Is Asset Demand Elasticity Set at the Household or Intermediary Level?*

Ehsan Azarmsa† Carter Davis‡

December 31, 2023

PRELIMINARY

Abstract

Household-based and intermediary-based asset pricing models disagree about the elasticity of the allocations to intermediaries. Household-based models (e.g., Lucas (1978); Campbell and Cochrane (1999); Bansal and Yaron (2004)) focus on households' risk-return trade-offs, implying that the allocation to intermediaries is so elastic that renders the intermediaries' portfolio behavior irrelevant. In contrast, intermediary-based models (e.g., He and Krishnamurthy (2013); Koijen and Yogo (2019); Haddad and Muir (2021)) emphasize households' inelastic allocations, leading to drastically different pricing predictions. We shed light on this discrepancy by examining households' allocations to intermediaries and estimating their price elasticity in the 13F data of institutional holdings. In a variance decomposition exercise, we find that households primarily respond to intermediaries' excess demand for stocks by rebalancing their direct stock holdings, while their allocation to intermediaries exacerbates the demand pressure by about 10%. Consistent with theory, allocations to some intermediary types, such as mutual funds and investment advisors, exhibit a negative and significant relationship with the price of their portfolio assets. However, the elasticity of these allocations is not large enough to have a first-order impact on the aggregate demand elasticity for assets. Our results support the central premise of intermediary-based asset pricing models: households do not reallocate enough to eliminate mispricings induced by intermediary-level frictions.

JEL Classification: G11, G12, G23

Keywords: Demand system, Intermediary asset pricing, Investor inattention

*We are grateful to Auditya Chaudhry (Discussant), Valentin Haddad, Zhiguo He, Paul Huebner, Ralph Koijen, Stefan Nagel, Jiacui Li, Kerry Siani (Discussant), Michael Weber, and conference participants at SFS Cavalcade NA 2023 and Yiran Fan Memorial Conference for helpful comments and discussions.

†College of Business Administration, University of Illinois Chicago; azarmsa@uic.edu

‡Kelley School of Business, Indiana University; ckd1@iu.edu

1 Introduction

The empirical asset pricing literature points to the importance of intermediary level frictions, such as balance-sheet constraints (Adrian et al., 2014; He et al., 2017), arbitrage limitations (Shleifer and Vishny, 1997; Brunnermeier and Pedersen, 2009), and investment mandates (Gabaix and Koijen, 2022), for asset returns, both at the cross-section and time-series. These observations led to the development of intermediary-based asset pricing models (He and Krishnamurthy, 2013). Yet, it is empirically challenging to rule out the possibility that intermediary-level variables are correlated with some macroeconomic-variables, so that the prices might still be efficient from the perspective of the representative investor under a correct specification of the preferences, beliefs, and endowment processes. In other words, households are the end owners of the assets, so they might still reallocate enough funds across the intermediaries to eliminate any mispricing induced by the frictions at the intermediary level.

As such, these two views depart in their implication for how actively households rebalance in response to non-fundamental price movements. The representative-investor view posits that household-level allocations are so elastic that renders the intermediary demand irrelevant. However, the intermediary view requires some inaction or market segmentation to make households' demand for assets inelastic. In this case, the aggregate demand elasticity is primarily set by the intermediaries, and would primarily depend on the size of arbitrage capital available to actively managed funds, investment mandates, and constraints faced by intermediaries. Certainly reality can be between these two extremes as well.

Two-layer Demand System Estimation

We contribute to this discussion by decomposing the aggregate demand elasticity into a direct component, which is the sum of the demand elasticity of the direct owners of assets, including the institutional holdings and households' direct holdings, and an indirect component, which captures the elasticity of the allocations to the institutional investors. Figures 1 and 2 clarify what different asset pricing models imply about the direct and indirect elasticities.

As suggested by Figures 1 and 2, the main distinction between household-based and intermediary-based asset pricing models is in their assumption about the indirect elasticity, that is, how elastic the capital flows to intermediaries are with respect to price movements. To understand this distinction, consider institution i that holds asset n. Instituition i's demand for this asset can

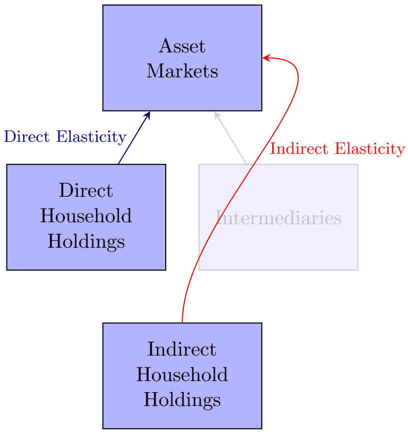


Figure 1: Classic Asset Pricing View

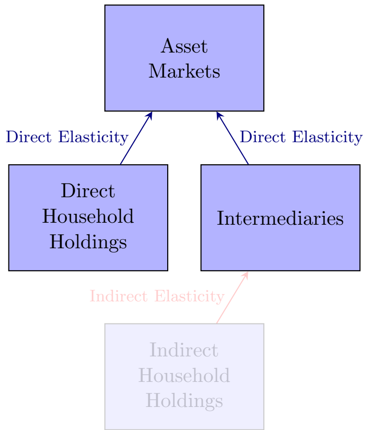


Figure 2: Demand System Asset Pricing View

be written as $Q_{i,t}(n) = A_{i,t}w_{i,t}(n)/P_t(n)$, where $A_{i,t}$ is the assets under management (AUM) of institution i, $w_{i,t}(n)$ is the portfolio weight of asset n, and $P_t(n)$ is the share price. The elasticity of institution i's demand is the derivative of log demand with respect to log price:

$$\eta_{i,t}(n) \equiv -\frac{\partial \log(Q_{i,t}(n))}{\partial \log(P_{i,t}(n))} = 1 - \underbrace{\frac{\partial \log(A_{i,t})}{\partial \log(P_t(n))}}_{\substack{\text{Indirect elasticity} \\ \text{(household-level)}}} - \underbrace{\frac{\partial \log(w_{i,t}(n))}{\partial \log(P_t(n))}}_{\substack{\text{Direct elasticity} \\ \text{(intermediary-level)}}} \tag{1}$$

This single equation nicely highlights the tension between household and intermediary-based models. In many households-based models, the indirect elasticity is so large that renders the intermediary-level elsaticity irrelevant. In these models, intermediaries are just veils through which households act. However, the most extreme intermediary-focused models assume that assets under management are exogenous to prices—i.e. $\partial \log(A_{i,t})/\partial \log(P_t(n)) = 0$ (e.g. Koijen and Yogo (2019)). In this world, investors have really *entrusted* capital to intermediaries, and while flows may be a function of *past returns*, they are not a function of *current prices or valuation ratios* of the fund's assets. In these models, frictions at the intermediary level are not undone by investors.

The purpose of this paper is to empirically disentangle the direct and indirect elasticity components of the aggregate asset demand. The relative contribution of these two terms is informative about whether households' reallocations undo the mispricings caused by intermediary-level frictions.

To this end, we develop a model featuring a representative investor who invests in some risky assets either directly, or indirectly through some intermediaries. The representative investor also has access to some outside funds (e.g., bonds) available to transfer to her direct or indirect holdings.

There are four types of reallocations in the model, which cancel out in the aggregate under the assumption that the asset supply is fixed: Intermediaries' rebalancing of their portfolios, the representative investor's rebalancing of her direct risky investments, the investor's transfers of funds in and out of her direct risky investments, and lastly, the investor's adjustments of her allocations to the intermediaries. Putting it differently, if intermediaries have an excess demand for an asset, the investor has to respond either by adjusting her direct holdings (via rebalancing her portfolio of risky assets or rebalancing across risky assets and outside assets), or by reducing her allocation to the intermediaries. We develop a reallocation identity that formalizes this idea and analyze the investors' response to such purchase (or selling) pressures by estimating the contribution of each term in a variance decomposition exercise.

We find that about 90% of the intermediary-level demand pressure is absorbed by rebalancings within the direct holdings, whereas the adjustments in the intermediated investments exacerbate the demand pressure by about 10%. The mutual fund flows account for 7% out of this 10%, consistent with the empirical evidence on the positive flow-performance relationship (Chevalier and Ellison, 1997; Lou, 2012)): An increase in the intermediary-level demand for an asset drives up its price, which attracts investors' flow. As such, the latent demand for intermediaries' equity is likely to be correlated with the prices, warranting an IV strategy to estimate the price elasticity of the capital allocations to intermediaries.

We assume that the capital allocation to each intermediary has an exponential-linear relationship with a weighted average of the log market equity and some exogenous characteristics of the portfolio assets, inspired by Koijen and Yogo (2019) (Henceforth, KY). In essence, we treat each intermediary as a basket of assets, thereby inheriting the characteristics of its portfolio assets. The assets are weighted by some proxy weights that resemble the portfolio weights. We do not use the portfolio weights for two reasons: First, it is due to the concern that the portfolio weights are endogenous and respond to price movements, complicating both our theoretical derivations and empirical analysis. Second, it is intuitive since households might not perfectly know or track the intermediaries' holdings. For instance, households understand energy funds primarily invest in energy companies, but they might not exactly know the portfolio weights. We use four different weighting schemes to demonstrate the robustness of our results.

We estimate the price elasticity of the capital flows to each intermediary type (Banks, Mutual funds, Investment advisors, etc) via a pooled regression of the intermediaries' log share onto the fund characteristics. Our coefficient of interest is the one in front of the average log market equity of the portfolio companies, which we denote by $\tilde{me}_i(t)$. To instrument for $\tilde{me}_i(t)$, a candidate is to use a weighted average of the price instruments employed by KY. However, it is not a suitable candidate since KY assumes that the allocation of funds among intermediaries is exogenous, while we relax this assumption.

To resolve this issue, we construct an instrument for $\tilde{me}_i(t)$ based on the following two-step procedure: For each intermediary, we assume that funds are allocated among the other intermediaries proportional to the number of assets in their investment space, as defined in KY, and those intermediaries allocate their funds to stocks in their investment space based on the book value of their holdings.[1] This first stage yields an instrument for stocks' log market capitalization that does not depend on the allocation of funds across intermediaries. In fact, the only source of variation among the instrumented market-to-book values across stocks is the set of investment spaces that they belong to and the book equity of other stocks in those investment spaces. In the second stage, the instrumented variable for $\tilde{me}_i(t)$ is constructed as the book equity-weighted average of the instruments for stocks' log market capitalization. Note that this instrument exploits intermediary-level frictions (i.e., the fact that all firms are not included in the investment space of all intermediaries) for identification. We confirm that KY's elasticity estimates would be quite similar when our IV is employed.

We find a positive and significant price elasticity (as defined in Equation 1) for the capital allocation to most intermediary types, including Mutual Funds and Investment Advisors. This result is robust across different choices of proxy weights. The results further reveal that the capital flows to mutual funds have the largest price elasticity, consistent with the notion that the transaction costs are the lowest for this intermediary type.

However, the price elasticities are not large enough to have a first-order impact on the aggregate demand elasticity. Our estimates indicate that the contribution of the indirect elasticity term (the term associated with cross-intermediary reallocations) is one to two orders of magnitude smaller than the direct elasticity term. As such, we conclude that the aggregate demand elasticity is primarily set at the intermediary level, and households' contribution to the aggregate demand elasticity is primarily through their direct holdings. This result is consistent with theories and earlier evidence on households' sluggish portfolio behavior (Ameriks and Zeldes,

[1] An intermediary's investment space is defined as the set of stocks that the intermediary has held with a positive weight over the past 36 months.

2004; Duffie, 2010; Abel et al., 2013; Giglio et al., 2021).

We also use the elasticity estimates to shed light on the transaction costs associated with the adjustments in the allocations to different intermediary types. Specifically, we consider mean-variance preferences for the representative investor and a quadratic adjustment cost for the investor's direct and indirect allocations. Theoretically, we find the transaction cost is inversely related to the price elasticity of the investor's allocation to an intermediary. Empirically, our estimates imply that households require a 1.1%-2.4% increase in the expected return for a mutual fund to increase their allocation by more than 1%. The corresponding range of values for investment advisers is between 3.7%-9.5%.

Our two-layer estimation methodology contributes to the fast-growing literature of demand-based asset pricing by modifying the methodology to estimate the aggregate demand elasticity. KY estimates the intermediary-level term in (1) as they assume an exogenous allocation of funds to intermediaries. This assumption biases their estimation of the aggregate demand elasticity downward, provided households have a downward-sloping demand. We find that the bias is statistically significant, but economically small. Another related study is Darmouni et al. (2022), which develops a two-layer demand system to study the role of capital flows in and out of institutional investors in the fragility of the corporate bond market. Thus, the focus is different from our study.

Furthermore, our study contributes to the literature on intermediary asset pricing by directly examining the magnitude of the "undoing effect" by households in response to intermediary-level frictions. Haddad and Muir (2021) develops and tests a theoretical framework with both direct and indirect holdings, where there is a quadratic cost is associated with direct holdings. Their model motivates a role for the risk-bearing capacity of intermediaries in the formation of equilibrium asset prices. In our model, households face a quadratic adjustment cost for both direct and indirect holdings. The adjustment costs determine the price elasticities of the direct and indirect allocations. Our method uncovers that cross-intermediary reallocations are at least one order of magnitude less responsive to non-fundamental price movements compared to intermediary-level reallocations. Thus, it is natural that frictions that impact intermediaries' portfolio behavior have a first-order impact on asset prices since allocations to intermediaries exhibit a weak response to price movements.

The rest of the paper is organized as follows. Section 2 presents our theoretical framework. Section 3 describes our methodology to estimate the aggregate demand elasticity. Section 4 explains the data sets used in our analysis. Section 5 provides the results. Section 6 reviews the variation in the elasticity estimates in the literature and discusses how our results compare. Section 7 concludes.

2 A Two-Layer Asset Demand Model

2.1 Setup

Consider an economy with infinite periods, $t = 1, 2, \ldots$. There are N risky assets, indexed by $n = 1, \ldots, N$, and an outside asset, indexed by $n = 0$. The outside asset has a deterministic one-period return of $R_0(t)$ at t. The excess return of the risky assets is normally distributed given the available information at time t, with a mean of $\mu(t) \in \mathbb{R}^N$ and a covariance matrix of $\Omega(t) \in \mathbb{R}^{N \times N}$. We consider a representative investor that invests in the risky assets either directly or indirectly through some intermediaries, indexed by $i = 1, \ldots, I$. In the case of 13F data, this representative investor represents the aggregate demand of some end investors, such as households and non-13F institutions (which, overall, we refer to as "households", following KY), that might provide capital to the 13F institutions.

Let $W(t)$ be the wealth of the representative investor at time t, which is invested in the assets either directly or indirectly through the intermediaries. $A_i(t)$ represents the dollar value of the investment through intermediary i at period t. To be consistent with our empirical analysis, let $A_H(t)$ represent the value of the investor's direct investment in risky assets. $Q_{H,0}(t)$ denotes the value of the investor's investment in the outside asset that is not through the intermediaries. Therefore,

$$W(t) = Q_{H,0}(t) + A_H(t) + \sum_{i=1}^{I} A_i(t). \tag{2}$$

Define $\alpha_H(t) = \frac{A_H(t)}{W(t)}$ and $\alpha_i(t) = \frac{A_i(t)}{W(t)}$, which respectively represent the fraction of direct holdings and the fraction of indirect investment through intermediary i at time t.

We represent the vector of the portfolio weights of risky assets for the representative investor and intermediaries, by $w_H(t) \in \mathcal{R}_{N\times 1}$ and $w_i(t) \in \mathcal{R}_{N\times 1}$, $i = 1, \ldots, I$, respectively. Let $P_n(t)$ be the unit price of risky asset n at period t. The unit price of the outside asset is normalized to one. We assume that risky assets have a fixed supply, which we normalize to one. The investor directly holds $Q_{H,n}(t)$ units of asset n, and likewise, $Q_{i,n}(t)$ represents the quantity of asset n held by intermediary i at period t. $p_n(t)$ and $q_{i,n}(t)$ represent the natural logarithm of the prices and quantities.

Define $X(t) \in \mathcal{R}_{N\times 1}$ as the risky portion of the effective portfolio held by the investor:

$$X(t) = \alpha_H(t) w_H(t) + \sum_{i=1}^{I} \alpha_i(t) w_i(t). \tag{3}$$

Since the investor is the ultimate holder of the assets, the portfolio weights in $X(t)$ are equal to the market portfolio weights:

$$X_n(t) = \frac{P_n(t)}{W(t)}, \qquad n = 1, \ldots, N. \tag{4}$$

Note that the investor can adjust her effective portfolio by adjusting her direct holdings (i.e., $Q_{H,n}(t)$, $n = 1, \ldots, N$), or her allocation to intermediaries (i.e., $\alpha_i(t)$, $i = 1, \ldots, I$). In the extreme case that the investor makes no adjustment at period t, the portfolio weights in her direct holdings only change due to price movements. Specifically, the portfolio weights in this benchmark case are:

$$w^0_{H,n}(t) = \frac{Q_{H,n}(t-1)P_n(t)}{A^0_H(t)}, \quad A^0_H(t) = \sum_{n=1}^{N} Q_{H,n}(t-1)P_n(t), \qquad n = 1, \ldots, N. \tag{5}$$

Thus, we can define the investor's reallocation in her direct holdings as:

$$\Delta w_H(t) := w_H(t) - w^0_H(t). \tag{6}$$

Likewise, if the investor makes no adjustment in her allocation to intermediary i at period t, the value of assets held by intermediary i only changes due to price movements between $t-1$ and t:

$$A^0_i(t) = Q_{i,0}(t-1)(1 + R_0(t-1)) + \sum_{n=1}^{N} Q_{i,n}(t-1)P_n(t), \qquad i = 1, \ldots, I. \tag{7}$$

Note that intermediary i might update its portfolio at period t, however, this rebalancing would not impact the value under its management when there is no capital flow from the investor. With these expressions, we obtain the passive benchmark for $\alpha_i(t)$:

$$\alpha^0_i(t) = \frac{A^0_i(t)}{Q_{H,0}(t-1)(1 + R_0(t-1)) + A^0_H(t) + \sum_{i=1}^{I} A^0_i(t)} = \frac{A^0_i(t)}{W(t)}. \tag{8}$$

In the equation above, we use the fact that $Q_{H,0}(t-1)(1 + R_0(t-1)) + A^0_H(t) + \sum_{i=1}^{I} A^0_i(t)$

is simply the sum of the market value of the assets at period t, which is equal to the investor's wealth.[2] The adjustment in the allocation to intermediary i is defined as $\Delta\alpha_i(t) := \alpha_i(t) - \alpha_i^0(t)$. $\Delta\alpha_H(t)$ is defined similarly.

Lastly, let $\Delta w_i(t)$ represent the adjustment in portfolio weights by intermediary i:

$$\begin{aligned} w_{i,n}^0(t) &= \frac{Q_{i,n}(t-1)P_n(t)}{A_i^0(t)}, \quad n = 1, \ldots, N \\ \Delta w_i(t) &:= w_i(t) - w_i^0(t), \quad i = 1, \ldots, I. \end{aligned} \tag{10}$$

2.2 A Reallocation Identity

The reallocations, as defined above, only result in transfers of ownership between the investor and intermediaries or among intermediaries. Consequently, these transfers should offset each other in the aggregate when there is no change in the asset supply. This section demonstrates this concept formally in our model and employs this fact to obtain an identity for reallocations, which we use in our empirical analysis to understand the magnitude of each type of allocation adjustment. To derive this identity, we first establish that if we define $X_n^0(t)$ as the effective portfolio prior to the reallocations (but after the price movements), it indeed equates to the vector of market portfolio weights, thereby being equal to $X(t)$:

$$\begin{aligned} X_n^0(t) &= \alpha_H^0(t) w_{H,n}^0(t) + \sum_{i=1}^{I} \alpha_i^0(t) w_{i,n}^0(t) \\ &= \frac{A_H^0(t)}{W(t)} \frac{Q_{H,n}(t-1)P_n(t)}{A_H^0(t)} + \sum_{i=1}^{I} \frac{A_i^0(t)}{W(t)} \frac{Q_{i,n}(t-1)P_n(t)}{A_i^0(t)} \\ &= \frac{Q_{H,n}(t-1)P_n(t) + \sum_{i=1}^{I} Q_{i,n}(t-1)P_n(t)}{W(t)} = \frac{P_n(t)}{W(t)} = X_n(t), \quad n = 1, \ldots, N. \end{aligned} \tag{11}$$

Equation 11 states that the effective benchmark portfolio is also the market portfolio since even in the benchmark case, the sum of direct and indirect investment in an asset should

[2] To see this, note that:

$$\begin{aligned} Q_{H,0}(t-1)(1+R_0(t-1)) + A_H^0(t) + \sum_{i=1}^{I} A_i^0(t) &= Q_{H,0}(t-1)(1+R_0(t-1)) + \sum_{n=1}^{N} Q_{H,n}(t-1)P_n(t) \\ &\quad + \sum_{i=1}^{I} \{Q_{i,0}(t-1)(1+R_0(t-1)) + \sum_{n=1}^{N} Q_{i,n}(t-1)P_n(t)\} \\ &= (1+R_0(t-1))Q_0(t-1) + \sum_{n=1}^{N} \underbrace{\left[Q_{H,n}(t-1) + \sum_{i=1}^{I} Q_{i,n}(t-1)\right]}_{=1} P_n(t) = W(t), \end{aligned} \tag{9}$$

where $Q_0(t-1) = Q_{H,0}(t-1) + \sum_{i=1}^{I} Q_{i,0}(t-1)$.

be its market value. By subtracting (11) from (3), we obtain the following identity for the reallocations:

$$0 = \underbrace{\sum_{i=1}^{I} \alpha_i^0(t)\Delta w_i(t)}_{\substack{\text{Within-intermediary} \\ \text{reallocation}}} + \underbrace{\sum_{i=1}^{I} \Delta\alpha_i(t) w_i(t)}_{\substack{\text{Cross-intermediary} \\ \text{reallocation}}} + \underbrace{\alpha_H^0(t)\Delta w_H(t)}_{\substack{\text{Reallocation in} \\ \text{direct holdings}}} + \underbrace{\Delta\alpha_H(t) w_H(t)}_{\substack{\text{Reallocation between} \\ \text{the outside and risky} \\ \text{assets}}} . \tag{12}$$

Equation 12 states that there are four types of reallocations, which cancel out in the aggregate: Within-intermediary reallocations, cross-intermediary reallocations by the investor, the investor's portfolio rebalancing in her direct holdings, and the investor's reallocation between the outside asset and the risky assets.

The intuition behind Equation 12 is as follows: When intermediaries collectively apply buying (selling) pressure on asset n, the investor must counterbalance this pressure either by adjusting her direct holdings (either by rebalancing the direct holdings, or by transferring funds between the risky and outside assets) or by reducing (increasing) her allocation to intermediaries that invest in asset n. In our empirical section, we utilize this identity to examine how households respond to asset demand shocks caused by intermediary-level rebalancings.

2.3 Optimal Reallocation

Next, we discuss what determines the magnitude of the investor's reallocations. To this end, we assume that the investor has mean-variance preferences over the excess return of her investment. Moreover, we assume that the investor faces a quadratic adjustment cost for both direct and indirect holdings. These adjustment costs represent the economic forces that result in households' sluggish portfolio behavior (e.g., See Ameriks and Zeldes (2004); Giglio et al. (2021)), such as inattention (Duffie, 2010) or inertia (Gabaix and Koijen, 2022). We employ this framework in Section 5.3 to understand how large the adjustment costs should be to justify our empirical results.

The quadratic cost for intermediary i is $\frac{1}{2}\frac{c_i}{\alpha_i^0(t)}\Delta\alpha_i^2$, where $c_i > 0$. The α_i^0 in the denominator of the quadratic cost implies that the marginal cost of the adjustment linearly increases with the percentage change in the position, instead of its absolute change. For instance, if the investor decides to increase her allocation to intermediary i by 1%, the marginal cost associated with this adjustment is c_i. Likewise, the quadratic adjustment cost in the direct holding of asset n is $\frac{1}{2}\frac{c_{H,n}}{\alpha_H^0 w_{H,n}^0(t)}(\alpha_H(t)w_{H,n}(t) - \alpha_H^0(t)w_{H,n}^0(t))^2$. Note that $\alpha_H w_{H,n}(t) = \frac{P_n(t)}{W(t)}Q_{H,n}(t)$ is the fraction of the investor's wealth invested in asset n. Thus, the marginal cost of adjustment linearly increases with the percentage change in the fraction of wealth invested directly in this asset.[3]

In particular, the effective portfolio at period t is the solution to the following optimization problem:

$$\max_{\Delta\alpha_i,\Delta\alpha_H,\Delta w_H} X(t)'\mu(t) - \frac{1}{2}\gamma X(t)'\Omega(t)X(t) - \sum_{n=1}^{N}\frac{1}{2}\frac{c_{H,n}}{\alpha_H^0 w_{H,n}^0(t)}(\Delta\alpha_H(t)w_{H,n}(t))^2 - \frac{1}{2}\sum_{i=1}^{I}\frac{c_i}{\alpha_i^0(t)}\Delta\alpha_i^2. \tag{13}$$

The first-order conditions imply that the following relationships hold for the investor's reallocations in relation to the effective portfolio:

$$\frac{\Delta\alpha_i}{\alpha_i^0(t)} = c_i^{-1}w_i(t)'(\mu(t) - \gamma\Omega(t)X(t)), \tag{14}$$

$$\frac{\alpha_H(t)w_{H,n}(t) - \alpha_H^0(t)w_{H,n}^0(t)}{\alpha_H^0(t)w_{H,n}^0(t)} = c_{H,n}^{-1}e_n'(\mu(t) - \gamma\Omega(t)X(t)). \tag{15}$$

Note that in the equations above, $\mu(t) - \gamma(t)\Omega(t)X(t)$ captures the distance between the effective portfolio and the first-best portfolio, i.e, $X^* = \gamma^{-1}\Omega^{-1}(t)\mu(t)$. e_n is a vector that its n'th element is one, and the other elements are zero. Equation 14 indicates that the size of reallocations in the indirect holdings depends not only on the deviation of the effective portfolio from the first-best portfolio, but also on the adjustment costs. For instance, if we find that the size of cross-intermediary reallocations is small, we cannot distinguish whether it is due to high transaction costs or high satisfaction of the investor in her effective portfolio after the intermediaries' reallocations. To distinguish these two channels, in our empirical section, we estimate the c_i's by exploiting Equation 14.

Note that this setting nests some special cases that have been analyzed before. In most household-based models (e.g., Lucas (1978)), the adjustment costs are assumed to be zero, which implies that the investor has a large price elasticity in her direct and indirect investments. The setting in Koijen and Yogo (2019) corresponds to the extreme case that $c_i = \infty$, meaning that the investor does not adjust her indirect holdings in response to price movements. Haddad and Muir (2021) considers a positive cost for direct investment and zero cost for indirect investment. We analyze the price elasticity of different types of reallocations to discern which setting yields

[3]The cost structure here is different from Haddad and Muir (2021), as it considers a quadratic cost of direct holding and no cost of indirect holding. However, we consider a quadratic adjustment cost for both direct and indirect holdings.

a more realistic response from households.

2.4 Aggregate Demand Elasticity

This section characterizes the aggregate demand elasticity. We drop the time index to simplify the notation. Note that the demand elasticity for asset n with unit price P_n is given by $\eta_n = -\frac{\partial \log Q_n^D}{\partial \log P_n}$, where Q_n^D denotes the aggregate demand of asset n. Since $Q_n^D = \frac{X_n W}{P_n}$, we have:

$$\eta_n = -\frac{\partial \log Q_n^D}{\partial \log P_n} = 1 - \frac{\partial \log X_n^D}{\partial p_n}. \tag{16}$$

Therefore, the aggregate demand elasticity for asset n with respect to its price given the vector of prices can be derived as follows:

$$\eta_n = 1 - \frac{1}{X_n}\frac{\partial X_n}{\partial p_n} = 1 - \frac{1}{X_n}\Big\{\sum_{i=1}^{I} \alpha_i \frac{\partial w_{i,n}}{\partial p_n} + \alpha_H \frac{\partial w_{H,n}}{\partial p_n} + \frac{\partial \alpha_H}{\partial p_n} w_{H,n} + \sum_{i=1}^{I} \frac{\partial \alpha_i}{\partial p_n} w_{i,n}\Big\}. \tag{17}$$

Equation 17 provides the four components that determine the elasticity of the aggregate demand for asset n. The first term inside the braces corresponds to the intermediaries' reaction to price movements. The three other terms correspond to the price elasticity of three types of allocation adjustments by the investor: The rebalancing of the direct allocation to risky assets (the second term), the reallocation between the outside asset and risky assets (the third term), and lastly, the adjustments in the allocation to the intermediaries (the fourth term).

The estimation method in Koijen and Yogo (2019) covers the first two terms. Gabaix et al. (2023) finds that the aggregate household demand for equity has a low price elasticity, implying that the third term is small. Our analysis complements the previous studies by estimating the last term, which we refer to as the "cross-intermediary reallocation" term.

3 Estimating the Demand Elasticity

We assume that the allocation to an intermediary depends on the characteristics of the assets in its portfolio. Specifically, we use $K = 120$ exogenous observable characteristics that are independent of the market prices, such as book equity and EBITDA.[4] Let $x_{n,k}(t)$ be the value of characteristic k for asset n at period t. Moreover, let $x_{n,0}(t) \equiv me_n(t)$ represent the logarithm of the market value of asset n, which corresponds to $p_n(t)$ in our model. The latent demand

[4] Table A.1 in Appendix provides the list of these variables, which are based on Jensen et al. (2023).

for intermediary i is denoted by $\epsilon_i(t)$, and its mean is normalized to one. We assume that the fraction of wealth invested through intermediary i has the following exponential-linear form:[5]

$$\alpha_i(t) = \exp\left(\sum_{k=0}^{K} \beta_{i,k} \left(\sum_{n=1}^{N} \nu_{i,n}(t) x_{k,n}(t)\right)\right) \epsilon_i(t), \quad i = 1, \ldots, I. \tag{19}$$

In this demand specification, the investor considers each intermediary as a basket of assets. That is, the investment in an intermediary depends on some weighted average of the characteristics of its portfolio assets, where $\beta_{i,k}$ reflects the sensitivity of the investor's allocation to intermediary i to its value of characteristic k. In our empirical analysis, we estimate $\beta_{i,k}$ for each intermediary type (Banks, Insurance companies, Mutual Funds, etc.) to account for the fact the allocation adjustment costs vary across these types.

To compute the intermediary-level characteristic values, we use some "proxy weights," denoted by $\nu_{i,n}(t)$. We do not use the portfolio weights for two reasons: First, if the portfolio weights were used, $\frac{\partial \alpha_i}{\partial p_n}$ would depend on how intermediary i's portfolio weights change in response to price movements (i.e., $\frac{\partial w_i}{\partial p_n}$), which would complicate our derivations. Second, it is intuitive that the investor does not perfectly know or track the intermediaries' portfolio weights. However, she understands which assets are relatively more represented in each intermediary's portfolio. For instance, the investor understands a fund specializing in the energy sector almost entirely invests in energy companies, but might not exactly know the portfolio weight of each company. In our analysis, we use four weighting schemes for $\nu_{i,n}(t)$:

- **log weights:** We run a regression of $\log w_i(t)$ on our exogenous asset characteristics in a regression that pools the intermediaries in each investor category (e.g., Banks, Insurance, Mutual Funds) for each period. Then, we apply the exponential function to the estimated value of the log portfolio weight for intermediary i to compute $\nu_{i,n}(t)$.

- **linear weights:** Similar to the previous weighting scheme, with the difference that we directly regress $w_i(t)$ on the characteristics and use the estimated portfolio weight to compute $\nu_{i,n}(t)$.

[5]This functional form is close to the demand specification in KY (See Equation 11). The key difference is that we do not consider the denominator term to simplify the derivations. Considering the denominator term would reduce our elasticity estimates. The equation below clarifies this point:

$$\alpha = \frac{x}{x+y} \Rightarrow -\frac{\partial}{\partial p} \log \alpha = -\frac{\partial}{\partial p} \log x + \frac{\partial}{\partial p} \log(x+y) < -\frac{\partial}{\partial p} \log x. \tag{18}$$

Thus, our estimates can be considered as upper bounds.

- **book weights:** $\nu_{i,n}(t)$ is equal to the book equity of asset n at period t divided by the sum of the book equity values of intermediary i's portfolio assets at period t, i.e., $\nu_{i,n}(t) = \frac{BE_n(t)}{\sum_{n'} BE_{n'}(t)\mathbb{I}\{w_{i,n'}(t)>0\}}$.

- **equal weights:** In this case, all assets in the portfolio receive the same weight, i.e., $\nu_{i,n}(t) = \frac{1}{\sum_{n=1}^{N} \mathbb{I}\{w_{i,n}(t)>0\}}$.

Note that in all these specifications, $\nu_{i,n}(t) = 0$ if $w_{i,n}(t) = 0$.

Given Equation 19, we have:

$$\frac{\partial \alpha_i(t)}{\partial p_n(t)} = \beta_{i,0}\nu_{i,n}(t)\alpha_i(t). \tag{20}$$

Therefore, we can derive the cross-intermediary reallocation term in the aggregate demand elasticity (Equation 17) as below:

$$-\frac{1}{X_n}\sum_{i=1}^{I}\frac{\partial \alpha_i(t)}{\partial p_n(t)}w_{i,n}(t) = -\frac{1}{X_n}\sum_{i=1}^{I}\beta_{i,0}\nu_{i,n}(t)\alpha_i(t)w_{i,n}(t). \tag{21}$$

To estimate this term in data, we need to estimate $\beta_{i,0}$'s in Equation 19. An approach is to run an OLS estimation for the following linear model of the log allocation to intermediary i:

$$\log \alpha_i(t) = [int.] + \beta_{i,0}\tilde{me}_i(t) + \sum_{k=1}^{K}\beta_{i,k}\tilde{x}_{k,t}(i) + \log \epsilon_i(t), \tag{22}$$

where

$$\begin{aligned} \tilde{me}_i(t) &= \sum_{n=1}^{N}\nu_{i,n}(t)me_n(t) \\ \tilde{x}_{k,i}(t) &= \sum_{n=1}^{N}\nu_{i,n}(t)x_{k,n}(t), \qquad k = 1,\ldots,K. \end{aligned} \tag{23}$$

The necessary identification assumption for this approach is that the latent demand is uncorrelated with the market prices, after accounting for the other factors.

$$\mathbb{E}[\epsilon_i(t)|\tilde{me}_i(t), \tilde{x}_{1,i}(t), \ldots, \tilde{x}_{K,i}(t)] = 1 \tag{24}$$

As pointed out by KY, the latent demand is likely to be correlated with the prices, which threatens the identification. Moreover, the latent demand and the proxy weights could be correlated. For instance, investors might select funds based on whether they are specialized in firms from a certain industry or with a certain set of characteristics.

Here, we propose an IV strategy to estimate the $\beta_{i,0}$s. This methodology helps us understand how the investor responds to price changes not caused by their own preference shocks, which is what $\beta_{i,0}$ captures. A plausibly exogenous source of variation in the prices is the changes in the investment universes.[6] Note that in this model, the investment universes of the intermediaries impose some restrictions, which have important pricing implications. For instance, in the extreme case that no fund holds the shares of a certain stock, that stock is only available through direct investment, which can be costly. Therefore, the inclusion of that stock in a fund would reduce the cost of investment in this stock, thereby pushing up its price.

Our goal is to exploit such exogenous variations in the intermediaries' investment universe to estimate the price elasticities. To this end, we construct the following instrumental variable:

$$\hat{me}_i(t) = \sum_{n=1}^{N} \nu_{i,n}(t) \hat{me}_{i,n}(t), \tag{25}$$

where $\hat{me}_{i,n}(t)$ is a proxy of the logarithm of market capitalization that is orthogonal to the allocation of funds across the intermediaries. Specifically,

$$\hat{me}_{i,n}(t) = \log\{\sum_{j \neq i} \frac{\mathbb{I}_{j,n}(t) BE_n(t)}{\sum_{n'=1}^{N} \mathbb{I}_{j,n'}(t) BE_{n'}(t)} \frac{N_j(t)}{\sum_{j' \neq i} N_{j'}(t)} \bar{ME}(t)\}. \tag{26}$$

In Equation 26, $\mathbb{I}_{j,n}(t)$ is an indicator function that is one when stock n is in the investment universe of intermediary j at period t. $\bar{ME}(t)$ is the overall market capitalization in the sample. Putting in words, $\hat{me}_{i,n}(t)$ is obtained from the following two-step procedure: First, suppose all funds are distributed among the other $I - 1$ intermediaries proportional to their number of assets in their investment universe. Then, those intermediaries allocate their funds to firms in their investment universe based on their book value. Therefore, different firms receive different instrumented market-to-book equity ($\hat{me}_{i,n}(t) - be_n(t)$) based on the number of investment universes they belong to, and the book value of the other stocks in those investment universes. This procedure obtains our price instruments for the stocks. The book-equity-weighted average of these price instruments is our instrument for the funds' unit value.[7]

[6] We follow KY in the definition of investment universes. KY shows that the intermediaries' investment universe is far from covering all available stocks. For instance, the number of stocks held by the median intermediary was 67 in 2015-2017.

[7] The results are similar across different choices of the stock-level price instrument ($\hat{me}_{i,n}(t)$), such as dis-

Therefore, the identification assumption in our IV specification is:

$$\mathbb{E}[\epsilon_i(t)|\hat{me}_i(t), \tilde{x}_{1,i}(t), \ldots, \tilde{x}_{K,i}(t)] = 1. \tag{27}$$

Note that the identification assumption under the IV specification is that the latent demand for investor i is uncorrelated with the investment universe of the other investors. In fact, once we control for book equities, the only source of cross-sectional variation in the instrumented market capitalization among stocks is the differences in the set of intermediaries that they belong to their investment universe. This variation is plausibly exogenous since investment universes are set based on mandates or investment style, and they are highly persistent. Thus, the shocks to latent demand are unlikely to affect the investment universes.

Our instrumental variable is different from the one employed by KY, primarily in the first stage of the construction of the stock-level IVs. KY assumes that the AUMs are exogenous, so they distribute each intermediary's AUM among its portfolio stocks. However, we relax this assumption in our analysis. Table 1 compares the elasticity estimates for KY's model with their and our choice of instruments. We see the estimates are quite close.

	Variable of Interest: Average Elasticity	
	Equal-Weighted	Value-Weighted
KY Instrument	0.426***	0.420***
	(0.012)	(0.008)
Our Instrument	0.373***	0.342***
	(0.007)	(0.006)
Observations	805684	805684
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 1: **Average Elasticity Estimate in KY.** This table compares the average elasticity estimates, across quarters and assets, in KY's model with their choice and our choice of IV.

A potential concern is that the instrument variable proposed here is weak for our estimation purposes. We address this concern by examining the t-statistics of the first-stage regressions, and demonstrating that the t-statistics well exceed the critical value suggested by Stock and Yogo (2002). The results, along with the estimation results obtained with this empirical strategy,

tributing $\bar{ME}(t)$ to intermediaries equally or proportionate to the overall book equity of their holdings.

are provided in the empirical section.

4 Data

Like KY, we use the 13F holdings data. The 13F data has long position holdings of institutions in the US. We use the same categories as KY—namely banks, insurance companies, investment advisors (which includes hedge funds), mutual funds, pension funds, and other. We follow their algorithm to correct institution type into these categories. The 13F data does not cover the complete holdings for every stock as KY document, so we label the residual holdings as "households' direct holdings." Importantly, these are not delegated investments through mutual funds or hedge funds. Thus, following KY, these households are placed at the intermediary level.

We also use characteristic data from Jensen et al. (2021). We take 120 stock characteristics that are not a function of prices, as shown in Table A.1, which we combine with the log of market equity. For the 120 characteristics, we use the inverse hyperbolic sine of each characteristic. We would take the log, but there are many characteristics with negative values or values around zero. For large values of x, $\sinh^{-1}(x) \approx \log(x) + \log(2)$, so this is an approximate log transformation for large positive values. For values of x close to zero, the $\sinh^{-1}(x) \approx x$. When we refer to the KY covariates, we mean the log of market equity, and the inverse hyperbolic sine of book equity (be), beta (beta_60m), operating profitability to book equity (ope_be), asset growth (at_gr1), and dividends to assets (div12m_at = div12m − at).

5 Results

The results can be summarized in a single sentence: Cross-intermediary reallocations are small and exhibit low sensitivity to prices, implying their minimal contribution to the aggregate demand elasticity. In this section, we initially utilize the reallocation identity (Equation 12) to conduct a variance decomposition analysis, which sheds light on how intermediary-level demand pressures are absorbed by households. Subsequently, we estimate and find a low price elasticity for cross-intermediary reallocations. Nonetheless, there exists significant variation in the price elasticity of the allocations to different intermediary types, consistent with the notion that transaction costs vary across these types.

5.1 How do Households Respond to Demand Pressure from Intermediaries?

We can move the within-intermediary reallocation term in Equation 12 to the left-hand side to obtain the following decomposition identity:

$$\underbrace{-\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_i(t)}_{\substack{\text{Within-intermediary}\\ \text{reallocation}}} = \underbrace{\sum_{i=1}^{I}\Delta\alpha_i(t)w_i(t)}_{\substack{\text{Cross-intermediary}\\ \text{reallocation}}} + \underbrace{\alpha_H^0(t)\Delta w_H(t)}_{\substack{\text{Rebalancing in}\\ \text{direct holdings}}} + \underbrace{\Delta\alpha_H(t)w_H(t)}_{\substack{\text{Reallocation between}\\ \text{the outside and risky}\\ \text{assets}}}. \tag{28}$$

The intuition for the equation above is as follows: If asset n is collectively demanded more by the intermediaries by tilting their portfolio weights toward this asset, the households can respond in two ways: They can absorb the demand by selling their direct holding of asset n and use the funds to purchase more of the outside asset or another risky asset. Or, they can withdraw funds from the intermediaries with a positive position in asset n to mitigate their excess demand.[8]

By taking the covariance of both sides with the left-hand side, we obtain:

$$\begin{aligned}1 = {} & \frac{COV(-\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_{i,n}(t), \sum_{i=1}^{I}\Delta\alpha_i(t)w_{i,n}(t))}{VAR(\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_{i,n}(t))} \\ & + \frac{COV(-\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_{i,n}(t), \alpha_H^0(t)\Delta w_{H,n}(t))}{VAR(\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_{i,n}(t))} + \frac{COV(-\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_{i,n}(t), \Delta\alpha_H(t)w_{H,n}(t))}{VAR(\sum_{i=1}^{I}\alpha_i^0(t)\Delta w_{i,n}(t))}.\end{aligned} \tag{29}$$

The first term in Equation 29 reflects the fraction of intermediaries' net demand that is absorbed by cross-intermediary reallocations. Likewise, the second and third terms show the fraction absorbed through adjustments in the direct holdings, either through rebalancings of direct holdings or reallocation between the outside asset and risky assets. Table 2 provides our estimates of these terms (in percentages) when we consider all periods and all assets in our data.

Table 2 reveals that approximately 90% of intermediaries' excess demand is absorbed through the rebalancing of direct holdings. In contrast, cross-intermediary reallocations actually exacerbate the demand pressure. This observation aligns with the return-chasing behavior of the capital flows, as documented in the empirical literature (e.g., See Chevalier and Ellison (1997),

[8] Recall that α_i's are the fraction of the total wealth managed by intermediary i, however, the data on the overall wealth is not available. We address this issue by multiplying all α's and $\Delta\alpha$'s in Equation 28 by $\frac{W(t)}{\bar{A}(t)}$, where $\bar{A}(t) = A_H(t) + \sum_{i=1}^{I} A_i(t)$. For instance, $\alpha_i^0(t)$ is replaced by $\frac{A_i^0}{\bar{A}(t)}$ and $\Delta\alpha_i(t)$ is replaced by $\frac{A_i(t)-A_i^0(t)}{\bar{A}(t)}$, where $A_i^0(t)$ is defined in Equation 7. This transformation is innocuous since it simply multiplies all α's and $\Delta\alpha$'s by a scalar.

	Across Intermediary	Direct Holdings (Rebalancing)
	(1)	(2)
Variance Component	-9.95***	89.66***
	(3.35)	(2.62)
Observations	802484	802484
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2: **Variance Decomposition.** This table provides the contribution (in percentage) of the cross-intermediary reallocation term and the term associated with the rebalancing of direct holdings in Equation 29. The residual is the term corresponding to the reallocation between the outside and risky assets.

	Banks	Insurance Companies	Investment Advisors	Mutual Funds	Pension Funds	Other	Direct Holdings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variance Component	-0.30***	-0.34***	-2.07***	-7.14***	0.80	-0.91***	89.66***
	(0.10)	(0.04)	(0.25)	(1.50)	(2.49)	(0.14)	(2.62)
Observations	802484	802484	802484	802484	802484	802484	802484
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 3: **Variance Decomposition By Intermediary Type.** This table decomposes the cross-intermediary reallocation term in Equation 29 by the intermediary type. The residual is the term corresponding to the reallocation between the outside and risky assets.

Lou (2012)).[9] An excess demand for asset n drives up its price, attracting capital flow toward intermediaries holding substantial positions in this asset. To shed light further on this observation, we decompose the cross-intermediary reallocation term based on intermediary types in Table 3. In line with this mechanism, we see that capital flows to mutual funds account for most of the positive relationship between the capital flows to intermediaries and their excess demand.

From the estimates presented in Table 2, we can infer that around 20% of the intermediary sector's excess demand is absorbed through the transfers of funds from the outside asset. Therefore, the rebalancing of direct holdings plays the most significant role in counterbalancing the purchasing or selling pressure exerted by intermediaries.

Overall, these observations imply that the capital flows are not orthogonal to within-intermediary portfolio adjustments. In particular, within-intermediary portfolio adjustments not only impact the prices, but also the latent demand for those intermediaries. As such, we employ the IV technique described in Section 3 to estimate the price elasticity of the allocations to intermediaries. The next section provides the results.

[9] This result is robust across time, as shown in Figure A.1.

5.2 Elasticity of Allocations to Intermediaries

First, we show that our empirical strategy does not suffer from the weak instrument bias. A standard methodology in the literature is to examine the t-statistics in the first-stage regressions, namely, the regression of the log market equities onto their instrumented value, and compare the results against the critical value proposed by Stock and Yogo (2002). Figure 3 provides the minimum t-statistics for each year. The minimum t-statistics well exceed the critical value of 4.05 for rejecting the null hypothesis of weak instrument at the five percent confidence level. Table 4 provides a similar result for the intermediary-level instruments.

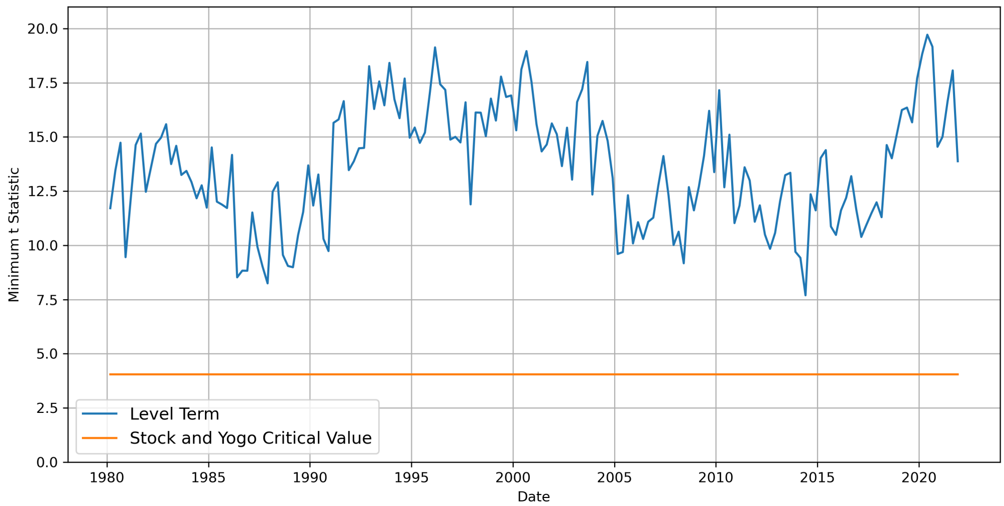


Figure 3: **First-stage minimum t-statistic on the instrumented log market equity** This plot shows the minimum t-statics of the first-stage regression of log-market equity on its instrumented value across intermediaries. The flat line displays the critical value to reject the null hypothesis of weak instrument at five percent level confidence (Stock and Yogo, 2002).

Table 5 provides the estimates of the price-elasticity ($\beta_{i,0}$) of the allocation to each intermediary type. The estimates are obtained from estimating the coefficients in Equation 22 with the IV specification. We see that the allocation to intermediaries is indeed negatively related to the prices, in contrast to what the results in Section 5.1 might suggest. This result is robust across intermediary types and the choice of weighting scheme for the proxy weights. The price elasticity is the largest for mutual funds, consistent with the idea that the adjustment cost is the lowest for this type. Moreover, the results are mixed for Insurance companies and Pension funds, in line with the inflexibility of the allocations to these types of intermediaries. In Section

	Banks	Insurance	Advisors	Mutual Funds	Pension	Other
log weights	21.650	18.820	13.740	15.760	27.010	7.230
linear weights	15.550	15.140	19.620	21.710	22.280	13.410
book weights	22.740	15.500	15.560	16.280	19.230	8.270
equal weights	15.330	8.280	15.130	16.590	15.200	7.340

Table 4: **First Stage t Statistics for Intermediaries.** This table provides the minimum t-statistic of the first-stage regression of the log-AUM on its instrumented value for each intermediary group. The t-statistics in each row correspond to a weighting scheme introduced in Section 3 for the proxy weights. All values are above the threshold values provided by Stock and Yogo (2002).

	Dependent variable: $\log(\alpha)$					
	Banks	Insurance	Advisors	Mutual Funds	Pension	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Log	-2.168^{**}	0.785	-0.890^{***}	-3.087^{***}	-0.766	-1.661^{*}
	(1.103)	(2.022)	(0.311)	(0.792)	(1.975)	(0.919)
Linear	-0.722	0.646	-0.631^{***}	-2.459^{***}	-2.214^{**}	-0.981^{***}
	(0.447)	(0.558)	(0.125)	(0.275)	(1.004)	(0.312)
Book	-0.456	-1.098	-1.531^{***}	-4.705^{***}	-3.050^{***}	-1.012^{***}
	(0.444)	(0.784)	(0.202)	(0.574)	(0.920)	(0.341)
Equal	-0.626	-3.664^{**}	-1.614^{***}	-5.484^{***}	2.080^{**}	-1.838^{***}
	(0.572)	(1.581)	(0.231)	(0.623)	(0.826)	(0.510)
Observations	20912	5608	218894	93910	4792	38788
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 5: **Price-elasticity of the allocation to intermediaries** This tables provides the IV estimates of $\beta_{i,0}$ in Equation 22 for each intermediary type. The regression pools all periods and all intermediaries for each intermediary type. The standard errors are double clustered by intermediary and quarter.

5.3, we use these elasticity estimates to compute the adjustment cost for each intermediary type.

Now, we compute the contribution of the cross-intermediary reallocation term to the aggregate demand elasticity by plugging in these estimates in Equation 21. Table 6 presents the weighted average elasticity of this term across stocks for each intermediary type. We see that the

	Average Intermediary Elasticity Effects (Times 100)						
	(1) Banks	(2) Insurance	(3) Advisors	(4) Mutual Funds	(5) Pension	(6) Other	(7) Combined
Log	0.0146***	-0.0032***	0.0258***	0.2741***	0.0126***	0.0216***	0.3455***
	(0.0033)	(0.0005)	(0.0024)	(0.0376)	(0.0022)	(0.0030)	(0.0460)
Linear	0.0114***	-0.0057***	0.0512***	0.4082***	0.0382***	0.0319***	0.5351***
	(0.0020)	(0.0006)	(0.0034)	(0.0316)	(0.0042)	(0.0036)	(0.0413)
Book	0.0076***	0.0103***	0.1611***	0.9294***	0.0721***	0.1440**	1.3245***
	(0.0021)	(0.0019)	(0.0174)	(0.1293)	(0.0130)	(0.0576)	(0.1713)
Equal	0.0038***	0.0197***	0.1205***	0.4072***	-0.0109***	0.2184**	0.7588***
	(0.0005)	(0.0018)	(0.0103)	(0.0497)	(0.0013)	(0.1033)	(0.1181)
Observations	805684	805684	805684	805684	805684	805684	805684
Note:							*p<0.1; **p<0.05; ***p<0.01

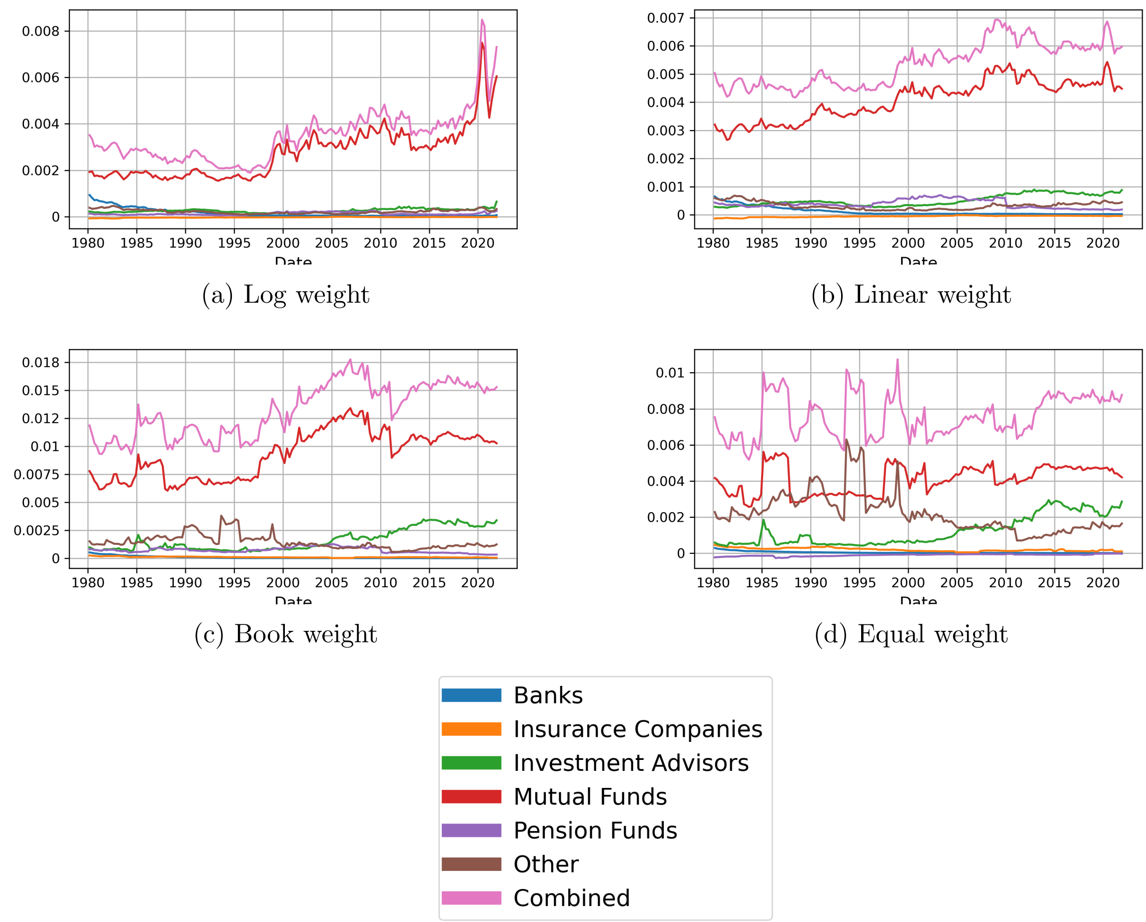
Table 6: **The contribution of the cross-intermediary reallocation term to the aggregate elasticity.** This table presents the overall price elasticity of allocations to each intermediary type, along with the combined effect. The reported numbers are 100 times larger than the estimated values. The combined effect provides an estimate for the contribution of the cross-intermediary reallocation term in Equation 17 to the aggregate demand elasticity.

estimated combined contribution to the aggregate demand elasticity is between $0.003 - 0.013$, depending on the choice of weighting scheme. These numbers are at least one order of magnitude smaller than the demand elasticity estimates in KY (See Table 1). The majority of this contribution is through adjustments in the allocations to mutual funds. Figure 4 displays the weighted average elasticity of the cross-intermediary reallocation term over time for different weighting schemes and different intermediary types.

Our results indicate that households' adjustments in the allocation to intermediaries play a weak role in the aggregate demand elasticity. It is an important observation since it implies that if intermediaries face some frictions that cause them to exert some buying or selling pressure on some assets, the pressure is unlikely to be undone by households' reallocations across intermediaries. It is in contrast with household-based asset pricing models that assign a weak role to intermediary-level portfolio decisions in the formation of equilibrium prices.

5.3 Adjustment costs

We can use the elasticity estimates in this section to shed light on the adjustment costs investors face in their allocations to intermediaries. In particular, Equation 14 shows how the adjustment cost impacts the sensitivity of the allocations to an intermediary with respect to expected return movements:



(a) Log weight

(b) Linear weight

(c) Book weight

(d) Equal weight

(e) Legend

Figure 4: **The contribution of the cross-intermediary reallocation term to the aggregate elasticity.** This figure presents the overall price elasticity of allocations to each intermediary type, along with the combined effect, over time. The combined effect provides an estimate for the contribution of the cross-intermediary reallocation term in Equation 17 to the aggregate demand elasticity.

$$\frac{\partial \alpha_i}{\partial \mu_n} = \alpha_i c_i^{-1} w_{i,n}. \tag{30}$$

In our empirical analysis, we estimate the price elasticity of α_i. To estimate c_i, we can apply the chain rule to the left-hand side and employ Equation 20:

$$\frac{\partial \alpha_i}{\partial \mu_n} = \frac{\partial \alpha_i}{\partial p_n} / \frac{\partial \mu_n}{\partial p_n} = \varphi^{-1} \beta_{i,0} \nu_{i,n} \alpha_i, \tag{31}$$

where $\varphi = \frac{\partial \mu_n}{\partial p_n}$ is the pass-through rate between the price change and the change in the expected return (Davis et al., 2022). By taking a summation of the right-hand sides in (30) and (31) over n, we derive the following equation for the transaction cost, assuming that $w_{i,0}$ is small:

$$c_i \simeq \beta_{i,0}^{-1} \varphi. \tag{32}$$

Equation 32 nicely relates the adjustment cost to the price-elasticity of the allocation to intermediary i. In line with intuition, there is a negative relationship between the adjustment cost and price elasticity. Recall that c_i is the marginal adjustment cost when the allocation is changed by 1%. That means the expected return of investing in intermediary i should increase by c_i to make investors increase their allocation by at least 1%.

Table 7 presents the estimates of the adjustment cost for different intermediary types. We see that the adjustment cost for mutual funds is estimated to be between 1.1%-2.4%. It implies that for a 1% increase in the allocation to a mutual fund, the expected return of the fund should increase by at least 2.4%, based on our most conservative estimate. We drop the estimates for Insurance companies and Pension funds due to the mixed results observed in Table 5.

6 Variation in the elasticity estimates across models

We found that the contribution of the cross-intermediary reallocation term to the aggregate demand elasticity ranges between 0.003-0.013 (See Table 6). In this section, we provide the estimates of the aggregate demand elasticity from various elasticity models and argue that the variation in these elasticity estimates across models is at least one order of magnitude larger than the elasticity of the cross-intermediary reallocation term.

The KY model is close to a standard isoelastic demand model, where portfolio weights have

	Transaction Costs			
	Banks	Advisors	Mutual Funds	Other
Log	0.028	0.067	0.019	0.036
Linear	0.083	0.095	0.024	0.061
Book	0.132	0.039	0.013	0.059
Equal	0.096	0.037	0.011	0.033

Table 7: **Adjustment Costs for Allocations to Intermediaries** This table presents the estimates of the adjustment cost for the adjustment of the allocation to each intermediary type. The numbers can be interpreted as the minimum increase in the expected return required to increase the allocation to an intermediary by 1%.

the following functional form:

$$w_{i,n}(t) = \exp(\hat{\varrho}_{i,n,1}(t) + \varrho_{i,2}(t)\log(ME_n(t)))$$

$$= \underbrace{0}_{\substack{\text{intercept}\\\text{term}}} + \underbrace{(\varrho_{i,n,1}(t))(ME_n(t))^{\varrho_{i,2}(t)}}_{\substack{\text{level}\\\text{term}}} + \underbrace{0 \cdot \log(ME_n(t))}_{\substack{\text{log}\\\text{term}}}, \tag{33}$$

where $\exp(\hat{\varrho}_{i,n,1}(t)) = \varrho_{i,n,1}(t)$ captures the exogenous covariates and latent demand (error term). With this specification, the elasticity of an investor is simply $1 - \varrho_{i,2}(t)$.[10] This isoleastic demand can be estimated by regressing the log of portfolio weights on the instrumented log of market equity and exogenous controls. We do this, using a KY instrument and the 120 exogenous controls. Note that the covariates constraints imposed by the KY GMM estimation method is not imposed. The first row in Table 8 shows that value-weighted elasticity across time periods, which is 0.41. This is quite similar to the model results from above. In other words, the KY elasticity results are quite robust to include many more controls, estimating a model with a more simple two-staged least-squares, and using only a simple isoleastic demand model.

Davis (2021) also considers a model where demand is linear is characteristics, the log of

[10]In the KY specification, the elasticity of an investor is actually $1 - \varrho_{i,2}(t)(1 - w_{i,n}(t))$. This $(1 - w_{i,n}(t))$ term comes from an adding up constraint for portfolio weights. Quantitatively, since portfolio weights tend to be small, this makes little difference.

Model	Covariates	Elasticity	Level	Log
Isoelastic	All	0.414***	0.586***	
		(0.045)	(0.045)	
Linear	Only Level	0.221***	0.779***	
		(0.006)	(0.006)	
Linear	Level & KY	0.275***	0.725***	
		(0.006)	(0.006)	
Linear	Level, KY, & Log	0.424***	0.754***	-0.178***
		(0.016)	(0.006)	(0.011)
Linear	All	0.646***	0.747***	-0.393***
		(0.030)	(0.005)	(0.026)
Observations		805684	805684	805684
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 8: **Elasticity Across Models** This table shows the elasticity estimates for the cross-intermediary reallocation term across models. There are results for both isoleastic and linear demand models, with covariates shown. KY covariates represent the covariates used in Koijen and Yogo (2019) use in their model. All covariates include all the 120 controls as discussed in the text (See Table A.1). The elasticity is one minus the log and level effects.

market equity, and market-portfolio weights:

$$w_{i,n}(t) = \underbrace{\beta_{i,n,0}(t)}_{\substack{\text{intercept}\\ \text{term}}} + \underbrace{\beta_{i,1}(t)\frac{ME_n(t)}{A(t)}}_{\substack{\text{level}\\ \text{term}}} + \underbrace{\beta_{i,2}(t)\log(ME_n(t))}_{\substack{\text{log}\\ \text{term}}}, \tag{34}$$

where $\beta_{i,n,0}(t)$ includes the error term (latent demand) and all the exogenous controls. Note that if $\beta_{i,1}(t) = 1$ and $\beta_{i,n,0}(t) = \beta_{i,2}(t) = 0$, this corresponds to a market-indexer with an elasticity of zero. If $\beta_{i,1}(t) < 1$ and $\beta_{i,2}(t) < 0$, then aggregate market demand for every asset is downward sloping and there is a unique equilibrium price. This demand function can be derived using log demand or mean-variance demand, similar to KY (see Davis, 2021). We briefly summarize some of the benefits and downsides of this demand function. The benefits are the following:

1. This demand function can hold short positions. Isoelastic demand of course can only accommodate long-only positions.

2. This demand function aggregates well. In other words, the demand function of many investors in aggregate has the same functional form as the individual investors. This is not true of isoelastic demand. This matters for estimating demand functions of the aggregate household and for institutions that are composed of many individual funds (e.g. BlackRock).

3. This linear demand function has a great deal of flexibility, with the level term fitting the degree to which an investor is a market-indexer and the log market equity term captures the sensitivity of investors to prices.

The downside of the model is that the elasticity will be very high for small positions. To see this, note that the elasticity of an investor with this demand function is given by:

$$1 - \frac{\beta_{i,1}(t)\frac{ME_n(t)}{A(t)} + \beta_{i,2}(t)}{w_{i,n}(t)}. \tag{35}$$

In aggregate, this means that the assets with small market-portfolio weights tend to have higher elasticity values. While there is some evidence that small stocks may have more elastic demand (Haddad et al., 2022), linear demand can at times make this effect too extreme. Since we look at value-weighted elasticity values in this section and for most of the paper, this right-skewed elasticity distribution is completely offset by using value-weighted averages.[11]

This demand function is estimated by regressing portfolio weights (not log portfolio weights) on characteristics and the two instrumented endogenous terms (level and log price term). The KY instrument is in logs, and thus it is natural to use the KY instrument and the exponentiated KY instrument as the two instruments. The minimum F statistic for these two instruments across pooled samples and all periods is above the Stock and Yogo (2002) critical value.[12]

The second row of table 8 shows the value-weighted elasticity across time when demand is estimated with only the level term. There are no other covariates used in this regression. This yields an average elasticity of 0.22. The third row shows the elasticity with the linear model only estimated with a linear term and the covariates used by KY, which increases the elasticity to 0.28. The next row uses the same specification except it also includes the log term, which increases the elasticity further to 0.42. This specification provides a similar elasticity to the isoelastic model and the KY model shown above. The prior specifications with this linear model

[11] Equal-weighted averages result in much higher elasticity values, as can be seen in Davis (2021).

[12] This critical value for two instruments and two endogenous covariates is 7.03, as can be seen in Table 5.2 of Stock and Yogo (2002).

miss obvious controls, while this model has a reasonable set of controls. Finally, the last row includes all 120 controls, which increases the elasticity further to 0.65.

The take-away from this section is that estimates of the aggregate market elasticity for individual assets vary from around 0.4 to 0.65 simply by using different functional forms and different controls. This increase in the average elasticity of 0.25 is much larger than the cross-intermediary effects, which range from 0.003 to 0.013. This is especially true when compared to the large variation in microeconomic elasticity values across different studies, as highlighted by Gabaix and Koijen (2022). In conclusion, the cross-intermediary elasticity term is very small to variation across demand models.

7 Conclusion

Both from a policy and investment perspective, it is crucial to understand the role of households and intermediaries in financial markets. Existing asset pricing models assume different allocations of the roles, which have resulted in drastic differences in their implications. For instance, how large is the impact of an asset purchase program on the prices? Different models have vastly different answers based on whose response is more salient in the "aggregate market's" response: Households or intermediaries. Our analyses are intended to shed light on this debate.

Do households undo? The answer to this question is what divides existing models; specifically, it separates household-based asset pricing models (Merton, 1973; Lucas, 1978; Campbell and Cochrane, 1999; Bansal and Yaron, 2004; Barro, 2006; Gabaix, 2012) from the more recent intermediary-based models (He and Krishnamurthy, 2013; Koijen and Yogo, 2019; Haddad and Muir, 2021; Gabaix and Koijen, 2022). The central premise of the former class of models is that enough rebalancing would take place to eliminate any mispricing induced by intermediary-level frictions. Thus, intermediary-level frictions, such as balance-sheet constraints and investment mandates, do not have a first-order effect on the prices.

In this paper, we directly test this central assumption by examining how sensitive cross-intermediary reallocations are with respect to non-fundamental price movements. We do this by developing and estimating a model in which a representative investor invests in some risky assets both directly and indirectly through some intermediaries. Then, we decompose the aggregate demand elasticity into a direct component, which captures the elasticity of the within-intermediary reallocations and the elasticity of the direct holdings, and an indirect component, which represents the elasticity of the allocations to intermediaries. Consistent with intermediary-based models, we find that the indirect elasticity component is substantially smaller than the direct component, meaning that intermediaries' portfolio behavior has a first-order impact on the aggregate demand elasticity.

Thus, we conclude that intermediary-level frictions play a first-order role in price movements since allocations to intermediaries exhibit a weak response to price dislocations.

References

Abel, Andrew B, Janice C Eberly, and Stavros Panageas, 2013, Optimal inattention to the stock market with information costs and transactions costs, *Econometrica* 81, 1455–1481.

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *The Journal of Finance* 69, 2557–2596.

Ameriks, John, and Stephen P Zeldes, 2004, How do household portfolio shares vary with age, Technical report, working paper, Columbia University.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The journal of Finance* 59, 1481–1509.

Barro, Robert J, 2006, Rare disasters and asset markets in the twentieth century, *The Quarterly Journal of Economics* 121, 823–866.

Brunnermeier, Markus K, and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *The review of financial studies* 22, 2201–2238.

Campbell, John Y, and John H Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of political Economy* 107, 205–251.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of political economy* 105, 1167–1200.

Darmouni, Olivier, Kerry Siani, and Kairong Xiao, 2022, Nonbank fragility in credit markets: Evidence from a two-layer asset demand system, *Available at SSRN* .

Davis, Carter, 2021, The elasticity of quantitative investment, *working paper* .

Davis, Carter, Mahyar Kargar, and Jiacui Li, 2022, An information-based explanation for inelastic demand, *Available at SSRN* .

Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, *The Journal of finance* 65, 1237–1267.

Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *The Quarterly journal of economics* 127, 645–700.

Gabaix, Xavier, and Ralph S. J. Koijen, 2022, In search of the origins of financial fluctuations: The inelastic markets hypothesis, *working paper* .

Gabaix, Xavier, Ralph SJ Koijen, Federico Mainardi, Sangmin Oh, and Motohiro Yogo, 2023, Asset demand of us households, *Available at SSRN 4251972* .

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five facts about beliefs and portfolios, *American Economic Review* 111, 1481–1522.

Haddad, Valentin, Paul Huebner, and Erik Loualiche, 2022, How competitive is the stock market? Theory, evidence from portfolios, and implications for the rise of passive investing, Working Paper, UCLA and Minnesota.

Haddad, Valentin, and Tyler Muir, 2021, Do intermediaries matter for aggregate asset prices?, *The Journal of Finance* 76, 2719–2761.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103, 732–70.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *The Journal of Finance* 78, 2465–2518.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen, 2021, Is there a replication crisis in finance?, Technical report, National Bureau of Economic Research.

Koijen, Ralph S. J., and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

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

Lucas, Robert E Jr, 1978, Asset prices in an exchange economy, *Econometrica: journal of the Econometric Society* 1429–1445.

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica: Journal of the Econometric Society* 867–887.

Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *The Journal of finance* 52, 35–55.

Stock, James H, and Motohiro Yogo, 2002, Testing for weak instruments in linear iv regression.

Appendix A: Additional Figures and Tables

cowc_gr1a	oaccruals_at	oaccruals_ni	taccruals_at
taccruals_ni	capex_abn	debt_gr3	fnl_gr1a
ncol_gr1a	nfna_gr1a	ni_ar1	noa_at
aliq_at	at_gr1	be_gr1a	capx_gr1
capx_gr2	capx_gr3	coa_gr1a	col_gr1a
emp_gr1	inv_gr1	inv_gr1a	lnoa_gr1a
mispricing_mgmt	ncoa_gr1a	nncoa_gr1a	noa_gr1a
ppeinv_gr1a	sale_gr1	sale_gr3	saleq_gr1
age	at_be	bidaskhl_21d	cash_at
ni_ivol	rd_sale	rd5_at	tangibility
beta_60m	beta_dimson_21d	betabab_1260d	betadown_252d
earnings_variability	ivol_capm_21d	ivol_capm_252d	ivol_ff3_21d
ivol_hxz4_21d	ocfq_saleq_std	rvol_21d	turnover_126d
zero_trades_21d	zero_trades_126d	zero_trades_252d	dsale_dinv
dsale_drec	dsale_dsga	niq_at_chg1	niq_be_chg1
niq_su	ocf_at_chg1	sale_emp_gr1	saleq_su
tax_gr1a	dolvol_var_126d	ebit_bev	ebit_sale
f_score	ni_be	niq_be	o_score
ocf_at	ope_be	ope_bel1	turnover_var_126d
at_turnover	cop_at	cop_atl1	dgp_dsale
gp_at	gp_atl1	ni_inc8q	niq_at
op_at	op_atl1	opex_at	qmj
qmj_growth	qmj_prof	qmj_safety	sale_bev
corr_1260d	coskew_21d	dbnetis_at	lti_gr1a
pi_nix	sti_gr1a	ami_126d	iskew_capm_21d
iskew_ff3_21d	iskew_hxz4_21d	rskew_21d	chcsho_12m
eqnetis_at	netis_at	netdebt	rd
at	be	debt	div12m
ebitda	eqnpo	eqpo	fcf
ival	ni	ocf	sale

Table A.1: **Exogenous Variable Names** This shows the 120 exogenous characteristic variables used in the demand estimation (Equation 19). The variables are obtained from Jensen et al. (2021).

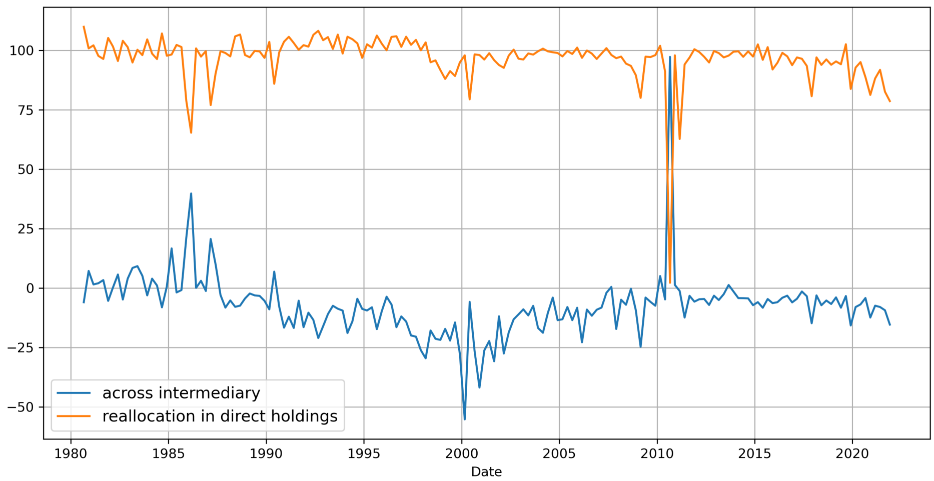


Figure A.1: **The decomposition of households' response to intermediaries' excess demand.** This figure decomposes the vector of intermediaris' excess demand at each period into three terms: The rebalancing of direct holdings (orange line), adjustments in the allocations to intermediaries (blue line), and transfers between the risky assets and outside assets. The terms add up to 100%. The last term is the residual.