Anatomy of the Treasury Market

Manav Chaudhary* Julie Zhiyu Fu† Haonan Zhou‡

First draft: November 15, 2024

This draft: November 9, 2026

[Click here for the latest version]

Abstract

The Treasury market's central role in the financial system rests on its liquidity and its safe-haven property. Factor regressions can track such properties but cannot say which investors drive them. We develop an equilibrium framework that decomposes yield-factor regressions into investor-level drivers, and an optimal estimator that recovers each investor's responses to yields and factors. Using sector-level flow data spanning 1986–2023, we show that both properties are eroding as private investors such as dealers and foreigners step back, and that the Federal Reserve has partially stepped in their place, moderating the aggregate decline.

Keywords: Treasury market; demand system; price elasticity; liquidity; safe haven; flight to safety; granular instrumental variables.

JEL codes: E43, E58, G11, G12, H63.

*Department of Finance, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom. Corresponding author. Email: M.Chaudhary6@lse.ac.uk.

†Olin Business School, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, United States. Email: z.fu@wustl.edu.

‡HKU Business School, The University of Hong Kong, Pokfulam Road, Hong Kong. Email: haonanz@hku.hk.

1 Introduction

The U.S. Treasury market is the bedrock of the global financial system. Two properties of this asset class help underpin that position. First, the market is deep and liquid: it can accommodate large shifts in investor demand without large price impact, which makes Treasuries the natural home for funds that investors worldwide may need to deploy quickly. Second, it is a safe haven: it provides insurance by appreciating in periods of economic stress. There is growing concern that both properties are eroding. Foreign and domestic private investors are stepping away from the market, and in March 2020, at the height of the COVID crisis, Treasuries fell in value just when a safe haven should rise. If these properties go, the central role of the Treasury market may come into question. Both properties are equilibrium outcomes: they reflect how the investors who hold Treasuries respond to prices and to risk. Whether these properties are eroding therefore depends on which investors sustain them, and on whether their behavior is changing.

Identifying those investors requires a framework that links the market's aggregate properties to the investors behind them. We develop one. It connects investors' portfolio choices to the aggregate pricing properties that standard factor regressions measure. A factor regression records how much Treasury prices move with a factor—for instance, how much they rise when market stress increases. Its loadings, however, blend two things: how much investors' demand shifts when the factor moves, and how elastically the market absorbs those shifts, which is its liquidity. The regression cannot tell the two apart, let alone say which investors are behind each. Our framework separates them and attributes each to the investor groups behind it. To estimate the model, we develop a flexible optimal estimator that needs only sector-level holdings and flows.

We apply the framework to four decades of data and find that both the liquidity and the safe-haven property are eroding. The erosion we observe is the net of two moves. A handful of private investor groups have stepped back, and the Federal Reserve (Fed) has partially stepped in. First, the market's liquidity has declined moderately. Private investors, foreign investors in particular, have pulled back considerably from providing liquidity, yet aggregate liquidity has fallen far less, because the Federal Reserve has filled much of the gap they left. Second, the safe-haven property has visibly faded. Treasuries became a safe haven only in the late 1990s and have since lost much of that property; dealers, not foreign investors, drove both the rise and the decline, and the Federal Reserve's response to stress slowed the fall without stopping it.

Our framework starts from the portfolio choice of each investor. We model each

sector's demand for Treasuries as a log-linear function of prices, economic factors, and sector-specific shocks. This form is a first-order approximation to a broad class of portfolio choice problems, so differences in investors' objectives, constraints, and beliefs show up as differences in their estimated price elasticities and factor loadings, without our having to model those differences explicitly. Because our question concerns the liquidity and safety of the Treasury market as a whole, we treat Treasuries as a single asset class and abstract from substitution along the yield curve. Aggregating these investor demands through market clearing gives us the standard price-factor regression, and shows what its loadings measure. When a factor moves, investors' demand shifts, and prices must move until the market absorbs the net shift. An aggregate price loading therefore combines the direct shift in demand with the price move needed to absorb it, and that price move is governed by the market's aggregate price elasticity, which is its liquidity.

Because we estimate demand investor by investor and link it to this aggregate pricing equation, we can decompose both the direct factor driven shift and subsequent liquidity provision into investor contributions. The aggregate direct response to an innovation in a factor is the size-weighted sum of investors' individual responses. This decomposition reveals which investors drive that response. For example, it shows which investors increase or reduce their Treasury demand during episodes of financial tightening. The aggregate price elasticity is likewise the size-weighted sum of investors' elasticities, so a large sector that does not respond to price adds little liquidity. Both are investor-level objects that we allow to change across regimes, so we can trace shifts in the aggregate price loadings, such as the rise and decline of the safe-haven property, to the changes in investor-level behavior that produced them.

Estimating these investor-level responses raises the central identification challenge. When a factor moves, investor holdings respond through two channels: directly, because the factor shifts their demand, and indirectly, because other investors' demand shifts move prices until the market clears. An investor may therefore sell Treasuries during a flight to safety not because its own response to stress is negative, but because others' purchases have raised prices. Distinguishing the two channels requires an instrument for the price in every sector's demand equation. Existing Treasury market instruments, such as regulatory changes and monetary policy surprises, cover particular sectors or short episodes. We instead develop an optimal generalized method of moments (GMM) estimator that uses the idiosyncratic component of each sector's demand as an instrument in the other sectors' equations. The identifying assumption is that these components are uncorrelated across sectors once common factors are accounted for. The estimator is an

optimal implementation of granular instrumental variables (GIV; Gabaix & Koijen, 2024). It has two advantages. First, it lets each sector have its own price elasticity and factor loadings. Second, by giving more weight to demand shocks from the sectors with greater influence on prices, it delivers much more powerful instruments. That statistical power makes it possible to estimate how price elasticities and factor loadings differ across sectors, and how they change over time.[1]

Our data are publicly available sector-level holdings of Treasuries, drawn from the Financial Accounts, Treasury International Capital data, and Call Reports, covering 1986 to 2023. The market's liquidity and its safe-haven property both change slowly, over years rather than quarters, and this long sample lets us follow them, and the investors behind them, through several distinct regimes. We take the two properties in turn.

The market's liquidity has declined moderately. Before the financial crisis, a sale of Treasuries equal to 1% of the market lowered their prices by about 0.9%, or equivalently raised their yields by about 16 basis points at an average duration of around 5.5 years; since then, it has lowered prices by about 1.2%, or 22 basis points in yields. This is a moderate decline, given how much the investor base has changed since the financial crisis.

Private investors have retreated from providing liquidity, and the Federal Reserve has filled much of the gap. Domestic institutions such as banks and pension funds held much of the market in the 1990s; by 2008, foreign investors held nearly two-thirds of it. After the financial crisis, however, the size of foreign holdings no longer translated into liquidity provision. Foreign holdings have declined steadily since 2008 but remain sizeable, at about a third of the market in 2023. Their price elasticity fell from 0.52 to 0.09, so these holdings now absorb little of the market's demand shifts. China, once the largest foreign holder, has cut its Treasury holdings by about 40% since 2013. The Federal Reserve filled this gap, moderating the aggregate decline in liquidity: its post-crisis holdings now account for about a third of the market's aggregate elasticity. Without them, and holding other investors' behavior fixed, the market's liquidity would have eroded far more: the same 1% sale would today lower prices by about 1.8% (33 basis points in yields), twice its pre-crisis impact. As a result, the Treasury market's liquidity now rests partly on the central bank rather than on private investors alone.

The safe-haven property emerged in the late 1990s and has faded since the early 2010s. Before 1997, Treasury prices did not rise when market stress increased, as mea-

[1] We make the estimator available as open-source packages for Julia (https://github.com/FuZhiyu/OptimalGIV.jl) and Python (https://github.com/FuZhiyu/optimalgiv).

sured by jumps in stock market volatility. From 1997 through the financial crisis, they did: a one-standard-deviation rise in volatility raised Treasury prices by about 1.5% within the quarter, pushing yields down by some 27 basis points. Since 2012, this response has largely faded, to less than a third of its peak by the end of our sample. The rise and fall echoes the well-documented shift in the stock–bond correlation and the concerns raised during the COVID-19 crisis about Treasuries' waning safety (Campbell, Pflueger, & Viceira, 2020; Antolín-Díaz, 2026; He, Nagel, & Song, 2022; Vissing-Jorgensen, 2021).

Foreign investors did not drive Treasuries' safe-haven property. A common view attributes the appreciation of Treasuries during periods of stress to foreign investors' flight to safety (Jiang, Krishnamurthy, & Lustig, 2021, 2024). Our demand system separates investors' demand responses to uncertainty from their responses to price changes. We find that foreign investors contributed little when the safe-haven property emerged in the late 1990s, and that since the financial crisis their demand has, if anything, declined with uncertainty. These findings support a precautionary saving view in which foreign investors accumulate Treasuries in normal times and draw on them for liquidity during periods of stress, benefiting from their insurance value without generating the countercyclical demand that sustains it (Eisenbach & Phelan, 2026; Bianchi, Hatchondo, & Martinez, 2018; Devereux, Engel, & Wu, 2023). Reduced-form evidence points in the same direction: after adjusting for growth in Treasury supply, foreign flows decline with uncertainty while domestic private flows rise, including when the sample is restricted to the global financial crisis.

Dealers drive both the rise and the decline of the safe-haven property. Before the late 1990s, dealers' demand for Treasuries showed little response to market stress. From then until the financial crisis, dealers bought Treasuries when stress rose, and afterwards this response fell back to about zero. After dealers stepped back, the safe-haven property survived for several more years, carried by the household sector—a group that includes hedge funds—and by the Federal Reserve's asset purchases in response to stress. After the household sector's responsiveness declined towards the end of our sample, the safe-haven property is almost entirely driven by the Fed's policy support.

Together, these findings show that the investors sustaining the Treasury market's central properties have changed, and that the Federal Reserve now stands behind both. It has offset much of the private retreat from liquidity provision, and its response to stress helped carry the safe-haven property after dealers withdrew. This poses a challenge for the design of monetary policy normalization strategies. If the Federal Reserve shrinks its balance sheet and private investors do not return, the market may see a further, and

larger, decline in its ability to accommodate investor demand shifts, and its safe-haven property may not return.

1.1 Related Literature

A burgeoning literature studies how investors' demand shapes the pricing of U.S. Treasuries. Our contribution is to decompose the market's two defining properties, its liquidity and its safe-haven role, into the contributions of investors and to follow those contributions over four decades.[2] Koijen, Koulischer, Nguyen, and Yogo (2017) employ a demand system for government bonds in the euro area to examine the portfolio rebalancing mechanism of quantitative easing; Jiang, Richmond, and Zhang (2024) and Jiang, Richmond, and Zhang (2025) combine cross-country bilateral portfolio holdings data with a global demand system to understand the exorbitant privilege the U.S. commands on its debt and 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 (2026) estimate sectoral elasticities for Treasuries using monetary policy surprises, and Jansen, Li, and Schmid (2024) study how sector-specific demand shocks propagate through the term structure via risk-averse arbitrageurs and habitat investors who substitute across maturities. We contribute to this literature on two fronts: we trace the market's liquidity to the investors that provide it and show how their roles have shifted over four decades, and we identify the investors that make Treasuries a safe haven.

Among demand-based studies of the Treasury market, our work is closest to Jansen et al. (2024) (henceforth JLS). The key difference is the primary focus. Using maturity-bucket holdings from 2011, JLS study substitution along the Treasury yield curve and the effects of monetary policy on term premia. We instead document the retreat of private investors from liquidity provision and the rise and fall of the safe-haven property.[3] These properties shift across regimes that span decades, so we follow the market back to the 1980s, and our optimal GIV estimator provides the power to estimate how investors' behavior changes between regimes. The trade-off is that the longer sample relies on coarser

[2]Many recent advances in this literature build on the demand-system approach to asset pricing first applied to the cross-section of stocks (Koijen & Yogo, 2019). Variants have since been applied to the equity market (Gabaix & Koijen, 2022), the corporate bond market (Chaudhary, Fu, & Li, 2026a; Bretscher, Schmid, Sen, & Sharma, 2026; Fang & Xiao, 2025), and exchange rates (Camanho, Hau, & Rey, 2022; Koijen & Yogo, 2026; An & Huber, 2024). We contribute to this broader literature by proposing a novel identification method, described below.

[3]Eren et al. (2026) also estimate sectoral demand elasticities, for several countries, using monetary policy surprises as instruments. Their application focuses on the interaction between investor composition and central bank asset purchases and sales rather than on the evolution of the market's liquidity and safe-haven property.

data, so we treat Treasury notes and bonds as a single asset class. The different focuses drive different methodological choices. To study how demand shocks propagate across the yield curve, JLS take a structural approach, placing arbitrageurs (dealers and hedge funds) at the center of their model alongside estimated demand curves for the other sectors. To decompose aggregate price loadings on factors into sector contributions, we take a more agnostic approach: we estimate every sector's demand with the same flexible specification, dealers included, and let price elasticities and factor loadings change across regimes. This flexibility matters for our results, since dealers turn out to be the largest contributor to both turns in the safe-haven property. The two papers thus answer complementary questions: JLS quantify how sector-specific demand shocks propagate across maturities and into term premia, while we document which investors have supplied the market's liquidity and its safe-haven property, and when they withdrew.

Our first contribution concerns the market's liquidity. Our aggregate estimate is in line with the literature that measures the price impact of demand or supply shifts through event studies, term-structure models, and auction bidding data (e.g., Bernanke, Reinhart, & Sack, 2004; Warnock & Warnock, 2009; Greenwood & Vayanos, 2010; Gagnon, Raskin, Remache, & Sack, 2011a; Krishnamurthy & Vissing-Jorgensen, 2012; Hanson & Stein, 2015; Kaminska & Zinna, 2020; Jansen et al., 2024; Somogyi, Wallen, & Xu, 2025); Appendix B.3.1 maps these estimates into macro multipliers and compares them with ours. We depart from this literature in two ways. First, we decompose the aggregate elasticity into the contributions of individual sectors, whereas existing work studies one investor group at a time, such as the Federal Reserve (Haddad, Moreira, & Muir, 2024; Krishnamurthy & Vissing-Jorgensen, 2011), banks (Jiang, Matvos, Piskorski, & Seru, 2024), mutual funds (Selgrad, 2023; Huang, Jiang, Liu, & Liu, 2025), insurance companies (Chaudhary, 2024; Koijen & Yogo, 2023), broker-dealers (Du, Hébert, & Li, 2023; Vayanos & Vila, 2021; Greenwood & Vayanos, 2014), and foreign investors (Beltran, Kretchmer, Marquez, & Thomas, 2013; Wolcott, 2020; Ahmed & Rebucci, 2024; Tabova & Warnock, 2026). Our equilibrium model estimates all sectors jointly and separates shifts in investors' demand from their responses to equilibrium price changes, which reduced-form analyses of observed holdings cannot disentangle. Second, because each sector's elasticity can change across regimes, we can date the decline in aggregate liquidity and attribute it. Somogyi et al. (2025) document that auction demand has become markedly less elastic since 2010 and point to the retreat of foreign investors as one possible reason. Our estimates show the same decline for the market as a whole and identify the sectors behind it. Foreign investors' elasticity fell far more than their holdings shrank, no private sector has taken their place, and the Federal Reserve's balance sheet now accounts for a substantial share

of the market's capacity to absorb demand shifts, which speaks to the literature on quantitative easing and tightening (Gagnon, Raskin, Remache, & Sack, 2011b; Krishnamurthy & Vissing-Jorgensen, 2011; Du, Forbes, & Luzzetti, 2026).

Our second contribution concerns the safe-haven property. A common view attributes the appreciation of Treasuries in periods of stress to foreign investors' flight to safety (Jiang et al., 2021, 2024). Because our framework decomposes the market's response to uncertainty into investors' direct demand responses, net of their responses to the price changes that others' purchases induce, we can test this view directly. We find little support for it. Foreign investors contributed little when the property emerged in the late 1990s, and since the financial crisis their demand for Treasuries has, if anything, declined as uncertainty rises. This cyclicality is consistent with a precautionary saving view of foreign holdings, in which foreign investors accumulate Treasuries in normal times and sell them to raise liquidity during risk-off episodes (Eisenbach & Phelan, 2026; Bianchi et al., 2018; Devereux et al., 2023, among others). Our estimates instead attribute the safe-haven property primarily to domestic investors, and above all to dealers, whose response to stress switched on in the late 1990s and off after the financial crisis. This links the safety of Treasuries to the literature on intermediaries' balance-sheet capacity and its role in Treasury market functioning (He et al., 2022; Du et al., 2023; Duffie et al., 2023), and to accounts in which the safety premium on Treasuries is time-varying and tied to their value in bad states (Acharya & Laarits, 2026; Brunnermeier, Merkel, & Sannikov, 2024). These findings have implications for the sources and sustainability of the special status of U.S. safe assets (Gourinchas & Rey, 2022; Du, Keerati, & Schreger, 2025).

Methodologically, our approach extends the GIV literature (Gabaix & Koijen, 2024). GIV has been increasingly applied in macro-finance settings where market-level instruments are difficult to find (Kundu & Vats, 2026; Kubitza, Sigaux, & Vandeweyer, 2024; Chodorow-Reich, Gabaix, Koijen, & Viviano, 2024; Adrian, Grinberg, Liang, Malik, & Yu, 2022; Camanho et al., 2022). Our extension accommodates heterogeneous price elasticities and factor loadings across investors without parameterizing them on observable characteristics, which is what allows the sector-level and regime-level variation that the two preceding paragraphs rely on. We also propose an economically motivated optimal GMM estimator that 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. It builds on other efficient GIV variants, such as Qian (2024) and Baumeister and Hamilton (2023), which characterize optimality statistically rather than economically.

1.2 Roadmap

The remainder of the paper is organized as follows. Section 2 describes our data and 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 documents the trends in the market's liquidity provision, and Section 5 the trends in its safe-haven property. Section 6 concludes. The Online Appendix develops the general estimation framework and its properties (Appendix A) and provides the data details, robustness checks, and additional results (Appendix B).[4]

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. Because our question concerns the liquidity and safe-haven property of U.S. government debt as a whole, we treat Treasury bonds as a single asset class and abstract away from modelling the entire yield curve. This lets us characterize how aggregate Treasury bond prices are shaped by investors' allocation and government issuance decisions. Our empirical framework, however, can accommodate multiple assets and maturities, and we formally discuss how accounting for substitutes affects the interpretation of our estimates in Section 2.2 and Appendix A.5.

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

[4]Additional results and discussions are available in the Supplementary Material at https://anatomyoftreasury.github.io/AnatomyOfTreasury/Treasury_GIV_supplement.pdf.

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 1986 to 2023, our sample period. Treasury bonds are held by five broad groups: (i) banks; (ii) asset managers, including mutual funds, insurers, and pension funds; (iii) households and other U.S. investors, such as non-financial institutions and state and local governments; (iv) the Federal Reserve; and (v) foreign investors. Over our sample period, the total supply of Treasury bonds increased from about $1.2 trillion to more than $20 trillion, rising from roughly 25% of GDP to over 70%.[6] Over the same period, investor composition has shifted substantially. In the first half of the sample period, domestic institutional investors such as banks and pension funds account for significant market shares, while holdings have shifted towards foreign investors and the Federal Reserve in the latter half. The foreign share peaked in 2008 and has declined since.

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. Average yields during our sample period fluctuate between 0.5 and 9.3 percent, while the average duration of the market portfolio ranges from 4.6 to 7.4 years.

2.2 An Equilibrium Model

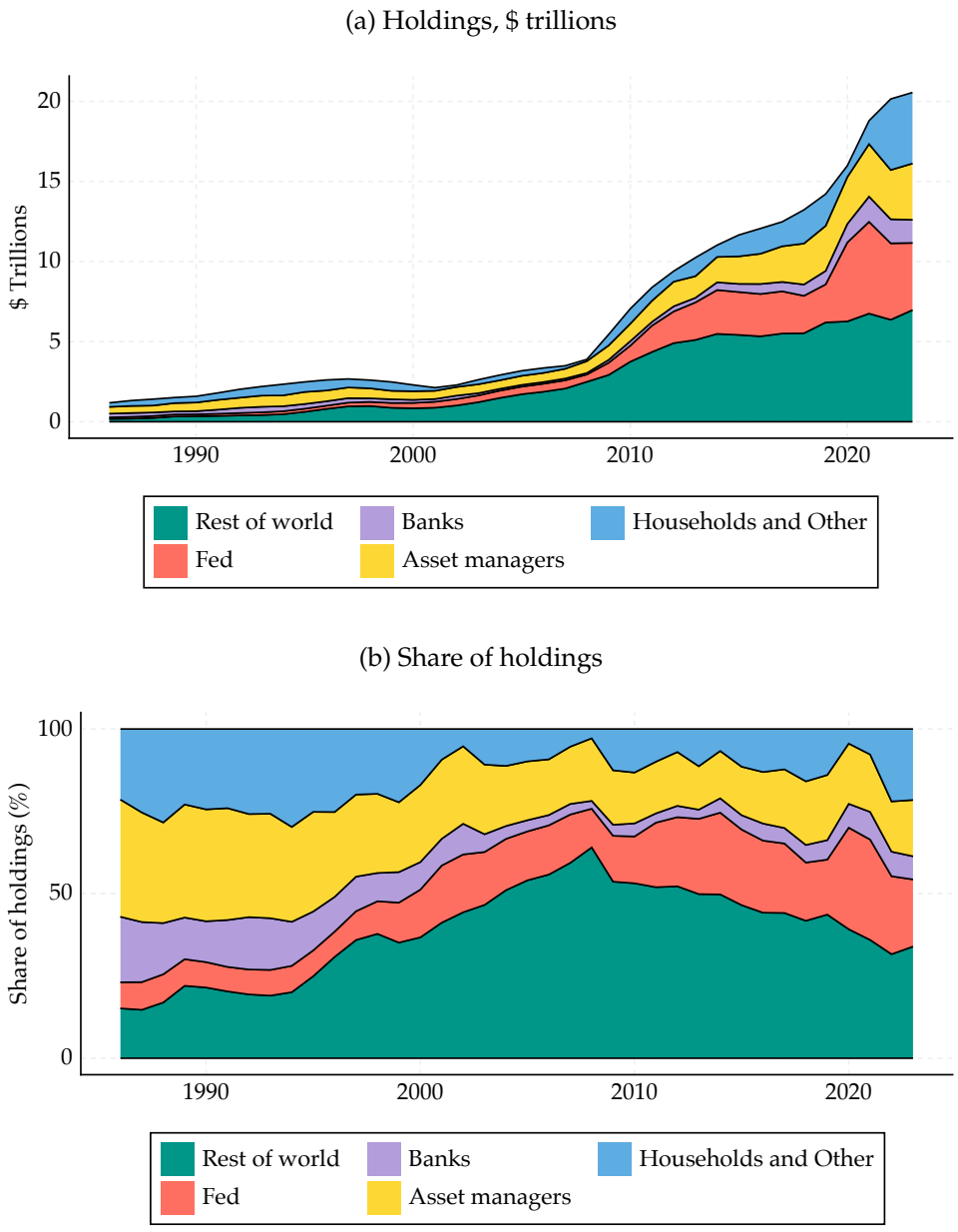
We outline our equilibrium model of the Treasury market. The model allows us to connect the yield we observe on Treasury bonds to the data we have collected on sectors' allocation decisions and the government's issuance decisions. The framework is parsimonious yet sufficiently flexible to incorporate sector-specific differences in objectives, constraints, beliefs, and substitution patterns across assets. Its flexibility also allows us to track the evolution of investors' liquidity provision and demand loadings on macro-financial factors over time to highlight the implications of structural shifts in the Treasury market.

Notation. 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

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



Note: This figure plots the year-end market value of U.S. Treasury notes and bonds holdings by investor groups in dollars (panel (a)) and as shares of the total (panel (b)). We aggregate investor sectors reported in Financial Accounts (Z.1) to five investor groups. *Asset managers* are insurers, pension funds, mutual funds, exchange-traded funds (ETFs), and dealers. *Households and Other* are the Households and Other categories. Appendix Table B.1 maps sectors to categories.

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 Trea-

sury 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:

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

Here, $\Delta q_{i,t}$ is the net purchase of Treasuries of sector i, normalized by a size weight. Δp_t is the percent change in the price of the Treasury market portfolio, $\zeta_{i,t}$ is sector i's price elasticity, $\overline{\Delta q}_i$ is a sector-specific 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.4, we discuss the structural interpretation of elasticity $\zeta_{i,t}$ 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.5, 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* $\zeta_{i,t}$, 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, $\zeta_{i,t}$ 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. The

reciprocal of the aggregate elasticity, $\zeta_{S,t}^{-1}$, corresponds to the overall price impact of a unit demand shock to cash 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]

The single-asset representation therefore aggregates the matrix of partial own- and cross-price elasticities into a single parameter $\zeta_{i,t}$, which reduces dimensionality and improves 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:

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

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 $\boldsymbol{\lambda}_{i,t}$. 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. Each sector's flow $\Delta q_{i,t}$ is weighted by a size weight $S_{i,t}$ that is allowed to be time-varying to arrive at the market clearing condition:

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

[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.

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:

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

where the S subscript denotes size-weighted aggregation across all sectors. For example, $\zeta_{S,t} \equiv \sum_i S_{i,t}\zeta_{i,t}$ is the aggregate elasticity and $\boldsymbol{\lambda}_{S,t} \equiv \sum_i S_{i,t}\boldsymbol{\lambda}_{i,t}$ 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, $\zeta_{S,t}^{-1}$. 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.5.

Once we obtain estimates for sector-specific elasticities $\zeta_{i,t}$ and factor loadings $\boldsymbol{\lambda}_{i,t}$, 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:

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

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 potentially time-varying elasticities $\zeta_{i,t}$ and factor loadings $\boldsymbol{\lambda}_{i,t}$. This presents a challenge: prices and quantities are jointly determined in equilibrium, creating simultaneity bias in standard estimation approaches. We lay out our identification strategy in the next section.

3 Identification and Estimation

3.1 Specification

We estimate (2.2) on nearly four decades of data. Rather than hold the demand parameters fixed over the whole sample, we allow them to shift across regimes. Letting $r(t)$ denote the regime that quarter t belongs to, we estimate

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

so that each sector's intercept, price elasticity and factor loadings can differ across regimes. Our baseline lets every parameter differ before and after the financial crisis; Section 4 reports these estimates.[8]

Sector definitions and groupings. Our sectoral definitions follow the Financial Accounts. Twenty-seven sectors hold Treasury securities during our sample period. We aggregate these sectors to 11 categories for interpretability, grouping similar investor types. For example, holdings of government pension funds and private pension funds are aggregated to a single "pension fund" category. The complete list of sectors and their mapping to the aggregated categories is provided in Appendix Table B.1. 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.

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 size weight $S_{i,t}$ times the total market values of

[8] Before 2009, the Fed's outright purchases of Treasury bonds were small in scale relative to quantitative easing (QE), mostly associated with reserve-management operations. We treat the Fed's holdings before 2009 as fully anticipated through the scheduled-purchase control, so that its unanticipated demand before QE is zero by construction. We estimate the Fed's price elasticity and factor loadings from 2009 onward. Appendix B.1.2 gives the details.

[9] 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) suggest that domestic and foreign hedge fund positions can account for more than 70 percent of household sector holdings of Treasury securities in recent years.

Treasury bonds outstanding in the last quarter A_{t-1}^{total}:

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

As such, the sign of $\Delta q_{i,t}$ follows that of the dollar flows $F_{i,t}$, with positive numbers indicating net buying and negative numbers indicating net selling. Supply enters the system as negative demand, so that a positive net issuance over a quarter is registered as negative $\Delta q_{i,t}$, allowing us to jointly model supply and demand in this framework.

For most sectors, the size weight $S_{i,t}$ is defined following Gabaix, Koijen, Mainardi, Oh, and Yogo (2026):

$$S_{i,t} = \frac{|A_{i,t-1} + F_{i,t}/2|}{A_{t-1}^{total}}.$$

where $A_{i,t-1}$ denotes the market value of sector i's Treasury holdings as of time $t-1$. Defining $S_{i,t}$ in this way leads to

$$\Delta q_{i,t} = \frac{F_{i,t}}{|A_{i,t-1} + F_{i,t}/2|}.$$

This measure ensures robustness of the flow against small values of $A_{i,t-1}$. When prices do not change from quarter to quarter, $F_{i,t} = \Delta A_{i,t}$, and $\Delta q_{i,t}$ falls back to the canonical Davis and Haltiwanger (1992) growth rate.

Three sectors depart from this definition. The supply sector's size weight is set to one, so that net issuance is measured relative to the size of the market.[10] Households, a sector with a large hedge-fund component, and dealers behave in part as arbitrageurs in the Treasury market. Both sectors hold long and short Treasury positions and move between being net long and net short, so their net positions are not a meaningful measure of their size, and we fix their size weights at one as well. With this normalization, their estimated elasticities measure their flows relative to the total market rather than to their own holdings, and are therefore not comparable one for one with those of other sectors. Their contributions to the aggregate elasticity, however, are the product of size and elasticity, which the normalization leaves unchanged; these contributions keep their interpretation and remain comparable with those of other sectors.

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

[10]The optimal GIV estimator to be introduced in Section 3.3 does not require the size weights to sum to one for it to be efficient, as one can always apply rescaling.

we will outline later, our identification strategy relies on properly controlling for common factors to isolate idiosyncratic demand shocks.

We incorporate three broad sets of observable factors in the estimation, guided by both economic theory and their empirical explanatory power for yields. Details on the construction of these variables are relegated to Appendix B.1. The first set collects key macro-financial factors: (i) innovations to the implied volatility of the U.S. stock market, which serves as a primary indicator of global risk sentiment; (ii) quarterly revisions in futures-implied expectations of the next quarter's Fed funds rate, to account for expected short-rate movements; (iii) quarterly AR(1) innovations to the USD broad index; and (iv) quarterly AR(1) innovations to realized CPI inflation.[11] We do not include other asset prices that could be endogenously affected by idiosyncratic shocks in the Treasury market, such as observed credit spreads.

In addition, we account for policy influences on Treasury demand. Two key policy entities, the Federal Reserve 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.

The final set of observable factors covers subjective expectations of key bond market participants. We extract unobserved factors from granular survey data. Specifically, we include four survey-based factors extracted from individual forecasts of the current-quarter Aaa corporate bond yield relative to the preceding quarter's realized yield, from the Survey of Professional Forecasters (SPF) panel data. Overall, the set of observable demand shifters explains more than 40% of variation of Treasury price changes over our sample period.[13]

[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.1.2 discusses the construction of the scheduled purchase series.

[13] See Figure B.1 in Appendix B.1.2, which also illustrates the contribution of each factor to the fit by progressively adding macro-financial factors and reporting the R^2.

3.2 Identification Assumption

The main challenge in estimating the model is the endogeneity of Treasury prices to investors' allocation decisions. The standard solution is to use idiosyncratic demand shifters from one sector to instrument the price in another sector's demand equation, e.g., mutual-fund flows used to estimate stock and corporate bond elasticities (Lou, 2012; Chaudhary et al., 2026a). Such instruments are valid if the two sectors' idiosyncratic shifters are orthogonal, and relevant if the shocks are large enough to move prices. Our setting needs such an instrument for every sector and every period, and typically used instruments only cover particular sectors or short episodes. Moreover, in a market as large as Treasuries, a single sector's shocks are often too small to move prices, so the instruments lack relevance.

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 the identifying assumption of the GIV approach (Gabaix & Koijen, 2024), and our estimator extends the standard GIV estimator with important differences.[14]

Strict independence of the idiosyncratic shocks in Assumption 1 requires that we capture all common variation in the data; despite our extensive controls, some residual

[14]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. Appendix A.1 discusses the differences in more detail.

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 bias-reduction 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 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 focus on the case without common factors. 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. We also maintain the assumption for illustrative purposes that the elasticities are fully heterogeneous across sectors and constant across regimes. We hold the size weight $S_{i,t}$ constant to economize on notation. Appendix A.1 states the properties of our estimator in a fully general demand system with time-varying elasticities, factor loadings and sector sizes. We show how to incorporate external common factors in Appendix A.2.

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 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 (Newey & Windmeijer, 2009; Han & Phillips, 2006).

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 system of sample moment equations:

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.

Theorem 1, a special case of Theorem A.1, shows 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

$$\mathbf{V}^{\zeta}=\zeta_S^2\times\mathrm{Inv}\left(\begin{bmatrix}\frac{1}{\sigma_1^2}\sum_{i\neq1}S_i^2\sigma_i^2 & \cdots & S_1S_N\\ \vdots & \ddots & \vdots\\ S_1S_N & \cdots & \frac{1}{\sigma_N^2}\sum_{i\neq N}S_i^2\sigma_i^2\end{bmatrix}\right), \tag{3.4}$$

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

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 stan-

dard 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.

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:

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

This is an IV estimator in which price changes are instrumented with the 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. Appendix A.1 shows that the estimator can equivalently be written as an OLS regression of quantities on prices plus a correction for the bias from the sector's own price impact, a form that also gives an alternative, and sometimes more stable, algorithm.

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. If the observable demand shifters do not account for all common demand movements, our estimates will be biased. In Appendix A.3, 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:

$$\hat{\zeta}_S^b \approx \zeta_S \sqrt{1 - \mathcal{R}^2} \tag{3.5}$$

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.[15] The higher the share of common comovement not controlled for, the stronger the downward bias in the estimated aggregate elasticity relative to the truth.

Based on Equation (3.5), 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 the factors extracted by principal component analysis (PCA), such a high level of unexplained common comovement seems unlikely.

To test this intuition directly, we develop an alternative set of moment conditions that take into account the factor structure for residuals (Appendix B.2). We discuss identifiability, and develop a method to jointly estimate "internal PCA" from holdings data along with other parameters of the model. Internally extracted factors provide little explanatory power for Treasury price changes beyond what observable controls provide: projecting price changes on one to three internal principal components gives an R^2 of at most 2.8%, with an adjusted R^2 of 0.9%, so the bias from residual comovement with prices is small by (3.5). Our key takeaways also remain unchanged when the internal components replace the survey-based factors in estimation (Appendix B.2, Figure B.4).

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

[15] 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.

the sensitivity of our results to particular sectors. The estimates are robust to excluding any single sector: the time-average aggregate elasticity ranges from 0.97 to 1.04 across the leave-one-out fits, against 1.01 in the full fit, and its decline across 2009Q1 keeps its sign and size in every one of them (Appendix B.2, Figures B.2 and B.3). Strategic interactions are therefore not a significant concern in our setting.[16]

Appendix B.2 collects these checks and two further diagnostics. The estimated idiosyncratic shocks $\hat{u}_{i,t}$ are close to uncorrelated across sectors (Figure B.5). China's extracted demand shocks line up with narrative evidence on its reserve management (Figure B.6). The aggregate shock $\hat{u}_{S,t}$ moves yields across the whole term structure in line with a duration-based extrapolation of the price impact on the market portfolio (Figure B.7), so our demand shocks mostly shift the level of the yield curve, consistent with our focus on the overall level of yields.

4 Shifting patterns of Treasury market liquidity provision

A deep and liquid Treasury market can accommodate large shifts in investor demand without large price impact. We measure this capacity by the market's aggregate price elasticity ζ_S, the size-weighted sum of investors' price elasticities, and by its reciprocal, the macro multiplier $M = 1/\zeta_S$: a demand shift equal to 1% of the market moves prices by M percent. This section documents how this capacity has evolved since 1986 and which investors provide it, against the backdrop of shifting investor composition (Figure 1).

4.1 The market's liquidity has declined moderately

Table 1 reports the aggregate price elasticity of the Treasury market before and after the global financial crisis, in panel (a), and the price elasticities of the sectors that make it up, in panel (b). The crisis brought profound structural changes across many sectors of the Treasury market. Among the most significant was the Federal Reserve's emergence as an active participant: it began large-scale purchases of Treasuries in March 2009, so we date the regime shift to 2009Q1. Every sector's elasticity, factor loadings and intercept differ before and after 2009Q1, so the two columns are two separate fits of the demand system on either side of the financial crisis. Before the crisis, a positive 1% demand shock raised Treasury prices by 0.88%, or equivalently, lowered yields by 16.0 basis points; since the

[16] Appendix A.1 points to further discussion of 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.

crisis, the same shock raises prices by 1.22% and lowers yields by 22.1 basis points. Compared with estimates for other asset classes in the literature, macro multipliers of 0.88 to 1.22 suggest that the U.S. Treasury market absorbs demand shocks far more effectively than 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., 2026a). 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., 2026a).

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 measures. 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 to obtain useful benchmarks. Appendix B.3.1 shows that our Treasury macro-multiplier estimates align with the magnitudes implied by existing price-impact studies.

The aggregate elasticity thus fell significantly across the financial crisis, from 1.14 to 0.82. The decline is moderate, however, given how completely the investor base turned over during the same period (Figure 1). The next section shows why: private investors withdrew from liquidity provision, and the Federal Reserve offset much of what they withdrew.

4.2 Private investors have retreated and the Fed has filled the gap

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. In panel (b) of Table 1, we set out to explain the moderate shift in aggregate elasticity by examining individual sectors. Figure 2 further quantifies each sector's contribution to the aggregate shift. The pre-2009 column establishes that there is substantial heterogeneity in the liquidity-provision capacity at the sector level, with banks by far the most elastic sector. ETFs, which enter the sample as a separate sector in 2009, have a negative point estimate, potentially reflecting return-chasing behavior by retail end-investors; due to their small market share, they contribute very little to the aggregate elasticity. Meanwhile, on the supply side, we estimate an elasticity close to zero. This result reflects the Treasury's commitment to a "regular and predictable" debt management approach. As we control

Table 1: **Price elasticity of the Treasury market and its sectors before and after 2009**

	(1) Pre-2009	(2) Post-2009	(2)−(1)
(a) Market aggregate			
Aggregate elasticity ζ_S	1.14***	0.82***	−0.32***
	[0.95, 1.33]	[0.69, 0.96]	[−0.55, −0.08]
Macro multiplier $M = 1/\zeta_S$	0.88***	1.22***	0.34***
	[0.73, 1.02]	[1.02, 1.42]	[0.09, 0.58]
Yield move from a 1% sale (bp)	16.0***	22.1***	6.1***
	[13.3, 18.6]	[18.5, 25.7]	[1.6, 10.6]
(b) Price elasticity by sector			
Households†	0.41***	0.25***	−0.16
	[0.23, 0.60]	[0.14, 0.36]	[−0.38, 0.05]
Rest of world	0.52***	0.09	−0.43***
	[0.32, 0.72]	[−0.06, 0.24]	[−0.68, −0.18]
Banks	1.69***	1.12***	−0.57**
	[1.32, 2.05]	[0.86, 1.37]	[−1.02, −0.12]
Dealers†	0.09*	0.03	−0.06
	[0.00, 0.18]	[−0.04, 0.09]	[−0.17, 0.05]
Fed		1.21***	
		[0.98, 1.43]	
Pension funds	0.54***	0.21**	−0.33**
	[0.28, 0.80]	[0.03, 0.39]	[−0.64, −0.01]
Other	−0.01	0.54***	0.55***
	[−0.28, 0.27]	[0.35, 0.73]	[0.21, 0.88]
Mutual funds	0.69***	0.43***	−0.25
	[0.23, 1.14]	[0.11, 0.76]	[−0.81, 0.31]
Insurance	0.47***	0.41***	−0.06
	[0.21, 0.73]	[0.23, 0.59]	[−0.37, 0.26]
ETFs		−0.20	
		[−0.54, 0.13]	
Supply†	0.13***	0.08***	−0.06
	[0.05, 0.21]	[0.02, 0.13]	[−0.15, 0.04]

Note: This table reports estimates from the specification in which every sector's price elasticity, factor loadings, and intercept differ before and after 2009Q1. Panel (a) reports the aggregate price elasticity $\zeta_S = \sum_i S_i \zeta_i$, the macro multiplier $M = 1/\zeta_S$, and the yield change implied by a demand shift of 1% of the market at a 5.5-year duration, around the sample average. Panel (b) reports the sector price elasticities ζ_i. Sectors marked † have unit size weights, so their coefficients measure demand responses relative to the total market; the other sectors' coefficients are relative to their own holdings. The Federal Reserve's pre-2009 purchases are treated as fully anticipated through the scheduled-purchase control, so its pre-2009 coefficients are not estimated and are left blank. ETFs enter the sample in 2009Q1. The sample is 1986Q1–2023Q4. 95% confidence intervals, from equation (A.5), are in brackets below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

for the preannounced supply schedules, there is little room for the supply to further respond to price changes.

The change column of panel (b) shows a broad-based decline in price elasticities across sectors. Among the sectors with statistically significant changes, banks' price elasticity fell by 34%, and pension funds' price elasticity fell by 61%. For banks, leverage requirements make expanding Treasury positions costly in terms of balance-sheet capacity and may limit their willingness to absorb additional securities when prices fall (Cochran, Infante, Petrasek, Saravay, & Tian, 2023). Liquidity regulation may also make the adjustment of Treasury holdings less flexible by tying Treasury demand to required liquidity buffers (Stein, 2013). For pension funds, the shift from defined-benefit to defined-contribution plans, greater delegation to investment funds and the associated changes in investment mandates may have reduced their role as direct, discretionary Treasury buyers (Ding, Fang, Hardy, & Lewis, 2026; Andonov, Jansen, & Rauh, 2026).

The main sector whose demand becomes more elastic is the Federal Reserve. Before the financial crisis, changes in the Fed's Treasury notes and bonds holdings are typically associated with its routine operational needs. We thus assume that all changes in the holdings in this period are perfectly anticipated and control for the Fed's scheduled purchases throughout, so that its Treasury demand would be perfectly inelastic.[17] After the crisis, however, as the Fed initiates large-scale asset purchases, its demand becomes more elastic, with an estimated elasticity of 1.21 (Table 1, panel (b)). Our analysis reveals the Fed is price elastic, purchasing additional Treasury bonds when the Treasury market experiences downward price pressures. This behavior aligns with the goal of Treasury market stabilization—one of the stated objectives of QE operations (Gagnon et al., 2011a).[18]

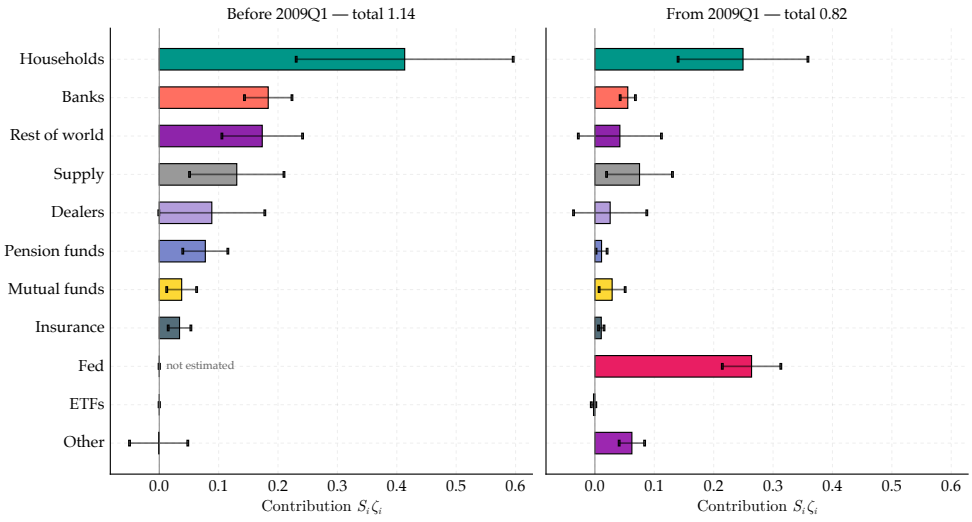
In Figure 2, Panel (a), we report the contributions of different sectors to the aggregate elasticity before and after 2009 by multiplying their respective size weights with the sector-level elasticities. Non-Fed investors' contributions to the aggregate elasticity fell by 0.52 across 2009Q1, led by households (-0.16), the rest of the world (-0.13), banks (-0.13), pension funds (-0.07), and dealers (-0.06); Treasury supply contributed a further -0.06. The household sector, whose holdings are mostly driven by hedge funds, remains the one with the largest relative contribution other than the Fed, reflecting its role as a nimble liquidity provider attributable to a looser regulatory constraint. The Fed-

[17] 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.1.2 for more details.

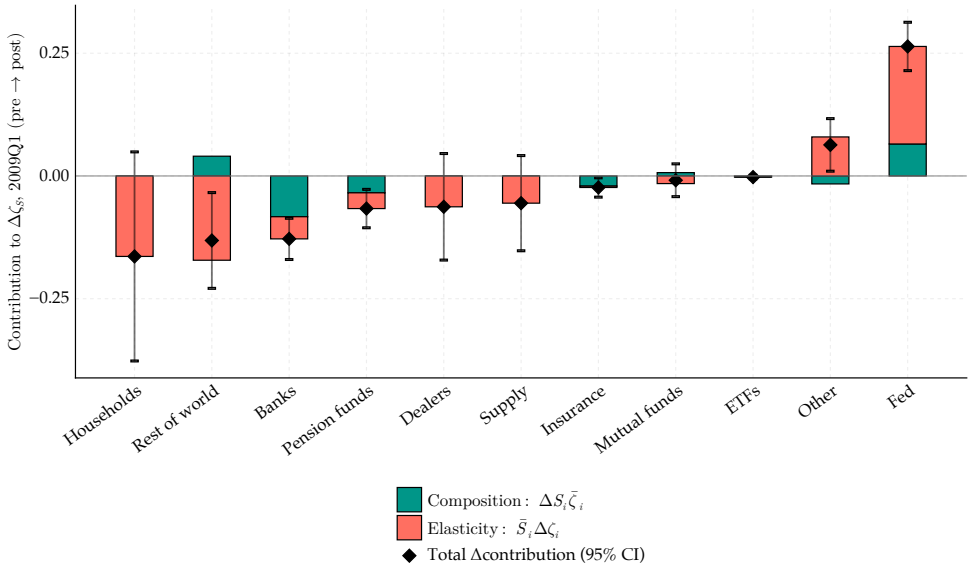
[18] The phrase "the smooth functioning of the Treasury market" appears frequently in Federal Open Market Committee minutes. For example, a 2020 New York Fed announcement stated: "The Desk stands ready to adjust the size and composition of its purchase operations as appropriate to support the smooth functioning of the Treasury market." Another contemporaneous announcement specified that "the Desk will conduct purchase...subject to reasonable prices."

Figure 2: **Contributions to aggregate elasticity before and after 2009**

(a) Contribution levels



(b) Change across 2009Q1: composition and elasticity



Note: This figure decomposes the aggregate elasticity of Table 1 into sector contributions. Panel (a) plots each sector's contribution $S_i\zeta_i$ to the aggregate elasticity, with 95% confidence intervals from equation (A.5); each regime's bars sum to ζ_S. Panel (b) splits the change in each contribution into a composition effect, $\Delta S_i\,\bar{\zeta}_i$, and an elasticity effect at the average size weight, $\bar{S}_i\,\Delta\zeta_i$. The two stacked components sum to the total change, marked by the diamond with its 95% confidence interval.

eral Reserve, on the other hand, becomes the biggest positive contributor to aggregate elasticity, offsetting about half of the negative impact from other investors and supplying around a third of the post-2009 aggregate elasticity. Holding other investors' behavior

and size weights fixed, removing the Fed's post-crisis contribution would increase the price impact of a sale equal to 1% of the market from 1.22% to about 1.79%, or from 22 to about 33 basis points in yields—roughly twice its pre-crisis level.

In Panel (b), we split the changes in the sectors' contributions into components driven by changes in their size weights that capture composition effect, and changes in their elasticities. Among the sectors showing significant elasticity decline, banks and pension funds experience a decline in both size and elasticity, while foreign investors' substantial negative contribution mainly comes from their demand becoming more price inelastic.

Taken together, the Treasury market's liquidity now rests partly on the central bank rather than on private and foreign investors. If the retreat of these investors is permanent, the market's ability to accommodate investor demand shifts could decline further during future policy tightening episodes when the Fed reduces its Treasury holdings.

4.3 Foreign holdings no longer translate into liquidity provision

Foreign investors are the clearest case of the retreat documented in Section 4.2. They held nearly two-thirds of the market in 2008 and still about a third in 2023, yet their contribution to the aggregate elasticity fell by 0.13 across 2009Q1, almost entirely because their demand became less price-sensitive rather than because their holdings shrank (Figure 2, panel (b)). On average, foreign investors' price elasticity declines from 0.52 to 0.09, with the latter not significantly different from zero (Table 1, panel (b), reproduced as panel (a) of Table 2). In order to understand the underlying drivers of this change, we keep the 2009 regime shift and allow the price elasticity within the foreign sector to vary across major country groups. We split countries into five groups: China, Japan, Europe, Americas, and a residual "other countries" category.[19]

Panel (b) of Table 2 reports region-specific elasticity shifts. Overall, allowing heterogeneous price elasticities across country groups barely changes the aggregate market elasticity estimates. Japan, China, and Europe all show significant elasticity declines. Their post-2009 price elasticities become statistically indistinguishable from zero. China's estimated price elasticity falls by approximately 82%, from 2.71 to 0.50, while Japan's falls from 0.59 to -0.05 and Europe's from 0.48 to -0.07. Overall, foreign investors now play a substantially smaller role in absorbing demand shocks in the Treasury market. Potential

[19]The Europe (Denmark, Finland, Germany, the Netherlands, Spain, Sweden, Switzerland, and the United Kingdom) and Americas (Canada, Brazil, Chile, and Colombia) groups contain countries that consistently appear in TIC data with substantial positions. Other countries in these continents are grouped in the residual category.

explanations of this retreat include a lower precautionary demand overall, more diversified use of foreign reserves, as well as geoeconomic tensions.

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

	1986Q1–2008Q4			2009Q1–2023Q4			Change	
Investor	ζ	Contrib.	Share (%)	ζ	Contrib.	Share (%)	$\Delta\zeta$	ΔContrib.
(a) Common foreign elasticity								
Rest of world	0.52^{***}	0.17^{***}	15.24	0.09	0.04	5.11	-0.43^{***}	-0.13^{***}
	[0.32, 0.72]	[0.11, 0.24]		[−0.06, 0.24]	[−0.03, 0.11]		[−0.68, −0.18]	[−0.23, −0.03]
Market aggregate	1.14^{***}			0.82^{***}			-0.32^{***}	
	[0.95, 1.33]			[0.69, 0.96]			[−0.55, −0.08]	
(b) Regional elasticities								
China	2.71^{***}	0.10^{***}	8.39	0.50	0.05	6.46	-2.21^{*}	−0.04
	[1.04, 4.37]	[0.04, 0.16]		[−1.15, 2.14]	[−0.12, 0.23]		[−4.55, 0.13]	[−0.23, 0.14]
Japan	0.59^{***}	0.06^{***}	4.98	−0.05	0.00	−0.51	-0.64^{**}	-0.06^{***}
	[0.20, 0.99]	[0.02, 0.10]		[−0.33, 0.24]	[−0.03, 0.02]		[−1.12, −0.15]	[−0.11, −0.01]
Europe	0.48^{**}	0.03^{**}	2.71	−0.07	0.00	−0.43	-0.55^{**}	-0.03^{**}
	[0.07, 0.89]	[0.00, 0.06]		[−0.37, 0.23]	[−0.02, 0.01]		[−1.06, −0.04]	[−0.07, 0.00]
Americas	0.46	0.01	0.48	0.33	0.01	1.23	−0.14	0.00
	[−0.42, 1.35]	[−0.01, 0.02]		[−0.31, 0.97]	[−0.01, 0.03]		[−1.23, 0.95]	[−0.02, 0.03]
Other countries	0.37^{**}	0.05^{**}	3.98	0.24^{*}	0.04^{*}	5.36	−0.13	0.00
	[0.02, 0.71]	[0.00, 0.09]		[−0.02, 0.50]	[0.00, 0.09]		[−0.56, 0.30]	[−0.07, 0.06]
Rest of world	0.71^{***}	0.24^{***}	20.55	0.21	0.10	12.11	-0.49^{**}	−0.14
	[0.45, 0.97]	[0.15, 0.32]		[−0.18, 0.61]	[−0.08, 0.28]		[−0.96, −0.02]	[−0.34, 0.07]
Market aggregate	1.15^{***}			0.83^{***}			-0.33^{**}	
	[0.96, 1.35]			[0.66, 1.00]			[−0.58, −0.07]	

Note: This table reports foreign price elasticities ζ_i from the specification of Table 1. Panel (a) reproduces the foreign sector's elasticity from Table 1; panel (b) re-estimates the same specification with separate elasticities for five country groups: China, Japan, Europe, Americas, and Other countries. Contrib. is the contribution $S_i\zeta_i$, averaged over the regime's quarters, and Share is the contribution divided by the aggregate elasticity ζ_S. In panel (b), Rest of world sums the five groups' contributions and reports their size-weighted elasticity. Market aggregate is ζ_S. Changes are post-2009 minus pre-2009. The sample is 1986Q1–2023Q4. 95% confidence intervals, from equation (A.5), are in brackets below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Foreign demand matters less for yields since the crisis. We quantify the contribution of each country to Treasury yield changes over time using the regional specification in Table 2, panel (b). By equation (2.4), group i contributes

$$\Delta p_{i,t} = \frac{1}{\zeta_{S,t}} S_{i,t} \left(\overline{\Delta q}_i + \boldsymbol{\lambda}_{i,t} \boldsymbol{\eta}_t + u_{i,t} \right)$$

to the price change: its demand shift at constant prices, scaled by the macro multiplier. The contributions sum to the observed price change. We convert them to yield changes at the lagged Treasury duration, so a negative contribution means the group's demand lowered yields. Figure 3 shows foreign countries' contribution to Treasury yield changes,

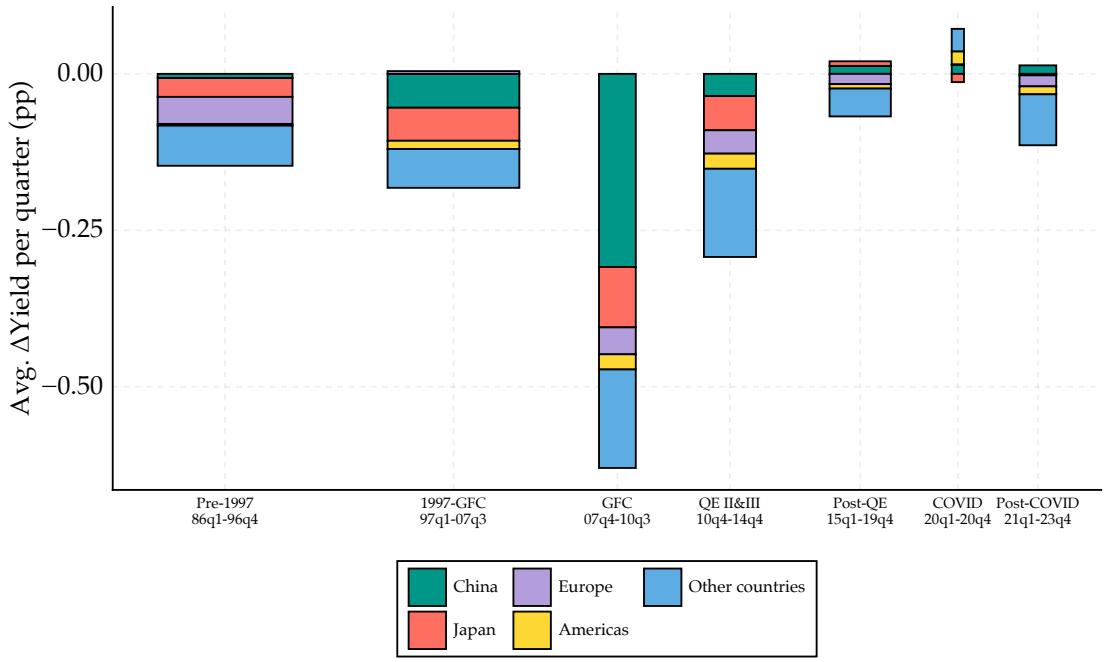
and Appendix Figure B.9 presents all investor groups. Foreign investors were the dominant driver of Treasury yields before the global financial crisis, contributing an average yield reduction of 28.9 basis points per quarter over 2003Q4–2007Q3. Their contribution peaked during the crisis, at 63.0 basis points per quarter over 2007Q4–2010Q3, and has waned since. Between 2003Q4 and 2007Q3, China, Europe, and Japan accounted for 56% of the foreign sector's impact on yields and lowered them by 0.65 percentage points per year; China alone depressed yields by a cumulative 1.4 percentage points. Chinese demand still lowered yields by about 31 basis points per quarter during the crisis (2007Q4–2010Q3), but its contribution all but disappeared after 2015, and foreign investors' gross contribution over 2020 was a 23 basis point increase in yields.[20] These estimates speak to the global savings glut debate: current-account-surplus countries channeled nearly 3 trillion USD into Treasuries during the 2000s, and the associated compression of long-term yields figured in debates surrounding Greenspan's "conundrum" and Bernanke's "savings glut" hypothesis (Greenspan, 2005; Bernanke, 2005; Caballero, 2006; Mendoza, Quadrini, & Ríos-Rull, 2009). Our decomposition shows that this compression was concentrated in the first decade of the 2000s and has not recurred.

The shrinking market share and diminished price impact of foreign investors also limit the disruption that unilateral foreign selling could cause today, say by a major holder such as China. Two forces have reduced China's footprint in the market: (i) rapid Treasury issuance since the crisis has lowered every foreign holder's share of the market; and (ii) China itself has withdrawn, cutting its holdings by about 40%, from $1.31 trillion in November 2013 to $0.80 trillion in December 2023. Consequently, China's share of Treasury notes and bonds fell from a 2010Q3 peak of 17.8% to 3.9% at the end of 2023. To quantify what this means for the most discussed scenario—a full liquidation of China's Treasury holdings within a single quarter[21]—we combine the regional specification's macro multiplier of 1.29 at end-2023 with China's 3.9% share of notes and bonds and a duration of 6.47 years: such a liquidation would raise Treasury yields by approximately 77 basis points, about 1.60 times the 48-basis-point standard deviation of quarterly yield changes over 1986Q1–2023Q4. The same liquidation at the end of 2008, when China held 17.1% of the market and the multiplier stood at 0.72, would have raised yields by 2.2 percentage points—a significant shock at the height of the crisis.

[20] China's reserve sales in 2015, prompted by depreciation pressure after the end of QE, had minimal price impact compared with its pre-crisis purchases.

[21] Reportedly proposed but rejected during the global financial crisis, the scenario has been characterized as China's economic "nuclear option" in trade disputes (Sorkin, 2018; Paulson, 2013).

Figure 3: **Important foreign holders' contributions to yields, percentage points per quarter**



Note: This figure plots the average quarterly contribution to Treasury yield changes, in percentage points, of the five foreign country groups in Table 2, panel (b), averaged within seven episodes; bar widths reflect episode lengths, and GFC denotes the global financial crisis. A group's contribution is $\Delta p_{i,t}$ in equation (2.4), converted to a yield change at the lagged Treasury duration; negative values mean the group's demand lowered yields.

5 Investors and the evolving safe-haven property

Alongside liquidity, the safe-haven property underpins the central role of the Treasury market: Treasuries tend to appreciate when global risk rises, most visibly during the global financial crisis, and this insurance is part of what makes them the natural home for the world's reserves and precautionary balances. Identifying the investors behind this response helps discipline models in which countercyclical dollar or Treasury demand shapes international risk sharing (Kekre & Lenel, 2024; Devereux et al., 2023). Linking the property to financial intermediation and market structure also speaks to a more practical question; namely, whether changes in the investor base can strengthen or erode the ability of Treasuries to insure global investors in future periods of stress.

We measure the safe-haven property as the response of Treasury prices to an innovation in uncertainty. The uncertainty measure is given by the implied U.S. stock-market volatility (VIX), extended back to 1986 using the VXO index.[22] By the pricing equation

[22]See Appendix B.1.2 for more detail. We directly use innovations in implied stock-market volatility as

(2.3), this response, β_t^{unc}, is the aggregate uncertainty loading divided by the aggregate elasticity:

$$\beta_t^{unc} = \frac{\lambda_{S,t}^{unc}}{\zeta_{S,t}} = \frac{1}{\zeta_{S,t}} \sum_i S_{i,t} \lambda_{i,t}^{unc}.$$

Investor i therefore contributes $S_{i,t}\lambda_{i,t}^{unc}$ to the aggregate demand loading and $S_{i,t}\lambda_{i,t}^{unc}/\zeta_{S,t}$ to the price response. Our estimates identify which investors' demand rises when uncertainty jumps, which investors' demand falls, and how these responses have changed over time.

To capture each investor's full demand response, we report *total* factor loadings. The total loadings include the influence that runs through responses of the other macro-financial factors to innovations to uncertainty. Increases in uncertainty may strengthen expectations of monetary policy easing, which in turn raises investors' demand. Because we control for short-rate expectations, the *direct* loading—the coefficient on uncertainty itself, holding the other factors fixed—excludes this channel, whereas the total loading includes it. In Appendix B.3.3, we formally define both loadings, and show that our takeaway remains robust when we use direct loadings to measure investors' response to uncertainty.

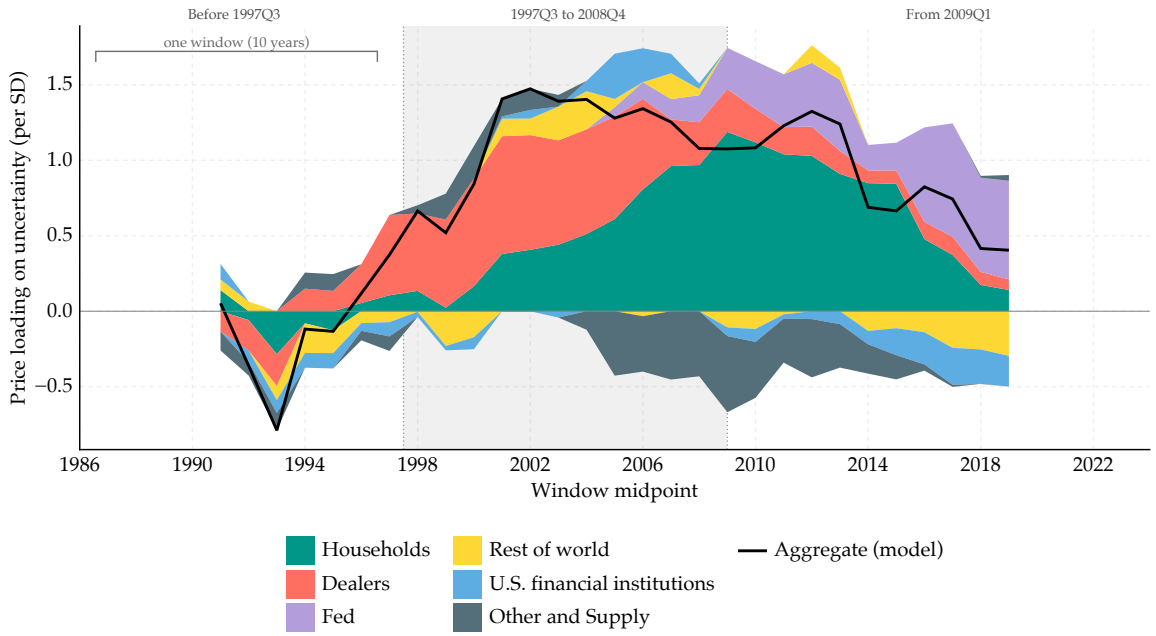
We first trace the safe-haven property over time by re-estimating the demand system in ten-year rolling windows. Figure 4 stacks the resulting contributions by investor group, with each window plotted at its midpoint; the black line is the model's aggregate price loading, the sum of the plotted contributions. It closely tracks the ten-year rolling regression of Treasury price changes on the uncertainty innovation (Appendix Figure B.10). The three regimes stand out. In windows centered before 1997, the price loading is close to zero or negative: Treasury prices did not rise when uncertainty jumped. It climbs through the late 1990s, peaks at about 1.5 in windows centered around 2002, and stays between 1.1 and 1.4 in the windows that span the financial crisis. It then declines once the windows are centered after 2013, and it ends the sample at 0.4, less than a third of its peak.

To test this profile formally, we split the sample into three regimes, 1986Q1–1997Q2, 1997Q3–2008Q4 and 2009Q1–2023Q4.[23] We re-estimate the demand system with every sector's price elasticity, uncertainty loading, and intercept allowed to differ across the

a barometer of global risk sentiment. Another related measure is stock–bond correlation, but it reflects a mixture of growth, monetary policy, and financial shocks (Cieslak & Pang, 2021; Antolín-Díaz, 2026).

[23] We place the first regime shift at 1997Q3, near the rise in the rolling price loading on uncertainty, and the second at 2009Q1, the same regime shift used in the liquidity analysis of Section 4. Appendix B.3.3 assesses the robustness of this timing choice using price regressions and demand-system estimates with alternative dates for the first shift, holding the second fixed at 2009Q1.

Figure 4: **Ten-year rolling contributions to the safe haven, by investor group**



Note: This figure plots 10-year rolling-window estimates of each group's contribution to the price loading on uncertainty, $\sum_i S_{i,t}\lambda_{i,t}^{unc}/\zeta_{S,t}$ over the group's sectors, using total loadings. U.S. financial institutions are banks, insurers, pension funds, mutual funds, and ETFs; Other and Supply are the Other sector and Treasury supply. For numerical stability, the household and dealer elasticities are fixed at their estimates from a single-regime specification, and all of the Federal Reserve's coefficients at their full-sample estimates before and after 2009Q1. The solid black line is the model aggregate, the sum of the contributions. Windows are plotted at their midpoints; dotted vertical lines mark the regime shifts at 1997Q3 and 2009Q1, and the shaded band marks windows whose midpoint falls between them.

three regimes, while the other factor loadings remain pooled. Table 3 reports each group's loading and its contribution to the aggregate.

Together, the rolling profile and the regime table establish three facts, which the rest of this section documents in turn: (i) the safe haven emerged in the late 1990s and has faded; (ii) foreign investors did not produce it; and (iii) dealers drove both its rise and its decline.

5.1 The safe haven emerged in the late 1990s and has faded since 2012

The aggregate row of Table 3 summarizes the three regimes. Before 1997Q3, the aggregate demand response to a one-standard-deviation rise in uncertainty is -0.35 and not significantly different from zero; over 1997Q3–2008Q4 it is 2.14, significant at the 1% level; and since 2009 it is 0.43, again indistinguishable from zero. Dividing by the regime's aggregate elasticity converts these demand responses into price responses: in the middle regime, a

Table 3: **Uncertainty loadings and contributions across three regimes**

	1986Q1–1997Q2		1997Q3–2008Q4		2009Q1–2023Q4	
	Loading	Contribution	Loading	Contribution	Loading	Contribution
Aggregate	−0.35	—	2.14***	—	0.43	—
	[−1.77, 1.07]		[1.02, 3.26]		[−0.70, 1.55]	
(a) Contributions to the aggregate						
Rest of world	0.28	0.06	0.10	0.04	−0.80	−0.37
	[−1.09, 1.64]	[−0.22, 0.34]	[−1.22, 1.42]	[−0.57, 0.66]	[−1.77, 0.17]	[−0.82, 0.08]
Households†	−0.02	−0.02	0.81**	0.81**	0.40**	0.40**
	[−0.63, 0.58]	[−0.63, 0.58]	[0.16, 1.47]	[0.16, 1.47]	[0.06, 0.74]	[0.06, 0.74]
Dealers†	−0.10	−0.10	1.31***	1.31***	0.03	0.03
	[−0.50, 0.31]	[−0.50, 0.31]	[0.88, 1.74]	[0.88, 1.74]	[−0.22, 0.29]	[−0.22, 0.29]
Remaining domestic	−0.20	−0.14	−0.20	−0.09	−0.42	−0.14
	[−1.33, 0.92]	[−0.91, 0.63]	[−1.17, 0.77]	[−0.50, 0.33]	[−1.07, 0.22]	[−0.34, 0.07]
Supply†	−0.15	−0.15	0.06	0.06	0.02	0.02
	[−0.48, 0.18]	[−0.48, 0.18]	[−0.30, 0.43]	[−0.30, 0.43]	[−0.19, 0.23]	[−0.19, 0.23]
Fed					2.18	0.48
					[−0.61, 4.97]	[−0.13, 1.09]
(b) Rest of world vs. domestic private						
Domestic private	−0.35	−0.26	3.94***	2.03***	0.80	0.30
	[−2.05, 1.36]	[−1.54, 1.02]	[2.20, 5.68]	[1.13, 2.93]	[−0.56, 2.17]	[−0.21, 0.81]
Difference	0.62	0.32	−3.84***	−1.99***	−1.60*	−0.67**
	[−1.27, 2.51]	[−0.92, 1.55]	[−6.05, −1.63]	[−3.09, −0.89]	[−3.25, 0.04]	[−1.34, 0.00]

Note: This table reports the total uncertainty loadings λ_i^{unc}, from the specification in which every sector's price elasticity, uncertainty loading, and intercept differ across the three regimes; the other factor loadings are constant. Contribution is $S_i\lambda_i^{unc}$, averaged over the regime's quarters; the contributions in panel (a) sum to the aggregate demand loading λ_S^{unc} in the Aggregate row. Sectors marked † have unit size weights, so their coefficients measure demand responses relative to the total market; the other sectors' coefficients are relative to their own holdings. Remaining domestic combines banks, insurers, pension funds, mutual funds, ETFs, and Other; Domestic private combines households, dealers, and remaining domestic. The Domestic private loading is its contribution normalized by the domestic private sectors' total share of Treasury holdings. The Difference row is the rest of the world minus domestic private. The Federal Reserve's pre-2009 purchases are treated as fully anticipated through the scheduled-purchase control, so its pre-2009 coefficients are not estimated. The sample is 1986Q1–2023Q4. 95% confidence intervals, computed as in Appendix B.3.3, are in brackets below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Appendix Table B.2 reports direct loadings.

one-standard-deviation rise in uncertainty raised Treasury prices by about 1.5%, equivalent to a fall in yields of roughly 27 basis points at an average duration of around 5.5 years, whereas in the other two regimes the price response is statistically zero. Within the post-2008 regime, the rolling estimates in Figure 4 show that the safe haven persisted until about 2012 and then declined; by the end of the sample, the price response to uncertainty is less than a third of its peak. The rise and fall we document echoes the change in the stock–bond correlation and the concerns raised during the COVID-19 crisis about Treasuries' waning safety (Baele, Bekaert, & Inghelbrecht, 2010; Campbell et al., 2020; Antolín-Díaz, 2026; He et al., 2022; Vissing-Jorgensen, 2021).

5.2 Foreign investors did not make Treasuries a safe haven

We next examine which investors have shaped the safe-haven property of Treasuries over time. Foreign investors are a natural candidate emphasized in the literature. The *flight-to-safety view* holds that foreign investors increase their demand for Treasuries during global financial tightening, supporting the countercyclical behavior of Treasury prices (Jiang et al., 2021, 2024). This view is central to models linking the dollar's special status to the United States' external funding advantage. The *precautionary saving view* holds instead that foreign investors accumulate Treasuries in normal times and draw them down to raise liquidity when international financial markets are disrupted (e.g., Eisenbach & Phelan, 2026; Bianchi et al., 2018; Devereux et al., 2023). Under this view, foreign investors benefit from the Treasury's hedging value during global downturns, but they do not lower Treasury yields by raising their demand in those downturns.

Observed patterns of foreign holdings and flows are seemingly consistent with both views. During the global financial crisis, foreign investors' Treasury holdings rose significantly. At the onset of the COVID-19 pandemic, on the other hand, foreign investors sold Treasuries, leading to a new round of debate over whether the special status of Treasuries has waned (He et al., 2022; Weiss, 2022; Vissing-Jorgensen, 2021). As holdings and prices are equilibrium outcomes shaped by the interaction between different investors, however, they do not directly reveal the source of the underlying demand shifts that drive them.

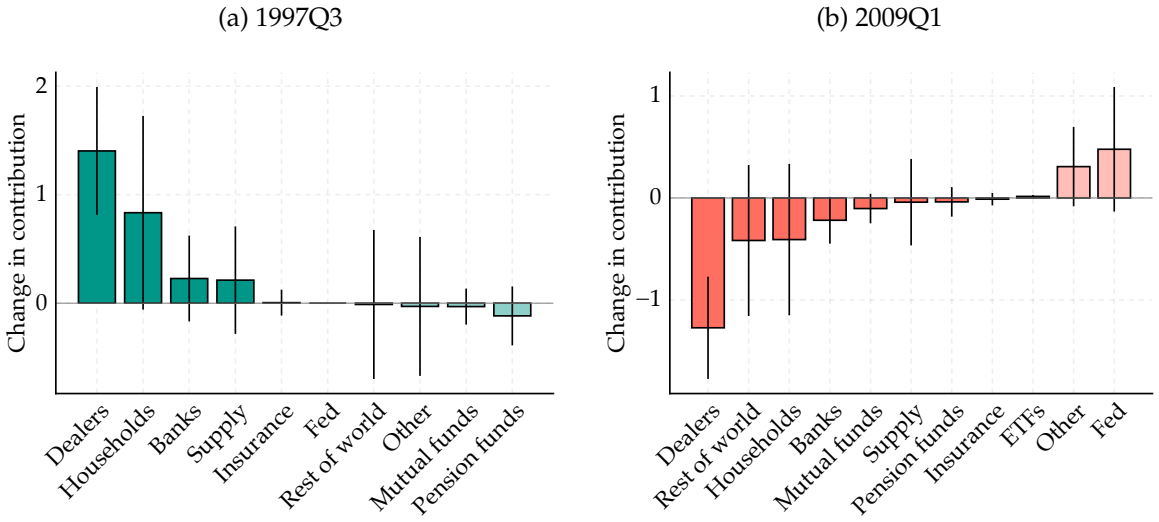
Our approach separates foreign investors' demand response to uncertainty from their response to equilibrium price changes. The estimates support the precautionary saving view. Foreign investors contributed little to the safe-haven property during 1997Q3–2008Q4: their contribution was 0.04, about 2% of the aggregate response of 2.14 (Table 3a). Since 2009, their demand has, if anything, declined as uncertainty rises, making them the largest negative contributor to the safe-haven property in this regime. Compared to domestic investors, the rest of the world's uncertainty loading is 3.84 percentage points below that of domestic private investors in 1997Q3–2008Q4, and 1.60 below since 2009 (Table 3b).

In Appendix Figure B.11, we illustrate the importance of distinguishing observed flows from underlying demand shifts. For both the full sample and the global financial crisis episode, we scale net purchases by foreign investors and by domestic private investors by each group's lagged holdings, then subtract the growth rate of privately held federal debt to account for supply-driven variation in observed flows. On this basis, higher uncertainty is strongly associated with lower foreign flows and higher domestic private flows. Foreign holdings can thus increase while foreign investors absorb propor-

tionately less of the available supply.

Our findings imply that models in which countercyclical demand for U.S. safe assets generates the dollar's exorbitant privilege need another source of countercyclical demand for Treasuries. We next show how domestic participants in the Treasury market such as dealers drive the rise and decline of the safe-haven property.

Figure 5: **Changes in contributions to the safe haven**



Note: This figure plots the change in each sector's contribution to the aggregate total demand loading on uncertainty at 1997Q3 (panel (a)) and 2009Q1 (panel (b)), from the specification of Table 3. Whiskers are 95% confidence intervals, computed as in Table 3.

5.3 Dealers drove both the rise and the decline

Figure 5 plots the change in each sector's contribution at the two regime shifts. Dealers account for the largest changes in the safe-haven property at both shifts. Their demand response to uncertainty rises to 1.31 during 1997Q3–2008Q4, approximately 60% of the aggregate response of 2.14, and is statistically indistinguishable from zero before and after this period. Both the increase in 1997Q3 and the decline in 2009Q1 are statistically significant. The same regime pattern appears in observed dealer flows: dealers' net purchases rise with uncertainty during the middle regime, but show no comparable relationship before or after (Appendix Figure B.12).[24] The household sector, which includes hedge

[24]Primary dealers' cash Treasury positions reported on Form FR 2004 show similar patterns. We estimate a loading on uncertainty at 0.53 over 1998–2008, and obtain estimates close to zero before and after this period. Dealers may hedge cash positions with futures, so a cash-only flow could overstate their response. The historical reports also include futures positions through 2000. Futures changes offset about

funds, exhibits changes in the same direction.

Positive contributions from households and the Federal Reserve sustained the safe-haven property for several years after the decline in dealers' contribution. The household sector contributes 0.40 to the aggregate uncertainty loading after 2009, significant at the 5% level. Still, the rolling estimates in Figure 4 show that its contribution also starts to decline after the early 2010s. After 2009, a one-standard-deviation increase in uncertainty is associated with an estimated increase in Federal Reserve demand equivalent to 0.48% of the Treasury market, although the estimate is imprecise. As the household contribution declines, the aggregate demand loading on uncertainty weakens, with the Federal Reserve accounting for almost all of the remaining response in the final rolling windows.

Because our approach is built to give a bird's-eye view of the whole market, our estimates do not identify why dealers' demand response to uncertainty changed across regimes, but the timing suggests a connection to their balance-sheet capacity. The increase in dealers' demand response to uncertainty in 1997Q3 coincides with the expansion of dealer balance sheets and leveraged relative-value trading in Treasuries, when they had both the capacity and the incentive to absorb Treasuries when risk rose and other investors' demand fell. The decline in dealers' demand response to uncertainty after 2009 coincides with tighter leverage and liquidity regulation, which raised the balance-sheet cost of warehousing Treasuries (Duffie et al., 2023; Cochran et al., 2023; Chaudhary, Fu, & Li, 2026b). This interpretation links the safe-haven property during 1997–2008 to intermediaries' capacity to absorb Treasuries during stress. Subsequent constraints on that capacity may help explain both the decline in the safe-haven property and the reduced private liquidity provision documented in Section 4.[25]

6 Conclusion

This paper asks whether the two properties that underpin the Treasury market's central role, its liquidity and its safe-haven property, are eroding, and which investors are behind the change. Both properties are equilibrium outcomes of investor behavior, so answering the question requires a framework that links the aggregate pricing properties measured by factor regressions to the investors who produce them. To that end, we develop a de-

30% of coupon cash changes, but adding them to cash Treasury positions leaves the estimated response to uncertainty after 1997Q3 nearly unchanged.

[25]Corroborating our finding is Antolín-Díaz (2026), who uses identified flight-to-safety shocks to find that broker-dealers' share of Treasury holdings increases following the tightening, while other sectors do not respond significantly.

mand system in which each sector's Treasury demand responds to prices and to economic factors, and in which market clearing aggregates these responses into the standard price-factor regression. An optimal granular instrumental variable estimator identifies each sector's price elasticity and factor loadings from publicly available holdings data, which lets us follow the market and its investors over four decades.

Applied to 1986 to 2023, the framework yields two conclusions. First, the market's liquidity has declined moderately, despite the significant change in the composition of the investor base. Private investors, in particular foreign investors, have pulled back from providing liquidity since the financial crisis; foreign holdings remain large, but their price elasticity has fallen to near zero, and no private investor has filled the gap. Aggregate liquidity has fallen far less than this retreat implies because the Federal Reserve's post-crisis holdings now account for about a third of the market's aggregate elasticity. Second, the safe-haven property has faded. It emerged only in the late 1990s, was strongest through the crisis, and has since largely faded. Dealers, not foreign investors, drove both the rise and the decline; foreign investors, if anything, sell Treasuries when stress rises, consistent with precautionary saving rather than flight to safety. After dealers stepped back, the household sector and the Federal Reserve's response to stress carried the property for several more years before it faded.

Both conclusions point to the Federal Reserve. Its post-crisis holdings supply about a third of the market's aggregate elasticity, and its purchases in response to stress slowed the decline of the safe-haven property without stopping it. Consequently, properties that private investors once supplied now depend on the size and conduct of the central bank's balance sheet. This dependence matters for balance-sheet normalization. A further reduction in the Fed's holdings removes liquidity that no private investor has shown signs of replacing, and leaves the safe-haven property with a smaller base still. Our estimates characterize these aggregate properties and their sector-level drivers; they do not model the mechanisms, such as regulation, balance-sheet constraints, or reserve management, behind each sector's behavior. Tracing those mechanisms, and asking whether private liquidity provision can be rebuilt, is a natural next step for future work.

Acknowledgments

We thank Philippe van der Beck, William Cassidy, Aditya Chaudhry, Greg Duffee, Zhengyang Jiang, Xiang Fang, Xavier Gabaix, Zhe Geng, Ruggero Jappelli, Iryna Kaminska, Rohan Kekre, Ralph Koijen, Lira Mota, Stefan Nagel, Andreas Neuhierl, Stijn van Nieuwer-

burgh, 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 Asset Pricing Working Group, IMIM Online Seminar, the Chicago Booth Treasury Markets Conference (2025), UCLA, Second UIC Finance Conference, the 2025 OFR Rising Scholar Conference, the 11th International Conference on Sovereign Bond Markets, CICF 2025, 2025 NBER Summer Institute Asset Pricing, EFA 2025, University of Virginia, the Fixed Income and Financial Institutions Conference, the 14th Term Structure Workshop at the Deutsche Bundesbank, the Federal Reserve Bank of New York, the Bank of England Asset Pricing Workshop, the CUHK-RAPS-RCFS Conference, and the Federal Reserve Board for useful interactions. We thank Marco (Erchuan) Zhang for excellent research assistance. All errors are ours.

Funding

The authors received no external funding for this research.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

Acharya, V. V., & Laarits, T. (2026). When Do Treasuries Earn the Convenience Yield?—A Hedging Perspective. *The Journal of Finance*.

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

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.

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

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

Andonov, A., Jansen, K. A., & Rauh, J. D. (2026, September 23). Natural Buyers? Pension Funds and Treasury Exposure.

Antolín-Díaz, J. (2026, September 30). How Did Government Bonds Become Safe? Retrieved from https://juanantolindiaz.com/docs/HowDidBondsBecomeSafe.pdf

Baele, L., Bekaert, G., & Inghelbrecht, K. (2010). The Determinants of Stock and Bond Return Comovements. *The Review of Financial Studies*, *23*(6), 2374–2428.

Barth, D., Beltran, D., Hoops, M., Kahn, J., Liu, E., & Perozek, M. (2025). The Cross-Border Trail of the Treasury Basis Trade. *FEDS Notes*.

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.

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.

Bretscher, L., Schmid, L., Sen, I., & Sharma, V. (2026). Institutional Corporate Bond Pricing. *The Review of Financial Studies*, *39*(3), 605–660.

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

Caballero, R. J. (2006, December). *On the macroeconomics of asset shortages*. 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. *The Review of Financial Studies*, *35*(11), 5228–5274.

Campbell, J. Y., Pflueger, C., & Viceira, L. M. (2020). Macroeconomic Drivers of Bond and Equity Risks. *Journal of Political Economy*, *128*(8), 3148–3185.

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.

Chaudhary, M. (2024, October 13). Regulator Beliefs.

Chaudhary, M., Fu, Z., & Li, J. (2026a). Corporate Bond Multipliers: Substitutes Matter. *The Review of Financial Studies*.

Chaudhary, M., Fu, J. Z., & Li, J. (2026b). Which Intermediary Costs Matter for Asset Prices?

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

Cieslak, A., & Pang, H. (2021). Common shocks in stocks and bonds. *Journal of Financial Economics*, *142*(2), 880–904.

Cochran, P., Infante, S., Petrasek, L., Saravay, Z., & Tian, M. (2023). Dealers' Treasury Market Intermediation and the Supplementary Leverage Ratio. *FEDS Notes*.

Davis, S. J., & Haltiwanger, J. (1992). Gross Job Creation, Gross Job Destruction, and Employment Reallocation. *The Quarterly Journal of Economics*, *107*(3), 819–863.

Devereux, M. B., Engel, C., & Wu, S. P. Y. (2023). *Collateral advantage: Exchange rates, capital flows and global cycles*. National Bureau of Economic Research. Retrieved from https://www.aeaweb.org/conference/2024/program/paper/3tYYE3aY

Ding, D., Fang, X., Hardy, B., & Lewis, K. K. (2026, June 3). *Global pension asset allocations and debt markets*. Bank for International Settlements. Retrieved from https://www.bis.org/publications/paper-172-global-pension-asset-allocations-and-debt-markets

Domanski, D., Shin, H. S., & Sushko, V. (2017). The Hunt for Duration: Not Waving but Drowning? *IMF Economic Review*, *65*(1), 113–153.

Drechsler, I., Savov, A., & Schnabl, P. (2017). The Deposits Channel of Monetary Policy. *The Quarterly Journal of Economics*, *132*(4), 1819–1876.

Du, W., Forbes, K., & Luzzetti, M. N. (2026). Quantitative Tightening Around the Globe: What Have We Learned? *Journal of Money, Credit and Banking*.

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

Du, W., Keerati, R., & Schreger, J. (2025, December). *Decoupling Dollar and Treasury Privilege*. Board of Governors of the Federal Reserve System.

Duffie, D., Fleming, M. J., Keane, F. M., Nelson, C., Shachar, O., & Van Tassel, P. (2023, August). *Dealer Capacity and U.S. Treasury Market Functionality*. Federal Reserve Bank of New York.

Eisenbach, T. M., & Phelan, G. (2026). Fragility of Safe Asset Markets. *The Review of Financial Studies*, *39*(5), 1310–1361.

Eren, E., Schrimpf, A., & Xia, F. D. (2026). The Demand for Government Debt. *Management Science*.

Fang, C., & Xiao, K. (2025, August 1). What Do $40 Trillion of Portfolio Holdings Say about Monetary Policy Transmission?

Fu, Z., Li, J., & Xie, Y. (2022). Convenience yield, inflation expectations, and public debt growth. *Inflation Expectations, and Public Debt Growth (February 5, 2022)*.

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

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

Gabaix, X., Koijen, R. S. J., Mainardi, F., Oh, S., & Yogo, M. (2026, May 9). Asset Demand of U.S. Households.

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.

Gourinchas, P.-O., & Rey, H. (2022, January 22). *Exorbitant Privilege and Exorbitant Duty*. Centre for Economic Policy Research. Retrieved from https://cepr.org/publications/dp16944

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.

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

Greenwood, R., & Vayanos, D. (2014). Bond Supply and Excess Bond Returns. *The Review of Financial Studies*, *27*(3), 663–713.

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, *70*(3), 393–408.

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.

Haddad, V., He, Z., Huebner, P., Kondor, P., & Loualiche, E. (2026, February 27). Causal Inference for Asset Pricing.

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.

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

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

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, *22*(3), 477–498.

Huang, S., Jiang, W., Liu, X., & Liu, X. (2025). Does Liquidity Management Induce Fragility in Treasury Prices? Evidence from Bond Mutual Funds. *The Review of Financial Studies*, *38*(2), 337–380.

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.

Jiang, Z., Krishnamurthy, A., & Lustig, H. (2021). Foreign Safe Asset Demand and the Dollar Exchange Rate. *The Journal of Finance*, *76*(3), 1049–1089.

Jiang, Z., Krishnamurthy, A., & Lustig, H. (2024). Dollar Safety and the Global Financial Cycle. *Review of Economic Studies*, *91*(5), 2878–2915.

Jiang, Z., Richmond, R., & Zhang, T. (2024). A portfolio approach to global imbalances. *The Journal of Finance*, *79*(3), 2025–2076.

Jiang, Z., Richmond, R. J., & Zhang, T. (2025). Understanding the strength of the dollar. *Journal of Financial Economics*, *168*, 104052.

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.

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

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.

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

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

Koijen, R. S. J., & Yogo, M. (2026). Exchange Rates and Asset Prices in a Global Demand System. *Journal of Political Economy*, *134*(8), 2441–2487.

Krishnamurthy, A., & Vissing-Jorgensen, A. (2011). The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy. *Brookings Papers on Economic Activity*, *2011*(2), 215–287.

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

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.

Kundu, S., & Vats, N. (2026). Banking Networks and Economic Growth: From Idiosyncratic Shocks to Aggregate Fluctuations. *American Economic Journal: Macroeconomics*, *18*(2), 332–377.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315–1335.

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

Lou, D. (2012). A flow-based explanation for return predictability. *The Review of Financial Studies*, *25*(12), 3457–3489.

Ludvigson, S. C., & Ng, S. (2009). Macro Factors in Bond Risk Premia. *Review of Financial Studies*, *22*(12), 5027–5067.

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.

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

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). Elsevier.

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). Elsevier.

Newey, W. K., & Windmeijer, F. (2009). Generalized Method of Moments With Many Weak Moment Conditions. *Econometrica*, *77*(3), 687–719.

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.

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

Somogyi, F., Wallen, J., & Xu, L. (2025, December 2). *What Treasury Auctions Reveal About Investor Demand*. Harvard Business School. Boston, MA. Retrieved from https://www.hbs.edu/faculty/Pages/item.aspx?num=68237

Sorkin, A. R. (2018). The Unknowable Fallout of China's Trade War Nuclear Option. *The New York Times: Business*.

Stein, J. C. (2013, April 19). *Liquidity Regulation and Central Banking*. Retrieved from https://www.federalreserve.gov/newsevents/speech/stein20130419a.htm

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.

Tabova, A. M., & Warnock, F. E. (2026). Preferred habitats and timing in the world's safe asset. *Journal of International Economics*, *161*, 104233.

van der Beck, P., Bretscher, L., & Fu, J. Z. (2025, June 13). A Bound on Price Impact and Disagreement.

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

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

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.

Weiss, C. R. (2022). Foreign Demand for U.S. Treasury Securities during the Pandemic.

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

Zhang, A. R., Cai, T. T., & Wu, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, *50*(1), 53–80.

Appendix A develops the estimation framework behind Section 3: time-varying elasticities, loadings, and sizes; common factors; omitted-factor bias; and structural interpretations. Appendix B describes the data and common factors, checks the robustness of identification, and reports additional results on multipliers, the yield decomposition, and the safe haven. Proofs, additional results, and further discussions are in the Supplementary Material (Appendices S1 to S4).

A Methodological Appendix

A.1 General Framework

Here we generalize the model of Section 2.2. Notation: for entities $i = 1, \ldots, N$ and periods $t = 1, \ldots, T$, 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:

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

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.2. 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 granular instrumental variables (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 Supplementary Appendix S4.1.

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, the optimal GIV estimator in this case is also asymptotically efficient. 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 $\mathbb{E}\left[u_{i,t}^2 \mid \mathrm{C}_t, \mathrm{S}_t, \mathbf{X}_t\right] = \sigma_i^2 > 0$.[26]
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.

To ensure the estimator is well-behaved, we make one additional assumption on the weighting matrix, where $\mathbf{W}_t \equiv \mathbf{W}_t(\boldsymbol{\zeta})$ denotes a weighting matrix evaluated at the true parameter.

Assumption A.2. $\mathbb{E}\left[\mathbf{W}_t{}'\mathbf{W}_t\right]$ *is non-singular.* $\mathbb{E}\left[\mathbf{W}_t{}'\boldsymbol{\psi}_t\left(\mathbf{z}\right)\right] = 0$ *if and only if* $\mathbf{z} = \boldsymbol{\zeta}$.[27]

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

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

[26] 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.

[27] For the optimal weighting matrix $\mathbf{W}_t^*$ and non-negative $\mathbf{C}_{i,t}$, the second part follows from the first; Supplementary Appendix S2.4.3 gives the argument. With unrestricted $\mathbf{C}_{i,t}$ multiple roots can arise, and uniqueness must be verified case by case.

for $T \to \infty$, where

$$\mathbf{V}_\zeta^{W*} = \left(\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^{*}\right]\right)^{-1}. \tag{A.5}$$

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

Proof. See Supplementary Appendix S4.1. □

Standard errors of nonlinear functions of the estimates, such as the macro multiplier, price responses, and total loadings, follow from the delta method. Leave-one-out estimates drop a sector's moment conditions, which breaks the market clearing behind (A.5); they use the sandwich variance, with $\mathbf{D}_t \equiv \partial\boldsymbol{\psi}_t(\boldsymbol{\zeta})/\partial\boldsymbol{\zeta}'$:

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

Bias-corrected OLS interpretation. In the simple case of the first example, the estimator can also be written as a bias-corrected OLS estimator. Market clearing implies $\hat{u}_{S(-i),t} = \hat{\zeta}_S p_t^{\varepsilon} - S_i\hat{u}_{i,t}$, where $\hat{\zeta}_S \equiv \sum_j S_j\hat{\zeta}_j$. Substituting into that example's moment condition gives

$$\hat{\zeta}_i = -\frac{\hat{\mathbb{E}}\left[q_{i,t}^{\varepsilon}p_t^{\varepsilon}\right]}{\hat{\mathbb{E}}\left[(p_t^{\varepsilon})^2\right]} + \frac{S_i\hat{\sigma}_i^2}{\hat{\zeta}_S\,\hat{\mathbb{E}}\left[(p_t^{\varepsilon})^2\right]}. \tag{A.7}$$

The first term is the OLS estimate from regressing quantities on prices. It is efficient for an infinitesimal sector but biased for a sector large enough to move the price: the sector's own demand shocks raise the price, so its demand appears less elastic than it is. The second term removes this bias using the sector's own price impact, and it vanishes when the sector is infinitesimal or the market is perfectly elastic. Equation (A.7) also gives an alternative, and sometimes more stable, algorithm. Supplementary Appendix S3.2 illustrates the correction for households and mutual funds.

Comparison with the standard GIV. The optimal GIV differs from the standard GIV of Gabaix and Koijen (2024) in two respects. First, it accommodates heterogeneous elasticities and factor loadings through $\mathbf{C}_{i,t}$ and $\mathbf{X}_{i,t}$, without parameterizing them on observable characteristics. Second, it weights the moment conditions by both size and precision and is therefore efficient, whereas the standard GIV instrument, a size-weighted minus precision-weighted average of quantities, loses power when larger entities have less volatile shocks. In exchange, the optimal GIV must be solved numerically, and it does not remove unobserved common factors by cross-sectional demeaning. Supplementary Appendix S2.2 formalizes the comparison, and Supplementary Appendix S2.4 discusses

missing sectors, measurement errors, and the uniqueness of the solution.

A.2 Common Factors and Identification of Factor Loadings $\boldsymbol{\lambda}_i$

We now consider external common factors. They can be directly observable, such as macrofinancial variables, or unobservable principal components extracted from granular data *outside of the model*. Supplementary Appendix S2.1 discusses the method to extract the unobserved common factors *within the model*.

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

$$\mathbb{E}\left[\mathbf{X}_{i,t}u_{i,t}\right] = \mathbb{E}\left[\mathbf{X}_{i,t}(q_{i,t} + p_t\mathbf{C}'_{i,t}\boldsymbol{\zeta} - \mathbf{X}'_{i,t}\boldsymbol{\beta})\right] = \mathbf{0}. \tag{A.8}$$

Hence, in principle, $\boldsymbol{\zeta}$ and $\boldsymbol{\beta}$ are estimated jointly in a generalized method of moments (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 principal component analysis (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:

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

   Proceed with the optimal weighting matrix defined in (A.3) to estimate $\hat{\boldsymbol{\zeta}}$.[28]
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 $\hat{\boldsymbol{\beta}}$.

[28] 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.

Proposition A.1. *The asymptotic variance formula for $\hat{\boldsymbol{\zeta}}$ in Theorem A.1 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:*

$$Var(\hat{\boldsymbol{\beta}}) = \frac{1}{T}\left[Var^{OLS}(\hat{\boldsymbol{\beta}}) + \boldsymbol{\beta}^{pC} Var(\hat{\boldsymbol{\zeta}}) \boldsymbol{\beta}^{pC\prime}\right] \tag{A.9}$$

$$Var^{OLS}\left(\hat{\boldsymbol{\beta}}\right) = \left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\right)^{-1} \left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\sigma_i^2\right) \left(\sum_i \mathbb{E}\left[\mathbf{X}_{i,t}\mathbf{X}_{i,t}'\right]\right)^{-1} \tag{A.10}$$

Proof. See Supplementary Appendix S4.2. □

A.3 Bias Analysis

Consider the following model:

$$\Delta q_{i,t} = -\zeta_i \Delta p_t + u_{i,t}, \tag{A.11}$$

$$\Delta p_t = \frac{1}{\zeta_S} u_{S,t}, \tag{A.12}$$

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:

$$\Delta q_{i,t} = -\zeta_i^b \Delta p_t + u_{i,t}^b \tag{A.13}$$

and estimates ζ_i^b using the following moment condition:

$$\mathbb{E}\left[u_{i,t}^b u_{j,t}^b\right] \equiv \mathbb{E}\left[\left(\Delta q_{i,t} + \zeta_i^b \Delta p_t\right)\left(\Delta q_{j,t} + \zeta_j^b \Delta p_t\right)\right] = 0, \tag{A.14}$$

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

$$\zeta_S^b \approx \zeta_S \sqrt{1 - \mathcal{R}^2}, \quad \mathcal{R}^2 \equiv \frac{\mathrm{Var}(\lambda_S \eta_t)}{\mathrm{Var}(u_{S,t})} = \frac{\lambda_S^2 \sigma_\eta^2}{\lambda_S^2 \sigma_\eta^2 + \sigma_{\epsilon,S}^2} \tag{A.15}$$

where $\sigma^2_{\epsilon,S} \equiv \sum_{i=1}^N S_i^2 \sigma^2_{\epsilon,i} \in (0,\infty)$ 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.15) 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.

Writing the optimal GIV as a bias-corrected OLS estimator (equation (A.7)) 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 is given in Proposition A.2 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, take the estimated aggregate elasticity $\hat{\zeta}^b_S = 1$. Equation (A.15) implies a true elasticity of 1.15 if an unrecognized common factor explains 25% of the residual variance of the demand shocks and 1.41 if it explains 50%; for the estimate to be off by 100%, the hidden factor would have to explain 75% of the residual variance. It seems unlikely to us that such a strong common factor would be unobservable and escape principal components.

Under the homogeneous case (homoskedasticity $\sigma_i = \sigma$, equal sizes $S_i = \frac{1}{N}$, and same loadings $\lambda_i = \lambda_S$), Equation (A.15) 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.2. *Let the true model be given by (A.11) and (A.12). 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* $\bar{\sigma}^2_\epsilon = \frac{\sum_i S_i^2 \sigma^2_{\epsilon,i}}{\sum_i S^2}$. *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 $\sigma^2_{\epsilon,S} = \sum_i S_i^2 \sigma^2_{\epsilon,i}$, $\sigma^2_{u,S} = \sigma^2_{\epsilon,S} + \lambda_S^2 \sigma^2_\eta$, $\mathcal{R}^2 = \frac{\lambda_S^2 \sigma^2_\eta}{\sigma^2_{u,S}} \in (0,1)$. *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. See Supplementary Appendix S4.3. □

A.4 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. Table A.1 maps four single-asset micro-foundations to the elasticity ζ_i and the demand shifter $\nu_{i,t}$ of the log-linear demand curve. Supplementary Appendix S2.3 derives each mapping, and Appendix A.5 extends the single-asset demand curve to a multi-asset setting.

The elasticity has a different source in each model. In the CRRA benchmark, the long-run expected return and risk-free rate pin it down, and at plausible values it lies far above our estimates. Learning from prices lowers it: the more weight investors put on the current price (ϕ_p), the less they trade against price changes. With convenience services, the curvature ψ_i of utility over Treasury holdings governs it. Under liability-driven investment, duration matching yields an elasticity below one when liabilities have longer duration than Treasuries. Wealth, risk aversion, interest rates, private signals, consumption, and liabilities enter the demand shifter $\nu_{i,t}$.

A.5 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.[29]

[29] 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.

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, P_t the Treasury market portfolio price, and $W_{i,t}$ the sector's wealth. 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. In the bond-in-utility model, $C_{i,t}$ is consumption (not the elasticity design $\mathbf{C}_{i,t}$ of Appendix A.1), ψ_i and θ_i are the curvatures of utility over Treasury holdings and consumption, and $\overline{cy}$ is the long-run convenience yield $r - y$. $\widehat{\mathbb{E}_t[p_{t+1}]}$ 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.

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:

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

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. (2026a).

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 $\mathcal{P}_{2\to 1} \equiv \frac{\zeta_{S,12}}{\zeta_{S,11}}$ is the price pass-through coefficient from the substitute asset to Treasuries, and $\mathcal{A}_1 \equiv \mathcal{P}_{1\to 2}\mathcal{P}_{2\to 1}$ 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. For most policy questions, such as the price impact of quantitative easing or of a foreign portfolio rebalancing, the object of interest is the total elasticity $\tilde{\zeta}_{i,1}$, which gives the response after substitute prices have adjusted; it is also directly estimable, whereas the full matrix of partial elasticities is high-dimensional.

B Empirical Appendix

B.1 Data

B.1.1 Holdings Data

Financial Accounts data. The main dataset we use for Treasury bonds and notes holdings comes from Table FU.210 and L.210 of the 2024Q2 Z.1 release of the Federal Reserve (Fed).[30] Our estimation sample spans 1986Q1–2023Q4. 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) Treasury securities by foreign investors. We use the benchmark-consistent estimates put together by Bertaut and Tryon (2007), Bertaut and Judson (2014), and Bertaut and Judson (2022). For July–December 2011, when the TIC methodology is in transition, we use SLT holdings and Form S transactions. From January 2012, we use the Bertaut–Judson SLT-based holdings and estimated transactions.

For individual U.S. bank holdings of Treasury notes and bonds, we use the Call Reports (FFIEC 031/041/051), downloaded from WRDS. We drop foreign branches (FFIEC 002 filers), identified with the relationship files of the NIC information system. We construct Treasury holdings separately for securities held to maturity (HTM), available for sale (AFS), and in the trading book. HTM and AFS amortized cost and fair value use consolidated values where reported and domestic-office values otherwise; trading holdings use domestic-office values.[31] Holdings are the sum of HTM, AFS, and trading fair values; our transaction proxy starts from the first difference of the sum of HTM amortized cost, AFS amortized cost, and trading fair value. After the maturity allocation described below,

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

[31] We use domestic-office trading holdings through 2018Q1. From 2018Q2, we approximate them from consolidated holdings: 75% of JPMorgan's, reflecting its historical foreign-office share of about one quarter; zero for Citibank, whose foreign-office share approaches 100% after 2012; and the full amount for other banks, whose earlier reports show little foreign-office Treasury trading.

we subtract estimated holding gains on the lagged trading position, using CRSP clean-price returns weighted by the bank's lagged long-maturity bucket shares. To separate bills from notes and bonds, we follow Jansen et al. (2024): using the residual-maturity breakdown of non-trading, non-mortgage debt securities, we impute notes and bonds holdings, assuming that the bill share is stable across residual-maturity buckets and that the buckets above one year approximate notes and bonds holdings.

We retain countries with nonzero holdings of Treasury notes and bonds and nonmissing transactions and lagged holdings throughout 1986Q1–2023Q4. Bank-level data begin in 1997Q3. For each bank, we keep its longest continuous spell with positive holdings, measured both at fair value and as non-trading amortized cost plus trading fair value. The spell must have mean total assets of at least $10 billion and at least 20 usable quarters in each regime it covers, before or after 2009Q1. Remaining holdings and transactions are absorbed into residual categories labelled "all other countries" and "all other banks," respectively.

Discrepancy. Because not all sectors report Treasury notes and bonds holdings separately, and direct estimates of notes and bonds holdings and transactions are unavailable, our construction leaves valuation and coverage discrepancies. We sweep the discrepancy component into the household sector (the residual sector in the Z.1 methodology).

We reallocate seven exceptional sector-quarter flows to the household residual to prevent reporting-related exits and re-entries and small sector entries from generating mechanically extreme normalized flows in the affected sectors.[32] We set each affected sector's flow to zero and transfer the same signed amount to households, preserving aggregate transactions and leaving holdings unchanged.

Sector classification and aggregation. Table B.1 reports our aggregation of Financial Accounts sectors to categories for GMM estimation and to groups for yield decomposition.

B.1.2 Observed Common Factors

We construct ten common factors for 1986Q1–2023Q4. Four describe macro-financial conditions, four summarize survey beliefs, and two control for policy flows. Except for the

[32]The affected observations are federal government defined-benefit retirement funds (1986Q2 and 1986Q3), federal government defined-contribution retirement funds (1988Q1), issuers of asset-backed securities (1999Q4, 2001Q1, and 2001Q4), and other financial business (2000Q1).

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

Sector	Category	Group
Banks in U.S.-affiliated areas	Banks	Banks and asset managers
Foreign banking offices in the U.S.	Banks	Banks and asset managers
U.S.-chartered depository institutions	Banks	Banks and asset managers
Exchange-traded funds (through 2008Q4)	Other	Households and Other
Exchange-traded funds (from 2009Q1)	ETFs	Banks and asset managers
Monetary authority	Fed	Fed
Households and nonprofit organizations	Households	Households and Other
Property-casualty insurance companies	Insurance	Banks and asset managers
Life insurance companies	Insurance	Banks and asset managers
Money market funds	Other	Households and Other
Closed-end funds	Mutual funds	Banks and asset managers
Mutual funds	Mutual funds	Banks and asset managers
Nonfinancial corporate business	Other	Households and Other
Nonfinancial noncorporate business	Other	Households and Other
State and local governments	Other	Households and Other
Government-sponsored enterprises	Other	Households and Other
Credit unions	Other	Households and Other
Other financial business	Other	Households and Other
Issuers of asset-backed securities	Other	Households and Other
Holding companies	Other	Households and Other
State and local government employee defined benefit retirement funds	Pension funds	Banks and asset managers
Federal government defined contribution retirement funds	Pension funds	Banks and asset managers
Federal government defined benefit retirement funds	Pension funds	Banks and asset managers
Private defined contribution pension funds	Pension funds	Banks and asset managers
Private defined benefit pension funds	Pension funds	Banks and asset managers
Rest of the world	Rest of world	Rest of world
Security brokers and dealers	Dealers	Banks and asset managers
Supply	Supply	Supply

Note: Italicized sectors indicate that we use data more granular than the sector-level information from the Financial Accounts. Banks and asset managers combines banks, insurers, pension funds, mutual funds, ETFs, and dealers; Households and Other combines the Households and Other categories.

futures-implied rate revision, we demean each factor and divide by its sample standard deviation over 1986Q1–2023Q4. The rate revision remains in percentage-point units. Figure B.1 reports the R^2 from regressions of Treasury price changes on these factors, added one block at a time; the full set explains more than 40% of the variation in Treasury price changes.

Uncertainty / risk sentiment (η_t^{unc}). We measure uncertainty using innovations in the logarithm of quarter-end implied equity volatility, combining the CBOE VXO and VIX indices (FRED series VXOCLS and VIXCLS). For each index, we estimate an AR(1) with an intercept on their common overlap, 1990Q1–2021Q3, and apply the estimated coefficients

over that index's available history. We use VIX residuals from 1990Q2 onward and VXO residuals before then.

Inflation. We use the seasonally adjusted consumer price index for all urban consumers (FRED series CPIAUCSL), taking the observation for the last month of each quarter. Annualized quarterly inflation is $400[\log(\mathrm{CPI}_t) - \log(\mathrm{CPI}_{t-1})]$. The inflation factor is the residual from an AR(1), including an intercept, estimated over 1986Q1–2023Q4.

Short-rate expectations. We construct quarterly revisions in the futures-implied expected federal funds rate using daily settlement prices of individually dated 30-day federal funds futures contracts sourced from LSEG. At each quarter-end cutoff, we use the latest available settlement on or before that date, treating settlements more than ten calendar days old as missing. For target month m and observation date d, the implied rate in percentage points is $100 - P^{\mathrm{fut}}_{m,d}$. We combine the three contracts for a target quarter using calendar days in each month as weights. Let $\widehat{r}_{t+1|t}$ denote this implied average rate for quarter $t+1$, measured at the end of quarter t. Our factor is

$$\widehat{r}_{t+1|t} - \widehat{r}_{t+1|t-1}.$$

Both observations use contracts for the same target quarter, so the difference measures repricing of the expected rate path during quarter t. For example, the first-quarter factor compares expectations for April–June at end-March with expectations for those same months at end-December.

Federal funds futures data are available from 1991Q1. For 1986Q1–1990Q4, we proxy the factor with the fitted value from a regression of futures revisions on quarterly changes in the effective federal funds rate, with an intercept, estimated over 1991Q1–2003Q3. The realized-rate regressor uses quarter-end observations from FRED series DFF.

Dollar. We construct an extended nominal dollar index from month-end observations of Federal Reserve trade-weighted indices. From January 1995 through December 2019, we use the broad goods index (FRED series DTWEXB); from January 2020, we extend it with monthly log changes in the broad goods-and-services index (DTWEXBGS). Before January 1995, we chain the index backward using log changes in the major-currencies goods index (DTWEXM). The early segment therefore reflects a different currency basket. We estimate an AR(1) with an intercept for the extended log index over January 1986–December 2023 and sum its three monthly residuals within each quarter to obtain the dollar innovation.

Survey factors. We use individual responses to the Philadelphia Fed's Survey of Pro-

fessional Forecasters (SPF) for the current quarter's average corporate bond yield, rather than Treasury-yield forecasts whose history does not span the full sample. For forecaster j, we subtract the previous quarter's realized average Moody's Aaa yield from the current-quarter forecast:

$$b_{j,t} = \widehat{y}_{j,t}^{\mathrm{SPF}} - \overline{y}_{t-1}^{\mathrm{Aaa}}.$$

The forecast is the survey's current-quarter observation, BOND2. The realized rate is taken from the quarterly average of monthly FRED series AAA. Thus $b_{j,t}$ measures a forecasted change from the previous quarter's realization.[33] We drop missing forecasts, retain forecasters with at least four observations during 1986Q1–2023Q4, and require at least seven forecasters in each quarter. We winsorize forecast levels at the 0.5% tails before subtracting the lagged realized yield. From the resulting forecaster-by-quarter panel, we extract four common factors using HeteroPCA (Zhang, Cai, & Wu, 2022), demeaning each forecaster's series and filling missing demeaned observations with zero. The estimation includes the four extracted factors.

Lagged Treasury supply. We control for the lagged normalized Treasury supply flow. Issuance carries the negative-demand sign used in Section 3.1.

Scheduled Federal Reserve purchases. 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 Federal Open Market Committee (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 four rounds of large-scale asset purchases including Treasuries and debt of government-sponsored enterprises, 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 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.[34]

[33]Before 1990Q4, the SPF corporate-yield forecast refers to the new, high-grade corporate bond yield; from 1990Q4, it refers to Moody's Aaa yield.

[34]One exception is the initial stage of the QE in response to the COVID pandemic. Due to the extremity

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 base the anticipated purchases on the FOMC's announcements in the previous quarter. For example, on 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 2009Q2 and 2009Q3 would be $150 billion each, until further policy changes by the FOMC.

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.[35] Because investors may use more information than the FOMC announcements alone, we also treat predictable changes as anticipated. For example, during the exit from QE3, the FOMC's reductions in purchases were not pre-scheduled in earlier quarters, but because they followed a predictable pattern and were well anticipated by investors, we treat them as scheduled. Purchases outside QE periods are rollovers of maturing Treasuries or serve reserve management; they are small, so alternative treatments have little effect on our results. 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. Supplementary Appendix S3.1 plots the realized and anticipated paths and tabulates them during active QE periods.

B.2 Robustness of Identification

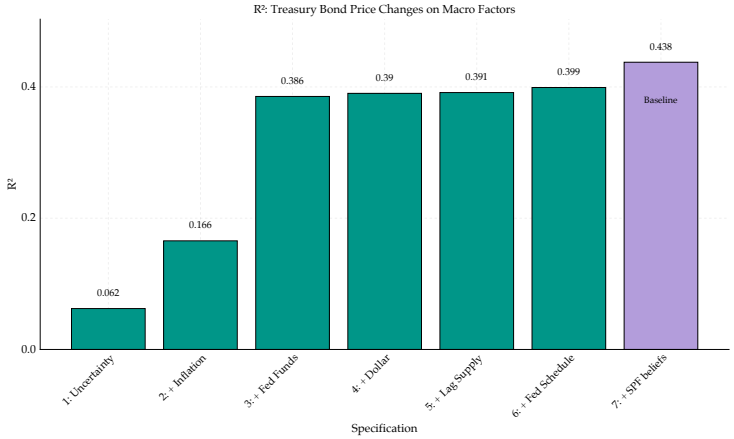
This subsection reports the checks on the identification strategy summarized in Section 3.4. We first drop one sector at a time from the moment conditions, then replace the survey-based factors with internally estimated principal components, then examine the correlations of the estimated idiosyncratic shocks across sectors, and finally check the shocks against narrative evidence and the term structure.

Leave-one-out estimates. Our identification strategy uses other sectors' shocks to instrument for the price in each sector's demand, so it relies on the shocks being idiosyncratic and independent across sectors. Section 3.4 discusses how correlated shocks could

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.

[35] 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 regular outright operations are an order of magnitude larger.

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



Note: This figure reports the R^2 from regressions of Treasury price changes on the observed factors of Appendix B.1.2, added in the order shown. The purple bar reports the R^2 using the set of factors included in our estimation.

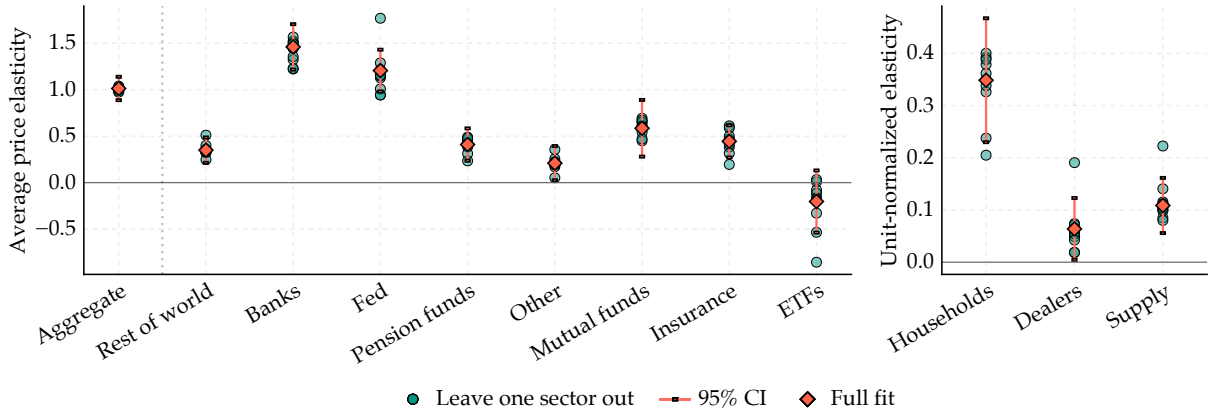
arise, for example from sophisticated investors who predict and front-run other sectors' demand shocks. We therefore re-estimate the model leaving out one sector at a time. Leaving out a sector reduces statistical power, but if the identification assumption holds, the estimated elasticities should remain stable. Figures B.2 and B.3 present time-average price elasticities and post-minus-pre-2009 changes from the specification with a 2009Q1 regime shift in every parameter; Supplementary Appendix S3.4 tabulates the estimates. The leave-one-out average aggregate price elasticity ranges from 0.97 to 1.04, very close to the full-fit average of 1.01, and its decline across 2009Q1 keeps its sign and size in every leave-one-out specification. The decline in the rest of the world's elasticity is also stable across the leave-one-out specifications. These results suggest that correlation between any single sector's demand shocks and those of the other sectors, whether from strategic behavior or other sources, does not drive the estimates.

Internal principal components and bias formula diagnostics. To allow for unobserved common factors, we replace the moment conditions (A.2) with

$$\mathbb{E}\left[\hat{u}_{i,t}(\boldsymbol{\zeta})\,\hat{u}_{j,t}(\boldsymbol{\zeta})\right] = \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\zeta})'\hat{\boldsymbol{\lambda}}_j(\boldsymbol{\zeta}), \qquad i \neq j, \tag{B.1}$$

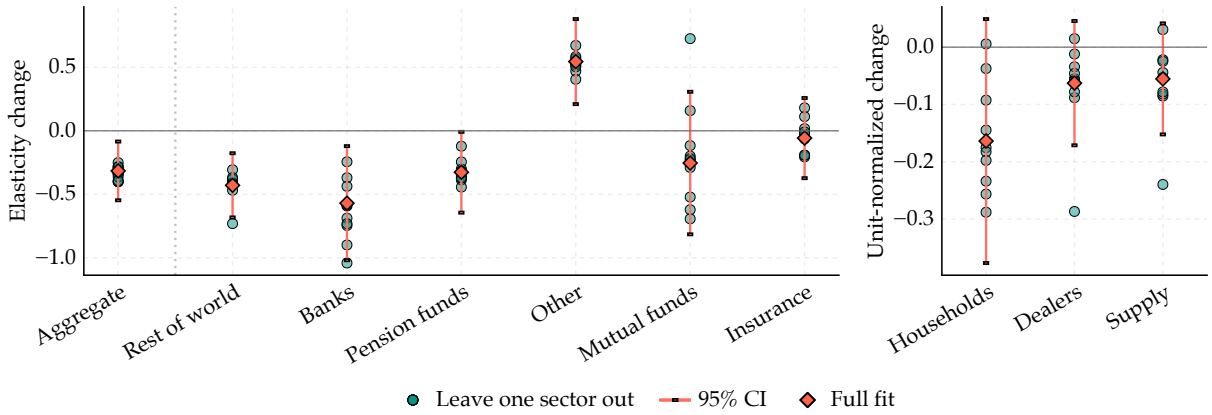
where the loadings come from a heteroskedasticity-robust PCA of the residuals at each trial $\boldsymbol{\zeta}$ (Supplementary Appendix S2.1). Figure B.4 shows that our key takeaways remain robust after replacing the survey-based common factors with these internally estimated principal components of flows (see the discussion in Section 3.4). Panel (a) plots our base-

Figure B.2: **Price elasticity by sector: Leave-one-out estimates**



Note: This figure plots average price elasticities from leave-one-out estimates. Each leave-one-out estimate re-estimates the specification of Table 1 without the moment conditions of one sector, and each sector's elasticity is averaged over its observed quarters. The averages for the Fed and ETFs cover 2009Q1–2023Q4. Red diamonds are the estimates using all sectors, with 95% confidence intervals from equation (A.5); teal circles are the estimates after leaving out one other sector. The aggregate elasticity combines the remaining sectors' re-estimated contributions with the left-out sector's contribution from the estimation using all sectors. Households, dealers, and supply have unit size weights and are shown in the right panel.

Figure B.3: **Change in price elasticities across 2009Q1: Leave-one-out estimates**



Note: This figure plots the post- minus pre-2009 change in price elasticities from the leave-one-out estimates of Figure B.2. Marker conventions and the construction of aggregate elasticity follow Figure B.2. The Fed is omitted because its elasticity is not estimated before 2009 and enters as zero, so its change is just its post-2009 elasticity. Households, dealers, and supply are shown in the right panel.

line estimates against estimates obtained with one to three internal principal components in place of the survey factors. Panel (c) reports the R^2 and adjusted R^2 obtained by projecting Treasury price changes onto the estimated internal principal components. The maximum R^2 is 2.8% with 3 internal PCs, and the adjusted R^2 is 0.9%. Guided by the approximate bias formula (3.5; Appendix A.3), we conclude that the bias arising from our baseline specification is small.

Correlations of idiosyncratic shocks $\hat{u}_{i,t}$. Pairwise sample correlations of the extracted demand and supply shocks, at the sector level ($\hat{u}_{i,t}$), are in general close to zero, including when the rest-of-world residual is split into China, Japan, and Europe (Figures B.5a and B.5b).

China as a case study. The estimated sector-level demand shocks $\hat{u}_{i,t}$ can be checked against narrative evidence. Figure B.6 shows that the extracted demand shocks for China are well aligned with narrative evidence on the role of China's domestic factors in influencing its Treasury demand.

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. Following the literature on the factor structure of yields (Litterman & Scheinkman, 1991, among others), in Figure B.7 we regress price changes implied in Treasury zero-coupon yields (Gurkaynak, Sack, & Wright, 2007) and term premia (Adrian, Crump, & Moench, 2013) across various tenors on the estimated aggregate demand and supply shocks $\hat{u}_{S,t}$. The price impact estimates increase with maturity and line up closely with a duration-based extrapolation of the price impact on the market portfolio that assumes a uniform impact on yields across maturities. This result suggests that our demand shocks mainly shift the level of the yield curve.

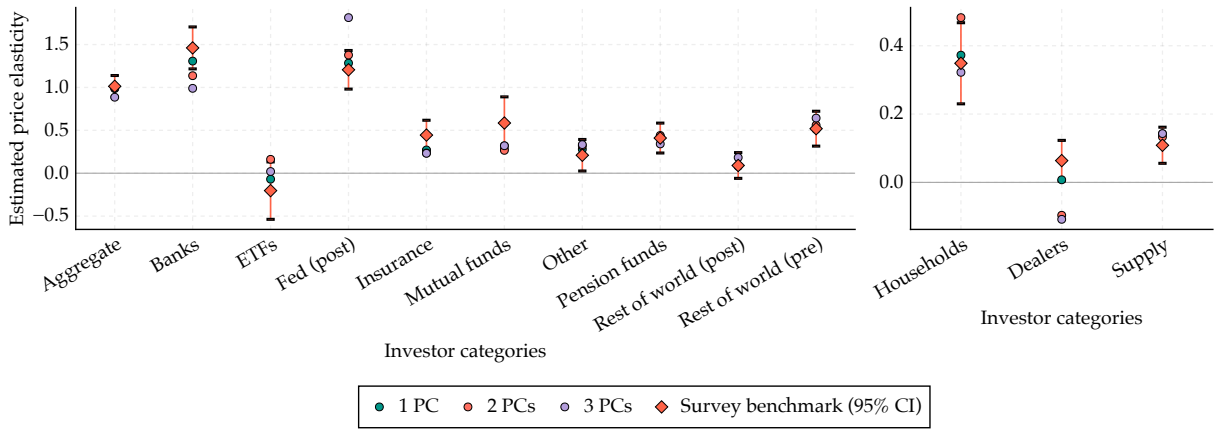
B.3 Additional Results

B.3.1 Multipliers Across the Literature

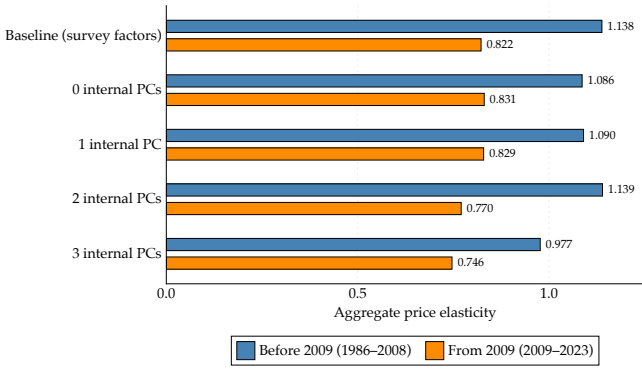
We compare our Treasury macro multiplier estimates with those from the literature. The focus on estimating multiplier parameters—the percentage change in prices to a percentage change in demand as a share of outstanding Treasury securities—is a recent development. However, the literature has long studied the price impact of demand shocks arising from foreign FX interventions, quantitative easing, Treasury repurchase operations, and

Figure B.4: **Robustness: Internal principal components in place of the survey factors**

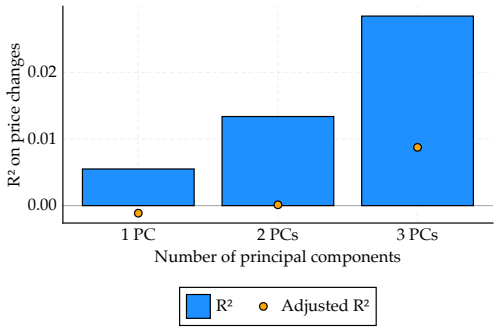
(a) Estimated price elasticities



(b) Average aggregate price elasticity before and after 2009



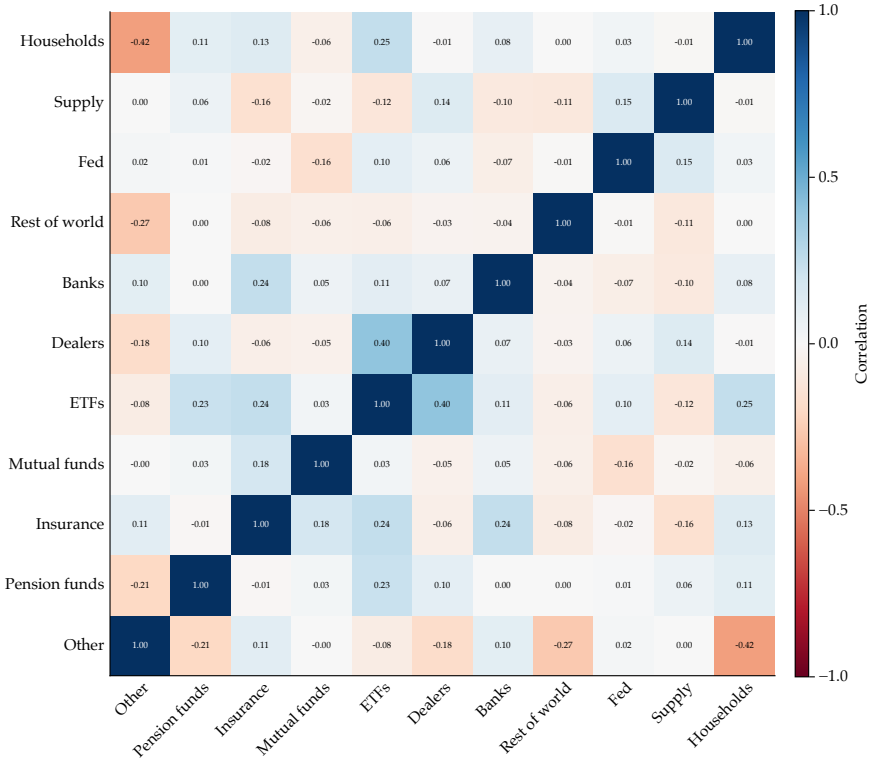
(c) Total price variation explained



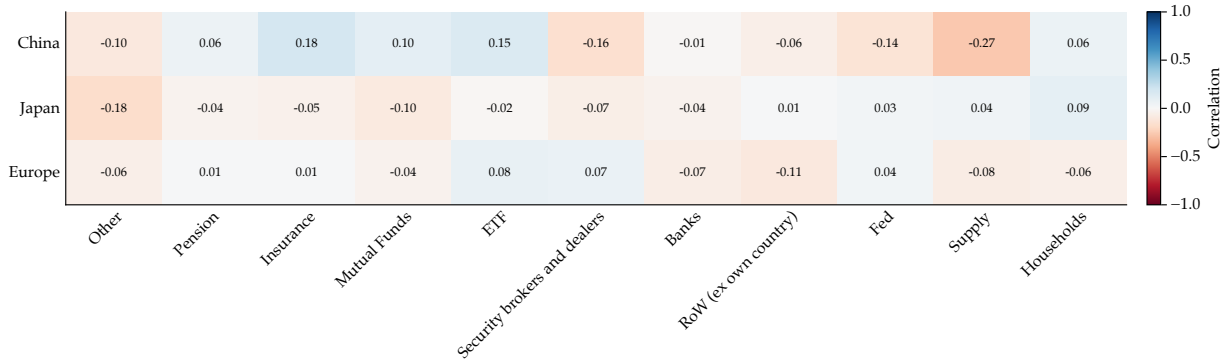
Note: This figure replaces the SPF survey-forecast factors with one to three internal principal components, estimated jointly with the model from flows over 1986Q1–2023Q4. Elasticities, intercepts, and loadings on the six other observed factors change in 2009Q1, as in Table 1; the loadings on the principal components do not. Panel (a) plots time-average sector elasticities. Red diamonds, with 95% confidence intervals from equation (A.5), are the estimates with survey factors; small diamonds are point estimates with principal components. Rest of world is shown separately before and after 2009, and the Fed after 2009 only. Households, dealers, and supply, which have unit size weights, are shown in the right panel. Panel (b) plots the average aggregate price elasticity over 1986Q1–2008Q4 and 2009Q1–2023Q4 with survey factors, without them, and with one to three principal components in their place; the specifications are otherwise as in panel (a). Each quarter's aggregate elasticity uses that quarter's size weights, and quarters are weighted equally. Panel (c) reports the full-sample R^2 and adjusted R^2 from regressing price changes on the principal components alone, with an intercept.

Figure B.5: **Correlations of estimated demand shocks across sectors**

(a) All sectors

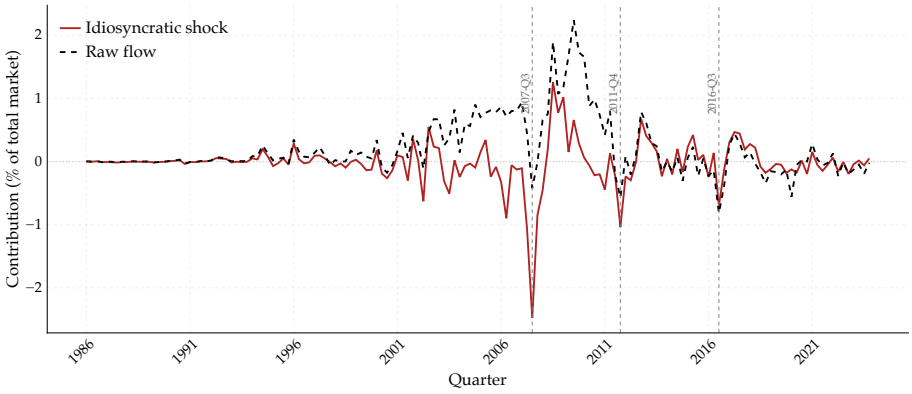


(b) China, Japan, and Europe versus other investor types



Note: Panel (a) plots pairwise sample correlations of the estimated idiosyncratic demand and supply shocks across sectors. Panel (b) plots sample correlations of the size-weighted idiosyncratic demand shocks of China, Japan, and Europe with those of other investor types over 1986Q1–2023Q4, using the quarters in which both are observed. In each row, the rest-of-world column excludes that row's region.

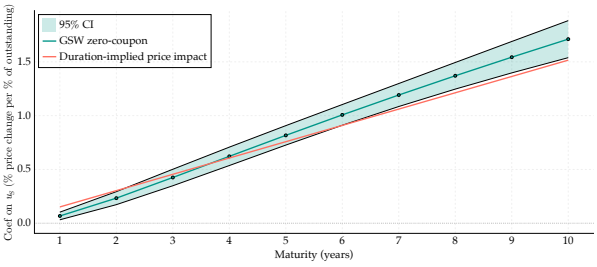
Figure B.6: **Interpreting estimated demand shocks: China as a case study**



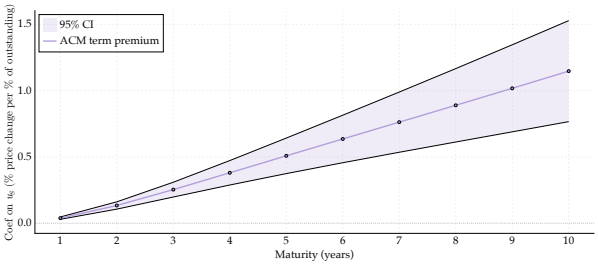
Note: This figure plots China's observed Treasury bond purchases and the idiosyncratic demand shocks estimated using optimal GIV. The large shocks line up with narrative evidence on domestic drivers of China's Treasury demand that are unrelated to global financial conditions or U.S. developments: the slowdown in China's purchases of U.S. securities in 2007Q3, read by analysts as the start of reserve diversification (USCC Economic Issue Brief); the slowdown in reserve accumulation in 2011Q4 as the RMB traded near the weak side of its band (U.S. Treasury); the sales to defend the RMB in 2016Q3 (Bloomberg); and the purchases of 2017 that stemmed appreciation pressure (Reuters).

Figure B.7: **Price impact of estimated demand shocks across maturity**

(a) Zero-coupon yields



(b) Term premia

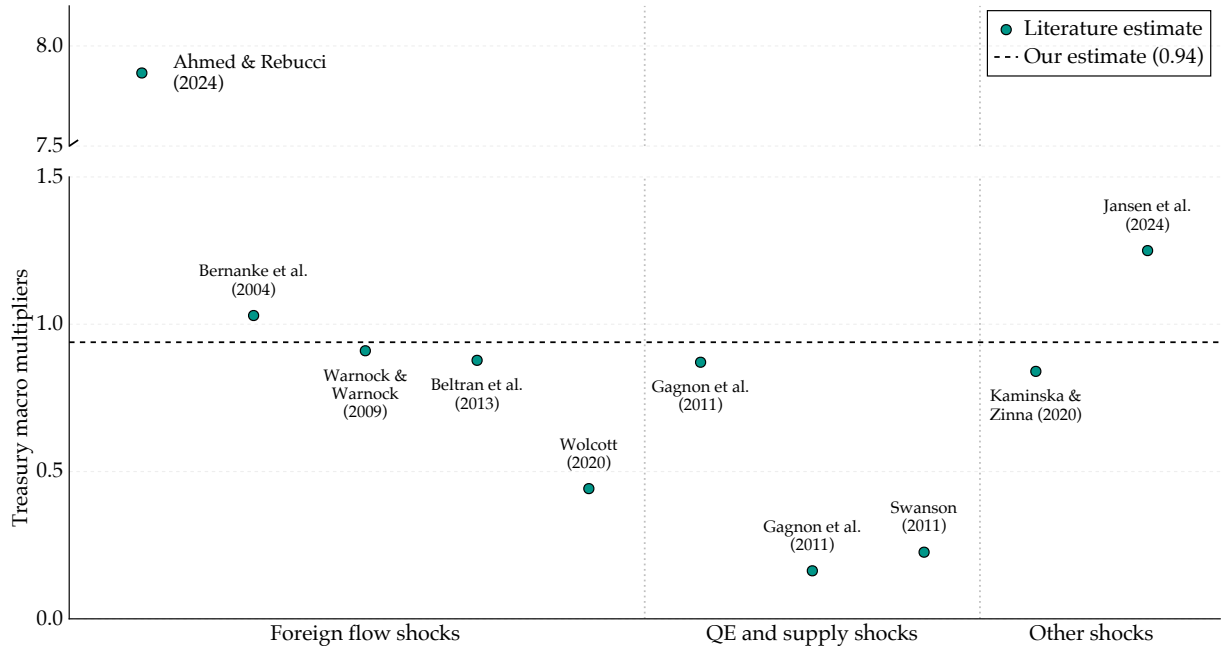


Note: This figure reports coefficients and heteroskedasticity-robust 95% confidence intervals from regressions of maturity-specific price changes on the estimated aggregate demand and supply shocks $\hat{u}_{S,t}$. Price changes are implied by zero-coupon yields (Gurkaynak, Sack, & Wright, 2007) in panel (a) and by term-premium estimates (Adrian, Crump, & Moench, 2013) in panel (b). The solid line in panel (a) is the price impact of a yield change that is uniform across maturities, scaled to match the estimated price impact on the market portfolio of notes and bonds at its average duration.

other policy interventions. By making a few assumptions about the implicit size of the quantity changes, we can approximately map these price impact estimates from the literature to multipliers.

Figure B.8 plots the comparison. 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. With the exception of a couple of outliers, our estimate of around 0.94 is in line with the macro multiplier estimates implied by the literature. Supplementary Appendix S3.3 details the mapping for each study.

Figure B.8: **Comparing Multiplier Estimates Across the Literature**

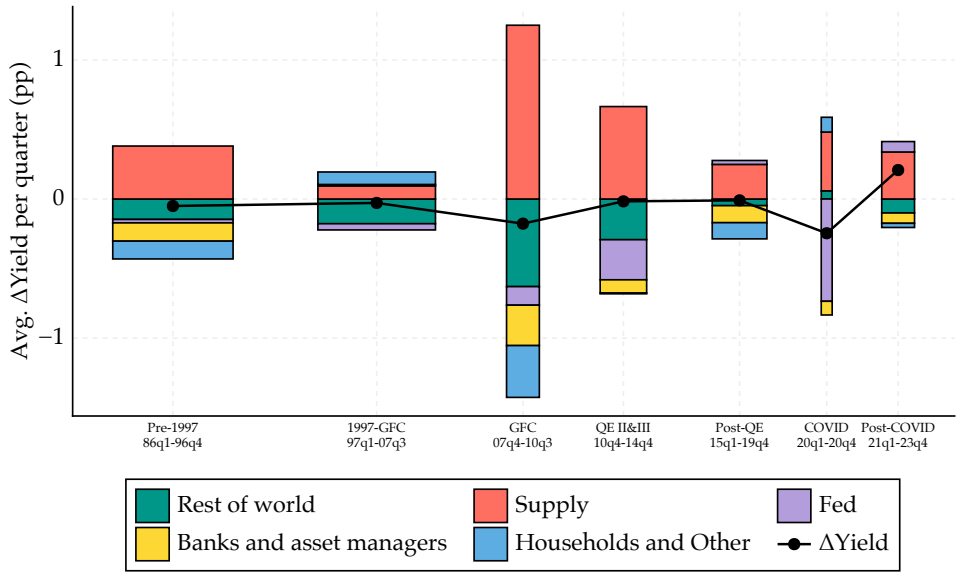


Note: This figure compares our Treasury macro multiplier estimates with those from the literature. When a paper directly estimates macro multipliers, we present them. When a paper only estimates price impact, we map it to a multiplier by making assumptions about the implied quantity change. Our benchmark is the macro multiplier from a single-regime specification, in which private-sector price elasticities and factor loadings are constant over the sample, computed as the inverse of full-sample average aggregate price elasticity over 1986Q1–2023Q4. The vertical-axis break omits multipliers between 1.5 and 7.5.

B.3.2 Yield Decomposition by Sector

Figure B.9 reports the decomposition in equation (2.4) of the main text by group, averaged within seven episodes, using the regional elasticity specification in Table 2, panel (b).

Figure B.9: **Contribution to Treasury yield fluctuations by investor and issuer groups**



Note: This figure plots each group's average quarterly contribution to Treasury yield changes within seven episodes over 1986Q1–2023Q4, with bar widths reflecting episode lengths; GFC denotes the global financial crisis. Each group's contribution is the size-weighted sum of its entities' factor responses, intercepts, and idiosyncratic shocks, divided by that quarter's aggregate price elasticity and converted to a yield change at the lagged Treasury duration. The five contributions sum to the observed yield change. Table B.1 maps sectors to groups.

B.3.3 Safe Haven: Direct Loadings, Raw Flows, and Timing

Total and direct loadings. In equation (3.1) of the main text, sector i's demand loads on the standardized uncertainty innovation η_t^{unc} with the regime-specific coefficient $\lambda_{i,r}^{unc}$ and on the other seven contemporaneous factors of Appendix B.1.2, η_t^k for $k = 1, \ldots, 7$, with coefficients λ_i^k pooled across regimes. The *direct* loading $\lambda_{i,r}^{unc}$ is the response to a one-standard-deviation uncertainty innovation holding the other factors fixed. The other factors move with uncertainty, however; when uncertainty rises, for example, investors revise down the expected path of the federal funds rate. The *total* loading adds the responses that run through these co-movements,

$$\lambda_{i,r}^{unc,\text{total}} = \lambda_{i,r}^{unc} + \sum_{k=1}^{7} b_{k,r}\, \lambda_i^k,$$

where $b_{k,r}$ is the regime-specific slope from a regression of η_t^k on η_t^{unc}, lagged Treasury supply, and scheduled Federal Reserve purchases. The last two are predetermined, so an uncertainty innovation cannot move them; they are held fixed in both loadings. Standard errors of total loadings add a heteroskedasticity-robust variance of $\hat{b}_{k,r}$ to the loading variance (A.9); the two are asymptotically uncorrelated because $u_{i,t}$ is mean-independent of the factors (Assumption A.1).

Direct-loading counterparts. Figure B.10 plots the ten-year rolling price regression with both loadings, and Table B.2 repeats Table 3 with direct loadings. The three-regime conclusions about dealers and foreign investors are the same: dealers dominate both transitions, and foreign investors make no positive contribution. Supplementary Appendix S3.5 gives the direct-loading versions of Figures 4 and 5 (Figures S3.4 and S3.5) and reports each sector's price elasticity and its direct and total uncertainty responses in the first and third regimes.

Raw flows. Observed flows show the same patterns before any estimation. Scaled by each group's lagged holdings and net of the growth of privately held debt, foreign flows fall and domestic private flows rise with uncertainty, both over the full sample and during the crisis (Figure B.11). Dealers' net purchases rise with uncertainty only in the middle regime (Figure B.12).

Timing of the first regime shift. Section 5 places the first regime shift at 1997Q3. We use the price regression to locate the window in which the safe-haven property emerged and examine how the demand-system results change as the first regime shift moves within that window. Figure B.13a re-estimates the price regression of Figure B.10 with two regime shifts, holding the second at 2009Q1 and moving the first from 1990Q1 to 2004Q4. The fit is best when the first shift falls in 1996–1998: the sum of squared residuals stays within 1% of its minimum over 1996Q2–1998Q3 for the total loading and over 1996Q2–1997Q3 for the direct loading. The rise in the price loading is also largest there; at 1997Q3 it is 1.65 percentage points for the total loading and 1.53 for the direct loading. Within this window, we place the shift at 1997Q3, when our bank-level flows in Treasury notes and bonds first become available. The Call Report maturity breakdown used to separate notes and bonds from bills is available from 1997Q2, and our flow measure requires two consecutive quarters.

The left and center panels of Figure B.13b re-estimate the three-regime demand system of Table 3 with the first regime shift at each quarter from 1995Q1 to 1999Q4 and

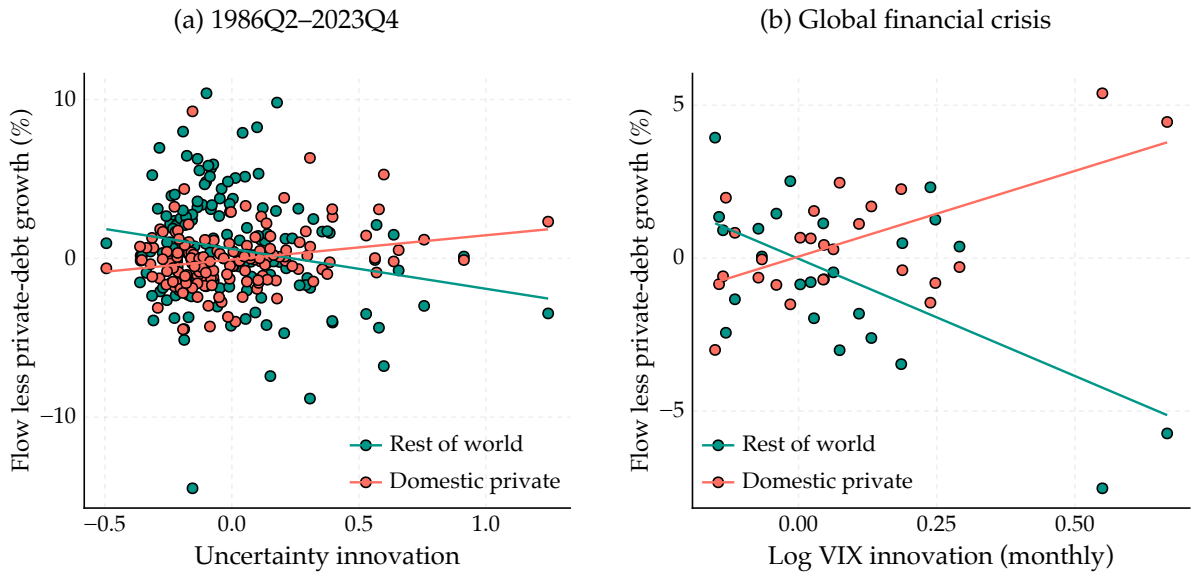
Figure B.10: **Ten-year rolling price loading on uncertainty innovations**



Note: This figure reports 10-year rolling OLS regressions of the quarterly price change of Treasury notes and bonds on the uncertainty innovation. The solid line is the total loading, from the univariate regression; the dashed line is the direct loading, controlling for the seven other contemporaneous factors and the two policy controls. Windows are plotted at their midpoints; dotted vertical lines mark the regime shifts at 1997Q3 and 2009Q1.

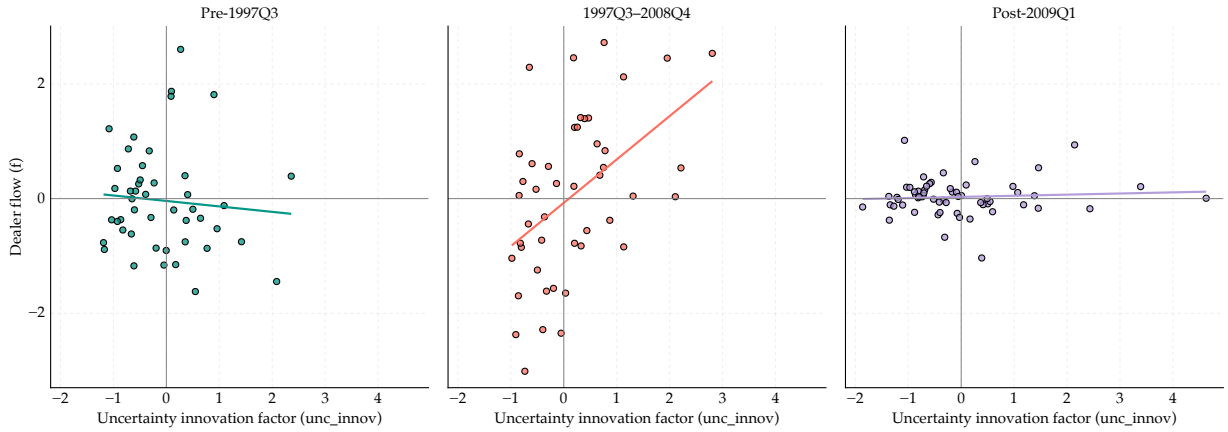
the second fixed at 2009Q1. The aggregate loading rises significantly at the first shift at 16 of the 20 dates; dealers' contribution rises significantly at every date through 1998Q4 and falls significantly at 2009Q1 in all 20 specifications. The right panel re-estimates the loadings on samples that begin at each candidate date and end in 2023Q4. The rest of the world's total loading is negative at every date, and domestic private investors' total loading exceeds it at every date by a margin significant at 5%.

Figure B.11: **Foreign and domestic private flows (net of private-debt growth) and uncertainty**



Note: This figure plots foreign and domestic private net flows into Treasury securities against uncertainty innovations. Domestic private holdings and flows equal privately held gross federal debt (Dallas Fed) less foreign Treasury holdings and flows. Each group's flow is divided by its own lagged holdings, and the growth rate of privately held debt is then subtracted. Panel (a) uses quarterly innovations in log implied volatility from 1986Q2 to 2023Q4; panel (b) uses monthly innovations in log VIX from December 2007 to December 2009.

Figure B.12: **Dealer flows and uncertainty innovations, by regime**



Note: This figure plots the quarterly dealer net flows used in the demand estimation, normalized by total Treasuries outstanding, against the standardized uncertainty innovation, 1986Q1–2023Q4, with one panel per regime. Lines are OLS fits.

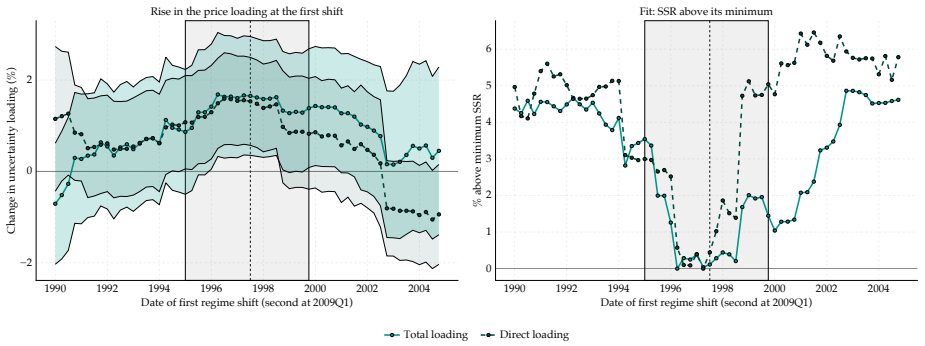
Table B.2: **Three-regime direct uncertainty responses**

	1986Q1–1997Q2		1997Q3–2008Q4		2009Q1–2023Q4	
	Loading	Contribution	Loading	Contribution	Loading	Contribution
Aggregate	-1.12^{**}	—	1.65^{***}	—	0.11	—
	[−2.20, −0.03]		[0.56, 2.73]		[−0.65, 0.88]	
(a) Contributions to the aggregate						
Rest of world	−0.16	−0.03	0.00	0.00	−0.61	−0.29
	[−1.56, 1.24]	[−0.32, 0.25]	[−1.37, 1.36]	[−0.64, 0.63]	[−1.66, 0.43]	[−0.77, 0.20]
Households†	−0.15	−0.15	0.72^{**}	0.72^{**}	0.42^{**}	0.42^{**}
	[−0.77, 0.46]	[−0.77, 0.46]	[0.09, 1.34]	[0.09, 1.34]	[0.05, 0.78]	[0.05, 0.78]
Dealers†	−0.13	−0.13	1.07^{***}	1.07^{***}	−0.07	−0.07
	[−0.52, 0.26]	[−0.52, 0.26]	[0.68, 1.47]	[0.68, 1.47]	[−0.31, 0.17]	[−0.31, 0.17]
Remaining domestic	-0.89^{*}	-0.61^{*}	−0.41	−0.18	−0.50	−0.16
	[−1.89, 0.12]	[−1.30, 0.08]	[−1.43, 0.60]	[−0.61, 0.26]	[−1.20, 0.20]	[−0.39, 0.07]
Supply†	−0.19	−0.19	0.04	0.04	0.04	0.04
	[−0.55, 0.18]	[−0.55, 0.18]	[−0.33, 0.40]	[−0.33, 0.40]	[−0.19, 0.26]	[−0.19, 0.26]
Fed					0.83	0.18
					[−0.27, 1.94]	[−0.06, 0.42]
(b) Rest of world vs. domestic private						
Domestic private	-1.19^{*}	-0.90^{*}	3.12^{***}	1.61^{***}	0.49	0.18
	[−2.51, 0.13]	[−1.89, 0.10]	[1.50, 4.75]	[0.77, 2.45]	[−0.83, 1.80]	[−0.31, 0.67]
Difference	1.03	0.86	-3.13^{***}	-1.61^{***}	−1.10	−0.47
	[−0.90, 2.96]	[−0.17, 1.90]	[−5.29, −0.97]	[−2.68, −0.54]	[−2.78, 0.59]	[−1.16, 0.23]

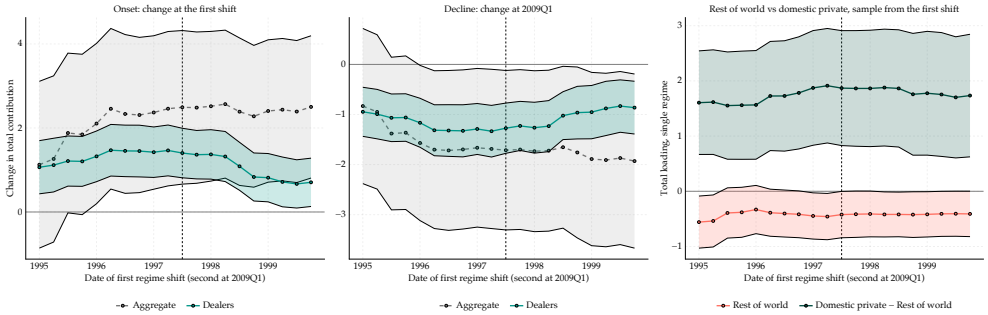
Note: This table repeats Table 3 with direct uncertainty loadings, which hold the other factors fixed, in place of total loadings. Rows, contributions, and the specification are as in Table 3. 95% confidence intervals, from equation (A.9), are in brackets below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Figure B.13: **Safe-haven estimates with alternative dates for the first regime shift**

(a) Treasury price regressions



(b) Demand-system uncertainty loadings



Note: Panel (a) re-estimates the price regression of Figure B.10 over 1986Q1–2023Q4, with regime shifts in slope and intercept at the first regime shift and at 2009Q1. The total loading comes from the univariate regression on uncertainty; the direct loading controls for the seven other contemporaneous factors and the two policy controls. Left: rise in the loading at the first shift. Right: sum of squared residuals relative to its minimum. The shaded span marks the first-shift dates of panel (b). Panel (b), left and center: changes in dealers' contribution and in the aggregate total loading at the first shift (left) and at 2009Q1 (center), from the specification of Table 3 with the first shift at the quarter shown. Panel (b), right: total loadings estimated on a single regime from the quarter shown to 2023Q4, with a 2009Q1 regime shift in foreign price elasticities; banks are aggregated because detailed bank data start in 1997Q3. Bands are 95% confidence intervals, heteroskedasticity-robust in panel (a) and computed as in Table 3 in panel (b); dashed lines mark 1997Q3.

This Supplementary Material complements the Online Appendix; references to Appendices A and B are to the Online Appendix. Appendix S1 decomposes Treasury yields by factor, Appendix S2 collects further methodological results, Appendix S3 reports additional empirical exhibits, and Appendix S4 contains the proofs.

S1 Yield Decomposition by Factors

This appendix decomposes Treasury yield fluctuations by factor, using the regional elasticity specification in Table 2, panel (b). Appendix B.3.2 reports the decomposition by sector, which follows equation (2.4) of the main text with the regime-specific parameters of equation (3.1); hats denote estimates, $r(t)$ is the regime of quarter t, and $\hat{\zeta}_{S,t} = \sum_i S_{i,t}\hat{\zeta}_{i,r(t)}$ is the quarterly aggregate elasticity.

Each sector's contribution combines the effects of several demand shifters. The factor decomposition separates them: it attributes yield changes to the economic factors that shift demand, such as risk sentiment, inflation, and monetary policy, and identifies the sectors that transmit each factor to prices. Like a reduced-form regression of prices on observed macroeconomic variables, it shows which factors drive Treasury yields; unlike such a regression, it also shows which sectors transmit them.

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

$$\Delta\hat{p}_{\eta(k),t} = \hat{\zeta}_{S,t}^{-1}\hat{\lambda}_{S,t}(k)\eta_t(k), \tag{S1.1}$$

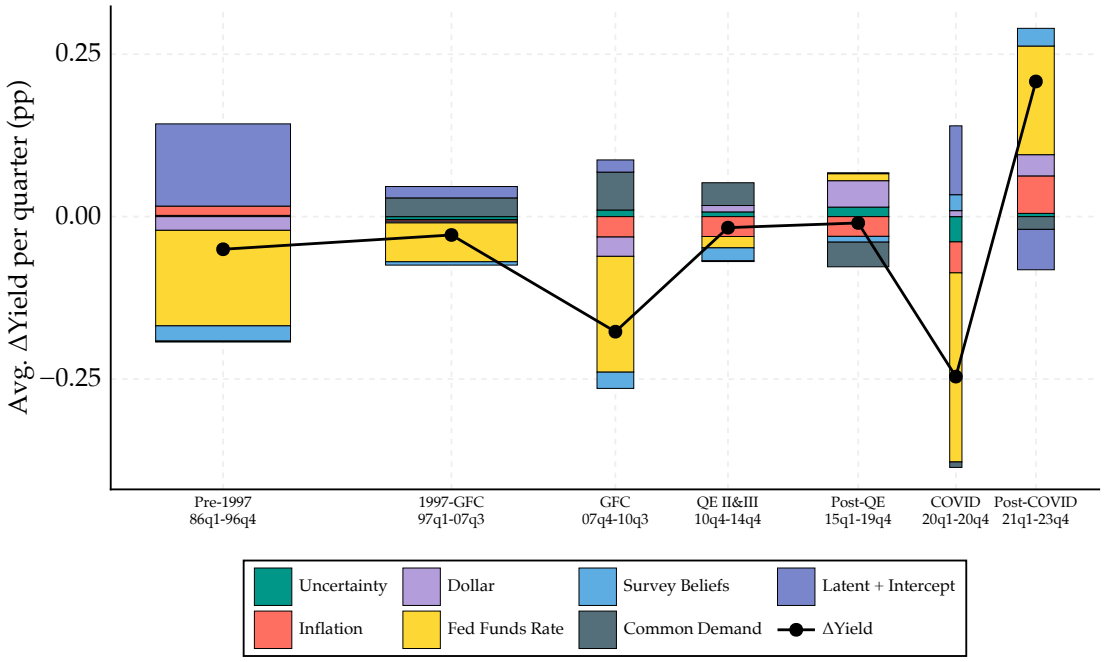
where $\hat{\lambda}_{S,t}(k) = \sum_i S_{i,t}\hat{\lambda}_{i,r(t)}(k)$ is the quarterly size-weighted sum of factor loadings across sectors. The contribution from idiosyncratic shocks equals $\hat{\zeta}_{S,t}^{-1}\hat{u}_{S,t}$, with $\hat{u}_{S,t} = \sum_i S_{i,t}\hat{u}_{i,t}$. Figure S1.1 combines this contribution with fitted intercepts.

Figure S1.1 illustrates the contribution of various factors to Treasury yield movements, expressed in quarterly averages (percentage points). Table S1.1 reports the full set of average factor loading estimates by sector. We distinguish uncertainty innovations, inflation innovations, U.S. dollar innovations, Federal Funds rate revisions, and four survey forecast factors. The survey factors appear as "Survey Beliefs." Scheduled Federal Reserve (Fed) purchases and lagged net issuance appear as "Common Demand." The decomposition includes sector-specific intercepts, grouped with the idiosyncratic shocks.[36]

The uncertainty component contributes a cumulative yield decline of 15.6 basis

[36]Changing the benchmark levels of the factors redistributes contributions between the factors and intercepts while preserving their total.

Figure S1.1: **Macro and idiosyncratic drivers of Treasury yields**



Note: Contributions are computed from equation (S1.1) using quarterly size weights and the quarterly aggregate price elasticity, then converted to yield changes using lagged Treasury duration. Bars report average contributions per quarter in percentage points; their widths reflect episode lengths, and GFC denotes the global financial crisis. "Survey Beliefs" combines the four principal components of survey forecast revisions; "Common Demand" combines scheduled Fed purchases and lagged net issuance. "Latent + Intercept" combines idiosyncratic shocks and fitted intercepts. All components sum to the observed yield change, shown in black.

points during 2020, but an increase of 12.2 basis points during 2007Q4–2010Q3. This component uses the direct uncertainty loading in the regional specification, with one regime shift at 2009Q1. Section 5 studies uncertainty responses with separate 1997Q3 and 2009Q1 regimes and includes transmission through co-moving macro factors.

The Federal Funds rate component contributes a cumulative yield increase of 56.2 basis points during 2004–2006, below one percentage point despite the substantial rise in short rates, consistent with the interest-rate "conundrum" discussed by Greenspan (2005). It contributes declines of 213.9 and 116.4 basis points during the global financial crisis and 2020, respectively, and an increase of 1.98 percentage points during 2022–2023.

Hanson and Stein (2015) emphasize a demand channel through which monetary policy affects long-term rates. Our framework traces the Federal Funds rate component to investor demand loadings. Banks' average loading is negative, associating positive rate revisions with lower Treasury demand, holding Treasury prices and other factors fixed. This sign is consistent with the deposit channel of monetary policy described by Drech-

sler, Savov, and Schnabl (2017).

CPI inflation innovations contribute a cumulative yield increase of 0.69 percentage points during 2021–2023 and a decline of 37.7 basis points during the global financial crisis. These contributions reflect inflation innovations, rather than inflation expectations. Households' average inflation loading is negative. The negative loading is consistent with households reducing nominal Treasury demand when inflation rises, as in Nagel and Yan (2022).[37]

[37]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.

Table S1.1: **Average sector parameter estimates**

Sector	Holdings share (%)	ζ	$\epsilon^{unc}_{(std.)}$	ϵ^{FFR}	Inf.	$\epsilon^{USD}_{(std.)}$	Lagged f_{supply}	Fed schedule
Aggregate		1.01^{***}	0.07	-3.99^{***}	-0.41	-0.46	-0.13	0.19
		[0.89, 1.14]	[−0.49, 0.63]	[−5.99, −2.00]	[−0.95, 0.14]	[−1.06, 0.14]	[−0.69, 0.44]	[−0.62, 1.00]
Supply†	101.01	0.11^{***}	-0.01	-0.13	0.03	0.04	1.74^{***}	0.15
		[0.06, 0.16]	[−0.18, 0.17]	[−0.72, 0.46]	[−0.13, 0.20]	[−0.14, 0.21]	[1.56, 1.91]	[−0.12, 0.42]
Rest of world	38.62	0.35^{***}	-0.38	0.13	-0.02	-0.54	-1.67^{***}	-0.46
		[0.21, 0.49]	[−1.14, 0.39]	[−2.67, 2.94]	[−0.76, 0.73]	[−1.39, 0.30]	[−2.43, −0.91]	[−1.50, 0.58]
Fed	21.88	1.21^{***}	0.84	-19.54^{***}	0.17	-0.37	0.06	4.17^{***}
		[0.98, 1.43]	[−0.27, 1.95]	[−24.53, −14.55]	[−1.00, 1.34]	[−1.77, 1.03]	[−1.01, 1.12]	[3.36, 4.99]
Other	15.22	0.21^{**}	-0.57	-0.41	0.39	0.05	0.06	-0.04
		[0.03, 0.39]	[−1.44, 0.29]	[−3.42, 2.59]	[−0.42, 1.21]	[−0.87, 0.96]	[−0.80, 0.91]	[−1.30, 1.23]
Pension funds	10.79	0.41^{***}	-0.38	-0.33	-0.01	-1.06^{***}	-0.74^{**}	-0.11
		[0.24, 0.58]	[−1.00, 0.25]	[−2.37, 1.71]	[−0.59, 0.58]	[−1.69, −0.43]	[−1.37, −0.11]	[−1.12, 0.90]
Banks	8.55	1.46^{***}	-0.27	-8.73^{***}	0.04	0.88	-3.08^{***}	0.59
		[1.22, 1.71]	[−1.73, 1.19]	[−13.87, −3.58]	[−1.38, 1.46]	[−0.70, 2.45]	[−4.55, −1.62]	[−1.51, 2.68]
Mutual funds	5.99	0.59^{***}	-0.92	0.01	0.29	0.08	-2.77^{***}	0.52
		[0.28, 0.89]	[−2.10, 0.27]	[−4.03, 4.05]	[−0.82, 1.41]	[−1.14, 1.31]	[−3.96, −1.58]	[−1.29, 2.33]
Households†	5.56	0.35^{***}	0.26^{*}	-0.69	-0.37^{**}	-0.23	-0.66^{***}	-0.18
		[0.23, 0.47]	[−0.04, 0.57]	[−1.75, 0.36]	[−0.66, −0.08]	[−0.55, 0.08]	[−0.97, −0.36]	[−0.65, 0.29]
Insurance	5.46	0.44^{***}	-0.48^{*}	-1.35	0.30	-0.42	-1.75^{***}	-0.29
		[0.27, 0.62]	[−1.01, 0.05]	[−3.18, 0.49]	[−0.20, 0.80]	[−0.97, 0.12]	[−2.28, −1.22]	[−1.10, 0.52]
ETFs	1.13	-0.20	1.23	2.16	-0.24	0.91	-1.00	-0.22
		[−0.54, 0.13]	[−0.30, 2.76]	[−4.75, 9.07]	[−1.85, 1.38]	[−1.02, 2.84]	[−2.47, 0.47]	[−1.34, 0.91]
Dealers†	1.07	0.06^{**}	0.13	-0.74^{**}	-0.14	0.06	0.02	-0.27^{*}
		[0.00, 0.12]	[−0.06, 0.32]	[−1.39, −0.10]	[−0.32, 0.04]	[−0.14, 0.25]	[−0.17, 0.21]	[−0.57, 0.02]

Note: This table reports price elasticities ζ_i and direct factor loadings from the specification of Table 1, averaged over each sector's observed quarters in 1986Q1–2023Q4; the rows for the Fed and ETFs cover 2009Q1–2023Q4. The columns are the observed factors of Appendix B.1.2: uncertainty ($\epsilon^{unc}_{(std.)}$), short-rate expectations (ϵ^{FFR}), inflation, the dollar ($\epsilon^{USD}_{(std.)}$), lagged Treasury supply, and scheduled Federal Reserve purchases; the survey-factor loadings are not shown. Holdings share is each sector's average share of Treasury holdings over the same quarters. Sectors marked † have unit size weights, so their coefficients measure demand responses relative to the total market; the other sectors' coefficients are relative to their own holdings. 95% confidence intervals, from equations (A.5) and (A.9), are in brackets below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

S2 Further Methodological Results

S2.1 Internal Common Factor Estimation

Appendix A.2 treats common factors that are observed or extracted outside the model. In this section we study identification when common demand shocks are unobservable and are 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 $\boldsymbol{\epsilon}_t := (\epsilon_{1,t}, \ldots, \epsilon_{N,t})'$:

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

where

$$\mathbb{E}[\boldsymbol{\eta}_t] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \mathbf{I}_K,\ \mathbb{E}[\boldsymbol{\epsilon}_t] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'] = \boldsymbol{\Sigma}_\epsilon = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_N^2),\ \mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\epsilon}_t'] = \mathbf{0}. \quad \text{(S2.1)}$$

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:

$$\psi_{ij,t}(\boldsymbol{\zeta}; \boldsymbol{\Lambda}) \equiv \hat{u}_{i,t}(\boldsymbol{\zeta}) \, \hat{u}_{j,t}(\boldsymbol{\zeta}) - \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_j, \qquad i \neq j. \quad \text{(S2.2)}$$

Stacking over (i, j) yields $\frac{N(N-1)}{2}$ moments. We can use the generalized method of moments (GMM) estimator for $\boldsymbol{\zeta}$ as in Appendix A.1, replacing the zero off-diagonal restriction with (S2.2) to capture the additional unobserved common variation.

Weighting. With latent factors, $\operatorname{Cov}\big(\psi_{ij,t}, \psi_{kl,t}\big)$ 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 $\mathbf{\Lambda}$ 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 $\widehat{\mathbf{\Sigma}}_{uu}(\boldsymbol{\zeta}) = T^{-1}\sum_t \hat{\mathbf{u}}_t(\boldsymbol{\zeta})\,\hat{\mathbf{u}}_t(\boldsymbol{\zeta})'$. We estimate the factor loadings $\widehat{\mathbf{\Lambda}}(\boldsymbol{\zeta})$ using heteroskedasticity-robust PCA methods (e.g., HeteroPCA or deflated HeteroPCA).

The off-diagonal moment restrictions become

$$\widehat{\psi}_{ij}(\boldsymbol{\zeta}) = \hat{u}_{i,t}(\boldsymbol{\zeta})\,\hat{u}_{j,t}(\boldsymbol{\zeta}) - \hat{\boldsymbol{\lambda}}_i(\boldsymbol{\zeta})'\,\hat{\boldsymbol{\lambda}}_j(\boldsymbol{\zeta}), \qquad i \neq j, \tag{S2.3}$$

S2.1.1 Identifiability

When we estimate $\mathbf{\Lambda}$ 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:

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

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:

$$\mathbf{\Sigma}_{qq} \;+\; \mathbf{\Sigma}_{qp}\,(\mathbf{C}\boldsymbol{\zeta})' \;+\; (\mathbf{C}\boldsymbol{\zeta})\,\mathbf{\Sigma}_{qp}' \;+\; \sigma_p^2\,(\mathbf{C}\boldsymbol{\zeta})(\mathbf{C}\boldsymbol{\zeta})' = \mathbf{\Lambda}\,\mathbf{\Lambda}' \;+\; \mathbf{\Sigma}_\epsilon. \tag{S2.4}$$

We denote $(\boldsymbol{\zeta}_\star, \mathbf{\Lambda}_\star, \mathbf{\Sigma}_\star)$ as the true parameters. Identification means $(\boldsymbol{\zeta}_\star, \mathbf{\Lambda}_\star, \mathbf{\Sigma}_\star)$ is the unique solution to the moment condition (S2.4). Equivalently, let $(\tilde{\boldsymbol{\zeta}}, \tilde{\mathbf{\Lambda}}, \tilde{\mathbf{\Sigma}})$ denote deviations from the truth.Identification requires that the only solution to the perturbed system (S2.5) is $\tilde{\boldsymbol{\zeta}} = \mathbf{0}$, $\tilde{\mathbf{\Lambda}} = \mathbf{0}$, and $\tilde{\mathbf{\Sigma}} = \mathbf{0}$.

$$(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})\,\mathbf{\Sigma}_{qp}' + \mathbf{\Sigma}_{qp}\,(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' + \sigma_p^2\Big((\mathbf{C}\,\tilde{\boldsymbol{\zeta}})(\mathbf{C}\,\boldsymbol{\zeta}_\star)' + (\mathbf{C}\,\boldsymbol{\zeta}_\star)(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' + (\mathbf{C}\,\tilde{\boldsymbol{\zeta}})(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})'\Big) = \mathbf{\Lambda}_\star\tilde{\mathbf{\Lambda}}' + \tilde{\mathbf{\Lambda}}\,\mathbf{\Lambda}_\star' + \tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}' + \tilde{\mathbf{\Sigma}}. \tag{S2.5}$$

To characterize identification more clearly, we study local identifiability around the true parameters by focusing on the first-order approximation to (S2.5). Dropping the

quadratic terms and refactoring, the first-order approximation is

$$(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})\left(\boldsymbol{\Sigma}_{qp}+\sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_\star\right)' + \left(\boldsymbol{\Sigma}_{qp}+\sigma_p^2\,\mathbf{C}\,\boldsymbol{\zeta}_\star\right)(\mathbf{C}\,\tilde{\boldsymbol{\zeta}})' - \tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Lambda}_\star\,\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Lambda}}\,\boldsymbol{\Lambda}_\star'. \tag{S2.6}$$

Proposition S2.1. *The parameters are first-order identified if and only if the only solution to the following equation is* $\tilde{\boldsymbol{\zeta}} = \mathbf{0}$*:*

$$\mathbf{Q}^\Lambda\,\mathbf{C}\,\tilde{\boldsymbol{\zeta}}\left(\mathbf{S}'\boldsymbol{\Sigma}_\star\right)\mathbf{Q}^\Lambda + \mathbf{Q}^\Lambda\left(\boldsymbol{\Sigma}_\star\,\mathbf{S}\right)\tilde{\boldsymbol{\zeta}}'\mathbf{C}'\,\mathbf{Q}^\Lambda = \zeta_S^\star\,\mathbf{Q}^\Lambda\,\tilde{\boldsymbol{\Sigma}}\,\mathbf{Q}^\Lambda, \tag{S2.7}$$

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 (S2.6) by $\mathbf{Q}^\Lambda$ on both sides and simplifying using $\mathbf{Q}^\Lambda\boldsymbol{\Lambda}_\star = \mathbf{0}$ yields (S2.7).

Sufficiency: Suppose (S2.7) holds for a nonzero $\tilde{\boldsymbol{\zeta}}$ and a diagonal $\tilde{\boldsymbol{\Sigma}}$. Define the left-hand side of (S2.6) as

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

By (S2.7) 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, (S2.6) 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 (S2.7) 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 (S2.7) 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.

S2.2 Comparison with the Standard GIV

The size weighting in the optimal granular instrumental variables (GIV) estimator resembles the original GIV method of Gabaix and Koijen (2024). The two estimators are closely linked but differ in ways that matter for statistical power. 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 in Appendix A.1, the standard GIV further imposes homogeneity in elasticities across entities.[38] To closely compare two estimators, we also

[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).

consider the case with homogeneous elasticity so that $\zeta_i \equiv \zeta$, and we also consider the case where the size vector sums to 1, $\sum_i S_i = 1$.

With homogeneity, the optimal GIV estimator pools the sample moment conditions (3.3) across i weighted by the inverse of residual variance $(\sigma_i^2)^{-1}$ (see Appendix 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:

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

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

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

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:

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

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

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

Importantly, the term ζp_t drops out from the second term in the expectation due to differencing. As it is linear in ζ, it 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 difference nevertheless matters for 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. Moreover, 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 $\sqrt{T}$.

The power comparison matters in practice. 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 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,

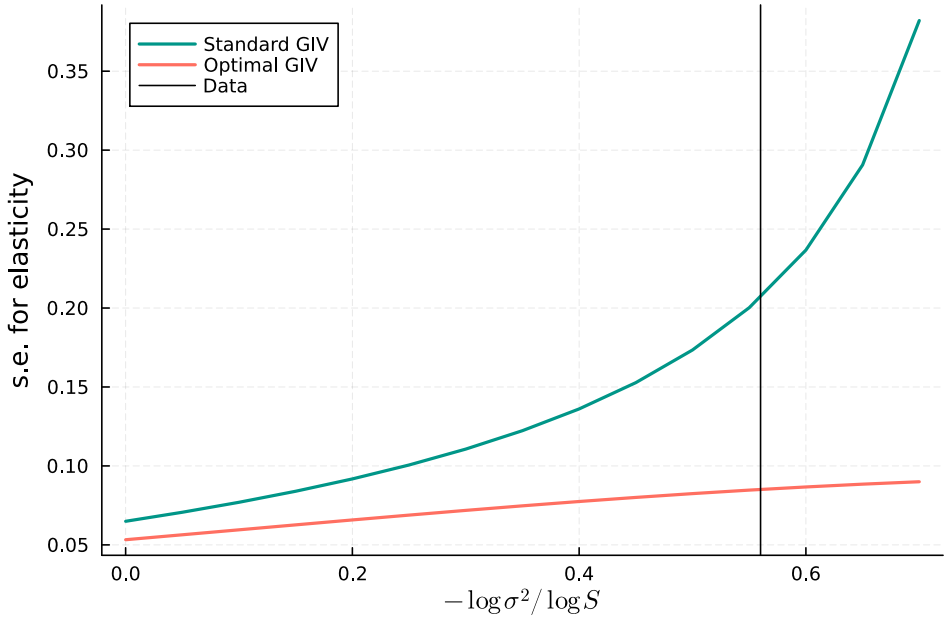
[39] See Appendix S4.1 for the derivation.
[40] See Gabaix and Koijen (2024) for the derivation.

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

Figure S2.1 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 Financial Accounts 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 S2.1. 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.

The higher statistical power of the optimal GIV comes with two trade-offs. 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 S2.1 describes how to extend the optimal GIV to account for unobserved common factors internally.

Figure S2.1: **Statistical power comparison: standard GIV vs. optimal GIV**



Note: This figure 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 $\sigma_p^2 = 2.5\%$.

S2.3 Structural Interpretations: Derivations

This subsection derives the micro-foundations summarized in Table A.1, using the notation of Appendix A.4.

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

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

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, with the notation of Appendix A.4, 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, Kyle (1985), Hellwig (1980), and Grossman and Stiglitz (1980).

Continuing from (S2.3), 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 (S2.3), 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 (Krishnamurthy & Vissing-Jorgensen, 2012; Nagel, 2016; Fu, Li, & Xie, 2022).

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):

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

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]

[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 $\kappa^L > \kappa^T$.

S2.4 Additional Discussions on Optimal GIV Estimator

S2.4.1 When only a subset of sectors are observed

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

$$\left.\begin{array}{rl} q_{i,t} & = -p_t \times \mathbf{C}_{i,t}'\boldsymbol{\zeta} + \mathbf{X}_{i,t}'\boldsymbol{\beta} + u_{i,t}, \\ q_{i,t}^{unobs} & = -\zeta_t^{unobs} p_t + \epsilon_t \\ 0 & = \sum_i S_{i,t} q_{i,t} + q_t^{unobs} \end{array}\right\} \implies \quad p_t = M_t \left[\mathbf{X}_{S,t}'\boldsymbol{\beta} + u_{S,t} + \epsilon_t\right], \tag{S2.8}$$

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' $\mathbf{C}_{i,t}'\boldsymbol{\zeta}$.

S2.4.2 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 $\text{Var}(\varepsilon_{i,t}) \propto \text{Var}(u_{i,t})$.

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 Financial Accounts data is a residual sector imputed using market clearing.

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

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

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

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

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

S2.4.3 Uniqueness of the optimal GIV solution

For the optimal weighting matrix $\mathbf{W}_t^*$, the second part of Assumption A.2 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 $\sigma_p^2 := \text{Var}(p_t \mid$

$\mathbf{C}_t, \mathbf{S}_t$). We have:

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{W}_t^{*\prime}\boldsymbol{\psi}_t\left(\tilde{\boldsymbol{\zeta}}\right)\right] &= \mathbf{0} \\
&= \mathbb{E}\left[M_t \sum_{i\neq j}\left(\frac{S_{jt}\mathbf{C}_{it}}{\sigma_i^2}+\frac{S_{it}\mathbf{C}_{jt}}{\sigma_j^2}\right)\left(u_{i,t}+p_t\mathbf{C}_{i,t}^{\prime}\Delta\boldsymbol{\zeta}\right)\left(u_{j,t}+p_t\mathbf{C}_{j,t}^{\prime}\Delta\boldsymbol{\zeta}\right)\right] \\
&= \mathbb{E}\left[M_t \sum_{i\neq j}\left(\frac{S_{jt}\mathbf{C}_{it}}{\sigma_i^2}+\frac{S_{it}\mathbf{C}_{jt}}{\sigma_j^2}\right)\left(M_t\left(S_{jt}\sigma_j^2\mathbf{C}_{i,t}^{\prime}+S_{it}\sigma_i^2\mathbf{C}_{j,t}^{\prime}\right)\Delta\boldsymbol{\zeta}+\sigma_p^2\Delta\boldsymbol{\zeta}^{\prime}\mathbf{C}_{i,t}\mathbf{C}_{j,t}^{\prime}\Delta\boldsymbol{\zeta}\right)\right] \\
&= \mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^{*}\right]\Delta\boldsymbol{\zeta}+\mathbb{E}\left[\sigma_p^2 M_t \sum_{i\neq j}\left(\frac{S_{jt}\mathbf{C}_{it}}{\sigma_i^2}+\frac{S_{it}\mathbf{C}_{jt}}{\sigma_j^2}\right)\left(\Delta\boldsymbol{\zeta}^{\prime}\mathbf{C}_{i,t}\mathbf{C}_{j,t}^{\prime}\Delta\boldsymbol{\zeta}\right)\right].
\end{aligned}
$$

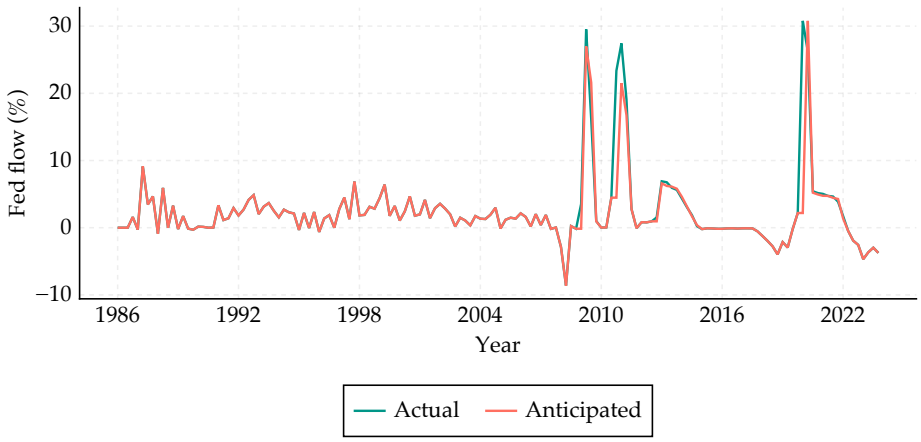
With non-negative $\mathbf{C}_{it}$ and non-singular $\mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{W}_t^{*}\right]$, the only solution to the equation above is $\Delta\boldsymbol{\zeta}=\mathbf{0}$. However, with unrestricted $\mathbf{C}_{i,t}$ potentially multiple roots can arise from this equation.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

S3 Additional Empirical Exhibits

S3.1 Scheduled Federal Reserve purchases: exhibits

Appendix B.1.2 describes the construction of the scheduled purchases. The exhibits below illustrate it. Figure S3.1 plots the path of the Fed's flows together with the one-quarter-ahead anticipated path sequence. For most quarters, the realized purchases closely track the anticipated purchases; the major deviations occur on the entry and exit of the quantitative easing (QE) programs. Table S3.1 reports the anticipated purchases versus the actual purchases of the Fed during active QE periods.

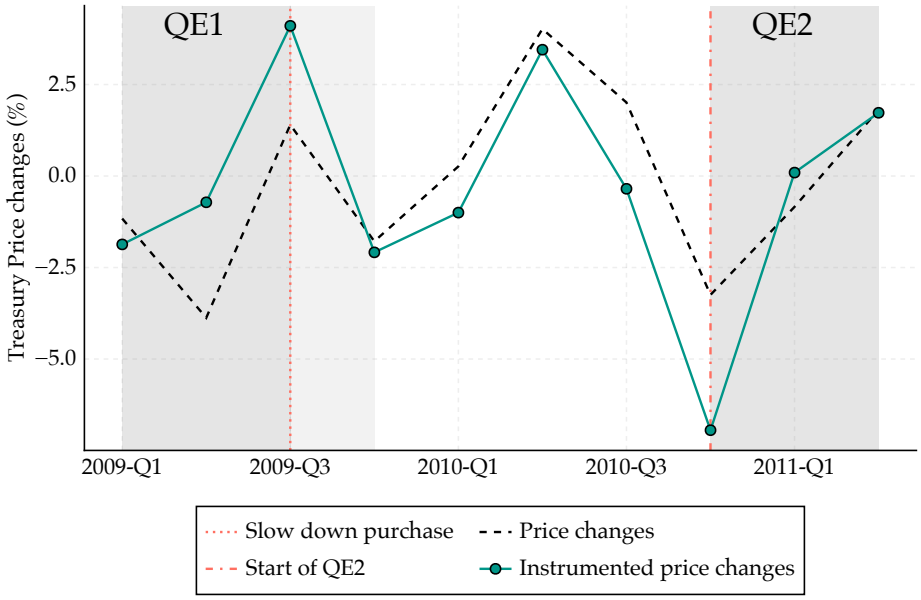
Figure S3.1: **Actual and anticipated Federal Reserve purchases**



Note: This figure plots actual Fed purchases of U.S. Treasury notes and bonds and the purchases anticipated one quarter ahead, normalized by total Fed holdings. Before 2009, anticipated purchases equal actual purchases; from the start of QE, they incorporate the New York Fed's pre-announced purchase schedule.

As a further illustration, in Figure S3.2 we focus on the end of the QE1 in 2009Q3 and the start of QE2 in 2010Q4. The original schedule for the first round of Treasury purchases ends at 2009Q3. In that quarter, the Federal Open Market Committee (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 S3.2: **Case Study: QE Entry and Exit between 2009 and 2010**



Note: This figure plots observed and predicted quarterly Treasury price changes from 2009 to 2011, covering the end of QE1 and the start of QE2. The dashed line shows observed price changes. The solid line shows instrumented price changes, predicted using the size-weighted idiosyncratic shocks of non-Fed sectors extracted from the GMM estimation.

Table S3.1: 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 relative to holdings
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	

S3.2 Case Study: Household and Mutual Fund Sectors

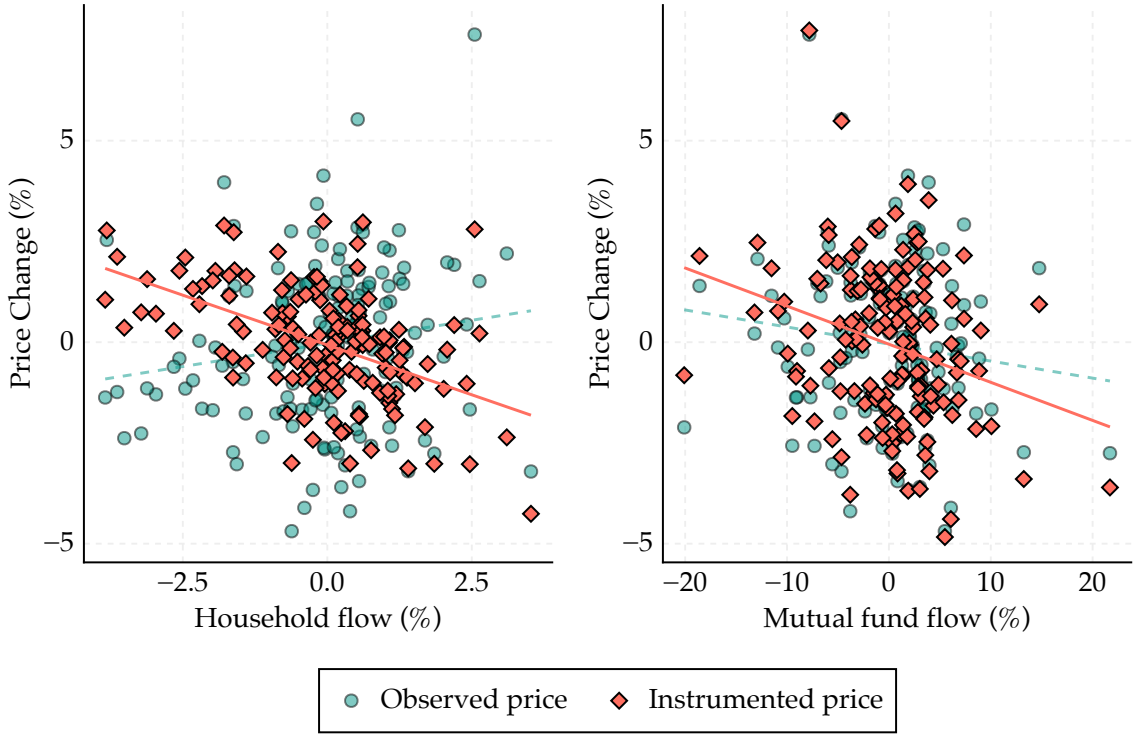
Figure S3.3 provides an illustration of our sector-specific estimates against the reduced-form OLS estimates. For the household sector and the mutual fund sector, we plot their "demand curves". We residualize sector inflow into Treasuries against common factors and plot it on the x-axis. The y-axis represents two notions of Treasury price changes. Using green markers, we plot the observed price changes, also residualized against the common factors. The observed-price relation slopes upward for households and slightly downward for mutual funds. 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 red, 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 in Section 3.3, the red lines summarize the relation between residualized flows and instrumented prices over the sample.[42] 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 (equation (A.7) in Appendix A.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 this and makes a larger correction for the household sector than for mutual funds.

[42] 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 corresponds to the leave-one-out estimator reported in Appendix Table S3.2.

Figure S3.3: **Optimal GIV vs. OLS: Two examples**



Note: This figure compares OLS and optimal GIV estimates of demand elasticities for households and mutual funds. In each panel, the horizontal axis shows the sector's net inflow into U.S. Treasury notes and bonds. The vertical axis shows quarterly Treasury price changes. Both flows and observed price changes are residualized against the common factors in the estimated specification. Green points use observed price changes; red points replace these with price changes predicted using size-weighted idiosyncratic shocks from the other sectors, extracted from the GMM estimation.

S3.3 Mapping Multiplier Estimates from the Literature

This appendix outlines how we map each paper's estimates to the multipliers plotted in Figure B.8.

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 (1994–2007). 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[43]) lowers the 10-year yield by 125 basis points. Scaling to a 1% shock gives a 1.21 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 (1961–1962) 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–Vila 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

[43] Financial Accounts table L.210, 2017Q3: total marketable notes-and-bonds liability less Federal Reserve SOMA holdings.

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 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).

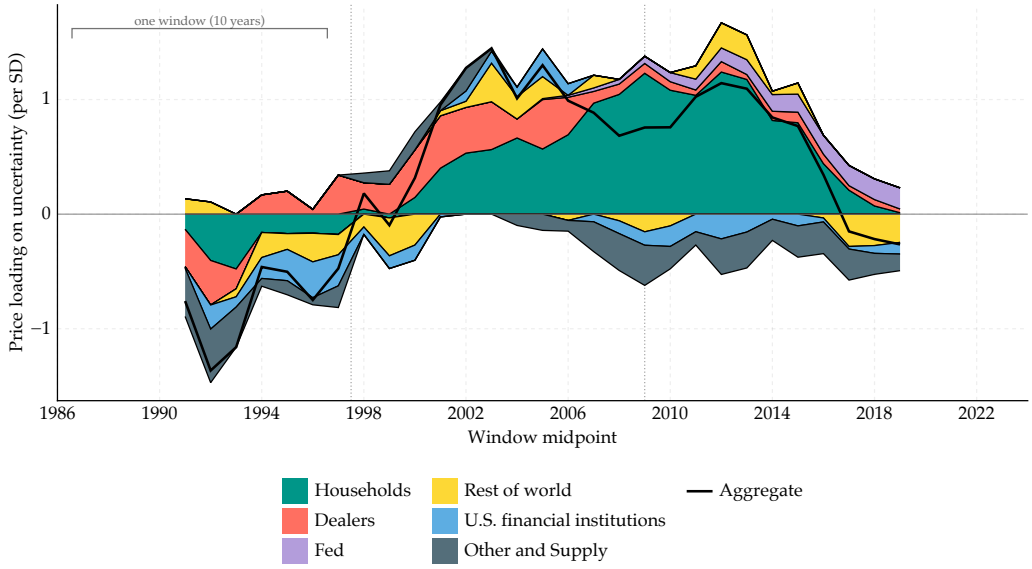
S3.4 Leave-one-out tables

Tables S3.2 and S3.3 report the estimates plotted in Figures B.2 and B.3.

S3.5 Safe haven: direct contributions and sector detail

Figures S3.4 and S3.5 are the direct-loading counterparts discussed in Appendix B.3.3. Table S3.4 reports each sector's price elasticity and its direct and total uncertainty responses in the first (panel a) and third (panel b) regimes.

Figure S3.4: **Ten-year rolling contributions to the safe haven: direct loadings**



Note: This figure plots investor groups' contributions to the price loading on uncertainty using direct loadings; see the note to Figure 4 for details on the rolling estimation and sector definitions. Windows are plotted at their midpoints; dotted vertical lines mark the regime shifts at 1997Q3 and 2009Q1.

S4 Proofs

This appendix proves the results stated in Appendix A and Theorem 1 of the main text.

S4.1 Proof of Theorem A.1

In this subsection, we proceed as if there were no exogenous variables $\mathbf{X}_{i,t}$. As we show in Appendix A.2, all results described here go through with observed common factors.

Table S3.2: **Price elasticity by sector: Leave-one-out Estimation**

Left out	Households†	Rest of world	Banks	Dealers†	Fed	Pension funds	Other	Mutual funds	Insurance	ETFs	Supply†	ζ_S
None	0.35***	0.35***	1.46***	0.06**	1.21***	0.41***	0.21**	0.59***	0.44***	−0.20	0.11***	1.01***
	(0.06)	(0.07)	(0.12)	(0.03)	(0.11)	(0.09)	(0.09)	(0.16)	(0.09)	(0.17)	(0.03)	(0.06)
Households	–	0.51***	1.22***	0.02	1.01	0.41**	0.20	0.54	0.47**	−0.16	0.10	0.98***
		(0.16)	(0.36)	(0.09)	(0.67)	(0.18)	(0.15)	(0.37)	(0.20)	(0.39)	(0.08)	(0.09)
Rest of world	0.39**	–	1.36***	0.04	1.22**	0.49**	0.05	0.67**	0.41*	−0.53	0.08**	0.97***
	(0.16)		(0.23)	(0.09)	(0.54)	(0.24)	(0.17)	(0.28)	(0.22)	(0.39)	(0.04)	(0.12)
Banks	0.39**	0.33***	–	0.07	0.94*	0.44**	0.17	0.69***	0.61**	−0.08	0.08*	1.03***
	(0.19)	(0.11)		(0.10)	(0.50)	(0.20)	(0.19)	(0.27)	(0.25)	(0.60)	(0.05)	(0.13)
Dealers	0.24	0.33**	1.50***	–	1.29**	0.48	0.20	0.47*	0.31	0.03	0.22	1.01***
	(0.19)	(0.13)	(0.28)		(0.61)	(0.30)	(0.15)	(0.24)	(0.21)	(0.44)	(0.26)	(0.09)
Fed	0.33**	0.36***	1.32***	0.07	–	0.46***	0.26*	0.45*	0.38**	−0.11	0.14***	1.02***
	(0.13)	(0.11)	(0.24)	(0.07)		(0.18)	(0.14)	(0.25)	(0.19)	(0.35)	(0.05)	(0.10)
Pension funds	0.34***	0.34***	1.45***	0.07	1.15**	–	0.19	0.61**	0.45**	−0.18	0.11**	1.01***
	(0.12)	(0.11)	(0.24)	(0.08)	(0.53)		(0.14)	(0.25)	(0.21)	(0.36)	(0.04)	(0.09)
Other	0.40***	0.25**	1.57***	0.02	1.15**	0.23	–	0.65**	0.50***	−0.33	0.12**	0.99***
	(0.14)	(0.11)	(0.25)	(0.06)	(0.54)	(0.16)		(0.27)	(0.17)	(0.31)	(0.05)	(0.12)
Mutual funds	0.38**	0.36***	1.51***	0.05	0.94**	0.39**	0.21	–	0.59***	0.01	0.10**	1.02***
	(0.16)	(0.10)	(0.24)	(0.07)	(0.47)	(0.18)	(0.15)		(0.22)	(0.56)	(0.05)	(0.12)
Insurance	0.36***	0.33***	1.52***	0.05	1.13**	0.42**	0.22	0.66***	–	−0.19	0.10**	1.00***
	(0.13)	(0.10)	(0.24)	(0.07)	(0.52)	(0.18)	(0.15)	(0.25)		(0.36)	(0.04)	(0.10)
ETFs	0.35***	0.35***	1.45***	0.06	1.15**	0.42**	0.20	0.60**	0.44**	–	0.11**	1.00***
	(0.13)	(0.10)	(0.23)	(0.07)	(0.53)	(0.18)	(0.14)	(0.25)	(0.19)		(0.04)	(0.10)
Supply	0.21	0.40***	1.23***	0.19	1.77*	0.31	0.36**	0.60**	0.20	−0.86***	–	1.04***
	(0.24)	(0.13)	(0.24)	(0.45)	(0.91)	(0.24)	(0.15)	(0.28)	(0.18)	(0.27)		(0.18)

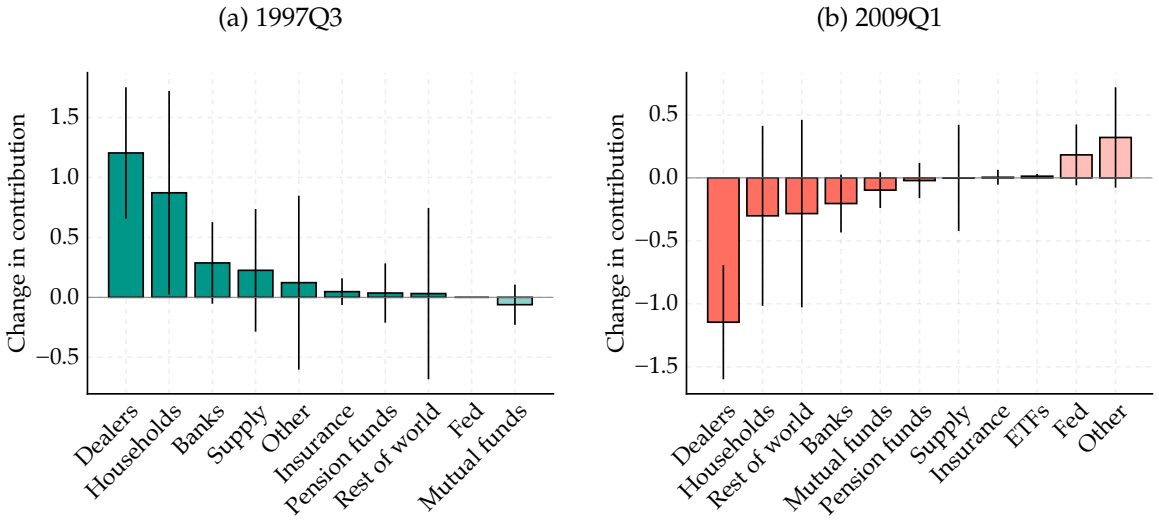
Note: This table reports leave-one-out estimates of the price elasticities in Table 1. Each row re-estimates the specification without the moment conditions of the named sector; each column reports a sector's price elasticity ζ_i, averaged over its observed quarters. The columns for the Fed and ETFs cover 2009Q1–2023Q4. Sectors marked † have unit size weights, so their coefficients measure demand responses relative to the total market; the other sectors' coefficients are relative to their own holdings. The first row uses all sectors and standard errors from equation (A.5); the other rows use the sandwich variance in equation (A.6). Standard errors are in parentheses below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table S3.3: **Change in price elasticities across 2009Q1: Leave-one-out Estimation**

	Elasticity change, post−pre 2009Q1										Aggregate ζ_S		
Left out	Households†	Rest of world	Banks	Dealers†	Fed	Pension funds	Other	Mutual funds	Insurance	Supply†	Pre	Post	Change
None	−0.16	−0.43***	−0.57**	−0.06	1.21***	−0.33**	0.55***	−0.25	−0.06	−0.06	1.14***	0.82***	−0.32***
	(0.11)	(0.13)	(0.23)	(0.06)	(0.11)	(0.16)	(0.17)	(0.29)	(0.16)	(0.05)	(0.10)	(0.07)	(0.12)
Households	–	−0.73**	−1.04	−0.01	1.01	−0.37	0.57*	0.73	−0.03	−0.09	1.14***	0.74***	−0.40**
		(0.30)	(0.80)	(0.15)	(0.67)	(0.33)	(0.29)	(0.76)	(0.37)	(0.14)	(0.11)	(0.13)	(0.17)
Rest of world	−0.29	–	−0.44	−0.06	1.22**	−0.12	0.50	−0.69	−0.06	−0.02	1.13***	0.73***	−0.40*
	(0.28)		(0.48)	(0.16)	(0.54)	(0.43)	(0.32)	(0.51)	(0.40)	(0.08)	(0.18)	(0.11)	(0.21)
Banks	−0.09	−0.38*	–	−0.09	0.94*	−0.30	0.55	0.16	0.02	−0.05	1.14***	0.86***	−0.28
	(0.41)	(0.21)		(0.16)	(0.50)	(0.38)	(0.38)	(0.54)	(0.52)	(0.09)	(0.15)	(0.23)	(0.28)
Dealers	−0.04	−0.42*	−0.73	–	1.29**	−0.44	0.54*	−0.12	0.11	−0.24	1.16***	0.79***	−0.36**
	(0.34)	(0.24)	(0.58)		(0.61)	(0.51)	(0.29)	(0.46)	(0.38)	(0.44)	(0.14)	(0.11)	(0.17)
Fed	−0.23	−0.41*	−0.90*	−0.03	–	−0.24	0.67**	−0.62	−0.21	0.03	1.13***	0.84***	−0.29
	(0.24)	(0.21)	(0.51)	(0.12)		(0.32)	(0.28)	(0.49)	(0.34)	(0.12)	(0.15)	(0.11)	(0.19)
Pension funds	−0.14	−0.41**	−0.58	−0.08	1.15**	–	0.52*	−0.22	−0.01	−0.08	1.14***	0.80***	−0.34*
	(0.23)	(0.21)	(0.51)	(0.13)	(0.53)		(0.28)	(0.47)	(0.38)	(0.08)	(0.13)	(0.13)	(0.18)
Other	−0.26	−0.31	−0.75	0.01	1.15**	−0.37	–	−0.20	−0.19	−0.02	1.10***	0.82***	−0.28
	(0.27)	(0.20)	(0.55)	(0.10)	(0.54)	(0.30)		(0.51)	(0.32)	(0.10)	(0.18)	(0.14)	(0.23)
Mutual funds	0.01	−0.47**	−0.37	−0.06	0.94**	−0.39	0.41	–	0.18	−0.08	1.12***	0.87***	−0.25
	(0.34)	(0.21)	(0.53)	(0.12)	(0.47)	(0.35)	(0.33)		(0.46)	(0.09)	(0.15)	(0.20)	(0.25)
Insurance	−0.18	−0.38*	−0.69	−0.05	1.13**	−0.37	0.47	−0.29	–	−0.04	1.14***	0.80***	−0.34*
	(0.25)	(0.20)	(0.51)	(0.12)	(0.52)	(0.33)	(0.29)	(0.47)		(0.08)	(0.15)	(0.13)	(0.20)
ETFs	−0.18	−0.42**	−0.59	−0.06	1.15**	−0.36	0.53*	−0.23	−0.05	−0.05	1.13***	0.80***	−0.33*
	(0.24)	(0.20)	(0.50)	(0.12)	(0.53)	(0.32)	(0.28)	(0.47)	(0.35)	(0.08)	(0.15)	(0.13)	(0.19)
Supply	−0.20	−0.37	−0.24	−0.29	1.77*	−0.32	0.58**	−0.52	−0.19	–	1.19***	0.80***	−0.39
	(0.41)	(0.25)	(0.55)	(0.74)	(0.91)	(0.42)	(0.28)	(0.52)	(0.32)		(0.28)	(0.13)	(0.31)

Note: This table reports the post- minus pre-2009 change in the sector price elasticities ζ_i from the leave-one-out estimates of Table S3.2, and the aggregate elasticity ζ_S before and after 2009Q1. Sectors marked † have unit size weights, so their coefficients measure demand responses relative to the total market; the other sectors' coefficients are relative to their own holdings. Standard errors are in parentheses below estimates, computed as in Table S3.2. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Figure S3.5: **Changes in direct uncertainty contributions**



Note: This figure plots changes in each sector's contribution to the aggregate direct uncertainty loading at 1997Q3 (panel (a)) and 2009Q1 (panel (b)), using the specification of Figure 5. Sectors, including supply, are shown separately and ranked by their changes. Whiskers are 95% confidence intervals, from equation (A.9).

Table S3.4: **Sector elasticities and uncertainty responses in the first and third regimes**

(a) 1986Q1–1997Q2

	Elasticity		Direct uncertainty		Total uncertainty	
	Loading	Contribution	Loading	Contribution	Loading	Contribution
Aggregate	1.04*** [0.76, 1.33]	—	−1.12** [−2.20, −0.03]	—	−0.35 [−1.77, 1.07]	—
Households†	0.33** [0.03, 0.64]	0.33** [0.03, 0.64]	−0.15 [−0.77, 0.46]	−0.15 [−0.77, 0.46]	−0.02 [−0.63, 0.58]	−0.02 [−0.63, 0.58]
Dealers†	−0.11 [−0.24, 0.03]	−0.11 [−0.24, 0.03]	−0.13 [−0.52, 0.26]	−0.13 [−0.52, 0.26]	−0.10 [−0.50, 0.31]	−0.10 [−0.50, 0.31]
Rest of world	0.64*** [0.35, 0.93]	0.13*** [0.07, 0.19]	−0.16 [−1.56, 1.24]	−0.03 [−0.32, 0.25]	0.28 [−1.09, 1.64]	0.06 [−0.22, 0.34]
Fed						
Banks	1.81*** [1.36, 2.27]	0.28*** [0.21, 0.35]	−0.84 [−2.61, 0.93]	−0.13 [−0.40, 0.14]	−0.09 [−2.32, 2.14]	−0.01 [−0.35, 0.33]
Pension funds	1.00*** [0.60, 1.40]	0.18*** [0.11, 0.25]	−0.40 [−1.64, 0.83]	−0.07 [−0.29, 0.15]	0.48 [−0.87, 1.84]	0.08 [−0.15, 0.32]
Insurance	0.21 [−0.18, 0.60]	0.02 [−0.02, 0.06]	−0.69 [−1.74, 0.36]	−0.06 [−0.16, 0.03]	−0.06 [−1.18, 1.06]	−0.01 [−0.11, 0.10]
Mutual funds	0.70** [0.02, 1.37]	0.04** [0.00, 0.08]	−0.17 [−2.39, 2.04]	−0.01 [−0.14, 0.12]	0.03 [−2.08, 2.15]	0.00 [−0.12, 0.12]
Other	−0.01 [−0.42, 0.41]	0.00 [−0.09, 0.08]	−1.62 [−4.38, 1.15]	−0.33 [−0.91, 0.24]	−1.00 [−3.45, 1.44]	−0.21 [−0.71, 0.30]
Supply†	0.17*** [0.05, 0.30]	0.17*** [0.05, 0.30]	−0.19 [−0.55, 0.18]	−0.19 [−0.55, 0.18]	−0.15 [−0.48, 0.18]	−0.15 [−0.48, 0.18]

(b) 2009Q1–2023Q4

	Elasticity		Direct uncertainty		Total uncertainty	
	Loading	Contribution	Loading	Contribution	Loading	Contribution
Aggregate	0.83*** [0.70, 0.97]	—	0.11 [−0.65, 0.88]	—	0.43 [−0.70, 1.55]	—
Households†	0.15*** [0.04, 0.26]	0.15*** [0.04, 0.26]	0.42** [0.05, 0.78]	0.42** [0.05, 0.78]	0.40** [0.06, 0.74]	0.40** [0.06, 0.74]
Dealers†	0.04 [−0.02, 0.11]	0.04 [−0.02, 0.11]	−0.07 [−0.31, 0.17]	−0.07 [−0.31, 0.17]	0.03 [−0.22, 0.29]	0.03 [−0.22, 0.29]
Rest of world	0.15* [0.00, 0.31]	0.07* [0.00, 0.14]	−0.61 [−1.66, 0.43]	−0.29 [−0.77, 0.20]	−0.80 [−1.77, 0.17]	−0.37 [−0.82, 0.08]
Fed	1.18*** [0.96, 1.40]	0.26*** [0.21, 0.31]	0.83 [−0.27, 1.94]	0.18 [−0.06, 0.42]	2.18 [−0.61, 4.97]	0.48 [−0.13, 1.09]
Banks	1.06*** [0.79, 1.32]	0.05*** [0.04, 0.07]	−0.91 [−3.07, 1.26]	−0.05 [−0.15, 0.06]	−0.10 [−2.05, 1.86]	0.00 [−0.10, 0.09]
Pension funds	0.17* [−0.01, 0.36]	0.01* [0.00, 0.02]	−1.07*** [−1.81, −0.34]	−0.06*** [−0.10, −0.02]	−1.35*** [−2.05, −0.65]	−0.07*** [−0.11, −0.03]
Insurance	0.28*** [0.09, 0.46]	0.01*** [0.00, 0.01]	−0.47 [−1.14, 0.19]	−0.01 [−0.03, 0.01]	−0.50 [−1.15, 0.16]	−0.01 [−0.03, 0.00]
Mutual funds	0.34* [0.00, 0.68]	0.02* [0.00, 0.05]	−2.51*** [−3.98, −1.05]	−0.17*** [−0.27, −0.07]	−1.98*** [−3.38, −0.59]	−0.13*** [−0.23, −0.04]
ETFs	−0.34** [−0.67, −0.01]	0.00** [−0.01, 0.00]	1.19 [−0.33, 2.70]	0.01 [0.00, 0.03]	1.36** [0.11, 2.60]	0.02** [0.00, 0.03]
Other	0.68*** [0.49, 0.87]	0.08*** [0.06, 0.10]	0.94 [−0.51, 2.39]	0.11 [−0.06, 0.28]	0.60 [−0.71, 1.92]	0.07 [−0.08, 0.22]
Supply†	0.15*** [0.09, 0.21]	0.15*** [0.09, 0.21]	0.04 [−0.19, 0.26]	0.04 [−0.19, 0.26]	0.02 [−0.19, 0.23]	0.02 [−0.19, 0.23]

Note: This table reports each sector's price elasticity ζ_i and its direct and total uncertainty loadings λ_i^{unc}, each with its contribution, $S_i\zeta_i$ or $S_i\lambda_i^{unc}$, from the specification of Table 3; each panel covers one regime. The Banks, Pension funds, Insurance, Mutual funds, ETFs, and Other rows make up Remaining domestic in Table 3; ETFs enter the sample in 2009Q1. Sectors marked † have unit size weights, so their coefficients measure demand responses relative to the total market; the other sectors' coefficients are relative to their own holdings. The Federal Reserve's pre-2009 purchases are treated as fully anticipated through the scheduled-purchase control, so its pre-2009 coefficients are not estimated and are left blank. 95% confidence intervals, computed as in Tables 1 and 3, are in brackets below estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

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.

S4.1.1 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:

$$\begin{aligned}\mathbb{E}\left[\psi_{ij,t}(\boldsymbol{\zeta}) \mid \mathbf{C}_t, \mathbf{S}_t\right] &= \mathbb{E}\left[\left(q_{i,t} + p_t \mathbf{C}'_{i,t}\boldsymbol{\zeta}\right)\left(q_{j,t} + p_t \mathbf{C}'_{j,t}\boldsymbol{\zeta}\right) \mid \mathbf{C}_t, \mathbf{S}_t\right] \\ &= \mathbb{E}\left[u_{i,t}u_{j,t} \mid \mathbf{C}_t, \mathbf{S}_t\right] = \mathbf{0}. \end{aligned} \tag{S4.1}$$

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:[44]

$$\mathbf{b}_T^{\mathbf{W}}\left(\hat{\boldsymbol{\zeta}}^W\right) \equiv \frac{1}{T}\sum_t W_t\left(\hat{\boldsymbol{\zeta}}^W\right)' \boldsymbol{\psi}_t\left(\hat{\boldsymbol{\zeta}}^W\right) = 0. \tag{S4.2}$$

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

[44] 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).

Assumption A.2 ensures the estimator is well-behaved.

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

Lemma S4.1 (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:*

$$\mathbf{V}_\zeta^W = \mathbb{E}\left[\mathbf{W}_t'\mathbf{D}_t\right]^{-1}\mathbb{E}\left[\mathbf{W}_t'\mathbf{V}^\psi\mathbf{W}_t\right]\mathbb{E}\left[\mathbf{D}_t'\mathbf{W}_t\right]^{-1} \tag{S4.3}$$

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

Proof. The consistency of the estimator directly follows the standard GMM argument. To 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 $\mathbf{D}_{W,t} = \frac{\partial(\mathbf{W}_t(\mathbf{z})'\boldsymbol{\psi}_t(\mathbf{z}))}{\partial \mathbf{z}}\,|_{z=\zeta}$. 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 (S4.3). □

S4.1.2 Semiparametric Efficiency Bound

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

Lemma S4.2 (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:*

$$\mathbf{W}_t^* = \left(\mathbf{V}^\psi\right)^{-1}\mathbf{D}_t, \tag{S4.4}$$

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

$$\mathbf{V}_\zeta^{W*} = \mathbb{E}\left[\mathbf{D}_t'\left(\mathbf{V}^\psi\right)^{-1}\mathbf{D}_t\right]^{-1} = \mathbb{E}\left[\mathbf{W}_t^{*\prime}\mathbf{V}^\psi\mathbf{W}_t^*\right]^{-1}. \tag{S4.5}$$

Proof. The proof for the optimality generally follows the strategy of optimal instrument

in conditional models, surveyed in Newey (1993).

To show Equation (S4.5), 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 (S4.5).

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 $\mathbf{\Delta} \equiv \left(\mathbf{V}_\zeta^{W^*}\right)^{-1} - \left(\mathbf{V}_\zeta^W\right)^{-1}$ is positive semi-definite:

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

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 $\mathbf{\Delta}$ as:

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

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

S4.1.3 Deriving the optimal GIV estimator

The remaining step is to show Equation (S4.4) 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:

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

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 (S4.4), we have:

$$\mathbf{W}^*_{ij,k,t} = \left(V_{ij}^{\psi}\right)^{-1} D_{ij,k,t} = \frac{1}{\mathbf{C}'_{S,t}\boldsymbol{\zeta}} \left(\frac{S_{j,t}C_{it,k}}{\sigma_i^2} + \frac{S_{i,t}C_{jt,k}}{\sigma_j^2}\right), \tag{S4.6}$$

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 $k, \ell \in \{1, \ldots, N\}$,

$$\left[\mathbb{E}[\mathbf{W}_t^{*\prime}\mathbf{V}^{\psi}\mathbf{W}_t^*]\right]_{k\ell} = \frac{1}{\zeta_S^2}\begin{cases} \dfrac{\Sigma}{\sigma_k^2} - S_k^2, & k = \ell, \\ S_k S_\ell, & k \neq \ell. \end{cases} \tag{S4.7}$$

Inverting this matrix and scaling by ζ_S^2 gives the asymptotic variance in equation (3.4) of the main text.

S4.2 Proof of Proposition A.1

Proof. The proof follows Newey and McFadden (1994). The estimator of Appendix A.2 first regresses $q_{i,t}$ and $p_t \mathbf{C}'_{i,t}$ on $\mathbf{X}_{i,t}$ and then applies the optimal GIV to the residuals. This 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 $\mathbf{g}^{OLS}(\mathbf{b})$ is the moment condition of the first step, which identifies $(\boldsymbol{\beta}^q, \boldsymbol{\beta}^{pC})$, and the second block contains the optimal GIV moment conditions, taking the OLS coefficients as given.[45] As the coefficients are just-identified, the solutions to the GMM estimation are identical to the two-step estimator, and hence we can utilize properties of the GMM estimator to study the asymptotic behaviors. Use $\boldsymbol{\theta}$ to denote the vector of $(\boldsymbol{\beta}, \boldsymbol{\zeta})$ and $\hat{\boldsymbol{\theta}}$ the corresponding estimator, we have:

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

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

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

To see $\mathbf{D}^g$ is diagonal, notice that the estimate for $\boldsymbol{\zeta}$ does not enter the moment con-

[45] 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.8). The derivation is slightly more complicated as the (A.8) involves both $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ in the moment conditions, but the results are equivalent.

ditions 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 $\mathbf{g}_t^{OLS}$ 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:

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

where the first part is equivalent to asymptotic variance of the OLS estimator of regressing $q_{i,t}+p_t\mathbf{C}_{i,t}'\boldsymbol{\zeta}$ on $\mathbf{X}_{i,t}$, which is given in (A.10). □

S4.3 Proof of Proposition A.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.11), it can be expressed as:

$$u_{i,t}^b=-\left(\zeta_i-\zeta_i^b\right)\Delta p_t+u_{i,t} \tag{S4.8}$$

Plugging Equation (A.12) into the misspecified residual, and denoting $b_i := \frac{\zeta_i-\zeta_i^b}{\zeta_S}$, we have:

$$u_{i,t}^b=-b_iu_{S,t}+u_{i,t} \tag{S4.9}$$

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

$$\begin{aligned}\mathbb{E}\left[u_{i,t}^b \sum_{j\neq i} S_j u_{j,t}^b\right] &= \mathbb{E}\left[u_{i,t}^b \left(u_{S,t}^b - S_i u_{i,t}^b\right)\right], \\ &= \mathbb{E}\Big[\big(-b_i u_{S,t} + u_{i,t}\big)\big((1-b_S+S_i b_i)u_{S,t} - S_i u_{i,t}\big)\Big]\end{aligned} \tag{S4.10}$$

Express it in terms of the moments of shocks:

$$-b_i(1-b_S+S_i b_i)\sigma_{u,S}^2 + (1-b_S+2S_i b_i)\sigma_{i,S} - S_i\sigma_{u,i}^2 = 0 \tag{S4.11}$$

where

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

using $\mathbb{E}[\eta_t\epsilon_{i,t}] = 0$ and $\mathbb{E}[\epsilon_{i,t}\epsilon_{j,t}] = 0$ for $i \neq j$. Equation (S4.11) 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 (S4.11) are symmetric, and hence $\bar{b}_i = \bar{b}_S = \bar{b}$.

Plugging in and simplifying, we have:

$$\bar{b}^2 - 2\bar{b} + \mathcal{R}^2 = 0 \tag{S4.12}$$

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.15).

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

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

We can express Equation (S4.11) as a function of deviations from the symmetric benchmark:

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

We fix $\sigma_{u,S}^2$ and λ_S because the perturbation in heterogeneity $(\omega_i, \delta_i, \kappa_i)$ does not enter λ_S and $\sigma_{u,S}^2$ 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:

$$(\partial_b F_i + \partial_{b_S} F_i)\ \Delta b_S + O(\xi^2) = 0 \tag{S4.13}$$

As we show below, $\partial_b F_i + \partial_{b_S} F_i \neq 0$. Hence, Equation (S4.13) 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:

$$\begin{aligned}
\frac{\partial F_i}{\partial\, \Delta b_i} &= \sigma_{u,S}^2\,(1-\bar{b})\,(2\bar{S}-1), \\
\frac{\partial F_i}{\partial\, \Delta b_S} &= -\sigma_{u,S}^2\,(1-\bar{b}), \\
\frac{\partial F_i}{\partial\, \omega_i} &= \sigma_{u,S}^2\,(1-\bar{b})^2\,\bar{b}\,(2\bar{S}-1), \\
\frac{\partial F_i}{\partial\, \delta_i} &= -\lambda_S^2\sigma_\eta^2\,(1-\bar{b})\,(2\bar{S}-1) = -\sigma_{u,S}^2\,\mathcal{R}^2\,(1-\bar{b})\,(2\bar{S}-1), \\
\frac{\partial F_i}{\partial\, \kappa_i} &= \sigma_{u,S}^2\,(1-\bar{b})^2\,\bar{b}\,(2\bar{S}-1).
\end{aligned} \tag{S4.14}$$

The derivatives with respect to ω_i and κ_i coincide because, to first order, both only scale $S_i\sigma_{\epsilon,i}^2$. Linearizing the moments gives $\partial\sigma_{i,S}/\partial\omega_i = \partial\sigma_{i,S}/\partial\kappa_i = \bar{S}\bar{\sigma}_\epsilon^2$, $\partial\sigma_{u,i}^2/\partial\omega_i = 0$, $\partial\sigma_{u,i}^2/\partial\kappa_i = \bar{\sigma}_\epsilon^2$, and $\partial S_i/\partial\omega_i = \bar{S}$; evaluating F_i at the symmetric benchmark and using $\bar{b}^2 - 2\bar{b} + \mathcal{R}^2 = 0$ reduces both derivatives to $\bar{S}\bar{\sigma}_\epsilon^2\bar{b}(2\bar{S}-1)$, which equals $\sigma_{u,S}^2(1-\bar{b})^2\bar{b}(2\bar{S}-1)$ because $\bar{S}\bar{\sigma}_\epsilon^2 = \sigma_{u,S}^2(1-\mathcal{R}^2)$ and $1-\mathcal{R}^2 = (1-\bar{b})^2$. 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) follows from $b_S = \sum_i S_i b_i$. (ii) $\sum_i \omega_i = 0$ by construction, so $\sum_i S_i\omega_i = \bar{S}\sum_i(1+\omega_i)\omega_i = \bar{S}\sum_i \omega_i^2 = O(\xi^2)$. (iii) $\sum_i S_i\delta_i = \lambda_S^{-1}\sum_i S_i\lambda_i - \sum_i S_i = 0$. (iv) The definition of $\bar{\sigma}_\epsilon^2$ gives $\sum_i S_i^2\kappa_i = \sum_i \bar{S}^2 - \sum_i S_i^2 = -\bar{S}^2\sum_i \omega_i^2 = O(\xi^2)$. Writing $S_i = S_i^2/\bar{S} - \bar{S}(\omega_i + \omega_i^2)$ then gives $\sum_i S_i\kappa_i = \bar{S}^{-1}\sum_i S_i^2\kappa_i - \bar{S}\sum_i(\omega_i+\omega_i^2)\kappa_i$; the first term is $O(\xi^2)$, and so is the second by Cauchy–Schwarz, since $\sum_i(\omega_i+\omega_i^2)^2 = O(\xi^2)$ and $\sum_i \kappa_i^2 \le C(N)\sum_i S_i^2\kappa_i^2 \le C(N)\xi^2$ for fixed N. □