV}$, with equality only in the degenerate case where a single entity holds the entire market share. 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) coincide, 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 (3.4), what matters for the statistical power is not only the size distribution but the distribution of each entity's

[39] See the Appendix A.2 for the derivation.

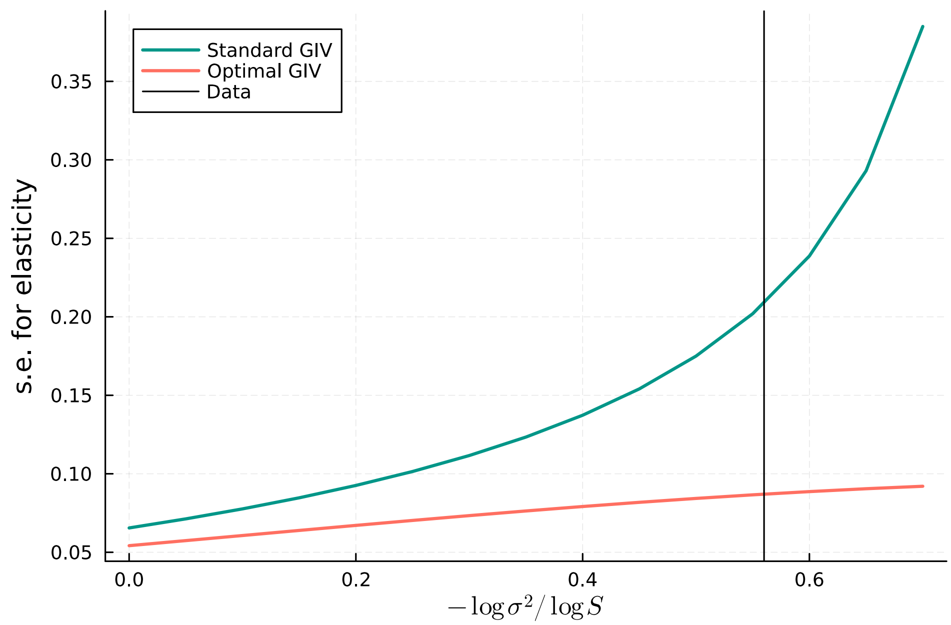
[40] See Gabaix and Koijen (2024) for the derivation.

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, $\sigma_i^2 \propto \frac{1}{S_i}$. In this case, the size vector again coincides with the precision weighting, and there is no variation in the size-minus-precision instruments.

Figure A.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 accompanying Figure A.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.

It should be noted that the higher statistical power of the optimal GIV does not come for free. There are two trade-offs when using the optimal GIV. First, the optimal GIV must be solved numerically, which adds computational complexity. As with any numerical scheme, multiple roots may exist or convergence may fail when the sample size is small. Second, the standard GIV automatically handles common factors with homogeneous loadings by cross-sectional demeaning. This step eliminates potential common factors not controlled for in the data and is, in this sense, more conservative, whereas the optimal GIV by default interprets all residuals as idiosyncratic shocks. Appendix A.3.2 describes how to extend the optimal GIV to account for unobserved common factors internally.

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



Note: Figure A.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 $\sigma_p^2 = 2.5\%$.

A.5 Bias Analysis

Consider the following model:

$$\Delta q_{i,t} = -\zeta_i \Delta p_t + u_{i,t}, \tag{A.23}$$

$$\Delta p_t = \frac{1}{\zeta_S} u_{S,t}, \tag{A.24}$$

Here, as a shorthand, we use the S subscript to denote the size-weighted sum, i.e., $\zeta_S = \sum_i S_i \zeta_i$ and $u_{S,t} = \sum_i S_i u_{i,t}$.

Importantly, the shocks $u_{i,t}$ are not orthogonal: they load on a common factor η_t, and $\epsilon_{i,t}$ is the true idiosyncratic component.

$$u_{i,t} := \lambda_i \eta_t + \epsilon_{i,t}$$

However, the econometrician fails to recognize the common factor η_t. Instead, the econometrician specifies the following model:

$$\Delta q_{i,t} = -\zeta_i^b \Delta p_t + u_{i,t}^b \tag{A.25}$$

and estimates ζ_i^b using the following moment condition:

$$\mathbb{E}\left[u_{i,t}^b u_{j,t}^b\right] \equiv \mathbb{E}\left[\left(\Delta q_{i,t} + \zeta_i^b \Delta p_t\right)\left(\Delta q_{j,t} + \zeta_j^b \Delta p_t\right)\right] = 0, \tag{A.26}$$

We can show that the estimated aggregate elasticity ζ_S^b is downward biased and approximately:

$$\zeta_S^b \approx \zeta_S \sqrt{1 - \mathcal{R}^2}, \quad \mathcal{R}^2 \equiv \frac{\mathrm{Var}(\lambda_S\, \eta_t)}{\mathrm{Var}(u_{S,t})} = \frac{\lambda_S^2\, \sigma_\eta^2}{\lambda_S^2\, \sigma_\eta^2 + \sigma_{\epsilon,S}^2} \tag{A.27}$$

where $\sigma_{\epsilon,S}^2 \equiv \sum_{i=1}^N S_i^2 \sigma_{\epsilon,i}^2 \in (0, \infty)$ is the size-weighted variance of the idiosyncratic shocks, and $\mathcal{R}^2$ is the share of the hidden factor in the total variance of the (residualized) price. Equation (A.27) states that if the econometrician fails to recognize the common factor, there would be a downward bias in the estimated aggregate elasticity, and the bias is increasing in the share of the unobserved comovement.

The bias-corrected OLS interpretation in (3.5) provides intuition for this bias formula. For an individual investor, the OLS estimator for their elasticity is downward-biased due to simultaneity—if their demand increases, this increases prices pushing demand in the opposite direction, dampening its effect. The optimal GIV corrects this bias by adjusting for the influence of their own demand shifts on price. Residual common factors lead to systematic under-correction of the downward bias, and the required correction increases with the influence of the hidden factors on price. The precise statement on the approximation and the proof are given in Proposition A.3 below, where we also further characterize the bias in individual elasticities using the individual loadings on the hidden factors and their individual contribution to the price.

To illustrate the magnitude of the bias, in the figure below we take our estimated aggregate elasticity $\hat{\zeta}_S^b = 1$, and plot the true aggregate elasticity ζ_S assuming we had missed out comovement in the residuals that explains $\mathcal{R}^2$ of the residual variance. It shows that for a medium level of residual comovement, the potential bias is quite small. In order for our estimates to be off by 100%, 75% of the residual variance in the demand shocks has to be explained by common factors. It seems unlikely to us that there exists such a strong hidden common factor that is not observable and cannot be picked up by PCA.

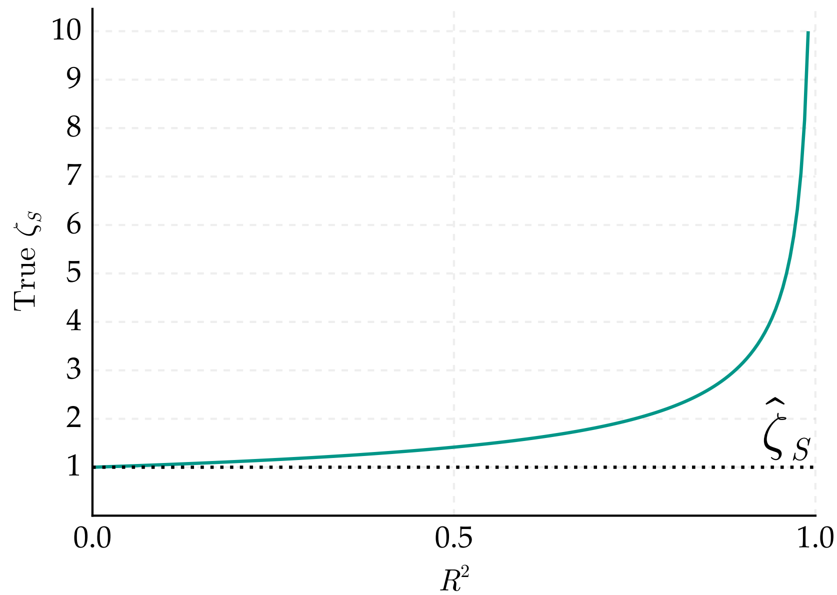


Figure A.2: Relationship between the true aggregate elasticity ζ_S and the share of hidden comovement $\mathcal{R}^2$. The horizontal dotted line represents the estimated aggregate elasticity $\hat{\zeta}_S = 1$. The curve shows the implied true aggregate elasticity for different values of $\mathcal{R}^2$.

Proposition and Proof

Under the homogeneous case (homoskedasticity $\sigma_i = \sigma$, equal sizes $S_i = \frac{1}{N}$, and same loadings $\lambda_i = \lambda_S$), Equation (A.27) is exact. When the size distribution is unequal, we show that heterogeneity only enters the bias formula in the second order. More precisely, we have the following proposition:

Proposition A.3. *Let the true model be given by (A.23) and (A.24). Define following deviations:*

$$\omega_i = \frac{S_i - \bar{S}}{\bar{S}}, \qquad \delta_i = \frac{\lambda_i - \lambda_S}{\lambda_S}, \qquad \kappa_i = \frac{\sigma^2_{\epsilon,i} - \bar{\sigma}^2_\epsilon}{\bar{\sigma}^2_\epsilon},$$

where $\bar{S} = \frac{1}{N}$ *and* $\bar{\sigma}^2_\epsilon = \frac{\sum_i S_i^2 \sigma^2_{\epsilon,i}}{\sum_i S_i^2}$. *Let the dispersion scale* ξ *be*

$$\xi \equiv \max \left\{ \sqrt{\sum_i \omega_i^2}, \sqrt{\sum_i S_i \delta_i^2}, \sqrt{\sum_i S_i^2 \kappa_i^2} \right\}.$$

Write $\sigma^2_{\epsilon,S} = \sum_i S_i^2 \sigma^2_{\epsilon,i}$, $\quad \sigma^2_{u,S} = \sigma^2_{\epsilon,S} + \lambda_S^2 \sigma^2_\eta$, $\quad \mathcal{R}^2 = \frac{\lambda_S^2 \sigma^2_\eta}{\sigma^2_{u,S}} \in (0,1)$. *Let* ζ_S *and* ζ_i *denote the* *true aggregate and sectoral elasticities, and ζ_S^b, ζ_i^b their biased estimates.*

Then, as $\xi \to 0$,

$$\zeta_S^b = \zeta_S\sqrt{1-\mathcal{R}^2} + O(\xi^2)$$

so dispersion affects the aggregate elasticity only at the second order. Moreover, denote $b = 1 - \sqrt{1-\mathcal{R}^2}$. For each sector i,

$$\frac{\zeta_i - \zeta_i^b}{\zeta_S} = b + \mathcal{R}^2\delta_i - b\big(1-b\big)(\omega_i + \kappa_i) + O(\xi^2).$$

Proof. Define $u_{i,t}^b := \Delta q_{i,t} + \zeta_i^b \Delta p_t$ as the misspecified residuals. Using the true model in (A.23), it can be expressed as:

$$u_{i,t}^b = -\left(\zeta_i - \zeta_i^b\right)\Delta p_t + u_{i,t} \tag{A.28}$$

Plugging Equation (A.24) into the misspecified residual, and denoting $b_i := \frac{\zeta_i - \zeta_i^b}{\zeta_S}$, we have:

$$u_{i,t}^b = -b_i u_{S,t} + u_{i,t} \tag{A.29}$$

Aggregating across sectors with size weights S_i, define $u_{S,t}^b \equiv \sum_i S_i u_{i,t}^b$, we obtain:

$$u_{S,t}^b = (1 - b_S)u_{S,t}$$

Here $b_S \equiv \sum_i S_i\, b_i$ denotes the size-weighted average of the sectoral coefficients b_i.

The optimal GIV estimator under the misspecified model solves:

$$\begin{aligned}\mathbb{E}\left[u_{i,t}^b \sum_{j\neq i} S_j u_{j,t}^b\right] &= \mathbb{E}\left[u_{i,t}^b\left(u_{S,t}^b - S_i u_{i,t}^b\right)\right], \\ &= \mathbb{E}\Big[\big(-b_i u_{S,t} + u_{i,t}\big)\big((1 - b_S + S_i b_i)u_{S,t} - S_i u_{i,t}\big)\Big]\end{aligned} \tag{A.30}$$

Express it in terms of the moments of shocks:

$$-b_i(1 - b_S + S_i b_i)\sigma_{u,S}^2 + (1 - b_S + 2S_i b_i)\sigma_{i,S} - S_i\sigma_{u,i}^2 = 0 \tag{A.31}$$

where

$$\sigma_{u,S}^2 := \mathbb{E}[u_{S,t}^2] = \mathbb{E}\Big[(\lambda_S\eta_t + \sum_j S_j\epsilon_{j,t})^2\Big] = \lambda_S^2\sigma_\eta^2 + \sum_j S_j^2\sigma_{\epsilon,j}^2,$$
$$\sigma_{i,S} := \mathbb{E}[u_{i,t}u_{S,t}] = \mathbb{E}\Big[(\lambda_i\eta_t + \epsilon_{i,t})(\lambda_S\eta_t + \sum_j S_j\epsilon_{j,t})\Big] = \lambda_i\lambda_S\sigma_\eta^2 + S_i\sigma_{\epsilon,i}^2,$$
$$\sigma_{u,i}^2 := \mathbb{E}[u_{i,t}^2] = \mathbb{E}\Big[(\lambda_i\eta_t + \epsilon_{i,t})^2\Big] = \lambda_i^2\sigma_\eta^2 + \sigma_{\epsilon,i}^2,$$

using $\mathbb{E}[\eta_t\epsilon_{i,t}] = 0$ and $\mathbb{E}[\epsilon_{i,t}\epsilon_{j,t}] = 0$ for $i \neq j$. Equation (A.31) gives N quadratic equations in b_i.

Symmetric benchmark. To proceed, we first construct a symmetric benchmark. Use upper bar to denote variables for the symmetric benchmark:

$$\bar{S}_i := \frac{1}{N}, \qquad \bar{\lambda}_i := \lambda_S, \qquad \bar{\sigma}_{\epsilon,i}^2 = \frac{\sum_i S_i^2\sigma_{\epsilon,i}^2}{\sum_i \bar{S}_i^2} \quad \forall i.$$

Notice that by construction, this symmetric benchmark has the same variance of the total demand shock $\bar{\sigma}_{u,S}^2$ and aggregate loading $\bar{\lambda}_S$, and hence the same $\mathcal{R}^2$ as the heterogeneous model.

Using the symmetry, we have:

$$\bar{\sigma}_{u,S}^2 = \sigma_{u,S}^2, \qquad \bar{\sigma}_{i,S} = \sigma_{u,S}^2, \qquad \bar{\sigma}_{u,i}^2 = \lambda_S^2\sigma_\eta^2 + \bar{\sigma}_\epsilon^2 = \lambda_S^2\sigma_\eta^2 + N\sigma_{\epsilon,S}^2.$$

In this case, the equation systems in (A.31) are symmetric, and hence $\bar{b}_i = \bar{b}_S = \bar{b}$. Plugging in and simplifying, we have:

$$\bar{b}^2 - 2\bar{b} + \mathcal{R}^2 = 0 \tag{A.32}$$

where $\mathcal{R}^2 \equiv \frac{\lambda_S^2\sigma_\eta^2}{\sigma_{u,S}^2}$. Focusing on the root $\bar{b} < 1$ so $\bar{\zeta}^b$ remains positive, we have $1 - \bar{b} = \sqrt{1-\mathcal{R}^2}$, hence

$$\zeta_S^b = \zeta_S\sqrt{1-\mathcal{R}^2}.$$

which is the exact symmetric-case formula in Equation (A.27).

First-order perturbation. Now we consider the first-order perturbation around the symmetric benchmark. Define the deviations from the symmetric benchmark as:

$$\Delta b_i = b_i - \bar{b}, \qquad \Delta b_S = b_S - \bar{b}, \qquad \omega_i = \frac{S_i - \bar{S}}{\bar{S}}, \qquad \delta_i = \frac{\lambda_i - \bar{\lambda}}{\bar{\lambda}}, \qquad \kappa_i = \frac{\sigma^2_{\epsilon,i} - \bar{\sigma}^2_{\epsilon}}{\bar{\sigma}^2_{\epsilon}},$$

We can express Equation (A.31) as a function of deviations from the symmetric benchmark:

$$F_i(\Delta b_i, \Delta b_S, \omega_i, \delta_i, \kappa_i \mid \sigma^2_{u,S}, \lambda_S) = 0.$$

We fix $\sigma^2_{u,S}$ and λ_S because the perturbation in heterogeneity $(\omega_i, \delta_i, \kappa_i)$ does not enter λ_S and $\sigma^2_{u,S}$ to the first order.

The first order approximation of F_i around the symmetric benchmark is then given by:

$$\partial_b F_i \Delta b_i + \partial_{b_S} F_i \Delta b_S + \partial_\omega F_i \omega_i + \partial_\delta F_i \delta_i + \partial_\kappa F_i \kappa_i + O(\xi^2) = 0.$$

We make the following claim, and prove it at the end of this section.

Claim 1. With $\bar{S} = 1/N$ and $\omega_i, \delta_i, \kappa_i$ as defined,

$$\sum_i S_i \Delta b_i = \Delta b_S, \qquad \sum_i S_i \omega_i = O(\xi^2), \qquad \sum_i S_i \delta_i = 0, \qquad \sum_i S_i \kappa_i = O(\xi^2).$$

With this claim, the last three terms vanish in the first order in the size-weighted summation of the first-order perturbation, which gives:

$$(\partial_b F_i + \partial_{b_S} F_i)\ \Delta b_S + O(\xi^2) = 0 \tag{A.33}$$

As we show below, $\partial_b F_i + \partial_{b_S} F_i \neq 0$. Hence, Equation (A.33) implies that to the first order in heterogeneity, the bias in aggregate elasticity is the same as the symmetric benchmark:

$$\Delta b_S = O(\xi^2) \implies b_S = \bar{b} + O(\xi^2)$$

Individual bias around the symmetric benchmark. To derive the first-order approximation of the bias for individual elasticity, we evaluate the partial derivatives of F_i around the symmetric benchmark.

As an intermediate step, we first linearize the moments:

$$\begin{aligned}
\sigma_{u,S}^2 &= \bar{\sigma}_{u,S}^2 + O(\xi^2) \quad \text{(no first-order change)},\\
\sigma_{i,S} &= \bar{\sigma}_{u,S}^2 + \bar{\lambda}^2\sigma_\eta^2\delta_i + \bar{S}\bar{\sigma}_\epsilon^2(\omega_i + \kappa_i) + O(\xi^2),\\
\sigma_{u,i}^2 &= \bar{\lambda}^2\sigma_\eta^2 + \bar{\sigma}_\epsilon^2 + 2\bar{\lambda}^2\sigma_\eta^2\delta_i + \bar{\sigma}_\epsilon^2\kappa_i + O(\xi^2).
\end{aligned}$$

With these quantities, we can derive the partial derivatives of F_i evaluated at the symmetric benchmark:

$$\begin{aligned}
\frac{\partial F_i}{\partial\,\Delta b_i} &= \sigma_{u,S}^2\,(1-\bar{b})\,(2\bar{S}-1),\\
\frac{\partial F_i}{\partial\,\Delta b_S} &= -\sigma_{u,S}^2\,(1-\bar{b}),\\
\frac{\partial F_i}{\partial\,\omega_i} &= \sigma_{u,S}^2\,(1-\bar{b})^2\,\bar{b}\,(2\bar{S}-1),\\
\frac{\partial F_i}{\partial\,\delta_i} &= -\lambda_S^2\sigma_\eta^2\,(1-\bar{b})\,(2\bar{S}-1) = -\sigma_{u,S}^2\,\mathcal{R}^2\,(1-\bar{b})\,(2\bar{S}-1),\\
\frac{\partial F_i}{\partial\,\kappa_i} &= \sigma_{u,S}^2\,(1-\bar{b})^2\,\bar{b}\,(2\bar{S}-1).
\end{aligned} \tag{A.34}$$

Derivation of $\partial_\omega F_i$ and $\partial_\kappa F_i$. It may come as a surprise that the effect of the size (ω_i) and the effect of the volatility κ_i enters the equation symmetrically. To see that, recall from Equation (A.31) that

$$F_i \;=\; -b_i(1-b_S+S_ib_i)\,\sigma_{u,S}^2 \;+(1-b_S+2S_ib_i)\,\sigma_{i,S} \;-S_i\,\sigma_{u,i}^2.$$

Linearizing the moments at the symmetric benchmark gives

$$\frac{\partial\sigma_{i,S}}{\partial\,\omega_i} = \frac{\partial\sigma_{i,S}}{\partial\,\kappa_i} = \bar{S}\,\bar{\sigma}_\epsilon^2, \qquad \frac{\partial\sigma_{u,i}^2}{\partial\,\omega_i} = 0, \quad \frac{\partial\sigma_{u,i}^2}{\partial\,\kappa_i} = \bar{\sigma}_\epsilon^2,$$

and $\frac{\partial S_i}{\partial\,\omega_i} = \bar{S}$, $\frac{\partial S_i}{\partial\,\kappa_i} = 0$. Evaluating at $b_i = b_S = \bar{b}$, $S_i = \bar{S}$, $\sigma_{i,S} = \sigma_{u,S}^2$, and $\sigma_{u,i}^2 = \lambda_S^2\sigma_\eta^2 + \bar{\sigma}_\epsilon^2$, we obtain

$$\begin{aligned}
\partial_\omega F_i &= \bar{S}\big(-\bar{b}^2\sigma_{u,S}^2 + 2\bar{b}\sigma_{u,S}^2 - \sigma_{u,i}^2\big) \;+(1-\bar{b}+2\bar{S}\bar{b})\,\bar{S}\,\bar{\sigma}_\epsilon^2,\\
\partial_\kappa F_i &= (1-\bar{b}+2\bar{S}\bar{b})\,\bar{S}\,\bar{\sigma}_\epsilon^2 \;-\bar{S}\,\bar{\sigma}_\epsilon^2.
\end{aligned}$$

Using $\bar{b}^2 - 2\bar{b} + \mathcal{R}^2 = 0$ (equivalently $-\bar{b}^2\sigma_{u,S}^2 + 2\bar{b}\sigma_{u,S}^2 - \lambda_S^2\sigma_\eta^2 = 0$) and $\sigma_{u,i}^2 = \lambda_S^2\sigma_\eta^2 + \bar{\sigma}_\epsilon^2$, the bracketed term in $\partial_\omega F_i$ simplifies to $-\bar{\sigma}_\epsilon^2$, so both derivatives reduce to

$$\partial_\omega F_i = \partial_\kappa F_i = \bar{S}\,\bar{\sigma}_\epsilon^2\,\bar{b}\,(2\bar{S}-1).$$

Intuitively, to first order both ω_i and κ_i only scale $S_i\sigma_{\epsilon,i}^2$, making their effects on F_i identical; any differences appear only at $O(\xi^2)$. Further using the identities $\bar{S}\bar{\sigma}_\epsilon^2 = \sigma_{u,S}^2(1-\mathcal{R}^2)$ and $1-\mathcal{R}^2 = (1-\bar{b})^2$, we can further express the partial derivatives as:

$$\partial_\omega F_i = \partial_\kappa F_i = \sigma_{u,S}^2\,(1-\bar{b})^2\,\bar{b}\,(2\bar{S}-1).$$

□

Plugging in the partial derivatives and $\Delta b_S = O(\xi^2)$, we solve for Δb_i as:

$$\Delta b_i = \mathcal{R}^2\delta_i - b\big(1-b\big)(\omega_i+\kappa_i) + O(\xi^2).$$

where we use $b = \bar{b}$ for simplicity. Or equivalently,

$$\frac{\zeta_i - \zeta_i^b}{\zeta_S} = b + \mathcal{R}^2\delta_i - b\big(1-b\big)(\omega_i+\kappa_i) + O(\xi^2).$$

□

Proof of Claim 1. (i) By $b_S = \sum_i S_i b_i$, $\sum_i S_i\Delta b_i = b_S - \bar{b} = \Delta b_S$.

(ii) Since $\omega_i = (S_i - \bar{S})/\bar{S}$, $\sum_i \omega_i = (\sum_i S_i - N\bar{S})/\bar{S} = 0$. Hence

$$\sum_i S_i\omega_i = \sum_i \bar{S}(1+\omega_i)\omega_i = \bar{S}\sum_i \omega_i^2 = O(\xi^2).$$

(iii) With $\delta_i = (\lambda_i - \lambda_S)/\lambda_S$ and $\lambda_S = \sum_i S_i\lambda_i$,

$$\sum_i S_i\delta_i = \frac{1}{\lambda_S}\sum_i S_i\lambda_i - \sum_i S_i = 1 - 1 = 0.$$

(iv) For κ_i, use $\bar{\sigma}_\epsilon^2 = \dfrac{\sum_j S_j^2 \sigma_{\epsilon,j}^2}{\sum_j \bar{S}^2}$ to get the exact identity

$$\sum_i S_i^2 \kappa_i = \frac{\sum_i S_i^2 \sigma_{\epsilon,i}^2}{\bar{\sigma}_\epsilon^2} - \sum_i S_i^2 = \sum_i \bar{S}^2 - \sum_i S_i^2 = -\bar{S}^2 \sum_i \omega_i^2 = O(\xi^2).$$

Also $S_i = \bar{S}(1+\omega_i) = \frac{S_i^2}{\bar{S}} - \bar{S}(\omega_i + \omega_i^2)$ (by algebraic rearrangement of $\omega_i = (S_i - \bar{S})/\bar{S}$), so

$$\sum_i S_i \kappa_i = \frac{1}{\bar{S}} \sum_i S_i^2 \kappa_i - \bar{S} \sum_i (\omega_i + \omega_i^2)\kappa_i =: T_1 + T_2.$$

Here $T_1 = O(\xi^2)$. For T_2, by Cauchy–Schwarz and fixed N,

$$|T_2| \le \bar{S} \Big(\sum_i (\omega_i + \omega_i^2)^2 \Big)^{1/2} \Big(\sum_i \kappa_i^2 \Big)^{1/2} = O(\xi) \cdot O(\xi) = O(\xi^2),$$

since $\sum_i (\omega_i + \omega_i^2)^2 = O(\xi^2)$ and $\sum_i \kappa_i^2 \le C(N) \sum_i S_i^2 \kappa_i^2 \le C(N)\xi^2$. Thus $\sum_i S_i \kappa_i = O(\xi^2)$. $\square$

A.6 Structural Interpretations of the Linear Demand System

In the main text, we estimate a log-linear demand system that can be interpreted as a first-order approximation of a generic portfolio choice problem. To aid interpretation, this section provides several micro-foundations for the log-linear demand curve. In this subsection, we focus on various single-asset portfolio choice problems; in Section A.7, we map the single-asset demand curve to a multi-asset setting.

Throughout this section, let $Q_{i,t}$ denote sector i's Treasury holdings and P_t the price of the Treasury market portfolio. We use bars $\bar{x}$ for long-run levels (or log-levels when a variable is in logs) and hats $\hat{x}_t$ for deviations from long-run levels (log-deviations when in logs). For example, if $q_{i,t} := \log Q_{i,t}$, then $\hat{q}_{i,t} = q_{i,t} - \bar{q}_i = \log(Q_{i,t}/\bar{Q}_i)$; for rates we use level deviations, e.g., $\hat{r}_t := r_t - \bar{r}$.

Table A.1 summarizes the mapping between the structural parameters and the reduced-form elasticities under different micro-foundations.

CRRA portfolio choices The canonical portfolio choice model features constant relative risk aversion (CRRA). Consider investor i choosing between the short-term risk-free asset and a long-term risky bond. The implied portfolio demand is

Table A.1: Mapping between Structural Parameters and Reduced-Form Elasticities

Model	Portfolio choice	Sources of ζ_i	Sources of $\nu_{i,t}$
CRRA portfolio choice (Vayanos & Vila, 2021)	$\frac{Q_{i,t}P_t}{W_{i,t}} = \frac{\mu_t - r_t}{\gamma_{i,t}\,\sigma_t^2}$	$\bar{\mu},\ \bar{r}$	$\widehat{\mathbb{E}_t[p_{t+1}]},\ \hat{w}_{i,t},\ \hat{r}_t,\ \hat{\gamma}_{i,t}$
Learning from price (Hellwig, 1980)	$\widehat{\mathbb{E}_t[p_{t+1}]} = \phi_p\,\hat{p}_t + \phi_s\,s_{i,t}$	$\phi_p,\ \bar{\mu},\ \bar{r}$	$\phi_s s_{i,t},\ \hat{w}_{i,t},\ \hat{r}_t,\ \hat{\gamma}_{i,t}$
Bond-in-utility (Krishnamurthy & Vissing-Jorgensen, 2012)	$\frac{(Q_{i,t}P_t)^{-\psi_i}}{C_{i,t}^{-\theta_i}} = r_t - y_t$	$\psi_i,\ \bar{r},\ \bar{y}$	$\frac{\theta_i}{\psi_i}\,\hat{c}_{i,t},\ -\frac{1}{\psi_i\overline{cy}}\,\hat{r}_t$
Liability-driven investment (Domanski, Shin, & Sushko, 2017)	$Q_{i,t}P_tD_t^T = L_{i,t}(y_t)D_{i,t}^L$	$\frac{\partial \log D_t^T}{\partial \log P_t},\ \frac{\partial \log D_{i,t}^L}{\partial \log P_t}$	$\hat{l}_{i,t}$

Note: Bars denote long-run levels (or log-levels when variables are in logs), and hats denote deviations from long-run levels (log-deviations when in logs). For instance, if $q_{i,t} := \log Q_{i,t}$ then $\hat{q}_{i,t} = q_{i,t} - \bar{q}_i$, while for rates we use $\hat{r}_t := r_t - \bar{r}$. $Q_{i,t}$ is sector i's Treasury holdings and P_t the Treasury market portfolio price. μ_t is the one-period expected return, r_t the short risk-free rate, y_t the Treasury yield, σ_t^2 the return variance, and $\gamma_{i,t}$ risk aversion. $\widehat{\mathbb{E}_t[p_{t+1}]}$ is the (log) deviation of the expected future price; $s_{i,t}$ is a private signal; ϕ_p and ϕ_s are learning-from-price weights. For liability-driven investment, D_t^T and $D_{i,t}^L$ are modified durations for Treasuries and liabilities, and $L_{i,t}$ is the present value of liabilities with $\hat{l}_{i,t}$ its deviation from the long-run level.

$$\frac{Q_{i,t}P_t}{W_{i,t}} = \frac{\mu_t - r_t}{\gamma_{i,t}\sigma_t^2}$$

where $1 + \mu_t \equiv \frac{C+\mathbb{E}_t[P_{t+1}]}{P_t}$ is the one-period expected return on the long-term bond (with C the coupon), r_t is the short-term risk-free return, $\gamma_{i,t}$ is risk aversion, σ_t^2 is the variance of Treasury returns, and $W_{i,t}$ is wealth.

Log-linearizing around long-run levels and using the notation above yields the demand curve:

$$\hat{q}_{i,t} = -\underbrace{\left(1 + \frac{1+\bar{\mu}}{\bar{\mu}-\bar{r}}\right)}_{\zeta_i}\hat{p}_t + \underbrace{\frac{1}{\bar{\mu}-\bar{r}}\,\widehat{\mathbb{E}_t[p_{t+1}]} + \hat{w}_{i,t} - \frac{1}{\bar{\mu}-\bar{r}}\,\hat{r}_t - \hat{\gamma}_{i,t} - \hat{\sigma}_t^2}_{\nu_{i,t}}.$$

Hence, in the standard CRRA case, the elasticity is pinned down by the average expected return and the risk-free rate. The demand shifter $\nu_{i,t}$ reflects fluctuations in wealth, the expected future price, the risk-free rate, risk aversion, and return volatility.

As well-known in the literature, a frictionless CRRA model typically implies unrealistically high elasticities (Gabaix & Koijen, 2022). As a numerical example, let the average expected return be $\bar{\mu} = 5\%$ and the risk-free rate $\bar{r} = 2\%$. Then $\zeta_i \approx 1 + \frac{1.05}{0.05-0.02} \approx 36$, far above our estimates. Several extensions can generate empirically plausible elasticities, as discussed below.

Learning-from-price models Under the CRRA benchmark we assume rational expectations, so expectations of future prices are independent of the current price. However, when investors receive noisy private signals about future prices, the current price aggregates dispersed private information and investors learn from it. A large literature on learning from prices analyzes this behavior; see, for example, Grossman and Stiglitz (1980), Hellwig (1980), and Kyle (1985).

Continuing from (A.6), suppose investor i receives a private noisy signal $s_{i,t}$ and understands that the current price $\hat{p}_t$ contains information about other investors' private signals. In this case, the investor's expectation of the future price is a linear combination of the private signal and the current price:

$$\widehat{\mathbb{E}_t[p_{t+1}]} = \phi_p \hat{p}_t + \phi_s s_{i,t}$$

The coefficients $\phi_p > 0$ and $\phi_s > 0$ depend on the information precision in the price and in the private signal, respectively, and are pinned down in equilibrium.

Plugging this into (A.6), the demand under learning from price is

$$\hat{q}_{i,t} = -\underbrace{\left(1 + \frac{1 + \bar{\mu} - \phi_p}{\bar{\mu} - \bar{r}}\right)}_{\zeta_i} \hat{p}_t + \underbrace{\frac{\phi_s}{\bar{\mu} - \bar{r}} s_{i,t} + \hat{w}_{i,t} - \frac{1}{\bar{\mu} - \bar{r}} \hat{r}_t - \hat{\gamma}_{i,t} - \hat{\sigma}_t^2}_{\nu_{i,t}}.$$

Under learning from price, investors infer future prices from the current price. When the current price rises, they infer that the market has priced in private signals indicating higher future prices, so they are less inclined to trade against price increases. Therefore, learning from price (larger ϕ_p) reduces the elasticity. If investors rely heavily on the current price, the demand may not be downward-sloping ($\zeta_i \leq 0$), consistent with passive or momentum trading.

Convenience yield models Beyond risk–return trade-offs, investors may hold Treasuries for their convenience yield. A large literature documents that, for the same payoff, investors are willing to pay a premium to hold Treasuries because they provide liquidity services and serve as safe assets (Fu, Li, & Xie, 2022; Krishnamurthy & Vissing-Jorgensen, 2012; Nagel, 2016).

Convenience services are often modeled via a bonds-in-utility specification. Consider an investor who derives utility from both consumption and Treasury holdings, which are typically treated as risk-free in this framework. The implied condition (see (Krishnamurthy & Vissing-Jorgensen, 2012) for the full setup and derivation) is

$$\frac{(Q_{i,t}P_t)^{-\psi_i}}{C_{i,t}^{-\theta_i}} = r_t - y_t \equiv cy_t$$

where $y_t \equiv \frac{F}{P_t} - 1$ is the yield of the Treasury bond, $C_{i,t}$ is the consumption, and r_t is the return from a risk-free bond without the convenience service.

Log-linearizing around long-run levels gives

$$\hat{q}_{i,t} = \underbrace{-\left(1 + \frac{1+\bar{y}}{\psi_i \overline{cy}}\right)}_{\zeta_i^{cy}} \hat{p}_t + \underbrace{\frac{\theta_i}{\psi_i}\,\hat{c}_{i,t} - \frac{1}{\psi_i \overline{cy}}\,\hat{r}_t}_{\nu_{i,t}^{cy}}\,.$$

Hence the elasticity is governed by the curvature parameter ψ_i associated with convenience services.

Liability-driven investment models Liability-driven investment is commonly adopted by insurance companies and pension funds (ICPFs). ICPFs typically have long-duration liabilities and hold long-term bonds to hedge the interest-rate risk of their liability exposures. Suppose ICPFs match the dollar duration of their assets and liabilities; equivalently, they match present value (PV) times modified duration on both sides (Domanski, Shin, & Sushko, 2017):

$$Q_{i,t}P_tD_t^T(P_t) = L_{i,t}(y_t)D_{i,t}^L(P_t)$$

where D_t^T and $D_{i,t}^L$ are modified durations (in years) of Treasuries and liabilities, respectively, and $L_{i,t}$ is the present value (PV) of liabilities. This equates dollar-duration exposures: $Q_{i,t}P_tD_t^T = L_{i,t}D_{i,t}^L$.

Log-linearize around a steady state. Taking logs of $Q_{i,t}P_tD_t^T = L_{i,t}D_{i,t}^L$ (and writing $d_{i,t}^X \equiv \ln D_{i,t}^X$) gives

$$q_{i,t} + p_t + d_t^T = l_{i,t} + d_{i,t}^L \quad\Rightarrow\quad \hat{q}_{i,t} = -\hat{p}_t - \hat{d}_t^T + \hat{l}_{i,t} + \hat{d}_{i,t}^L.$$

Hence the LDI demand is

$$\hat{q}_{i,t} = -\underbrace{\left(1 + \kappa^T - \kappa_i^L\right)}_{:=\zeta_i^{LDI}} \hat{p}_t + \underbrace{\hat{l}_{i,t}}_{\nu_{i,t}^{LDI}}.$$

where we define the elasticities of duration with respect to price as:

$$\kappa^T := \frac{\partial d_t^T}{\partial p_t} = \frac{\partial \log D_t^T}{\partial \log P_t}, \qquad \kappa_i^L := \frac{\partial d_{i,t}^L}{\partial p_t} = \frac{\partial \log D_{i,t}^L}{\partial \log P_t}.$$

Typically, liabilities for ICPFs have much longer duration than Treasury assets and similar dispersion in cash flows, so $\kappa_i^L > \kappa^T$ and hence $\zeta_i^{LDI} < 1$. [41]

A.7 Single-Asset Representation of Multi-Asset Demand

A multi-asset system can be collapsed to a single-asset system with a total elasticity, which embeds the general-equilibrium feedback from cross-asset substitution. This appendix formalizes this mapping and discusses why the total elasticities, rather than the partial elasticities, are typically more relevant for policy analysis.[42]

Two-Asset System Let asset 1 be Treasuries and asset 2 be a close substitute; the logic extends naturally to an N-asset system. For exposition, assume only sector-specific demand shifters (no common shifters). Sector i's demands are:

$$q_{i,1,t} = -\zeta_{i,11} p_{1,t} + \zeta_{i,12} p_{2,t} + u_{i,1,t},$$
$$q_{i,2,t} = -\zeta_{i,22} p_{2,t} + \zeta_{i,21} p_{1,t} + u_{i,2,t},$$

[41] To see this, consider price movement due to a parallel shift in the yield curve. Using $\frac{\partial \ln P_t}{\partial y_t} = -D_t^T$ and $\frac{\partial \ln D_t^X}{\partial y_t} = D_t^X - \frac{C_t^X}{D_t^X}$ with respect to the underlying yield y_t (where $C_t^X := \frac{1}{P_t^X} \frac{\partial^2 P_t^X}{\partial y_t^2}$ is convexity and $X \in \{T, L\}$), the chain rule implies

$$\kappa^T = -1 + \frac{C_t^T}{(D_t^T)^2}, \qquad \kappa^L = -\frac{D_t^L}{D_t^T} + \frac{C_t^L}{D_t^L D_t^T} = \frac{D_t^L}{D_t^T}\Big(-1 + \frac{C_t^L}{(D_t^L)^2}\Big).$$

For any cash-flow stream, $C_t^X = E_w[t^2]$ and $D_t^X = E_w[t]$ for the cash-flow time distribution w, so $-1 + C_t^X/(D_t^X)^2 = \mathrm{Var}_w[t]/(E_w[t])^2 \geq 0$. Hence κ is nonnegative and increasing in the dispersion/convexity of cash flows, and tends to be larger for longer-duration portfolios with similar dispersions. In particular, if liabilities typically have longer duration than Treasury assets ($D_t^L > D_t^T$) and comparable dispersion, then $\kappa^L > \kappa^T$.

[42] Our methodology yields price multipliers estimated using time-series variation. In related work, Haddad, He, Huebner, Kondor, and Loualiche (2026) discuss the interpretation of price elasticities of demand using cross-sectional instruments.

where $\zeta_{i,11}$ and $\zeta_{i,22}$ are the own-price elasticities, $\zeta_{i,12}$ and $\zeta_{i,21}$ are the cross-price elasticities, and $u_{i,1,t}$ and $u_{i,2,t}$ are the idiosyncratic demand shocks.

Treasury market impact on substitute prices: Market clearing for the substitute asset, $\sum_i S_{i,2}q_{i,2,t} = 0$, implies:

$$p_{2,t} = \underbrace{\mathcal{P}_{1\to 2}}_{:=\zeta_{S,21}/\zeta_{S,22}} p_{1,t} + \frac{1}{\zeta_{S,22}} u_{S,2,t}$$

where $\mathcal{P}_{1\to 2} \equiv \frac{\zeta_{S,21}}{\zeta_{S,22}}$ is the price pass-through coefficient from Treasuries to the substitute asset. It captures the equilibrium percentage change in the substitute asset price induced by a 1% change in Treasury prices—for a detailed discussion on pass-through coefficients see Chaudhary et al. (2023).

Single-asset representation Substituting p_2 into the demand for Treasuries we get,

$$q_{i,1,t} = -\underbrace{(\zeta_{i,11} - \zeta_{i,12}\mathcal{P}_{1\to 2})}_{:=\tilde{\zeta}_{i,1}} p_{1,t} + \frac{\zeta_{i,12}}{\zeta_{S,22}} u_{S,2,t} + u_{i,1,t}$$

where $\tilde{\zeta}_{i,1} \equiv \zeta_{i,11} - \zeta_{i,12}\,\mathcal{P}_{1\to 2}$ is the total elasticity, which comprises two parts. The partial elasticity $\zeta_{i,11}$ captures the direct response to Treasury price changes, holding p_2 fixed. The second term, $\zeta_{i,12}\,\mathcal{P}_{1\to 2}$, captures the indirect effect: Treasury price adjustments cause p_2 changes (by $\mathcal{P}_{1\to 2}$), which in turn induce additional Treasury demand changes due to cross-asset substitution. Intuitively, the total elasticity $\tilde{\zeta}_{i,1}$ measures the equilibrium effect of a 1% exogenous change in Treasury prices, after all prices have adjusted. Note that mapping the multi-asset system to a single-asset equation introduces an aggregate demand shifter $\frac{\zeta_{i,12}}{\zeta_{S,22}}u_{S,2,t}$ capturing allocation shifts induced by aggregate demand shocks to the substitute asset.

Macro multiplier and amplification Applying the market clearing for asset 1, we can express the Treasury macro multiplier M_1 in terms of the aggregate total elasticity and its underlying components,

$$M_1 = \frac{1}{\tilde{\zeta}_{S,1}} = \frac{1}{\zeta_{S,11}(1 - \underbrace{\mathcal{P}_{1\to 2}\mathcal{P}_{2\to 1}}_{:=\mathcal{A}_1})}$$

where $\mathcal{P}_{2\to1} \equiv \frac{\zeta_{S,12}}{\zeta_{S,11}}$ is the price pass-through coefficient from the substitute asset to Treasuries, and $\mathcal{A}_1 \equiv \mathcal{P}_{1\to2}\mathcal{P}_{2\to1}$ is the amplification factor capturing how Treasury price changes impact substitute prices, which then feed back to impact Treasury prices. This feedback effect is geometric: a change in Treasury prices impacts substitute prices, which then feed back to impact Treasury prices, and so on. Consequently, the multiplier M_1 (which captures the total equilibrium effect on Treasury prices from a 1% change in Treasury demand) depends on: (i) the partial effect of the demand shock holding other prices fixed, $1/\zeta_{S,11}$, which is then scaled by (ii) the total amplification due to equilibrium price adjustments, given by the geometric sum $\sum_{k=0}^{\infty} \mathcal{A}_1^k = \frac{1}{1-\mathcal{A}_1}$.

Policy-relevant parameter Which parameters are relevant for researchers depends on the research objective. For many policy questions, the total elasticity $\tilde{\zeta}_{i,1}$ is usually the relevant parameter. When the Federal Reserve conducts quantitative easing or foreign investors rebalance portfolios, Treasury prices change and other asset prices adjust simultaneously in response to the shock. The total impact of the policy after all equilibrium adjustments (rather than just the partial adjustment holding other prices fixed) is typically of interest. In other words, the total elasticity (or multiplier) is the object of direct policy relevance. Additionally, in practice, the high dimensionality of the full partial elasticity matrix makes direct estimation empirically challenging or infeasible. In contrast, estimating the total elasticity directly is both more robust to misspecification and empirically feasible.

A.8 Additional Discussions on Optimal GIV Estimator

A.8.1 When only a subset of sectors are observed

Consider the case where only a subset of sectors are observed. In this case,

$$\left.\begin{array}{rl} q_{i,t} & = -p_t \times \mathbf{C}_{i,t}'\boldsymbol{\zeta} + \mathbf{X}_{i,t}'\boldsymbol{\beta} + u_{i,t}, \\ q_{i,t}^{unobs} & = -\zeta_t^{unobs} p_t + \epsilon_t \\ 0 & = \sum_i S_{i,t} q_{i,t} + q_t^{unobs} \end{array}\right\} \implies p_t = M_t \left[\mathbf{X}_{S,t}'\boldsymbol{\beta} + u_{S,t} + \epsilon_t\right], \tag{A.35}$$

where $M_t = \left(\mathbf{C}_{S,t}'\boldsymbol{\zeta} + \zeta_t^{unobs}\right)^{-1}$ is the inverse of the sum of the observed and unobserved sectors' demand elasticities. In this case, whether we can correctly identify the elasticity depends on the covariance structure of the error terms.

Consistency Notice that to identify ζ_i for observed sectors, the original moment conditions $\mathbb{E}\left[u_{i,t}u_{j,t}\right] = 0$ suffice, even if $u_{i,t}$ can be correlated with the unobserved sector's error term $\epsilon_{i,t}$. Using the language of the IV interpretation, the IV is valid regardless of how the instruments affect the endogenous variable p_t, as long as the instruments affect the outcome variable $q_{i,t}$ only through p_t.

The aggregate elasticity, or the multiplier M_t, cannot be identified simply by $\mathbf{C}'_{S,t}\boldsymbol{\zeta}$, due to the elasticity provision of the unobserved sector. If we have $\mathbb{E}\left[u_{i,t}\epsilon_t \mid \mathbf{X}_{i,t}\right] = 0$, i.e., the demand shifts from the unobserved sectors are orthogonal to idiosyncratic shocks conditional on common factors, we can recover M_t by regressing p_t on $u_{S,t}$. Alternatively, if we can assume that the observed sectors are representative of the unobserved sectors, then we can impute the ζ_t^{unobs} using the observed sectors' $\mathbf{C}'_{i,t}\boldsymbol{\zeta}$.

A.8.2 The "Missing Intercept" Problem

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:

$$\lambda_i^q \equiv \frac{\mathbb{E}\left[q_{i,t}\eta_t\right]}{\mathbb{E}\left[\eta_t^2\right]} = \frac{\mathbb{E}\left[\left(-\zeta_i p_t + \lambda_i \eta_t\right)\eta_t\right]}{\mathbb{E}\left[\eta_t^2\right]} = \lambda_i - \frac{\zeta_i}{\zeta_S}\lambda_S. \tag{A.36}$$

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 elasticities are homogeneous ($\zeta_i = \zeta_S$), we can infer their demand is lower than the market average ($\lambda_i < \lambda_S$). Nevertheless, without knowing the price elasticities and the aggregate loading, 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.

A.8.3 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 (2.3).

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 (3.2) 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 $\mathrm{Var}\left(\varepsilon_{i,t}\right) \propto \mathrm{Var}(u_{i,t})$.

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:

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

If one is not interested in the aggregate market elasticity, one can simply exclude the residual sector from 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.

B Empirical Appendix

B.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.[43] 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 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 Judson (2022), Bertaut and Judson (2014), and Bertaut and Tryon (2007). 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.[44] To get

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

[44] 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 directly use

the breakdown between bills and other securities, we use an imputation method similar to Jansen et al. (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 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.[45]

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 B.6). We sweep the discrepancy component into the household sector (the residual sector in the Z.1 methodology) in our baseline estimation. In robustness checks, we explore alternative treatments of the discrepancy, by first reassigning a fixed proportion (75%) of the holdings of the "other sector" (dominated by potential mismeasurement of Treasury holdings from state and local governments) to discrepancy, and then assigning the adjusted discrepancy to the household sector.[46]

RCFD3531 to impute missing RCON3531 numbers.

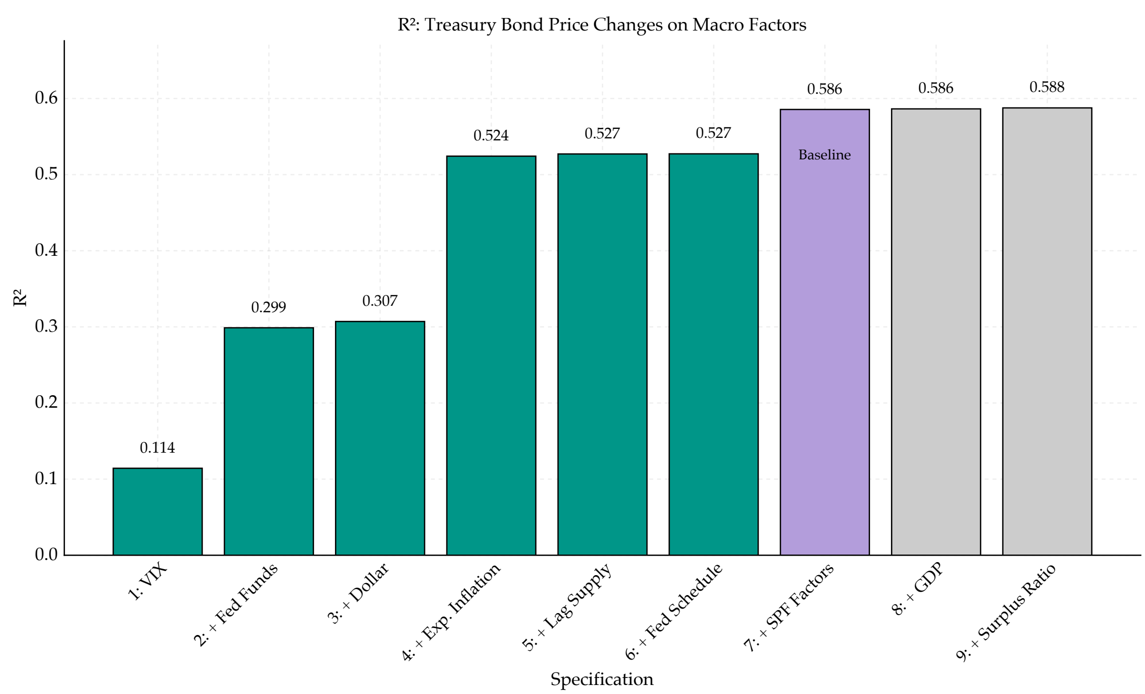
[45]More specifically, we require countries to hold non-zero Treasury bonds and notes throughout from 2003Q1 to be included individually, and banks to have a non-zero position starting from 2001Q4 to be included. As countries and banks with positive holdings tend to be large, our country- and bank-level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

[46]In the financial accounts, four sectors do not report breakdown of Treasury holdings into bills versus notes and bonds: households, pension funds, ETFs and the "other" sector, mostly comprised of state and local governments. A number of states consistently report large holdings of Treasury bills and low durations of their entire Treasury portfolios (see California 2024, Table 4 and Oregon 2024, Page 77 for examples). For this reason and the fact that the "other" sector has considerable Treasury holdings, mismeasurement is likely most relevant in the "other" sector, motivating this alternative discrepancy adjustment strategy.

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), 10-year inflation expectations (Cleveland Fed), 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). In addition, we extract common factors from Survey of Professional Forecasters (SPF) forecasts of 10-year Treasury yield.

Sector classification and aggregation Table B.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 report and TIC data to break down sector aggregates into individual bank/country holdings.

Figure B.1: **Explanatory power of macro-financial factors on Treasury price changes**



Note: Figure B.1 reports the R^2 measures from fitting time-series regressions of Treasury price changes on the observed macro-financial factors (VIX, effective federal funds rate, 10-year inflation expectations, nominal broad dollar index, real GDP growth and U.S. government surplus to debt ratio) and the first four principal components extracted from SPF forecasts of 10-year Treasury yields. We add factors progressively. The purple bar corresponds to the set of factors used in the our baseline estimation.

B.2 Predictability of Policy Sectors

B.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. From 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 a 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.[47]

To control for the scheduled Fed purchases, we construct a sequence of anticipated purchases by the Fed in each quarter. Outside of QE regimes, we treat all purchases as anticipated; during QE regimes, we construct the anticipated purchases 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.

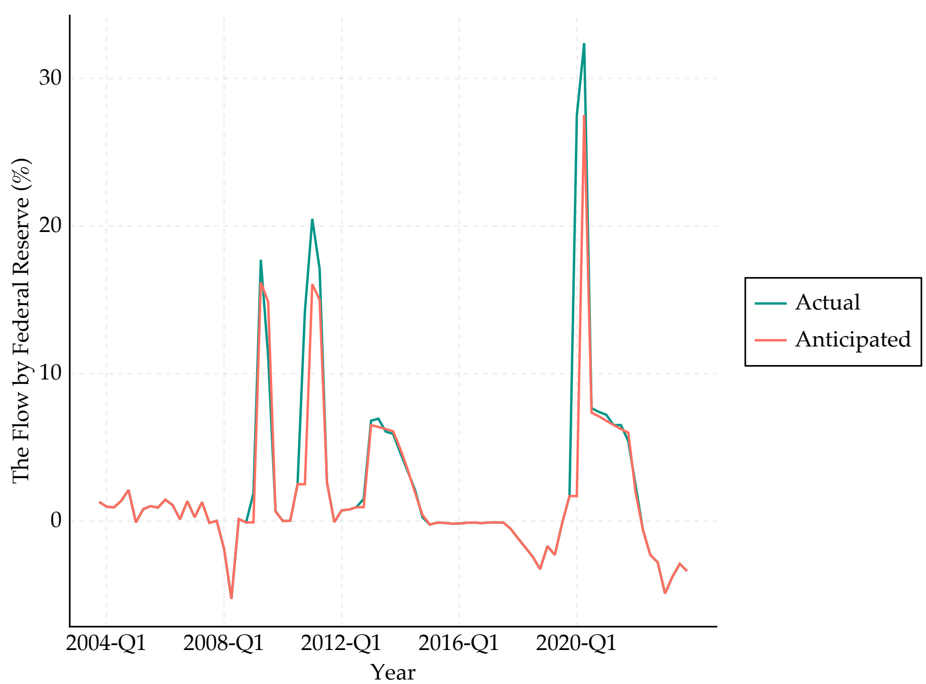
Table B.2 reports the anticipated purchases versus the actual purchases of the Fed during active QE periods. When the QE is not pre-scheduled in that quarter (typically at the start of the QEs), we use lagged purchases by the Fed from the previous quarter as the expected purchase.[48] As investors may use more information than the FOMC announce-

[47]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.

[48]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 deviations of the QE purchase from the

ments alone, we take a conservative approach 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 pattern and is 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 is small, alternative treatments have little effect on our results. Figure B.2 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.

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

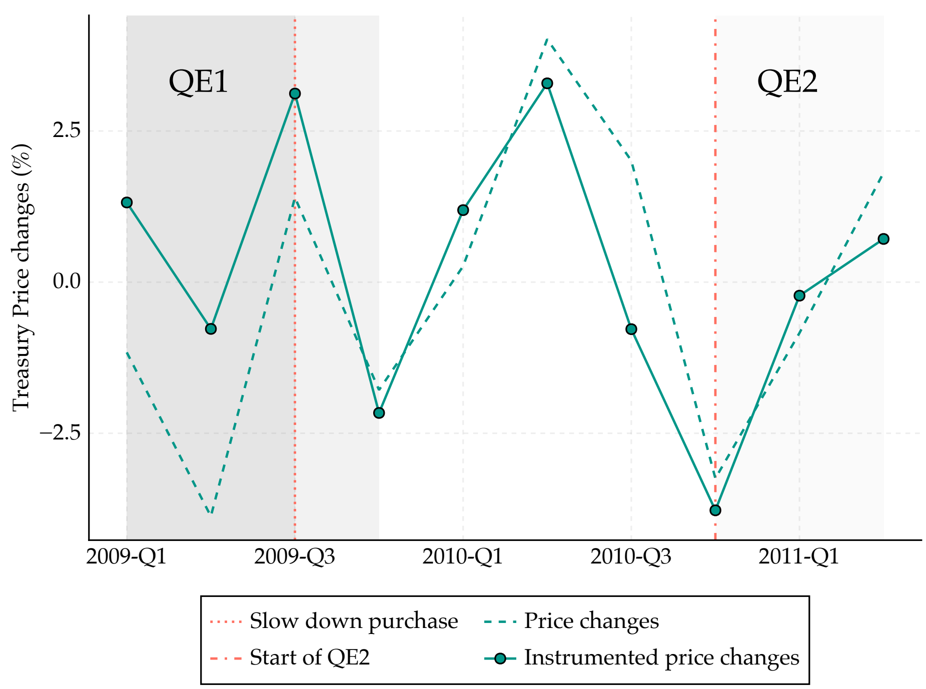


Note: Figure B.2 plots the actual purchases (normalized by total Fed holdings) of U.S. Treasury notes and bonds against the constructed one-quarter-ahead anticipated Fed purchase measure in green. The anticipated path is generated by combining the actual Fed purchase path before 2009 with the pre-announced purchase schedule by the New York Fed starting from the beginning of the QE operation.

regular outright operations are an order of magnitude larger.

As a further illustration, in Figure B.3 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 B.3: **Case Study: QE Entry and Exit between 2009-2010**



Note: Figure B.3 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 of non-Fed sectors extracted from the GMM estimation.

B.3 Case Study: Household and Mutual Fund Sectors

Figure B.4 provides an illustration of our sector-specific estimates against the reduced-form OLS estimates. 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 exactly to the negative reciprocal of the elasticity coefficients reported in Table 1.[49] Using 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.

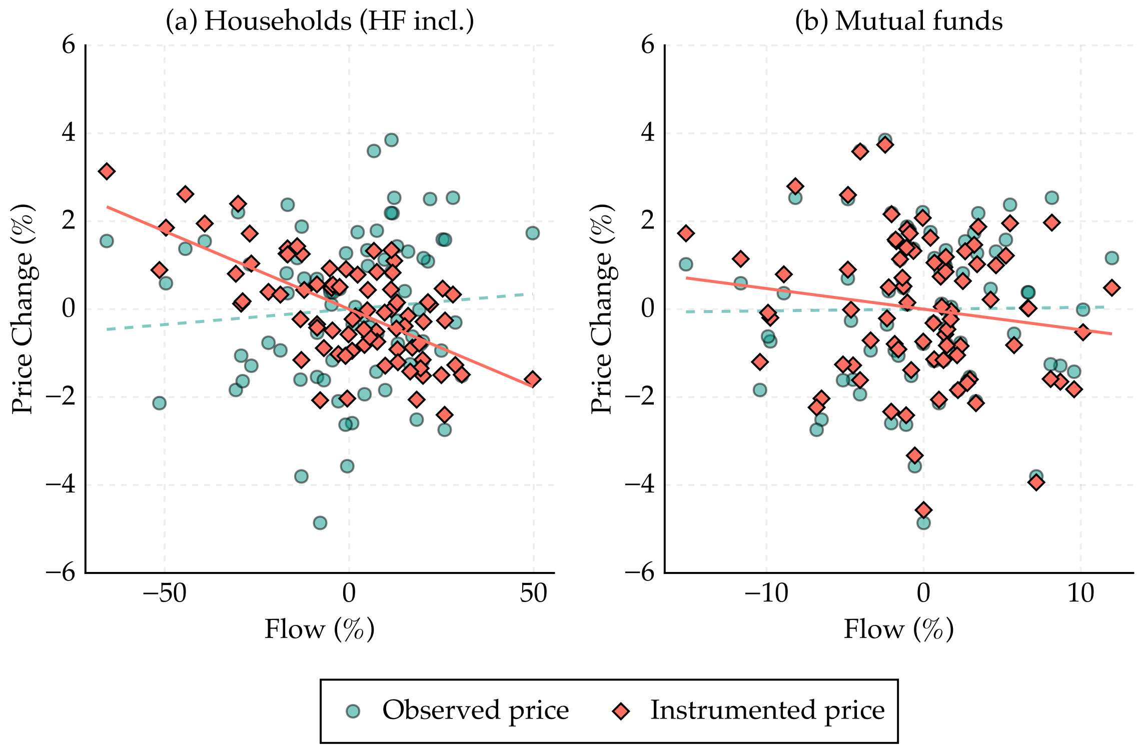
The bias-corrected OLS interpretation of our optimal GIV estimator (see Lemma 1) helps interpret these differences. Though the mutual fund sector and the household sector are comparable in terms of sizes, the household sector has much more volatile idiosyncratic shocks than the mutual fund sector, a 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.

B.4 Robustness of Elasticity Estimation

Table B.7 reports the full set of estimated parameters from the baseline specification. Table B.5 demonstrates the robustness of our price elasticity estimation to alternative samples, controls, or dependent variables. Table B.3 shows that our finding remains robust if we remove any sector from the construction of moment conditions and reestimate the model.

[49] 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 in Appendix Table B.3.

Figure B.4: **Optimal GIV vs. OLS: Two examples**



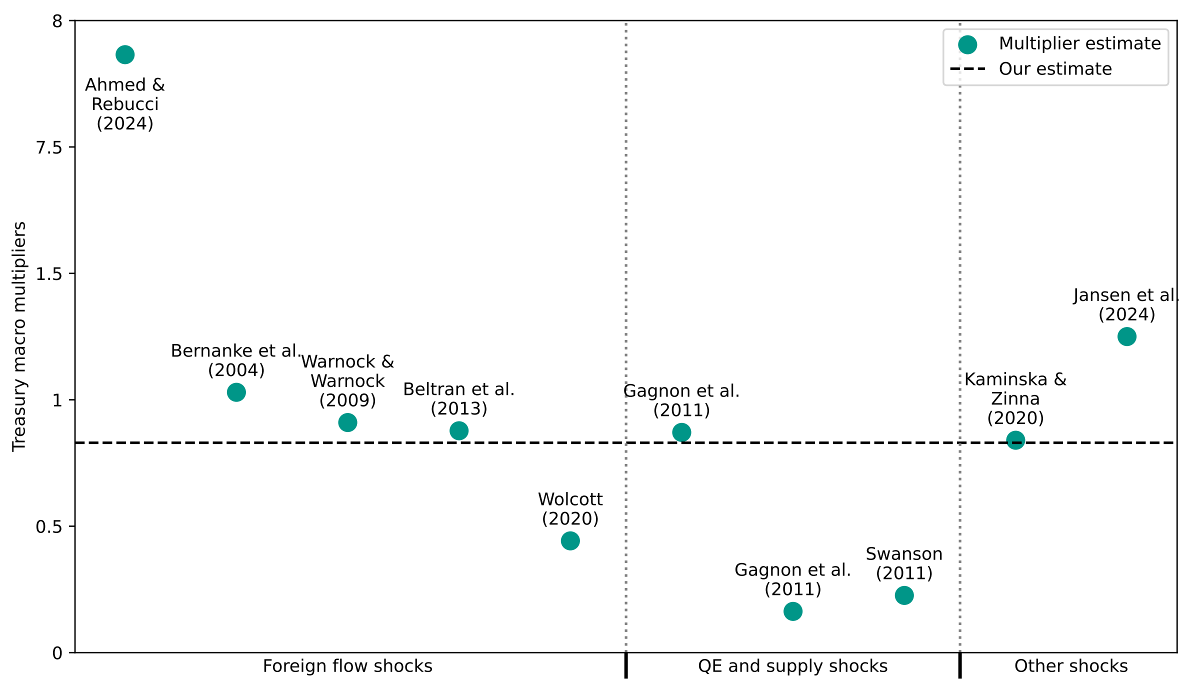
Note: Figure B.4 compares the estimated price elasticities of demand using our identification approach to the OLS estimates, using the household sector and the mutual fund sector as case studies. In both panels, the green dots relate the sector inflow to U.S. Treasury notes and bonds to observed quarterly 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.

B.5 Comparing Multipliers Estimates Across the Literature

In this section, we compare our Treasury macro multiplier estimates with those from the literature. The focus on estimating multiplier parameters—the percentage change in prices to a percentage change in demand as a share of outstanding Treasury securities—is a recent development. However, the literature has long studied the price impact of demand shocks arising from foreign FX interventions, quantitative easing, Treasury repurchase operations, and other policy interventions. By making a few assumptions about the implicit size of the quantity changes, we can map these price impact estimates from the literature to multipliers. While these mappings are not perfect, they provide a useful benchmark for our estimates.

Figure B.5 compares our Treasury macro multiplier estimates with those from the literature. The y-axis shows the macro multiplier—defined as the percentage change in prices relative to the percentage change in demand as a share of outstanding Treasury securities. The x-axis groups estimates by the different types of demand shocks and methodologies they used to measure price impact. These include foreign flow shocks, quantitative easing (QE) and supply shocks, and structural methods that estimate Vayanos and Vila (2021)-style models to derive macro multipliers. The details of this mapping are discussed below. With the exception of a couple of outliers, our estimate of around 0.83 is in line with the macro multiplier estimates implied by the literature.

Figure B.5: **Comparing Multipliers Estimates Across the Literature**



Note: Figure B.5 compares our Treasury macro multiplier estimates with those from the literature. When the paper directly estimates macro multipliers, we present them. When paper only estimate the price impact, we map them to multipliers by making a few assumptions about the implied size of the quantity changes.

Below we outline how we map each paper's estimates to multipliers.

Bernanke et al. (2004) examine how 10-year Treasury yields respond to Japanese Ministry of Finance interventions in the foreign exchange market between 2000 and 2004. Specifically, they study three-day windows around these interventions and find that a $1 billion purchase of dollars (subsequently invested in Treasuries) reduces the 10-year yield by 0.66 basis points. To express this result as a macro multiplier, we proceed in two steps.

First, we scale their semi-elasticity to a 1% demand shock as a share of outstanding. The average amount of Treasury notes and bonds outstanding during the sample period was $2.4 trillion. Thus, a 1% demand shock corresponds to a $24 billion purchase. Since a $1 billion purchase lowers yields by 0.66 basis points, a 1% demand shock lowers yields by $0.66bp/100 \times 24 = 0.158$ percentage points. Second, we now map this semi-elasticity to a price multiplier. If we interpret the yield change as a level shift of the curve, we can approximate the price level impact on the Treasury market using the fact $\Delta P_t/P_t \approx -\text{Duration} \times \Delta y_t$. The average duration of the Treasury market is around 6.5 years. Thus, the implied macro multiplier is 1.03.

Warnock and Warnock (2009) They regress monthly 10-year Treasury yields on official inflows (as a share of lagged GDP) over the 1984-2005 period. Their estimates imply that a 1% flow shock (relative to GDP) lowers yields by 25 basis points. During this sample, outstanding Treasuries averaged 56% of GDP, so a 1% demand shock relative to outstanding supply corresponds to a decline in 10-year yields of $(25bp/100) \times 0.56 = 0.14$ percentage points. Assuming the yield change reflects a parallel shift of the yield curve, and using an average Treasury duration of 6.5 years, this implies a price impact multiplier of approximately 0.91.

Beltran et al. (2013) use instruments (Japanese FX interventions, oil shocks, Chinese balance-of-payments surprises) to estimate the effect of official foreign purchases on the five-year term premium (19942007). A 1% demand shock (of outstanding) lowers the five-year term premium by 0.135 percentage points. Assuming this corresponds to a parallel shift in the yield curve and applying the duration approximation, this estimate maps to a macro multiplier of 0.88.

Wolcott (2020) estimate a VAR with macro variables, foreign flows, and yields, imposing no-arbitrage. Using Cholesky identification, they find a one-standard-deviation foreign flow shock (0.3% of outstanding) lowers the five-year yield by 2 basis points. Scaling to a 1% shock gives a 0.067 percentage point yield effect. Using the duration approximation this implies a macro multiplier of 0.44.

Ahmed and Rebucci (2024) identify exogenous foreign official flow shocks via heteroskedasticity in a structural VAR. An unexpected $100 billion purchase (1.03% of the $9.7 trillion stock of Treasury notes and bonds held by the public at the end of the sample period[50]) lowers the 10-year yield by 125 basis points. Scaling to a 1% shock gives a 1.21

[50]Flow of Funds table L.210, 2017Q3: total marketable notes-and-bonds liability less Federal Reserve

percentage point yield move. Applying the duration approximation, this implies a macro multiplier of 7.87.

Gagnon et al. (2011b) provide two estimates. First, an event study of QE1 announcements (2008–2009) finds that eight windows lowered the 10-year yield by 91 basis points. If the full $300 billion (6.8% of the $4.41 trillion stock) was a surprise revealed in these eight windows, this is a 0.134 percentage point yield effect per 1% shock. Using the duration approximation, the multiplier is 0.87. Second, a monthly regression of yields on supply (assumed exogenous) finds a 1% increase in supply-to-GDP raises the term premium by 4.4 basis points. With average debt/GDP of 57%, this is 0.025 percentage point per 1% shock (of outstanding). With the duration approximation, the multiplier is 0.16. The challenge with these supply shocks is that Treasury supply decisions are typically pre-announced so the price impact tends to happen before the actual supply change, resulting in understating the size of the multiplier.

Swanson (2011) uses the Kennedy administration's Operation Twist announcements (19611962) as exogenous shocks. The first four announcements lowered the 10-year yield by 16 basis points, with purchases equal to 4.6% of total marketable Treasury debt outstanding. Thus, a 1% shock moves yields by 0.035 percentage points. With the duration approximation this implies a macro multiplier of 0.23.

Kaminska and Zinna (2020) estimate the Vayanos–Villa preferred-habitat model via Bayesian MCMC and simulate the three LSAP waves. The implied multipliers are $M_1 = 0.76$, $M_2 = 1.32$, and $M_3 = 0.45$, with an average of 0.84.

Jansen et al. (2024) estimate investor demand (pension funds, mutual funds, insurance companies, etc.) for short-, medium-, and long-term Treasuries, and embed these in a Vayanos and Vila (2021)-style model, treating broker-dealers and hedge funds as arbitrageurs. They estimate a macro multiplier of 1.25 for a permanent 1% demand shock (focusing on medium- and long-term Treasuries for comparability). We focus on their permanent demand shock, as our analysis also considers permanent changes in holdings.

A few notable papers examine the effects of exogenous quantity changes in the Treasury market, though their results are not directly comparable to ours.

For example, Krishnamurthy and Vissing-Jorgensen (2012) study how Treasury supply affects the convenience yield. They find that, between 1969 and 2008, a 1% increase in

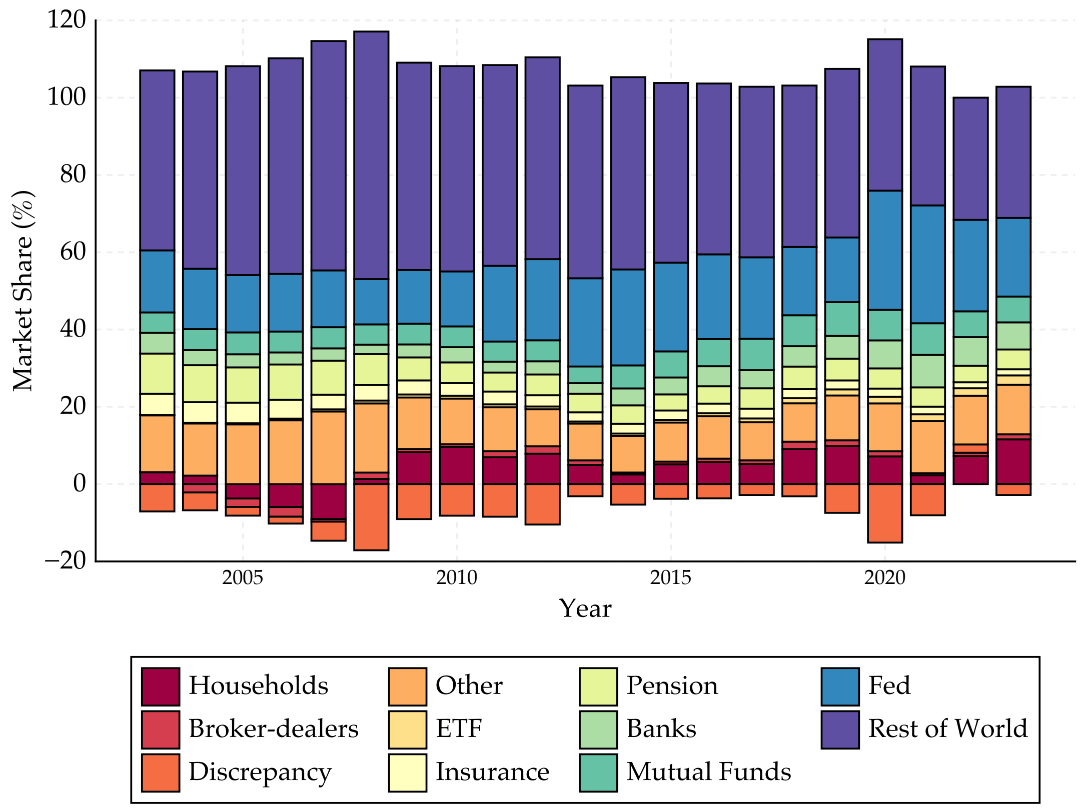
SOMA holdings.

the debt-to-GDP ratio reduces the convenience yield by 0.87 basis points (measured using 10-year yields). Adjusting for the average debt-to-GDP ratio during this period and using the duration approximation, this implies a multiplier of 0.02-0.03. This smaller multiplier reflects two key differences: their dependent variable is the convenience yield (not the level of yields), and their analysis is conducted at an annual frequency (rather than quarterly). When we estimate the impact of our demand shocks on the convenience yield at an annual frequency, we obtain a similar estimate.

Similarly, Greenwood, Hanson, and Stein (2015) show that short-term debt yields rise when Treasury bill supply increases, consistent with a downward-sloping demand curve for T-bills. However, their object of interest differs from ours: they focus on the effect of Treasury bill supply on short-term debt yields, while we study the effect of demand on Treasury Notes and Bonds yields (long-term debt).

B.6 Additional Tables and Figures

Figure B.6: **Investor composition of U.S. Treasury notes and bonds**



Note: Figure B.6 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 B.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 B.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

Note: Italicized sectors indicate that we use data more granular than the sector-level information from the Financial Accounts.

Table B.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 B.3: **Price elasticity by sector: Leave-one-out Estimation**

Sector	Baseline	Other	Households	Pension	Insurance	Funds	ETF	Dealers	Fed	Banks	RoW	Supply
Aggregate	1.2	1.16	0.66	1.34	1.38	1.45	1.36	1.28	1.16	1.38	1.4	1.27
	(0.9, 1.5)	(0.73, 1.59)	(0.34, 0.99)	(0.86, 1.82)	(0.86, 1.91)	(0.83, 2.06)	(0.87, 1.85)	(0.55, 2.02)	(0.73, 1.58)	(0.73, 2.03)	(0.4, 2.4)	(0.63, 1.91)
Other	0.05		0.04	0.7	0.69	0.66	0.71	0.89	0.79	0.72	0.62	0.83
	(-0.22, 0.32)		(-0.44, 0.52)	(0.38, 1.02)	(0.37, 1.0)	(0.35, 0.98)	(0.39, 1.02)	(0.53, 1.25)	(0.45, 1.13)	(0.38, 1.05)	(0.29, 0.95)	(0.48, 1.17)
Households	10.54	9.87		12.54	13.68	15.61	12.87	13.28	10.94	14.46	19.1	13.49
	(4.68, 16.41)	(1.76, 17.99)		(3.34, 21.74)	(3.36, 24.0)	(3.45, 27.77)	(3.39, 22.34)	(0.31, 26.25)	(2.19, 19.68)	(0.96, 27.97)	(-0.16, 38.35)	(1.11, 25.88)
Pension	-0.09	-0.05	-0.06		-0.01	0.0	-0.0	0.0	0.04	0.01	0.0	-0.02
	(-0.28, 0.09)	(-0.27, 0.17)	(-0.38, 0.26)		(-0.22, 0.2)	(-0.2, 0.21)	(-0.21, 0.21)	(-0.22, 0.22)	(-0.18, 0.26)	(-0.2, 0.23)	(-0.2, 0.21)	(-0.23, 0.2)
Insurance	0.6	-0.05	0.47	-0.06		-0.08	-0.06	-0.22	-0.09	-0.09	-0.02	-0.14
	(0.24, 0.97)	(-0.46, 0.37)	(-0.17, 1.1)	(-0.46, 0.33)		(-0.47, 0.31)	(-0.45, 0.34)	(-0.64, 0.19)	(-0.5, 0.33)	(-0.5, 0.32)	(-0.42, 0.38)	(-0.55, 0.27)
Mutual Funds	0.18	0.19	0.82	0.23	0.21		0.21	0.07	0.01	0.3	0.24	0.18
	(-0.39, 0.74)	(-0.48, 0.86)	(-0.3, 1.93)	(-0.42, 0.88)	(-0.43, 0.86)		(-0.43, 0.86)	(-0.6, 0.75)	(-0.65, 0.66)	(-0.38, 0.98)	(-0.43, 0.9)	(-0.49, 0.85)
ETF	-0.75	-0.44	-0.65	-0.47	-0.48	-0.48		-0.45	-0.38	-0.42	-0.56	-0.49
	(-1.19, -0.31)	(-0.96, 0.08)	(-1.41, 0.1)	(-0.96, 0.03)	(-0.98, 0.01)	(-0.97, 0.01)		(-0.97, 0.07)	(-0.9, 0.14)	(-0.92, 0.09)	(-1.05, -0.07)	(-1.0, 0.02)
Dealers	3.36	7.18	-6.45	1.85	1.34	0.59	2.0		4.49	2.65	-1.48	0.02
	(-6.48, 13.21)	(-5.92, 20.28)	(-21.26, 8.35)	(-9.19, 12.89)	(-9.67, 12.36)	(-10.31, 11.49)	(-9.07, 13.07)		(-7.88, 16.87)	(-9.56, 14.86)	(-12.53, 9.56)	(-11.28, 11.32)
Fed	0.61	0.67	0.24	0.57	0.55	0.47	0.57	0.63		0.49	0.49	0.67
	(0.21, 1.02)	(-0.1, 1.44)	(-0.61, 1.09)	(-0.15, 1.3)	(-0.16, 1.26)	(-0.21, 1.14)	(-0.15, 1.29)	(-0.23, 1.49)		(-0.22, 1.21)	(-0.31, 1.29)	(-0.2, 1.54)
Banks	0.95	0.98	0.64	0.96	0.94	0.98	0.94	0.92	0.86		0.91	0.91
	(0.73, 1.16)	(0.72, 1.25)	(0.25, 1.03)	(0.7, 1.21)	(0.69, 1.19)	(0.73, 1.23)	(0.69, 1.2)	(0.65, 1.18)	(0.59, 1.12)		(0.66, 1.16)	(0.65, 1.17)
RoW	0.54	0.46	1.0	0.51	0.5	0.51	0.51	0.31	0.48	0.49		0.46
	(0.32, 0.76)	(0.14, 0.77)	(0.64, 1.36)	(0.21, 0.82)	(0.2, 0.81)	(0.21, 0.81)	(0.2, 0.81)	(-0.37, 0.99)	(0.18, 0.79)	(0.17, 0.81)		(0.14, 0.78)
Supply	0.12	0.13	0.12	0.11	0.1	0.09	0.11	0.1	0.15	0.1	0.08	
	(0.05, 0.2)	(0.03, 0.23)	(-0.03, 0.28)	(0.02, 0.2)	(0.02, 0.19)	(0.01, 0.18)	(0.02, 0.2)	(0.0, 0.19)	(0.04, 0.25)	(0.01, 0.19)	(-0.01, 0.17)	

Note: Table B.3 reports the estimated 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 and report the estimated elasticities of other sectors. 95% confidence intervals are reported based on standard errors derived in Theorem A.1.

Table B.4: **VIX loadings by sector: Leave-one-out Estimation**

Removed Sector	Aggregate Loading	RoW Loading
Baseline	0.42	-0.4
	(-0.41, 1.25)	(-1.12, 0.33)
Other	0.56	-0.48
	(-0.28, 1.4)	(-1.23, 0.28)
Households	-0.21	-0.13
	(-0.86, 0.44)	(-0.94, 0.67)
Pension	0.56	-0.37
	(-0.33, 1.45)	(-1.13, 0.38)
Insurance	0.55	-0.37
	(-0.37, 1.47)	(-1.13, 0.38)
Funds	0.72	-0.37
	(-0.24, 1.68)	(-1.13, 0.38)
ETF	0.5	-0.38
	(-0.4, 1.4)	(-1.14, 0.37)
Dealers	0.67	-0.2
	(-0.31, 1.64)	(-1.27, 0.87)
Fed	0.08	-0.43
	(-0.68, 0.85)	(-1.19, 0.32)
Banks	0.58	-0.36
	(-0.37, 1.53)	(-1.12, 0.4)
Supply	0.58	-0.42
	(-0.34, 1.5)	(-1.17, 0.34)

Note: Table B.4 reports the estimated sectoral demand / supply loadings on changes in the VIX index, leaving one sector out of the moment conditions in each run. 95% confidence intervals are reported based on standard errors derived in Theorem A.1.

Table B.5: **Price elasticity by sector: Alternative Specifications**

Sector	**Baseline**	**More factors**	**1970-2023**	**Including Bills**	**Alt. Disc. Adj.**	**Alt. Mom. Combo**
Aggregate	1.2 (0.9, 1.5)	1.11 (0.84, 1.38)	1.25 (1.06, 1.43)	2.29 (1.53, 3.05)	1.19 (0.91, 1.48)	1.22 (0.89, 1.55)
Other	0.05 (-0.22, 0.32)	0.17 (-0.07, 0.41)	0.21 (0.04, 0.37)	0.66 (0.29, 1.03)	0.08 (-0.2, 0.36)	-0.3 (-0.57, -0.03)
Households	10.54 (4.68, 16.41)	9.67 (4.39, 14.94)	10.03 (7.03, 13.04)	5.02 (0.67, 9.38)	13.63 (6.31, 20.94)	12.85 (6.13, 19.58)
Pension	-0.09 (-0.28, 0.09)	-0.1 (-0.26, 0.06)	-0.17 (-0.32, -0.01)	0.13 (-0.14, 0.4)	-0.1 (-0.29, 0.09)	-0.13 (-0.31, 0.05)
Insurance	0.6 (0.24, 0.97)	0.61 (0.28, 0.93)	-0.61 (-0.87, -0.36)	-0.31 (-0.91, 0.29)	0.6 (0.22, 0.97)	0.83 (0.46, 1.19)
Mutual Funds	0.18 (-0.39, 0.74)	0.26 (-0.29, 0.82)	0.44 (0.2, 0.67)	0.21 (-0.55, 0.97)	0.2 (-0.39, 0.78)	-0.08 (-0.62, 0.47)
ETF	-0.75 (-1.19, -0.31)	-0.9 (-1.3, -0.5)	0.18 (-0.01, 0.37)	-0.52 (-1.14, 0.09)	-0.65 (-1.1, -0.2)	-0.84 (-1.27, -0.42)
Dealers	3.36 (-6.48, 13.21)	6.07 (-3.81, 15.95)	-2.0 (-8.63, 4.63)	-1.1 (-13.12, 10.92)	3.41 (-6.71, 13.53)	3.52 (-5.98, 13.02)
Fed	0.61 (0.21, 1.02)	0.5 (0.17, 0.84)	0.02 (-0.07, 0.11)	0.29 (-0.16, 0.74)	0.62 (0.19, 1.04)	0.51 (0.13, 0.89)
Banks	0.95 (0.73, 1.16)	0.8 (0.6, 1.01)	0.96 (0.48, 1.43)	1.05 (0.79, 1.32)	0.94 (0.72, 1.17)	0.94 (0.73, 1.15)
RoW	0.54 (0.32, 0.76)	0.52 (0.32, 0.72)	0.36 (0.24, 0.49)	0.68 (0.43, 0.92)	0.54 (0.32, 0.77)	0.46 (0.25, 0.67)
Supply	0.12 (0.05, 0.2)	0.08 (0.01, 0.15)	0.25 (0.18, 0.32)	1.41 (0.46, 2.37)	0.13 (0.05, 0.2)	0.12 (0.05, 0.19)
MMF				3.42 (0.95, 5.88)		

Note: Table B.5 demonstrates the robustness of our price elasticity estimation by augmenting the model with 10 estimated principal components from forecasts on one-year and ten-year yield changes, five factors each (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). In column 5, we explore an alternative adjustment strategy (discussed in Section B.1) to deal with the reported discrepancy in notes and bonds holdings, by assigning 75% of the "other" sector holdings to the discrepancy first before sweeping the discrepancy into the household sector. In Column 6 ("Alt. Mom. Combo"), guided by Figure B.13, we exclude "other" sector and foreign countries, as well as Fed and mutual funds, from the pair of moment conditions that enter into the GMM estimation for price elasticities. 95% confidence intervals are reported based on standard errors derived in Theorem A.1.

Table B.6: **VIX loadings by sector: Alternative Specifications**

Sector	**Baseline**	**More factors**	**Including Bills**
Aggregate	0.42	0.42	1.03
	(-0.41, 1.25)	(-0.47, 1.31)	(0.03, 2.03)
Rest of World	-0.4	-0.2	0.21
	(-1.12, 0.33)	(-0.94, 0.54)	(-0.47, 0.89)
Domestic demand	1.02	1.15	0.74
	(-1.77, 3.82)	(-1.8, 4.11)	(-0.13, 1.61)
Supply	-0.04	-0.2	0.52
	(-0.23, 0.16)	(-0.4, -0.0)	(-0.36, 1.4)

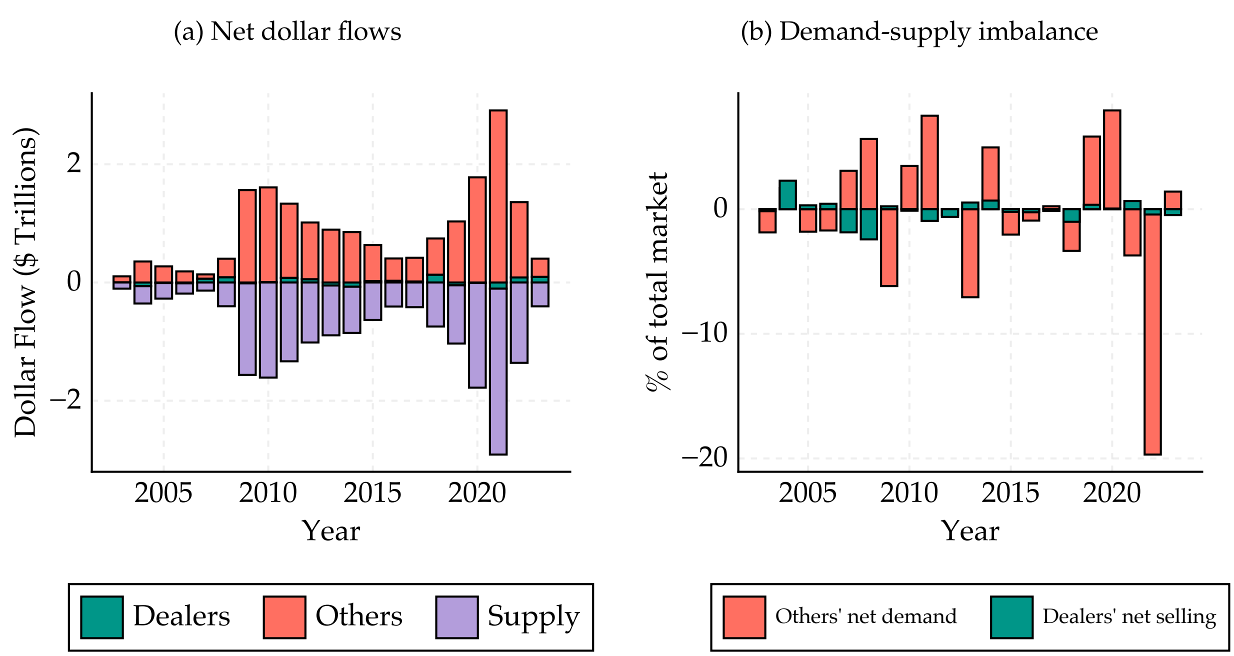
Note: Table B.6 demonstrates the robustness of our VIX loading estimation by augmenting the model with 10 estimated principal components from forecasts on one-year and ten-year yield changes, five factors each (column 2), and by including Treasury bills in the estimation (column 3). The coefficients for "domestic demand" are obtained by size-weighting the sector-specific VIX loadings for non-foreign investor sectors. Standard errors for "domestic demand" are constructed using the variance-covariance matrix of the sector-specific VIX loadings.

Table B.7: **Baseline Estimates: Full Table**

Sector	$S(\%)$	ζ	$\epsilon^{VIX}_{(std.)}$	**Δ FFR**	$Inf.$	$\epsilon^{USD}_{(std.)}$	**Lagged** f_{supply}	**Fed schedule**
Aggregate		1.2	0.42	-2.68	-1.46	-0.29	0.1	-0.05
		(0.9, 1.5)	(-0.41, 1.25)	(-4.16, -1.21)	(-2.23, -0.69)	(-0.99, 0.42)	(-0.56, 0.76)	(-0.85, 0.75)
Supply	100.0	0.12	-0.04	-0.03	-0.12	-0.06	1.89	-0.19
		(0.05, 0.2)	(-0.23, 0.16)	(-0.4, 0.33)	(-0.31, 0.07)	(-0.23, 0.11)	(1.73, 2.04)	(-0.35, -0.03)
RoW	44.45	0.54	-0.4	-0.55	-0.49	-0.22	-2.2	-0.65
		(0.32, 0.76)	(-1.12, 0.33)	(-1.84, 0.73)	(-1.13, 0.16)	(-0.87, 0.43)	(-2.77, -1.63)	(-1.24, -0.06)
Fed	22.08	0.61	1.74	-4.58	-1.05	-0.46	0.71	5.4
		(0.21, 1.02)	(0.16, 3.32)	(-7.29, -1.87)	(-2.35, 0.26)	(-1.68, 0.77)	(-0.55, 1.96)	(3.43, 7.36)
Other	11.78	0.05	-1.45	-1.54	0.3	1.18	-0.57	0.79
		(-0.22, 0.32)	(-2.63, -0.26)	(-3.69, 0.61)	(-0.77, 1.38)	(0.13, 2.22)	(-1.52, 0.38)	(-0.18, 1.76)
Mutual Funds	6.75	0.18	-2.18	-0.1	0.91	0.75	-2.37	-0.7
		(-0.39, 0.74)	(-3.82, -0.53)	(-3.15, 2.94)	(-0.66, 2.48)	(-0.7, 2.2)	(-3.68, -1.06)	(-2.05, 0.64)
Households	5.74	10.54	11.3	-18.93	-12.98	-4.83	-10.39	-15.45
		(4.68, 16.41)	(1.48, 21.12)	(-38.13, 0.28)	(-23.68, -2.29)	(-13.37, 3.71)	(-17.94, -2.83)	(-23.14, -7.75)
Pension	5.42	-0.09	-0.86	1.84	0.6	-0.41	-1.27	-0.59
		(-0.28, 0.09)	(-1.86, 0.14)	(0.03, 3.65)	(-0.29, 1.5)	(-1.29, 0.47)	(-2.08, -0.47)	(-1.42, 0.23)
Banks	5.26	0.95	-0.17	-2.43	-0.12	1.18	-1.02	2.38
		(0.73, 1.16)	(-1.88, 1.53)	(-5.51, 0.65)	(-1.64, 1.4)	(-0.33, 2.69)	(-2.41, 0.36)	(0.97, 3.8)
Insurance	2.59	0.6	-0.05	2.48	0.39	-0.55	-2.62	-0.69
		(0.24, 0.97)	(-1.08, 0.98)	(0.57, 4.39)	(-0.6, 1.38)	(-1.46, 0.35)	(-3.44, -1.79)	(-1.53, 0.15)
ETF	1.18	-0.75	1.4	3.92	1.05	0.86	-0.22	-0.25
		(-1.19, -0.31)	(0.16, 2.64)	(1.62, 6.22)	(-0.13, 2.24)	(-0.23, 1.95)	(-1.21, 0.76)	(-1.26, 0.76)
Dealers	0.81	3.36	-5.47	-25.17	-36.36	3.08	5.94	0.77
		(-6.48, 13.21)	(-30.99, 20.05)	(-72.7, 22.36)	(-61.13, -11.59)	(-19.33, 25.48)	(-14.32, 26.19)	(-19.92, 21.46)

Note: Table B.7 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 of the parameters are taken for sectors that are assumed to have different price elasticities before and after 2009Q1 (the sectors include Federal Reserve and Rest of World investors). We further assume that Federal Reserve's loadings on common factors can be regime-dependent. $\varepsilon^{VIX}_{(std.)}$ 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 quarterly changes in long-term (10-year) inflation expectation. 95% confidence intervals are reported, with the standard errors given by Theorem A.1 and Proposition A.1.

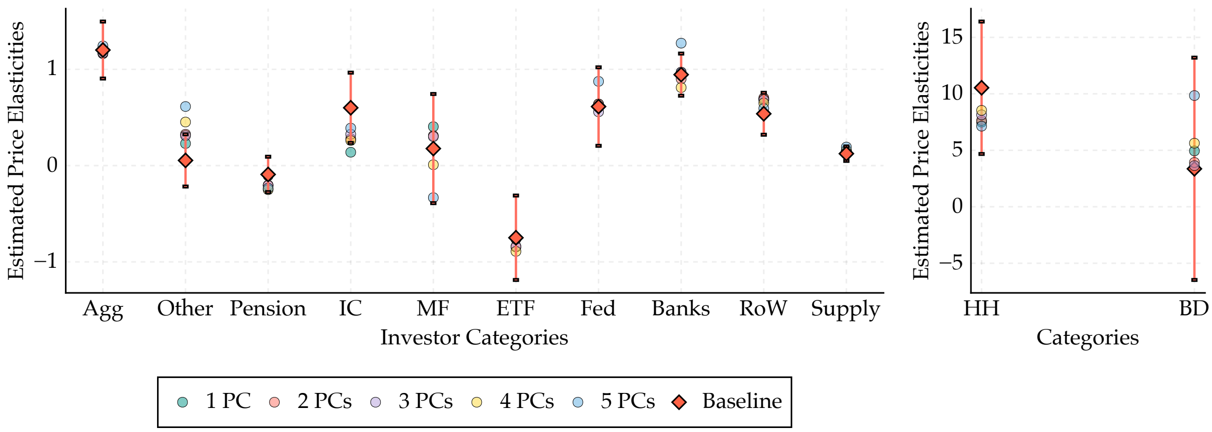
Figure B.7: **The role of security broker-dealers**



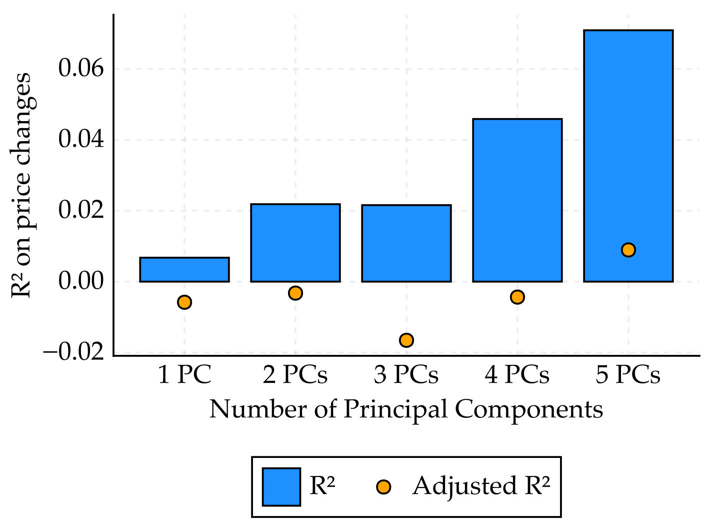
Note: Figure B.7 examines the role of security broker-dealers in affecting U.S. Treasury yields at low frequency. In Panel (a), we compare net annual flow into the Treasury market due to broker-dealers (in green) with those of other holders and the net issuance. In Panel (b), we plot a measure of demand-supply 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, $\Delta q_{it} + \zeta_{ir(t)}\Delta p_t$ 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.

Figure B.8: **Robustness: Adding internal principal components**

(a) Estimated price elasticities

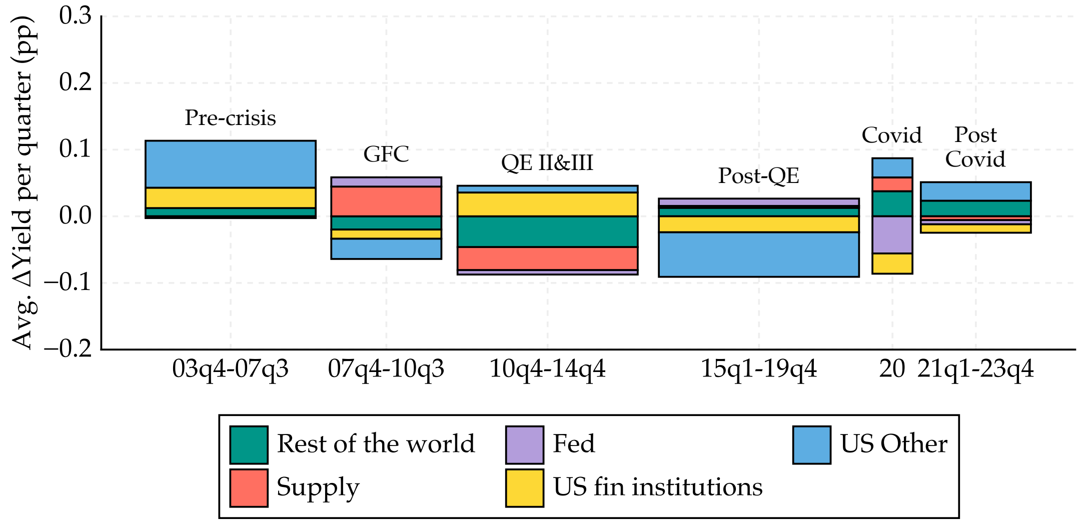


(b) Total Price Variation explained by internal principal components



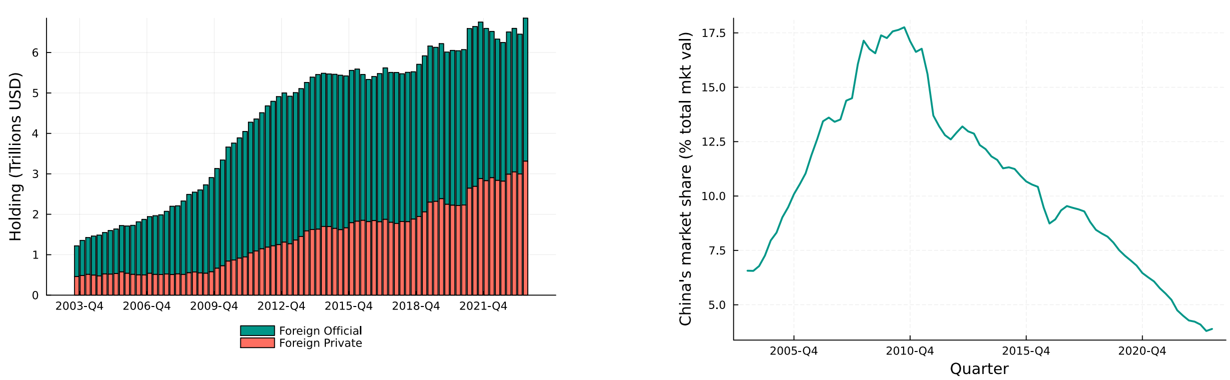
Note: Figure B.8 examines the robustness of our price elasticity estimates to adding internal principal components extracted from the flow data. Panel (a) reports the estimated price elasticities by sector. The red diamonds correspond to our baseline estimates, and the dots corresponds to the estimates when 1 through 5 principal components are estimated jointly. 95% confidence intervals are plotted. Panel (b) reports the R^2 and adjusted R^2 obtained from regressing price changes on the estimated principal components. The fit measures suggest that potential bias arising from unobserved factor structures are likely small.

Figure B.9: **Sectoral decomposition of idiosyncratic demand and supply shocks**



Note: Figure B.9 further breaks down the contribution of latent demand and supply shocks (plotted in Figure 6) 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 B.10: **Foreign holdings of Treasury notes and bonds: Official holdings and China**

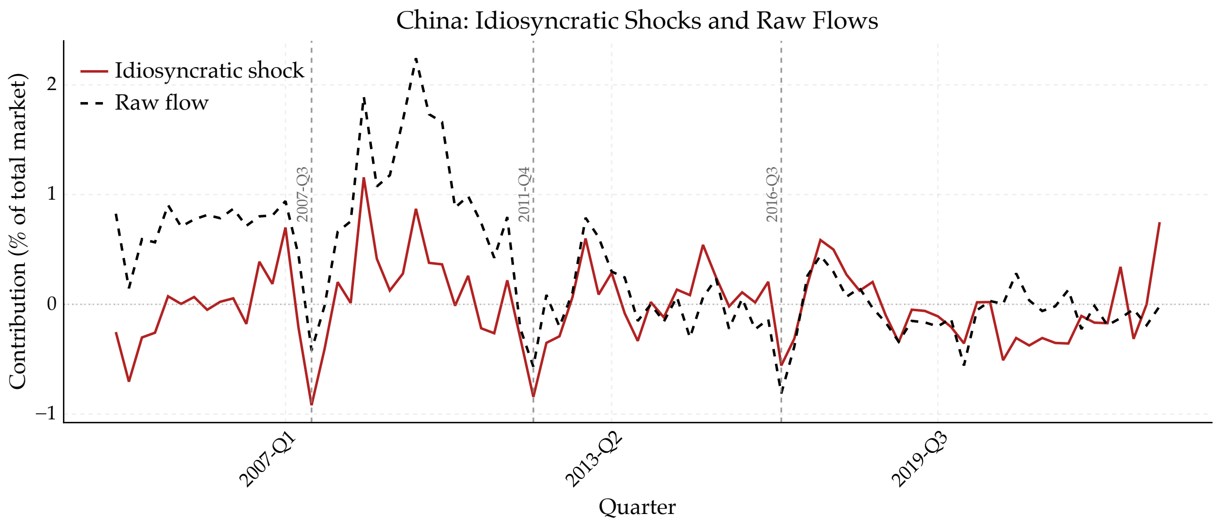


(a) Official vs. Private investors (level)

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

Note: Figure B.10 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 plots its market share evolution over the sample period.

Figure B.11: **Interpreting estimated demand shocks: China as a case study**



Note: Figure B.11 tracks the estimated idiosyncratic Treasury bond demand shocks and raw Treasury bond purchases for China estimated from our optimal GIV procedure. Narrative evidence supports the idea that our estimated demand shocks capture idiosyncratic demand changes in China driven by domestic factors that are less related to global financial conditions or U.S. macro-financial advances. The significant negative demand shock in 2007Q3 corresponds to the first year since 2003 that China's pace of investment in U.S. securities started to moderate, interpreted by market analysts as a sign that China was starting to diversify its foreign reserves away from U.S. Treasuries (see, for example, USCC Economic Issue Brief). The Department of Treasury's Report to Congress on International Economic and Exchange Rate Policies points out that in 2011Q4, reserve accumulation significantly slowed, with capital flows into China slowing down, and the RMB exchange rate trading at the weaker side of the trading band. The significant negative shock of China's Treasury holdings in 2016Q3 follows from the substantial foreign exchange intervention to stabilize the CNY exchange rate (Bloomberg), while the positive shocks that followed in 2017 are interpreted as an attempt to stem the appreciation pressure of RMB since the end of 2016 (Reuters).

Table B.8: **Foreign Official/Private Sector: Estimation Results**

$Sector$	ζ	ζ **Share (%)**
Foreign non-official (03-08)	0.75 (-1.22, 2.72)	5.57
Foreign non-official (09-23)	2.52 (1.33, 3.7)	19.65
Foreign official (03-08)	2.28 (1.45, 3.11)	33.91
Foreign official (09-23)	0.6 (0.11, 1.09)	9.46

(a) Regime shift in price elasticities

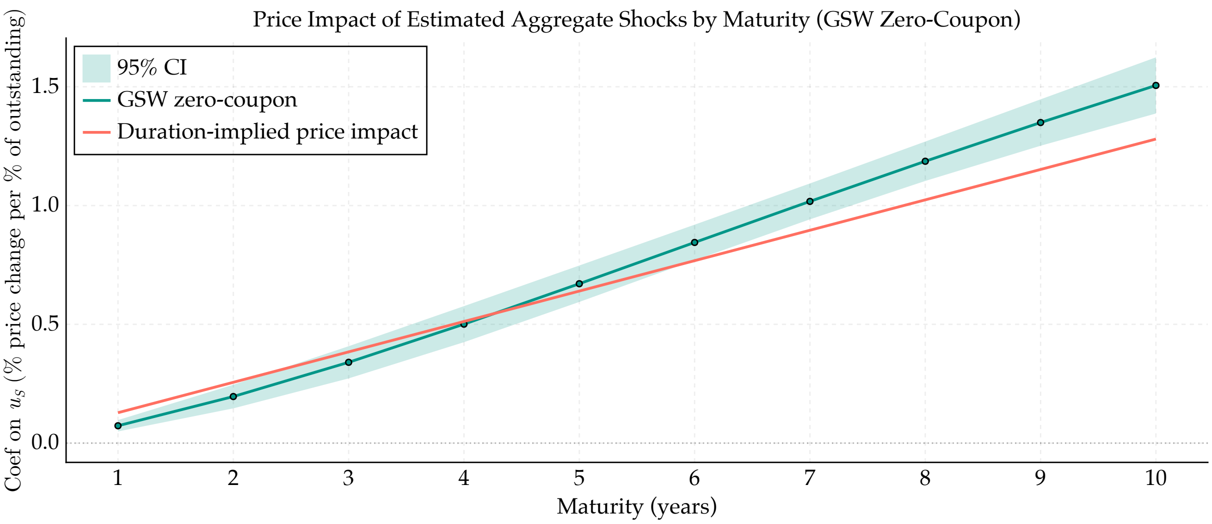
$Sector$	$S(\%)$	$\epsilon^{VIX}_{(std.)}$	Δ **FFR**	$Inf.$	$\epsilon^{USD}_{(std.)}$
Foreign official	29.66	-0.23 (-1.22, 0.77)	-2.29 (-4.08, -0.5)	-1.7 (-2.64, -0.76)	-0.36 (-1.22, 0.5)
Foreign non-official	14.79	0.51 (-1.8, 2.82)	-1.66 (-5.85, 2.53)	-1.25 (-3.46, 0.97)	-0.93 (-2.94, 1.08)

(b) Loadings on common factors

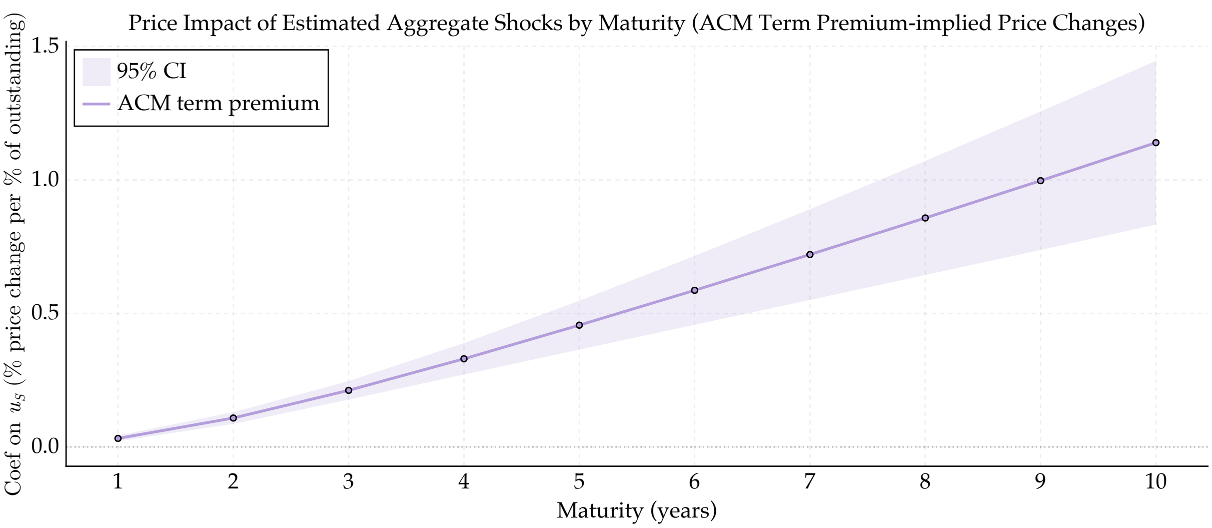
Note: Table B.8 reports the estimated price elasticities and loadings on common factors for the foreign official and private sectors based on TIC data, allowing for a regime shift in price elasticities after 2009Q1. 95% confidence intervals are reported based on standard errors derived in Theorem A.1.

Figure B.12: **Price impact of estimated demand shocks across maturity**

(a) Zero-coupon yields

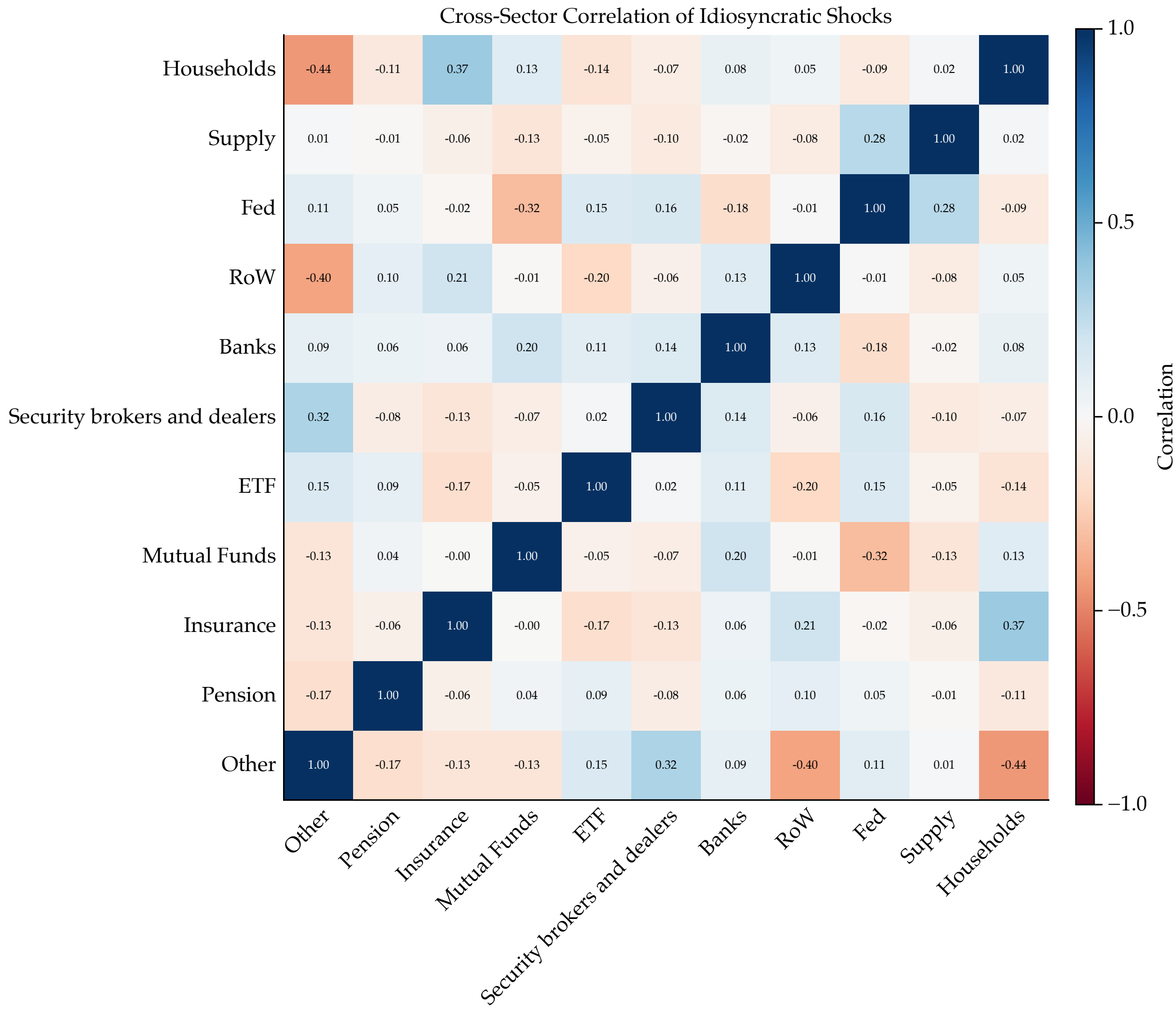


(b) Term premia



Note: Figure B.12 report the coefficients and 95% confidence intervals for a series of contemporaneous price impact regressions (see Equation (4.1)) of maturity-specific implied price changes using zero-coupon yields (Panel (a), (Gurkaynak, Sack, & Wright, 2007)) and term premia estimates (Panel (b), (Adrian, Crump, & Moench, 2013)) on the estimated aggregate demand/supply shocks, $\hat{u}_{S,t}$. The red line plots the implied price impact of a "level shift" yield compression equal to 12.8 basis points across all maturities, corresponding to our baseline estimated price multiplier of 0.83 for the market portfolio of Treasury notes and bonds. The fact that the estimated price impact is close to the level shift line suggests that the demand shocks in our framework are most relevant in moving the "level factor" of Treasury yield curves.

Figure B.13: **Estimated demand shocks: Correlations across sectors**



Note: Figure B.13 plots the pairwise sample correlations of the estimated idiosyncratic demand/supply shocks across sectors. In Table B.5, we exclude sector pairs shown in this figure to exhibit moderate sample correlations (other vs. foreign, Fed vs. mutual funds) from the moment conditions used in the GMM estimation, and show that our price elasticity estimates remain robust.