# Why is Asset Demand Inelastic?*

Carter Davis† Mahyar Kargar‡ Jiacui Li§

January 14, 2024

**Abstract**

Classic asset pricing models predict high demand elasticities: investors trade aggressively against price deviations. In contrast, empirical estimates are three orders of magnitude lower. To explain this gap, we use a portfolio choice framework to show that demand elasticity depends on two components: "price pass-through", which measures how price movements forecast returns, and "unspanned returns", which reflects an asset's distinctiveness relative to others. Contrary to classic model assumptions, we find evidence for low price pass-throughs and high unspanned returns. By considering these two channels, we can largely reconcile the difference between theoretical predictions and empirical estimates of demand elasticity.

**Keywords**: Price elasticity, price pass-through, spanning, demand system asset pricing, shrinkage.

**JEL Classification**: G11, G12, G14.

*We thank Aditya Chaudhry, Olivier Darmouni, Zhiyu Fu, Valentin Haddad, Christian Heyerdahl Larsen, Kristy Jansen, Tim Johnson, Ralph Koijen, Daniel Neuhann, Dejanir Silva, and conference and seminar participants at AFA 2024, Indiana University, University of Utah, and 2022 Wabash River Finance Conference for valuable feedback. All errors are our own.

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

‡Gies College of Business, University of Illinois Urbana-Champaign; kargar@illinois.edu

§David Eccles School of Business, University of Utah; jiacui.li@eccles.utah.edu

# 1 Introduction

In many frictionless asset pricing models, investor demand curves are virtually flat, implying very high demand elasticities (e.g., Petajisto, 2009; Gabaix and Koijen, 2021). Such models featuring very elastic demand fail to generate many salient asset pricing facts such as the price impact of fund flows and retail trades, excess volatility puzzle, etc. In stark contrast to these theories, empirical studies overwhelmingly find demand elasticities to be much smaller, usually lower by three orders of magnitude.[1] How can we reconcile this large gap? Can some of the differences be accounted for in a frictionless framework, or is the entire difference a "puzzle" and requires frictional or behavioral explanations?[2]

In this paper, we shed light on the key determinants of demand elasticity by introducing a novel model-free approach that decomposes it into two components. Demand elasticity is high if and only if both components are high. Guided by this framework, we show that theories predict high demand elasticities because they assume both components to be high. Once we plug in empirically estimated values, most of the gap between theory predictions and empirical estimates is explained away.

Our decomposition comes from the simple insight that demand elasticity for assets is fundamentally different from that for consumption goods. In consumer theory, preferences are defined over goods, and demand elasticities are determined by preferences. In finance, however, investor preferences are defined over consumption or wealth, not the assets themselves. Therefore, investors do not care about assets *per se*, but rather about the returns these assets generate. Therefore,

[1] For a theoretical model with demand elasticities of 5,000, 1% change in shares supplied only moves prices by 0.02 basis points. In contrast, in empirical studies, estimated demand elasticities in the stock market are around 2, implying price impacts of 0.5% for a 1% change in supply (Koijen and Yogo, 2019; Gabaix and Koijen, 2021).

[2] To generate lower demand elasticities, some models include additional frictions or biases. For example, Hong and Stein (1999) argue that contrarian strategies that lessen the momentum trader effects in their model are unlikely given payoffs to contrarian strategies, thus offering a behavioral account for low demand elasticities.

investor demand elasticity of assets should depend primarily on how price changes affect expected returns. Guided by this perspective, we decompose investor demand elasticities — changes in the log number of shares held ($Q_i$) to changes in the log asset price ($P_i$) — into one plus the product of two components:

$$\eta_i = -\frac{\partial \log(Q_i)}{\partial \log(P_i)} \approx 1 + \overbrace{\underbrace{\frac{\partial \log(w_i)}{\partial \tilde{\mu}_i}}_{=((1-\omega_{i,t})\tilde{\mu}_{i,t})^{-1} \text{ for MV}}}^{\text{weight responsiveness}} \times \overbrace{\left(-\frac{\partial \tilde{\mu}_i}{\partial \log(P_i)}\right)}^{\text{price pass-through}} \tag{1}$$

where the second term, "price pass-through", measures the change in next-period expected return $\tilde{\mu}_i$ in response to price changes. The first term, "weight responsiveness", measures the response of the investor's log portfolio weight to changes in expected returns. As such, our decomposition focuses on understanding demand shifts in response to changes in expected returns. In principle, demand may also respond to price-induced changes in risk or other variables, but we show that those effects are second-order in typical portfolio choice settings.[3]

While the price pass-through is intuitive and depends on price dynamics, a key question remains: what determines weight responsiveness? To offer intuition, we focus on investors who form mean-variance (MV) efficient portfolios — not because it is realistic, but to gain insight into investor behavior in the absence of frictions or biases. We show that, in the MV setting, weight responsiveness is inversely related to the size of "unspanned returns", $(1 - \omega_i)\tilde{\mu}_{i,t}$, where $\omega_i$ in Equation (1) is the fraction of expected return spanned by the other assets. Intuitively, this measures how *substitutable* the asset is. For instance, if an asset has nearly perfect substitutes and

[3]To study the importance of other drivers of portfolio weights (e.g., risk, hedging), in Appendix C, we estimate the components of a more general portfolio choice problem with Epstein-Zin preferences, following Campbell, Chan, and Viceira (2003). We decompose the return-unrelated components of demand elasticity into a covariance component, a variance component, and a consumption-to-wealth ratio hedging component. We find that these components are small relative to the component associated with changes in expected returns.

thus exhibits low unspanned returns, investors are likely to significantly increase their positions in this asset and correspondingly decrease their positions in a hedging portfolio when its expected return increases. Conversely, if an asset lacks close substitutes, investor portfolio responses should be more subdued.

Guided by our decomposition framework, we show that using empirical estimates for price pass-throughs and unspanned returns explains most of the gap between theory and estimated demand elasticities (Figure 1). Classic models assume a high price pass-through of approximately one (so all non-fundamental price movements revert in one period) and that stocks are almost perfect substitutes. As a consequence, they predict stock-level demand elasticities of approximately 7,000, as illustrated in the leftmost bar. If we use empirical estimates of price pass-through which is around 0.06, demand elasticity drops to approximately 600. We empirically estimate that around half of stock-level expected returns are not spanned, and accounting for this large unspanned return leads to elasticity estimates of around 12. These two steps are illustrated in the next two bars in Figure 1. Therefore, without appealing to frictions or biases, a significant part of the "demand elasticity gap" is explained away.

We now explain how we estimate the two components of demand elasticity. To estimate price pass-throughs, we consider the implied beliefs of arbitrageurs who predict stock returns using stock characteristics. To understand the logic, consider a fund that predicts returns using $\tilde{\mu}_{i,t} = \pi \cdot \log(B_{i,t}/M_{i,t})$, where $\pi$ is a coefficient and $\log(B_{i,t}/M_{i,t})$ is the log book-to-market ratio. This implies a pass-through of $-\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(M_{i,t})} = \pi$. In the paper, we specify a model where both expected returns and covariances are functions of stock characteristics, and we use maximum likelihood to estimate model parameters. We find that the implied pass-through is approximately 0.06 at a monthly horizon — i.e., each 1% of price drop translates into a 6 basis point higher expected returns in the following month. As described above, this empirically estimated price

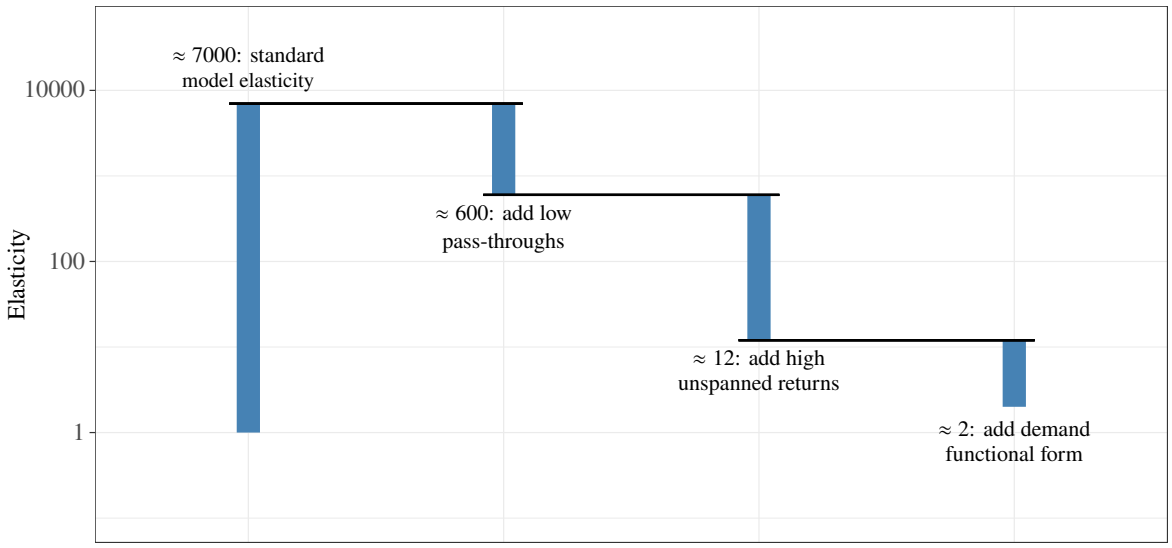



**Figure 1. Demand elasticity decomposition**
This figure shows how we explain the gap between high demand elasticity values in standard asset pricing models and low values in empirical estimates. The leftmost bar illustrates predictions from standard asset pricing models with perfect pass-through and low unspanned returns Petajisto (2009) (≈ 7,000). Taking into account a price pass-through of 0.06 instead of 1 reduces elasticity by a factor of 12 (as shown by the second bar from left) to approximately 600. Taking into account our empirically estimated unspanned returns of 0.7% per month, we arrive at elasticity estimates of around 12, further reducing elasticity by a factor of 50, as shown by the third bar from the left. Finally, incorporating exponential-linear demand as in Koijen and Yogo (2019) gets the elasticity down by a factor of 6 (the rightmost bar) to about 2, in line with the empirical estimates. The $y$-axis is in log scale.

pass-through lowers demand elasticity prediction from 7,000 to approximately 600. To alleviate the concern that this result relies on a parametric model of expected returns, we also use the price instrument from Koijen and Yogo (2019) to predict returns and find broadly similar price pass-through estimates.

To estimate unspanned returns, we model the covariance of stock returns as linear functions of stock characteristics. The results indicate that approximately half of the expected return of stocks is unspanned. Incorporating this finding further lowers demand elasticity prediction from 600 to 12. To ensure that our findings are not merely the result of our parametric restrictions, we also directly estimate the full covariance matrix from daily returns and find similar conclusions. Applying the

commonly used shrinkage method of Ledoit and Wolf (2004) also has a minimal impact. Our findings further indicate that, even with flexible combinations of all other stocks, the spanning is still far from perfect. Our results contrast with theoretical models that rely on the CAPM being true (e.g., Berk and van Binsbergen, 2022). If the CAPM is true, then the expected return of every stock is almost perfectly spanned by the other stocks.

Our analysis so far suggests that considering realistic price pass-through and unspanned returns significantly bridges the gap between theoretical and empirical elasticity estimates. However, a portion remains unexplained (from 12 to 2). What explains the remainder? There could be numerous reasons — rigid investment mandates, trading frictions, inattention, or a combination of these. We provide a tentative suggestion that using more realistic functional forms to model investor portfolios may be an explanation. The MV investor considered so far has portfolio weights that are linear in stock characteristics. In contrast, empirical evidence suggests that institutional portfolios are closer to being exponential-linear in stock characteristics (Koijen and Yogo, 2019). Adopting this specification further reduces weight responsiveness (the sensitivity of portfolio weights to expected returns) and lowers demand elasticity from 12 to 2, as shown in the rightmost bar in Figure 1. We emphasize that our approach is one possible way to address the remaining discrepancy and alternative explanations for this gap exist.[4]

It is worth clarifying that we do not intend to offer a new empirical estimate of demand elasticity, and therefore, we do not present a new instrument. Instead, we derive theoretical predictions of what demand elasticities of a hypothetical mean-variance investor *should* be, based on estimates of price pass-through and unspanned returns. Once we use realistic estimates of these components, the predicted demand elasticities are much closer to empirical estimates from the existing literature.

Our primary contribution is presenting the decomposition framework, showing that demand

[4]There may be other explanations for why exponential-linear specifications can better fit portfolio data. For instance, the exponential-linear specification prohibits short-selling, which can be motivated by short-sale frictions.

elasticity depends on price pass-through and the size of unspanned returns. We show that using empirically estimated values for these two components can explain most of the gap between theoretical predictions and empirical estimates of demand elasticity. In addition to this application, our decomposition also makes it clear that demand elasticity is not a stable structural parameter, but rather it could vary depending on the investor, asset, and time period. For instance, our framework predicts that demand elasticities should be lower when considering portfolios at higher levels of aggregation, which are less substitutable (e.g., at the style level). This prediction is consistent with the results in Li, Fu, and Chaudhary (2022) and Li and Lin (2023). Furthermore, high-credit quality corporate bonds with short maturities necessarily have high price pass-throughs, as price deviations must revert by the maturity date, while longer maturity bonds may have lower price pass-throughs. Thus, our framework predicts higher demand elasticities for shorter maturity bonds, consistent with the findings in Li et al. (2022).

Two clarifications are in order. First, some readers may argue that demand elasticity is only well-defined when price pass-through is exactly equal to one, as typically assumed in one-period models.[5] We note that it is quite common for price pass-throughs to be lower than one in empirical settings. Moreover, nearly all empirical estimates of demand elasticity rely on price shocks that do not immediately revert. In the paper, we show that demand elasticity when pass-through is less than one is also well defined. Our aim is not to argue over semantics. Should one prefer to define demand elasticities with one-to-one pass-through, our paper can be understood as explaining why the empirical estimates of demand elasticities do not represent "true demand elasticities". This difference is due to a multiplicative factor equal to the price pass-through.

Second, our paper studies portfolio optimization problems for investors who take prices as given. We use empirical estimates to show that price pass-throughs are low and unspanned returns

[5]Dynamics models typically do not adhere to this assumption. For example, Johnson (2006) studies demand elasticities in dynamic economies. Price pass-throughs are significantly lower than one in all his applications.

are high, but we do not attempt to provide a general equilibrium explanation of how they arise in the first place. The main benefit of our analysis is that it relies on relatively weak assumptions, making it more broadly applicable. In particular, our two-component decomposition of demand elasticity into price pass-through and weight responsiveness can be applied to all portfolio choice settings.

## Related literature

Classic asset pricing theories predict high demand elasticities (e.g., Petajisto, 2009; Gabaix and Koijen, 2021). Our paper shows that their prediction arises from assuming low price pass-throughs and high substitutability between stocks. These models are not just "bookshelf models"; they influence how economists think. For instance, recently, Berk and van Binsbergen (2022) use a model with similar features to quantify the price impact of the rise of impact investing.

There is a long literature that estimates demand elasticities in stocks using price impacts of demand or supply shocks: index exclusion (Shleifer, 1986; Chang, Hong, and Liskovich, 2015; Pavlova and Sikorskaya, 2023), dividend payments (Schmickler, 2020), mutual fund flows (Lou, 2012), and trade-level price impacts (Frazzini, Israel, and Moskowitz, 2018; Bouchaud, Bonart, Donier, and Gould, 2018). More recently, there are also structural approaches that estimate demand elasticities using asset demand systems (Koijen and Yogo, 2019; Haddad, Huebner, and Loualiche, 2022). The estimates of micro price multipliers (inverse of micro elasticities) range from 0.3 to 15, much higher than what existing models predict.[6] Therefore, demand curves are much more inelastic compared to theory predictions. In this paper, we provide a microfoundation for inelastic demand.

Our finding that most stock price movements exhibit small pass-throughs to future expected

[6] See Table 1 and Figure 2 of Gabaix and Koijen (2021) for more details.

returns is consistent with extant findings in the cross-section of stock returns. Stock returns typically exhibit reversals within a month (Jegadeesh, 1990), momentum over quarterly to annual frequency (Jegadeesh and Titman, 1993), and reversals over multiple years (De Bondt and Thaler, 1985). These effects are much less than one-for-one pass-throughs and consistent with our estimates. The innovation of our paper is not in estimating these pass-throughs but rather showing their critical role in demand elasticity. In our framework, weak price pass-throughs leads to a low demand elasticity. Our estimated low pass-throughs are also consistent with the literature that explores discount rate and cash flow variations (e.g., Vuolteenaho, 2002; Cochrane, 2008).

Our finding of sizeable unspanned returns is consistent with Lopez-Lira and Roussanov (2023) who show that expected stock returns are not well spanned by factors. This low degree of spanning and abundance of "anomaly" alphas has been justified by a high degree of complexity (Martin and Nagel, 2022), among other explanations. We contribute to the literature by showing that high unspanned returns are a significant factor in determining asset demand elasticity. Additionally, we provide empirical estimates for the degree to which an asset is spanned by other assets, rather than just a variety of factors.

The remainder of the paper has three parts. In Section 2, we present our decomposition of demand elasticity into price pass-through and weight responsiveness. In Section 3, we estimate the components of demand elasticity and present evidence of low pass-throughs and high unspanned returns. Section 4 concludes.

# 2 A Decomposition of Demand Elasticity

In this section, we show that demand elasticities can be decomposed into two terms, price pass-through and weight responsiveness. The former is determined by the dynamic properties of

prices, and in a mean-variance setting, the latter depends on the distinctiveness of assets. We then apply our decomposition framework to shed light on why classic models deliver very high demand elasticities.

Our decomposition is motivated by the simple insight that demand for assets is fundamentally different from that of consumption goods, e.g., bananas. A consumer's elasticity of demand for bananas is simply determined by their primitive preference for bananas. In contrast, investors do not care about assets per se, but rather about the *returns* those assets generate, which ultimately contribute to their consumption or wealth. Therefore, how an investor adjusts her portfolio in response to price changes should primarily depend on her perception of how returns have changed.

## 2.1 A two-component decomposition

Consider an investor who holds $Q_{i,t}$ shares of asset $i$ at time $t$. The demand elasticity $\eta_{i,t}$ is defined as:

$$\eta_{i,t} \equiv -\frac{\partial \log(Q_{i,t})}{\partial \log(P_{i,t})}, \tag{2}$$

where $P_{i,t}$ represents the per-share asset price. For example, an elasticity of 4 means that a 1% drop in the price leads to an increased demand by 4% of shares invested.

To express demand elasticity using portfolio weights $w_{i,t}$, we can write $Q_{i,t} = A_t w_{i,t}/P_{i,t}$, where $A_t$ is the assets under management (AUM) or investor's wealth. Plugging this into Equation (2) and assuming that $A_t$ is exogenous, we have:[7]

$$\eta_{i,t} = 1 - \frac{\partial \log(w_{i,t})}{\partial \log(P_{i,t})}. \tag{3}$$

[7]In Equation (3), the first term, "one", is included because demand elasticity is defined in terms of shares held, $Q_{i,t}$, rather than portfolio weights, $w_{i,t}$. Concretely,

$$-\frac{\partial \log(Q_{i,t})}{\partial \log(P_{i,t})} = -\frac{\partial [\log(A_t) + \log(w_{i,t}) - \log(P_{i,t})]}{\partial \log(P_{i,t})} = 1 - \frac{\partial \log(w_{i,t})}{\partial \log(P_{i,t})}.$$

If the change in prices mainly affects the asset's expected return, $\tilde{\mu}_{i,t}$, we can use the chain rule to arrive at our decomposition (also stated in Equation (1)):

$$\eta_{i,t} \approx 1 + \underbrace{\frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}}}_{\text{weight responsiveness}} \cdot \underbrace{\left(-\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})}\right)}_{\text{price pass-through}}. \tag{4}$$

The first term, $\frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}}$, which we refer to as "weight responsiveness", describes how responsive the investor's portfolio weights are to expected returns. The second term, $-\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})}$, which we refer to as "price pass-through", measures the change in expected return in response to price movements. If a price increase of 1% lowers next-period expected returns by 0.4%, then the price pass-through would be 0.4. As these two terms are the focus of this paper, we denote them using $(\theta_{i,t}, \psi_{i,t})$ to simplify notation:

$$\theta_{i,t} \equiv \frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}} \quad \text{and} \quad \psi_{i,t} \equiv -\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})}, \quad \Rightarrow \quad \eta_{i,t} = 1 + \theta_{i,t}\psi_{i,t}. \tag{5}$$

Our decomposition only considers the demand response to changes in expected returns, even though demand can also respond to other drivers of demand (e.g., risk, hedging, etc.) which may also change with prices. We justify this choice in Appendix C, where we show that those other effects are small in a flexible portfolio choice model with Epstein-Zin utility. Therefore, we focus on the return-based effects for now.

This can also be intuitively understood using a passive investor who holds the market. A passive indexer has $\frac{\partial \log(w_{i,t})}{\partial \log(P_{i,t})} = 1$, meaning that a 1% increase in prices causes a 1% increase in portfolio weights and dollars invested in the asset. Thus the elasticity is zero (= 1 − 1), which aligns with the very idea of passive investing: price fluctuations do not lead to trading.

## 2.2 Weight responsiveness depends on unspanned returns

It is easy to see that the pass-through is determined by price dynamics. If a price movement is expected to mostly revert in the next period, then pass-through is approximately one. One the other hand, if the price movement is more persistent, the pass-through will be lower.

The determinants of weight responsiveness are less clear. In this section, we shed light on these factors by focusing on a mean-variance (MV) investor. It is important to note that we are not suggesting that the MV model realistically describes all investor behavior; we study more realistic models in Section 3.5. Rather, we use the MV model as a benchmark to understand how an optimizing investor without additional frictions or biases would behave. Focusing on MV investors also allows us to derive analytical expressions that help explain the underlying economic mechanism: weight responsiveness is high if and only if the asset is well-spanned by other assets.

**Model set up.** We consider a standard CARA utility model with multivariate normally distributed returns which delivers MV behavior. Suppose there are $N$ assets and consider an investor with CARA utility. Their maximization problem is:

$$\max_{w_t} \tilde{\mathbb{E}}_t \left[ -\exp\{-\gamma A_t \left( w_t' r_{t+1} + R_{f,t} \right)\} \right], \tag{6}$$

where $\gamma$ is the absolute risk aversion, $w_t$ is an $N$ dimensional vector of portfolio weights, $r_t$ is an $N$ dimensional vector of excess returns, and $\iota$ is an $N$ dimensional vector of ones. $R_{f,t}$ is the gross risk-free rate and $A_t$ is the investor's wealth, both of which are exogenous. The first-order condition gives the optimal portfolio weights:

$$w_t = \frac{1}{\gamma A_t} \tilde{\Sigma}_t^{-1} \tilde{\mu}_t, \tag{7}$$

where $\tilde{\Sigma}_t$ is the covariance matrix of returns. Thus, taking derivatives, we can write:

$$\frac{\partial w_{i,t}}{\partial \tilde{\mu}_{i,t}} = \frac{\tilde{\tau}_{i,t}}{\gamma A_t}, \tag{8}$$

where $\tilde{\tau}_{i,t}$ is the $i^{th}$ term along the diagonal of the precision matrix $\tilde{\Sigma}_t^{-1}$.

**The role of unspanned returns.** It turns out that weight responsiveness in Equation (8) depends on the degree to which asset $i$'s expected return is spanned by other assets. Formally, we define $\tilde{\beta}_{-i,t} = \tilde{\Sigma}'_{-i,t}\tilde{\Sigma}^{-1}_{-i,-i,t}$ as the $(N-1)$ row vector of betas (regression coefficients) of asset $i$ on all other assets, where $\tilde{\Sigma}_{-i,t}$ is the $i^{th}$ column vector excluding the $i^{th}$ row term, and $\tilde{\Sigma}_{-i,-i,t}$ is the $(N-1)\times(N-1)$ covariance matrix excluding the terms associated with asset $i$. Let $\tilde{\mu}_{-i,t}$ be the $(N-1)$ column vector of expected excess returns excluding the $i^{th}$ asset. Then we can define the following scalar which we call asset $i$'s unspanned returns:

$$\tilde{\alpha}_{-i,t} \equiv \tilde{\mu}_{i,t} - \beta'_{-i,t}\,\tilde{\mu}_{-i,t}, \tag{9}$$

which measures how well the asset's expected return is spanned by the other assets. Intuitively, if assets are almost perfect substitutes, then we would anticipate $\tilde{\alpha}_{-i,t}$ to be small. If assets are rather distinctive, we would anticipate large $\tilde{\alpha}_{-i,t}$. It is worth clarifying that $\tilde{\alpha}_{-i,t}$ is different from standard asset pricing alpha in models such as the CAPM. In CAPM, the explanatory variable — the market portfolio — is formed using all assets, whereas in Equation (9), the explanatory variables do not include asset $i$ itself.[8]

To emphasize the role of spanning, and to avoid confusion with the standard asset pricing alpha,

[8]Therefore, even in a world where CAPM holds, $\tilde{\alpha}_{-i,t}$ can be non-zero for all assets even though the CAPM alpha is zero by definition.

we can write $\tilde{\alpha}_{-i,t}$ in terms of the unspanned returns:

$$\tilde{\alpha}_{-i,t} = (1 - \omega_{i,t})\tilde{\mu}_{i,t}, \quad \text{with} \quad \omega_{i,t} \equiv \frac{\beta_{-i,t}\tilde{\mu}_{-i,t}}{\tilde{\mu}_{i,t}}, \tag{10}$$

where $\omega_{i,t}$ is the fraction of the return spanned by the other assets. Proposition 1 shows that the weight responsiveness for asset $i$ is equal to the inverse of $\tilde{\alpha}_{-i,t}$.

**Proposition 1.** *If portfolio weights take the form of* $w_{i,t} = \frac{1}{\gamma A_t}\tilde{\Sigma}_t^{-1}\tilde{\mu}_t$*, then for positive weights* $w_{i,t} > 0$*, weight responsiveness is given by:*

$$\theta_{i,t} = \frac{1}{(1 - \omega_{i,t})\tilde{\mu}_{i,t}}. \tag{11}$$

*Thus, it follows that the demand elasticity takes the following form:*

$$\eta_{i,t} = 1 + \frac{\psi_{i,t}}{(1 - \omega_{i,t})\tilde{\mu}_{i,t}}. \tag{12}$$

*Proof.* See Appendix A. □

Equation (12) follows immediately from Equations (5) and (11). In words, Equation (12) shows that the investor portfolio sensitivity to changes in expected return is inversely proportional to the asset's unspanned return. Investor demand is more inelastic for assets that are more distinctive and have larger unspanned returns.

One may wonder why the unspanned expected return enters Equation (12) but the unspanned (idiosyncratic) volatility does not. It is true that for assets with higher idiosyncratic volatility, the *level* of investor portfolio weight is lower. However, demand elasticity is defined in log terms, so the influence of idiosyncratic volatility drops out. We delve into this in greater detail in Appendix E.

## 2.3 Why do classic models predict extremely high demand elasticities?

Using our decomposition framework, we show that many theoretical models predict high demand elasticities because they assume high price pass-through and low unspanned returns.

**Price pass-through.** Many models assume that price pass-through is approximately one. This is often because they use static models where payoffs realize in the next period, so all current period price movements must revert by assumption (e.g., Petajisto, 2009; Berk and van Binsbergen, 2022). Even in dynamic settings, if one assumes that all future prices are fixed — so the current period price effect is fully temporary — then pass-through is also approximately one. Appendix B.1 provides further details.

**Unspanned returns.** We provide a calibration of the CARA-normal model in Section 2.2. The results closely resemble those in Petajisto (2009). We use these results to illustrate how classic models often yield high demand elasticities by assuming low unspanned returns. To begin with, assume that price pass-through is equal to one. To simplify the calibration, we assume all $N$ assets have the same expected returns, standard deviations, and correlations. We set the expected annual excess return at 0.06, the average correlation at 0.3, and the volatility at 0.3. These parameters are approximately similar to those observed in annual stock returns. We consider $N$ = 1,000 assets and assume that the investor holds an equal-weighted portfolio.[9] We set the CARA risk aversion coefficient times wealth, $\gamma A_t$, to 2.2 because this allows the portfolio weights to sum to one. Note that the exact value of $\gamma A_t$ only impacts the *level* of portfolio holdings, without affecting demand elasticities, which are defined using log changes. We set the risk-free rate to zero. Under these parameter assumptions, the $i^{th}$ term along the diagonal of the precision matrix is $\tilde{\tau}_i \approx 7.2$. This

[9]It is important to note that this assumption does not imply that we are only considering an equal-weighted portfolio; rather, we are calculating the elasticity for an asset with average weights.

calibration yields an elasticity of approximately 7,000 which is also plotted in the leftmost bar in Figure 1.[10]

Why is the demand elasticity so high? Apart from having a high price pass-through (assume to be one here), crucially, unspanned returns are close to zero in this calibration. Because correlation is uniform across all pairs of assets, asset returns share a single common factor. Thus, every asset $i$ has a beta of almost one to the equal-weighted portfolio formed using all other assets. Furthermore, the return of that equal-weighted portfolio is the same as asset $i$ itself, so most of the expected return of asset $i$ is fully spanned. In our calibration of the model above, the weight responsiveness is $\theta_{i,t} \approx 7{,}000$, indicating that $(1 - \omega_{i,t})\tilde{\mu}_{i,t} \approx 0.00014$. In other words, the unspanned return is merely 1.4 basis points.

This is not specific to our calibration; it is also true in other calibrations such as Petajisto (2009) and Berk and van Binsbergen (2022). In general, when a model, similar to the ones in these papers, assumes that a low dimensional factor model can explain expected returns well, it will generate low unspanned returns and thus high demand elasticities.

## 2.4 Why would price pass-through be less than one?

In Section 3, we empirically estimate the price pass-through and find it to be much less than one across many settings. What does it mean for the pass-through to be low? To better understand this, recall the Campbell-Shiller identity for the log dividend-to-price ratio, $dp_t$, omitting asset

[10] Naturally, this elasticity fluctuates as the parameter values change, particularly when considering assets with higher and lower weights. Experimenting with a range of parameter values typically produces an average elasticity across assets that is at least three orders of magnitude above unity. It is important to note that the average elasticity tends to increase with the number of assets $N$, due to the increased substitutability resulting from a larger number of assets.

subscript, $i$, to simplify notation:

$$dp_t \approx a + \sum_{h=0}^{\infty} \rho^h (\mu_{t+1+h} + r^f_{t+1+h}) - \sum_{h=0}^{\infty} \rho^h \tilde{E}_t(\Delta d_{t+1+h}),$$

where $a$ is a constant, $(\mu_t + r^f_{f,t})$ is the log return inclusive of risk-free rate, $\Delta d_t$ is the log change in dividend, and $\rho \approx 0.97$ is a discount factor. Therefore, a cash flow (dividend)-unrelated change in the current price must predict changes in the stream of future expected returns $(\mu_{t+1}, \mu_{t+2}, ...)$. Using the Campbell-Shiller identity, we note that a price pass-through less than one simply means that the effect is not solely on the immediate next-period return, $\mu_{t+1}$, but also on the expected returns in future periods. This realization also implies that, unless further restrictions are imposed, price deviations that are unrelated to future dividends do not necessarily lead to a price pass-through of one. This insight raises two possible questions.

First, why would price pass-through be less than one? In Appendix B.2, we theoretically show that this occurs naturally in several settings. For instance, consider the case where the current price movement is caused by an uninformed supply shock — the typical empirical setting used in estimating demand elasticities. In general equilibrium models with long-lived assets, as long as the supply shock does not revert immediately, we should expect the impact on expected returns to also be persistent (e.g., Johnson, 2006; Gabaix and Koijen, 2021). To see this, consider a permanent increase in the supply of shares. In most models, this leads to a permanently higher risk premium in all future periods. Assuming the asset is infinitely lived like a stock, the price pass-through for an annual horizon is only $1 - \rho \approx 0.03$. If the supply increase reverts slowly, price pass-through would be higher, but still not as high as one (e.g., An and Zheng, 2023). Furthermore, investors may also perceive (or fear) that price movements contain private information about future prices, which also naturally leads to lower-than-one pass pass-throughs. There may also be other possibilities. In this

paper, we do not take a stance on what price pass-through should be but empirically estimate it in Section 3.

Second, one may insist that demand elasticities are only defined when the price pass-through is one. This may be motivated by the desire to "hold value fixed and only vary prices". Specifically, if only $P_t$ varies, but all future prices $\{P_{t+1}, P_{t+2}, ...\}$ are held fixed, then pass-through is indeed approximately equal to one. While this is justified when modeling the demand for consumption goods such as bananas whose utility value does not change, this definition is less obvious for modeling the demand of assets. The value of an asset, to an investor, primarily lies in its return. As mentioned above, from the Campbell-Shiller identity, any dividend-unrelated price change *must* be associated with changes in future expected returns, so it is simply impossible to "hold value fixed". Restricting attention to price pass-through of one amounts to requiring that all adjustments happen via the next-period return, which is somewhat arbitrary.

That said, we do not wish to argue on semantics. If one insists on using one-to-one price pass-throughs to define "true demand elasticities", our paper simply points out that most empirical studies estimate demand responses using price shocks with lower pass-throughs. According to our decomposition in Equation (1), one can always adjust for the effect of price pass-through to arrive at the "true demand elasticities".

# 3 Empirical Estimates of Demand Elasticities

In Section 2, we showed that price pass-through and unspanned returns are the key determinants of demand elasticity. In this section, we empirically estimate these two channels to determine what demand elasticity values are plausible. We start by modeling stock returns and covariances using stock characteristics, as is done by many quantitative funds, and then derive the resulting

mean-variance (MV) efficient demand. We then show that our results are robust to alternative specifications for the data-generating process for stock returns. Finally, we also explore alternative functional forms for investor demand.

## 3.1 Data

We use the standard CRSP-Compustat merged dataset for returns and asset characteristics. We follow Koijen and Yogo (2019) in calculating the stock characteristics of profitability, book-to-market ratios, investment, dividend-to-book ratios, and beta. To form returns for the quarterly and annual frequencies, we cumulate returns from the monthly stock data. To form weekly frequency stock data, we cumulate returns from the daily frequency.

We use Treasury bill rates from Ken French's website as the risk-free rate for monthly, weekly, and daily frequencies. For the quarterly and annual frequencies, we use the 3-month and 1-year Treasury bill rates from the Federal Reserve Economic Data (FRED), identified by their respective codes TB3MS and GS1. We also obtain daily, weekly, and monthly Fama and French (2015) 5-factor and momentum returns from Ken French's website. We cumulate these returns to obtain quarterly and annual factor returns.

We use the quarterly log consumption-to-wealth deviations, defined in Lettau and Ludvigson (2001), from Martin Lettau's website. Finally, we obtain quarterly 13F institutional holdings data from Thomson Reuters when replicating the Koijen and Yogo (2019) price instrument in Section 3.4.

## 3.2 Demand elasticity with characteristics-based models of stock returns

In this section, we focus on a price-taking MV investor who, similar to a quantitative fund, estimates a model of stock prices where both expected returns and covariances depend on stock

characteristics. In academic research, such models have been estimated by Kelly, Pruitt, and Su (2019), among others. Similarly, in the industry context, the Barra risk model has long utilized stock characteristics to estimate covariances. We use monthly stock return data, where excess returns are calculated as the stock returns minus the return on the one-month T-bill rate. Our primary focus is on whether the estimated price pass-through and unspanned returns can justify inelastic demand.

We model the mean vector and covariance matrix of the excess returns of $N$ stocks ($\tilde{\mu}_t$ and $\tilde{\Sigma}_t$) as linear functions of stock characteristics:

$$\tilde{\mu}_t = Z_t^{\mu}\pi, \quad \tilde{\Sigma}_t = \Gamma_t\Gamma_t' + \zeta I_N, \quad \text{and,} \quad \Gamma_t = Z_t^{\Gamma}\xi, \tag{13}$$

where $Z_t^{\mu}$ and $Z_t^{\Gamma}$ are matrices of characteristics, including functions of prices, with $N$ rows and the number of columns corresponding to the number of characteristics in each matrix.[11] $\pi$ and $\xi$ are column vectors of parameters, $\zeta > 0$ is a scalar that controls the variance and ensures that the covariance matrix is positive definite, and $I_N$ is the identity matrix. Let $Z_{i,j,t}^{\mu}$ and $Z_{i,j,t}^{\Gamma}$ denote the term in the $i^{th}$ row and $j^{th}$ column of $Z_t^{\mu}$ and $Z_t^{\Gamma}$, respectively. With fixed parameters $\pi$, we can calculate the price pass-through as:

$$\psi_{i,t} = -\frac{\partial\tilde{\mu}_{i,t}}{\partial\log(P_{i,t})} = -\sum_j \frac{\partial Z_{i,j,t}^{\mu}}{\partial\log(P_{i,t})}\pi_j. \tag{14}$$

For the matrix of characteristics used to model the mean return, $Z_t^{\mu}$, we use a set of characteristics similar to those in Koijen and Yogo (2019). The price-related characteristics include the most recent log price change, $\Delta p_{i,t} = p_{i,t} - p_{i,t-1}$, the log book-to-market ratio, $b_{i,t} - p_{i,t}$, and the log market capitalization cross-sectionally normalized every period, $(p_{i,t} - \bar{\mu}_t^p)/\bar{\sigma}_t^p$. We also use additional

[11] Koijen and Yogo (2019) have the similar functional form for the mean and covariance matrix of returns. We also follow Kelly et al. (2019) and many others in modeling expected returns and factor loadings as linear functions of the observable asset characteristics.

characteristics that are not functions of prices including profitability, investment, dividend-to-book ratio, and market beta. We also add a column vector of ones to estimate an intercept term.

When stock prices change, the price-related characteristics change their values and their prices of risk ($\pi_j$) determine the price pass-throughs. Let $\pi_{rev}$, $\pi_{bp}$, and $\pi_{size}$ denote the elements of $\pi$ corresponding the reversals variable, book to market ratio, and size variable, respectively — the three price-related characteristics. Then the price pass-through is:

$$\psi_{i,t} = -\pi_{rev} + \pi_{bp} - \frac{\pi_{size}}{\sigma_t^p}. \tag{15}$$

In the matrix of characteristics that controls the covariance terms ($Z_t^\Gamma$), we only incorporate the price-unrelated characteristics. In robustness checks, we find that including price-related characteristics makes little difference. In fact, Section 3.3 shows that estimating covariance matrices using daily data without parametric restrictions also makes little difference.

We estimate the parameters, $\pi$, $\xi$, and $\zeta$ using maximum likelihood, assuming that returns are multivariate normally distributed. The parameter estimates are presented in Table 1. Once we have these parameters, we have a complete data-generating process for prices, so we can calculate the pass-through, $\psi_{i,t}$, and the unspanned return, $(1 - \omega_{i,t})\tilde{\mu}_{i,t}$. This allows us to determine the MV theory-predicted demand elasticity for each asset under optimal demand in Equation (12).

Table 2 presents demand elasticities and their components. We calculate the demand elasticity only for portfolio positions with strictly positive weights since these are the only elasticity values that are well-defined (Koijen and Yogo, 2019). For consistency, all terms reported in Table 2 are those with strictly positive portfolio weights. We present the median demand elasticity in the first row of Table 2; the standard errors are obtained via bootstrap.[12] The median price pass-through

[12]Specifically, we first bootstrapped by selecting random months with replacement, followed by selecting random stocks with replacement, and then computing the median value in this bootstrapped sample. This procedure is repeated 10,000 times, and the resulting standard deviation of the calculated median values is reported as the standard error.

is 0.062, meaning that every 1% price decline translates into a 6.2 basis point higher expected return in the subsequent month. This value is positive, indicating that an investor using such a return-forecasting model would be trading against observed price movements. However, it is substantially lower than one. The median unspanned monthly return is 0.5% which is almost half of the total expected return of 1.1%. This means that, contrary to the assumptions in standard cross-sectional asset pricing factor models, spanning is far from perfect.[13] The resulting median demand elasticity across all stocks is approximately 12. If we plug in the median values of pass-through and unspanned returns, we obtain something similar with $1 + 0.062/0.005 \approx 13.4$.[14]

We report median values because simple averages are distorted by a few stocks with close-to-zero portfolio weights. This happens to stocks with marginally positive unspanned returns. Because demand elasticities are computed in logs, for those positions, small changes in portfolio weights can lead to large changes in log portfolio weights, generating very large demand elasticities. If we take a simple average over all stocks, the result will be distorted by those outliers that are not economically important — they only constitute a small fraction of the total portfolio value. Consistent with that, the second row reports that the average demand elasticity is approximately 81.

We then verify that this is indeed due to outliers with small portfolio weights. In subsequent rows of Table 2, we report averages across stocks after dropping outliers. For instance, in the third row, we recompute averages using 99.9% of portfolio positions, dropping those with the lowest weights that constitute 0.1% of the portfolio by value. The average demand elasticity drops to around 19. The subsequent rows progressively take more restrictive subsamples and the average demand elasticity drops further. Finally, the last row presents the value-weighted average and obtains an average demand elasticity of approximately 10.

[13] The median return is higher than the typical stock return because we are reporting results for stocks with positive weights.

[14] There is a small difference, as the median of the ratio between two variables is not necessarily the ratio of the medians.

Overall, our results indicate that taking into account realistic pass-throughs and unspanned returns generates demand elasticities in the range of 10. While it is still higher than the typical empirical estimate of around 2, it is substantially lower than the theoretical predictions of around 7,000. If we conduct a back-of-the-envelope exercise, we find that both channels are important to our result. The estimated median pass-through is about 0.06. Taking this into account, lowers the predicted demand elasticity to approximately 600, as shown in Figure 1. The remaining drop is attributed to higher unspanned returns.

We address two possible concerns. First, one may suspect that high unspanned returns indicate unrealistically high Sharpe ratios. This isn't the case. Using our model of stock returns, we find that the maximum Sharpe ratio is 1.45 with a standard error of 0.059. This is calculated in-sample; the out-of-sample Sharpe ratio is likely lower. This Sharpe ratio is modest when compared to modern asset pricing factor models (Gu, Kelly, and Xiu, 2021). Second, our findings are robust to using a wide range of empirical models for expected returns. Davis (2023) calculates the demand elasticity of 13 portfolio choice models from the literature using a much broader set of asset characteristics and finds similar elasticity values.

## 3.3 Robustness: Unrestricted covariance matrix

Section 3.2 imposes a parametric restriction on the covariance matrix using stock characteristics. There might be concerns that the high unspanned returns are specific to this restriction. To alleviate these concerns, we show that our results are robust to directly estimating the covariance matrix without restrictions.

Following Lopez-Lira and Roussanov (2023), we estimate the stock-level covariance matrix on a rolling one-year basis using daily data in each month. Because we have more stocks than the number of days in a year, the sample covariance matrix is not full-rank. Unlike Lopez-Lira

and Roussanov (2023) who do not require a full-rank covariance matrix, we need one to construct the mean-variance optimal portfolio. To get a full-rank covariance matrix, we use a standard covariance matrix shrinkage method, similar to Ledoit and Wolf (2004), which is popular among finance practitioners. Specifically, we use $\Sigma = (1-h)\hat{\Sigma} + h\overline{\Sigma}$, where $\hat{\Sigma}$ is the sample covariance matrix estimate, $\overline{\Sigma}$ is the shrinkage target, and $h$ is the scalar shrinkage weight. The shrinkage target matrix has the average stock-specific return variance along the diagonal ($\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{i,i}$) and the average covariance on the off-diagonal ($\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\hat{\Sigma}_{i,j}$).

Like Lopez-Lira and Roussanov (2023), we consider asset expected excess returns that are a function of characteristics based on the same rolling year, using monthly data.[15] We accomplish this simply by regressing excess returns on stock characteristics. The issue with rolling regressions is that they sometimes lead to negative price pass-throughs, which would imply negative demand elasticities. While including such negative estimates makes it easier to explain low demand elasticities, we prefer a more conservative approach. So, we estimate constrained least squares regressions where we constrain the coefficients to have a positive price pass-through.[16]

Our sample is monthly from January 1971 to July 2019. In each month, we form mean-variance optimal portfolios using the shrinkage-based covariance matrix estimated from the previous twelve months, as well as the predicted excess returns from the characteristics and the regression coefficients estimated over the previous twelve months. Table 3 presents the results of this exercise. We consider a large range of shrinkage weights ranging from $h = 0.05$ to 0.95. In all specifications, the median demand elasticity ranges from 7.6 to 8.3, while the weighted average

[15] Note that while we can avoid using characteristics for the covariance matrix, we still need to use characteristics to model the expected returns to infer price pass-throughs.

[16] Specifically, we constrain the coefficient on price reversals, $\Delta p_t$, minus the coefficient on book to market, $b_t - p_t$, to be less than or equal to zero. Moreover, given that the size variable, $(p_t - \mu_t^p)/\sigma_t^p$, has a slightly time-varying derivative, it is easier to exclude this variable from the set of characteristics. It is important to note that including the size variable has minimal impact on the elasticity results. Otherwise, we use the same characteristics mentioned earlier.

elasticity is in the range of 7.5 to 9.2. The unspanned returns are similar to those in Section 3.2 and lie in the range of 0.6% to 0.7% per month. Therefore, the results are not very sensitive to shrinkage weights. Overall, we conclude that our earlier finding — that unspanned returns are large — is not due to imposing specific parametric restrictions on the covariance matrix.

## 3.4 Price pass-through with the Koijen and Yogo (2019) instrument

One may also be concerned that our price pass-through estimate is constrained by the characteristics-based specification. To alleviate this concern, we directly estimate the price pass-throughs associated with the price instrument in Koijen and Yogo (2019) and find that the implied pass-through is even lower.

We estimate the following panel regression for multiple horizons $h$:

$$r_{i,t+1\to t+h} = \beta_0 + \beta_1 \log\left(\frac{M_t}{B_t}\right) + (X_t^c)'\beta + \epsilon_{i,t+1}, \tag{16}$$

where the main independent variable, log market-to-book ratio, $\log(\frac{M_t}{B_t})$, is instrumented using $\log(\frac{\hat{M}_t}{B_t})$, where $\hat{M}_t$ is the instrument in Koijen and Yogo (2019). The coefficient estimate of $(-\beta_1)$ reveals the implied price pass-through. We also control for the same vector of stock characteristics as in Koijen and Yogo (2019) and control for time-fixed effects. We calculate Driscoll-Kraay standard errors (Driscoll and Kraay, 1998) with 8 quarterly lags which control for both time-series and cross-sectional correlations.

The results are reported in Table 4. The first four columns report equal-weighted results for forecasting returns for the subsequent one, two, four, and eight quarters. We find pass-through coefficients of 0.131 and 0.352, respectively, at the quarterly and annual frequencies, which are statistically significant but also much lower than the $-\frac{\partial \tilde{\mu}}{\partial \log(P)} = 1$ benchmark in frictionless models.

To compare to the price pass-through in Section 3.2, we convert these estimates to monthly pass-through estimates of 0.044 and 0.029, respectively. Therefore, the implied price pass-through from the Koijen and Yogo (2019) instrument is even lower than the 0.06 estimate in Section 3.2.

Further, the implied price pass-through is even weaker in larger stocks. To obtain more economically relevant results from the perspective of large institutions, columns (5) to (8) estimate value-weighted regressions. To account for the fact that market size has grown dramatically over time, we normalize the sum of weights in each period to one. The results are now only statistically significant at the 5% level, but not the 1% level. In addition, the degrees of reversion are much smaller, with $-\frac{\partial \tilde{\mu}}{\log(P)}$ only 0.018 and 0.062 at the quarterly and annual frequencies. These imply monthly price pass-throughs of only 0.005 to 0.006, much less than one and even lower than what we found earlier. In summary, while there is evidence of positive price pass-throughs associated with the Koijen and Yogo (2019) instrument, the pass-throughs are quite weak, especially when considering large-cap stocks that institutions more heavily hold.

## 3.5 Exponential-linear demand specification

As explained in Section 3.2, when considering realistic price pass-through and unspanned returns, demand elasticities of MV investors drop to approximately 12. This is much lower than the classic theory predictions of 7,000, but still higher than typical empirical estimates which are on the order of 1 to 2. What accounts for the remaining gap? There are many possibilities — rigid investment mandates, trading frictions, inattention, behavioral biases, or a combination of those.

In this subsection, we provide a tentative suggestion that using more realistic demand function specifications may explain a significant part of the gap. Koijen and Yogo (2019) and Koijen, Richmond, and Yogo (2023) note that the MV model does not fit empirical observations of 13F institutional holdings data. Most importantly, portfolio weights exhibit fat tails and are more

closely described by lognormal distributions, while MV portfolio weights are typically closer to being normally distributed. Furthermore, MV portfolio weights often contain many short positions while many institutions tend to shy away from shorting. Therefore, to better fit data, Koijen and Yogo (2019) consider demand functions that are exponential-linear in stock characteristics instead of the MV demand which, when implemented using specifications in Section 3.2, is linear in stock characteristics. In this section, we find that switching to this exponential-linear functional form lowers demand elasticities by nearly another order of magnitude.

Following Koijen and Yogo (2019), we assume portfolio weights are exponential-linear in characteristics $\{Z^{\mu}_{i,j,t}\}$ with coefficients $\{b_j\}_j$:

$$w_{i,t} = \frac{\delta_{i,t}}{\sum_k \delta_{k,t}} \quad \text{where} \quad \delta_{i,t} \equiv \exp\left(\sum_j Z^{\mu}_{i,j,t} b_j\right) \tag{17}$$

The demand elasticity is calculated as:

$$\eta_{i,t} = 1 - (1 - w_{i,t}) \left(\sum_j \frac{\partial Z^{\mu}_{i,j,t}}{\partial \log(P_{i,t})} b_j\right) \tag{18}$$

Unlike the MV case where there is an analytical solution for the maximum Sharpe ratio portfolio, in the exponential-linear specification, we need to numerically solve for it by varying coefficients $\{b_j\}_j$. We follow Brandt, Santa-Clara, and Valkanov (2009) in conducting this optimization. Table 5 shows the resulting parameter estimates. Standard errors are calculated using the usual formula for extremum estimators (Newey and McFadden, 1994). The first few columns report the demand response coefficients to stock characteristics (the $\{b_j\}_j$ in Equation (17)). The average elasticity in the last column, calculated according to Equation (18) above, is 2.6. We illustrate this using the rightmost bar in Figure 1.

How can we understand exponential-linear demand in light of our two-component decomposition? Intuitively, if the price-related stock characteristics are linear in log price, then exponential-linear demand is approximately:

$$w_{i,t} \approx \exp(-\beta \log(P_{i,t}) + \upsilon_{i,t}) = e^{\upsilon_{i,t}} P_{i,t}^{-\beta}, \tag{19}$$

where $\upsilon_{i,t}$ is the sum of all price-unrelated demand determinants. The resulting demand elasticities:

$$\eta_{i,t} = 1 - \frac{\partial \log w_{i,t}}{\partial \log(P_{i,t})} = 1 + \beta \tag{20}$$

Equating this to our decomposition in Equation (4) yields:

$$\theta_{i,t} \cdot \psi_{i,t} = \beta.$$

Therefore, imposing exponential-linear demand amounts to locking together price pass-through $\psi_{i,t}$ and weight responsiveness $\theta_{i,t}$. Because we are using the same set of stock characteristics as in Section 3.2, insofar as we are holding pass-through fixed, the further drop in demand elasticity when using exponential-linear demand is due to lower weight responsiveness. This can be seen from the weight-responsiveness just being proportional to the inverse of the pass-through: $\theta_{i,t} = \beta/\psi_{i,t}$. Appendix B.3 presents further discussions.

Overall, this exercise suggests that using more realistic specifications for demand may explain a significant part of the remaining gap. Of course, there can be alternative explanations based on frictions or biases. In fact, the exponential-linear specification itself may be motivated by frictions or biases. For instance, the exponential-linear demand rules out shorting which may be motivated

by short-sale constraints. All portfolio weights are positive in this exercise, alleviating any concerns about conditioning only on positive portfolio weights in the previous section.

# 4 Conclusion

In this paper, we present a novel decomposition of demand elasticity into (one plus) the product of two components: "price pass-through" and "weight responsiveness". The former measures the extent to which prices predict returns, and the latter captures the magnitude of portfolio weight changes in reaction to shifts in expected returns. We show that for a mean-variance optimizer, the weight responsiveness is inversely proportional to "unspanned returns" which measures the amount of asset return not spanned by other assets. The more distinctive is an asset, the less its returns are spanned by others, and the lower is weight responsiveness.

We show that incorporating realistic price pass-throughs and unspanned returns can explain most of the gap between the high demand elasticity values in classic theories and the low estimates found in the data. When we use a model of stock returns and covariances based on stock characteristics, as is often done by quantitative funds, we find price pass-through to be much lower and unspanned returns much higher than that typically assumed in classic models. Using our estimates, we find that even for an investor who holds a mean-variance efficient portfolio, the implied demand elasticity is only around 12, significantly lower than the 7,000 predictions from class theories. Our findings are robust to alternative ways to estimate price pass-through and unspanned returns.

Therefore, our results imply that low demand elasticity is consistent with investor rationality. Even though frictions and biases can — and likely do — play a role in explaining the inelastic demand observed in real life, a significant fraction of the difference with theoretical predictions is accounted for without frictions or biases. Moreover, our decomposition of demand elasticity

into price pass-through and weight responsiveness is broadly applicable across various models of portfolio choice.

# References


An, Yu, and Zeyu Zheng, 2023, An dynamic factor model of price impacts, Working Paper, Johns Hopkins.

Berk, Jonathan, and Jules H. van Binsbergen, 2022, The impact of impact investing, Working Paper, Stanford GSB and Wharton.

Bouchaud, Jean-Philippe, Julius Bonart, Jonathan Donier, and Martin Gould, 2018, *Trades, Quotes and Prices: Financial Markets Under the Microscope* (Cambridge University Press).

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns, *Review of Financial Studies* 22, 3411–3447.

Campbell, John Y., Yeung Lewis Chan, and Luis M. Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression discontinuity and the price effects of stock market indexing, *Review of Financial Studies* 28, 212–246.

Chaudhry, Aditya, 2023, Do subjective growth expectations matter for asset prices?, Working Paper, OSU.

Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Davis, Carter, 2023, Mispricing and the elasticity of quantitative investment, Working Paper, Indiana.

De Bondt, Werner F.M., and Richard Thaler, 1985, Does the stock market overreact?, *Journal of Finance* 40, 793–805.

Driscoll, John C., and Aart C. Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, *Review of Economics and Statistics* 80, 549–560.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1 – 22.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, 2018, Trading costs of asset pricing anomalies, Working Paper, AQR and Yale.

Gabaix, Xavier, and Ralph S.J. Koijen, 2021, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Working Paper 28967, National Bureau of Economic Research.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2021, Autoencoder asset pricing models, *Journal of Econometrics* 222, 429–450.

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.

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

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets., *Journal of Finance* 54, 2143.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Johnson, Timothy C., 2006, Dynamic liquidity in endowment economies, *Journal of Financial Economics* 80, 531–562.

Kelly, Bryan T., Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return., *Journal of Financial Economics* 134, 501 – 524.

Koijen, Ralph S.J., Robert J. Richmond, and Motohiro Yogo, 2023, Which investors matter for equity valuations and expected returns?, *Review of Economic Studies* Forthcoming.

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

Ledoit, Olivier, and Michael Wolf, 2004, Honey, I shrunk the sample covariance matrix, *Journal of Portfolio Management* 30, 110–119.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.

Li, Jiacui, and Zihan Lin, 2023, Prices are less elastic for less diversifiable demand, Working paper.

Li, Jian, Zhiyu Fu, and Manav Chaudhary, 2022, Corporate bond elasticities: Substitutes matter, Working Paper, Columbia and Chicago.

Lopez-Lira, Alejandro, and Nikolai L. Roussanov, 2023, Do common factors really explain the cross-section of stock returns?, Working Paper, Florida and Wharton.

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

Martin, Ian W.R., and Stefan Nagel, 2022, Market efficiency in the age of big data, *Journal of Financial Economics* 145, 154–177.

Milgrom, Paul, and Nancy Stokey, 1982, Information, trade and common knowledge, *Journal of Economic Theory* 26, 17–27.

Newey, Whitney K, and Daniel McFadden, 1994, Large sample estimation and hypothesis testing, *Handbook of econometrics* 4, 2111–2245.

Pavlova, Anna, and Taisiya Sikorskaya, 2023, Benchmarking intensity, *The Review of Financial Studies* 36, 859–903.

Petajisto, Antti, 2009, Why do demand curves for stocks slope down?, *Journal of Financial and Quantitative Analysis* 44, 1013–1044.

Schmickler, Simon, 2020, Identifying the price impact of fire sales using high-frequency surprise mutual fund flows, Working Paper, Princeton.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41, 579–590.

Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns?, *Journal of Finance* 57, 233–264.

**Table 1. Maximum Likelihood Estimation.** In this table we show the parameter estimates of $\pi$, $\xi$, and $\zeta$ in Equation (13). The estimates of $\pi$ are shown in the first row labeled "Mean". The estimates of $\xi$ and the scalar $\zeta$ are shown in the next row labeled "Covariance". The parameter estimates are labeled by their respective covariates in $Z_t^{\mu}$ and $Z_t^{\Gamma}$ above the estimates. We estimate the model using maximum likelihood, assuming returns in each period have a multivariate normal distribution. Standard errors are shown in parentheses.

| | intercept | $\Delta p_t$ | $b_t - p_t$ | $\frac{p_t - \mu_t^p}{\sigma_t^p}$ | profit. | invest. | div.-to-book | beta | |
|---|---|---|---|---|---|---|---|---|---|
| Mean | $0.008^{***}$ | $-0.060^{**}$ | 0.001 | $-0.002^{***}$ | $0.011^{***}$ | $-0.019^{***}$ | −0.003 | 0.007 | |
| | (0.001) | (0.036) | (0.003) | (0.000) | (0.003) | (0.000) | (0.040) | (0.281) | |
| | intercept | | | $\frac{b_t - \mu_t^b}{\sigma_t^b}$ | profit. | invest. | div.-to-book | beta | $\zeta$ |
| Covariance | 0.036 | | | −0.006 | −0.020 | 0.003 | −0.168 | 2.174 | $0.027^{***}$ |
| | (0.598) | | | (0.036) | (0.241) | (0.034) | (3.975) | (3.979) | (0.003) |
| *Note:* | | | | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

**Table 2. Implied Optimal Demand Elasticities and the Components.** Using the estimated parameters values shown in Table 1, we compute the demand elasticity ($\eta_{i,t}$), price pass-through ($\psi_{i,t}$), unspanned return ($(1-\omega_{i,t})\tilde{\mu}_{i,t}$), expected return ($\tilde{\mu}_{i,t}$), and the component of the return spanned by all other assets ($\omega_{i,t}\tilde{\mu}_{i,t}$) for every stock and in every period where the optimal portfolio weight is positive. Optimal portfolio weights are calculated as $w_t = \frac{1}{\gamma A_t}\tilde{\Sigma}_t^{-1}\tilde{\mu}_t$. The first row shows the median values. The next row shows the mean. The $(1-\omega_{i,t})\tilde{\mu}_{i,t}$ term is sometimes very close to zero, which creates near infinite elasticity values (Equation 12), so we consider various cuts of the data that exclude outliers. The "Top Percent of Value Invested" column, if not empty, indicates that we use a subsample of the positive-weight observations when calculating the averages. For example, a 99.9% filter means we screen out the bottom 0.1% of positions (in value weights) with the smallest holdings. The last row shows the portfolio weight-weighted average. Standard errors are shown in parentheses. For the averages in rows 2 to 6, these standard errors are double clustered by stock and month. For the median in row 1, standard errors are calculated via bootstrap as described in the text. The in-sample maximum annual Sharpe ratio for this model is 1.45, with a standard error of 0.059.

| Statistic | Top Percent of Value Invested | Demand Elasticity | Price Pass-Through | Unspanned Return | Total Return | Spanned Return | Observations |
|---|---|---|---|---|---|---|---|
| | | $\eta_{i,t}$ | $\psi_{i,t}$ | $(1-\omega_{i,t})\tilde{\mu}_{i,t}$ | $\tilde{\mu}_{i,t}$ | $\omega_{i,t}\tilde{\mu}_{i,t}$ | |
| Median | | 11.784*** | 0.062*** | 0.006*** | 0.011*** | 0.005*** | 1,344,642 |
| | | (0.014) | (0.000) | (0.000) | (0.000) | (0.000) | |
| Average | | 81.423*** | 0.062*** | 0.007*** | 0.013*** | 0.005*** | 1,344,642 |
| | | (10.484) | (0.000) | (0.000) | (0.000) | (0.000) | |
| Average | 99.9% | 19.145*** | 0.062*** | 0.008*** | 0.013*** | 0.005*** | 1,299,830 |
| | | (0.173) | (0.000) | (0.000) | (0.000) | (0.000) | |
| Average | 99.0% | 13.939*** | 0.062*** | 0.008*** | 0.013*** | 0.005*** | 1,200,188 |
| | | (0.114) | (0.000) | (0.000) | (0.000) | (0.000) | |
| Average | 97.5% | 11.973*** | 0.062*** | 0.009*** | 0.014*** | 0.005*** | 1,113,961 |
| | | (0.093) | (0.000) | (0.000) | (0.000) | (0.000) | |
| Average | 95.0% | 10.518*** | 0.062*** | 0.009*** | 0.015*** | 0.005*** | 1,015,555 |
| | | (0.078) | (0.000) | (0.000) | (0.000) | (0.000) | |
| Weighted Avg. | | 10.100*** | 0.062*** | 0.013*** | 0.019*** | 0.006*** | 1,344,642 |
| | | (0.080) | (0.000) | (0.000) | (0.000) | (0.000) | |

*Note:* *p<0.1; **p<0.05; ***p<0.01

**Table 3. Optimal Demand Elasticities with Rolling Estimates of Expected Return and Covariance.** This table shows the out-of-sample performance of the portfolio in Section 3.3. Each column represents a different portfolio with a different degree of shrinkage; the "shrinkage weights" refers to the fraction of weight given to the shrinkage target as opposed to the in-sample covariance matrix estimate. The first row shows the resulting portfolio Sharpe ratio. The remaining rows report stock-level statistics including expected return, unspanned return, as well as median and various versions of average stock-level demand elasticities. The stock-level statistics are shown for stocks receiving positive portfolio weights. The average elasticity values are also shown for subsamples of assets with the largest portfolio weights. For example, the top 99.9% subsample labeled below represents the assets that compose 99.9% of the value invested (in terms of portfolio weights), dropping the smallest weights first. To conserve space, we do not report results for all shrinkage weights, but Table A.1 in the appendix shows results for a fuller set of shrinkage weights in $\{0.05, 0.1, 0.15, ..., 0.95\}$.

| Shrinkage Weight: | 0.05 | 0.25 | 0.50 | 0.75 | 0.85 | 0.95 |
|---|---|---|---|---|---|---|
| Sharpe ratio | 0.957 | 1.017 | 1.106 | 1.219 | 1.267 | 1.268 |
| Expected Return | 0.015 | 0.016 | 0.016 | 0.016 | 0.017 | 0.017 |
| Unspanned Return | 0.006 | 0.006 | 0.006 | 0.006 | 0.007 | 0.007 |
| Median Elasticity | 8.331 | 8.369 | 8.408 | 8.328 | 8.147 | 7.567 |
| Avg. Elasticity (all pos. weights) | 93.580*** | 196.684* | 68.124*** | 68.808*** | 71.234*** | 74.132*** |
| | (23.721) | (101.356) | (3.977) | (4.628) | (7.180) | (7.815) |
| Avg. Elasticity (top 99.9%) | 18.792*** | 18.782*** | 18.740*** | 17.932*** | 17.131*** | 15.241*** |
| | (0.673) | (0.670) | (0.664) | (0.610) | (0.562) | (0.468) |
| Avg. Elasticity (top 99%) | 13.185*** | 13.228*** | 13.162*** | 12.664*** | 12.121*** | 10.834*** |
| | (0.460) | (0.462) | (0.455) | (0.421) | (0.388) | (0.324) |
| Avg. Elasticity (top 97.5%) | 11.130*** | 11.166*** | 11.104*** | 10.718*** | 10.257*** | 9.194*** |
| | (0.383) | (0.384) | (0.378) | (0.352) | (0.324) | (0.270) |
| Avg. Elasticity (top 95%) | 9.632*** | 9.659*** | 9.610*** | 9.295*** | 8.901*** | 7.999*** |
| | (0.326) | (0.327) | (0.323) | (0.300) | (0.276) | (0.230) |
| Weighted Avg. Elasticity | 9.208*** | 9.223*** | 9.149*** | 8.775*** | 8.384*** | 7.518*** |
| | (0.323) | (0.322) | (0.314) | (0.286) | (0.262) | (0.218) |
| *Note:* | | | | | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 4. Predicting Price Pass-throughs with the Koijen and Yogo (2019) instrument.** This table presents estimated coefficients from the regression in Equation (16). The main independent variable $\log(M_t/B_t)$, is instrumented using $\log(\widehat{M}_t/B_t)$ where $\widehat{M}_t$ is computed using the Koijen and Yogo (2019) instrument. We follow Koijen and Yogo (2019) to control for several other characteristics that are cross-sectionally transformed to be uniformly distributed between 0 and 1. All regressions control for time fixed effects, and we compute Driscoll-Kraay standard errors with 8 lags, which controls for both time-series and cross-sectional correlations. Horizon $h$ over which the returns are calculated is shown in different columns. Columns (1) to (4) estimate equal-weighted forecasting regressions. Columns (5) to (8) estimate value-weighted regressions where, to account for the fact that total market size went up over time, we standardized the sum of weights in each period to unity.

| | *Dependent variable*: $r_{t+1\to t+h}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal-weighted | | | | Value-weighted | | | |
| | (1) $h=1$ | (2) 2 | (3) 4 | (4) 8 | (5) $h=1$ | (6) 2 | (7) 4 | (8) 8 |
| $\log(M_t/B_t)$ | $-0.131^{***}$ | $-0.224^{***}$ | $-0.352^{***}$ | $-0.537^{***}$ | $-0.018^{**}$ | $-0.033^{**}$ | $-0.062^{**}$ | $-0.114^{**}$ |
| | (0.010) | (0.015) | (0.027) | (0.061) | (0.008) | (0.014) | (0.025) | (0.049) |
| beta | $0.035^{***}$ | $0.059^{***}$ | $0.088^{***}$ | $0.127^{***}$ | 0.000 | −0.001 | −0.004 | −0.013 |
| | (0.008) | (0.015) | (0.027) | (0.047) | (0.004) | (0.007) | (0.012) | (0.021) |
| investment | 0.000 | 0.001 | 0.004 | 0.017 | −0.002 | −0.003 | −0.005 | 0.001 |
| | (0.002) | (0.004) | (0.007) | (0.010) | (0.003) | (0.005) | (0.007) | (0.012) |
| profitability | $0.077^{***}$ | $0.133^{***}$ | $0.218^{***}$ | $0.346^{***}$ | $0.024^{**}$ | $0.043^{***}$ | $0.076^{**}$ | $0.132^{**}$ |
| | (0.007) | (0.013) | (0.024) | (0.052) | (0.009) | (0.016) | (0.031) | (0.063) |
| div/book | 0.007 | 0.009 | 0.003 | −0.026 | $-0.013^{**}$ | $-0.023^{**}$ | $-0.032^{**}$ | $-0.053^{*}$ |
| | (0.005) | (0.008) | (0.014) | (0.023) | (0.005) | (0.009) | (0.016) | (0.032) |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Obs. | 602,377 | 590,955 | 562,106 | 506,480 | 602,377 | 590,955 | 562,106 | 506,480 |
| $R^2$ | 0.011 | 0.021 | 0.031 | 0.041 | 0.001 | 0.011 | 0.011 | 0.021 |
| *Note:* | | | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

**Table 5. Exponential Brandt et al. (2009) Estimation.** In this table, we show the parameters estimates ($b_j$ parameters) associated with the Brandt et al. (2009) style optimizer with portfolio weights that are exponential linear in characteristics, as shown in Equation (17). These parameters are calculated to maximize the Sharpe ratio, as discussed in the paper. Standard errors, calculated via the usual extremum estimator formula, are shown in parentheses. The average elasticity across stocks and time is shown in the last column. The standard error for this average elasticity is calculated via the delta method. The in-sample Sharpe ratio is 1.2, with a standard error of 0.054.

| | Coefficients | | | | | | | Elasticity |
|---|---|---|---|---|---|---|---|---|
| | $\Delta p_t$ | $b_t - p_t$ | $\frac{p_t - \mu_t^p}{\sigma_t^p}$ | profit. | invest. | div.-to-book | beta | $\eta$ |
| Estimates | $-0.535^{***}$ | $-0.349^{***}$ | $-2.878^{***}$ | $1.590^{***}$ | $-0.557^{***}$ | $8.392^{***}$ | $-82.361^{***}$ | $2.590^{***}$ |
| | (0.002) | (0.000) | (0.001) | (0.001) | (0.001) | (0.010) | (0.250) | (0.001) |
| Obs. | 2,217,863 | 2,217,863 | 2,217,863 | 2,217,863 | 2,217,863 | 2,217,863 | 2,217,863 | 2,217,863 |
| *Note:* | | | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

# Appendix

## A Proof of Proposition 1

In this proof, we drop the $t$ subscripts and tildes for notational simplicity. Recall that for positive portfolio weights, $w_i > 0$, we have:

$$\theta_i \equiv \frac{\partial \log(w_i)}{\partial \mu_i} = \frac{\tau_i}{\iota_i' \Sigma^{-1} \mu} \tag{A.1}$$

where $\tau_i$ is the $i^{th}$ term along the diagonal of the precision matrix (inverse of the covariance matrix).

Without loss of generality, just consider the last asset, asset $N$. Subdivide the matrix into blocks

$$\Sigma = \begin{bmatrix} \Sigma_{-N,-N} & \Sigma_{-N} \\ \Sigma_{-N}' & \sigma_N^2 \end{bmatrix} \tag{A.2}$$

Using the block diagonal matrix formula, note that:

$$\Sigma^{-1} = \begin{bmatrix} \cdots & \cdots \\ -\tau_N \Sigma_{-N}' \Sigma_{-N,-N}^{-1} & \tau_N \end{bmatrix} \tag{A.3}$$

where we fill in "..." above because these terms are ultimately not used below. In other words, we care only about this single row of the precision matrix. Then, calculating $\theta_N$ we have:

$$\theta_N = \frac{\tau_N}{-\tau_N \Sigma_{-N}' \Sigma_{-N,-N}^{-1} \mu_{-N} + \tau_N \mu_N} \tag{A.4}$$

$$= \frac{1}{\mu_N - \Sigma_{-N}' \Sigma_{-N,-N}^{-1} \mu_{-N}} \tag{A.5}$$

Recall that $\beta_{N,-N} = \Sigma_{-N}' \Sigma_{-N,-N}^{-1}$, which is the beta of asset $N$ with all other assets as factors. Also recall that the corresponding alpha is defined as:

$$\alpha_{N,-N} \equiv \mu_N - \beta_{N,-N} \mu_{-N} \tag{A.6}$$

Thus

$$\theta_i = \alpha_i^{-1}. \tag{A.7}$$

Notice that this formula goes back to using $i$, since the proof above is without loss of generality.

In words, this formula shows that how much investors trade on expected returns is just the inverse of how big they think the alpha is of that asset compared to all others. If they think the assets has fairly large alpha, then they will be relatively inelastic. If the investors think that alpha will be quite small, the elasticity will be quite large.

# B Supporting Derivations

## B.1 Price pass-through is close to one in classic models: the dynamic case

We now show that, in a dynamic setting, assuming all future prices are unchanged leads to a price pass-through that is approximately one. Recall that one-period expected return is:

$$\begin{aligned}\tilde{\mu}_{i,t} &= \frac{\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]}{P_{i,t}} - R_{f,t} \\ &= \left(\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]\right) \times \exp\left(-\log(P_{i,t})\right) - R_{f,t}\end{aligned} \tag{B.1}$$

Suppose that future expectations of payouts are fixed, which amounts to assuming that the current period price change is only associated with short-term (i.e., the next period) discount rate change. Under these assumptions, we can write the price pass-through as:

$$\begin{aligned}\psi_{i,t} = -\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})} &= \left(\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]\right) \times \exp\left(-\log(P_{i,t})\right) \\ &= \frac{\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]}{P_{i,t}} \\ &= R_{f,t} + \tilde{\mu}_{i,t}\end{aligned} \tag{B.2}$$

which is slightly above one in typical calibrations.

## B.2 Why can price pass-through be less than one?

As shown in section 2.3, classic models assume a one-for-one price pass-through. However, section 3 shows that empirical estimates of price pass-throughs are usually much lower than one.

In this section, we provide three modeling frameworks in which price pass-through is less than one. We do not claim any of them are the "correct" explanation of prices. Our goal is simply to give some concrete reasons why the price pass-through may be less than one from a theoretical perspective.

### B.2.1 Partial equilibrium flow-based framework

In this framework, an investor cannot observe the trading flows of other investors but can observe prices. When an investor observes a price change that is not dividend-related, she correctly infers that it was driven by flows. If the flows constitute a permanent investment that does not reverse the next period, then the investor correctly infers that the price of the asset in the next period will also be higher. From the investor's perspective, a price change is caused by flows, which have implications for the next-period price.

Consider a very simple partial equilibrium framework with only one asset for simplicity. This is similar in style to models such as those found in Gabaix and Koijen (2021) and Chaudhry (2023). Let $\mu_{t,d} = \mathbb{E}_t[D_{t+1}]$ denote the conditional expectation of the dividend. To make sure price movements are unrelated to cash flows, we assume that the expectation of dividends is fixed through time (e.g. $\mu_{t,d} = \mu_{t+1,d}$). The supply of the asset is fixed and normalized to one. Demand is isoelastic in shares demanded (not portfolio weights), denoted as $Q_t = (1 + f_t)a_t^{-1}P_t^{-1/\lambda_t}$, where $a_t$ and $\lambda_t$ are exogenous to prices, and $f_t$ represents a noisy flow that is exogenous to prices and dividends. Importantly, we assume:

$$a_{t+1}^{-1} = (1 + f_t)^{\varphi_t} a_t^{-1}, \tag{B.3}$$

where $\varphi_t \in [0, 1]$ represents how mean-reverting the period $t$ flow is for demand. If $\varphi_t = 0$, then the flow reverses the next period completely and does not constitute a higher level of demand going forward, and if $\varphi_t = 1$ then the flow is a permanent investment that does not reverse. Note that in period $t + 1$, there is a new noisy flow initiated with its own persistence, but we assume that $\mathbb{E}_t[f_{t+1}] = 0$.

Now consider a rational atomistic investor with an information set $I_t = \{P_t, \lambda_t, \lambda_{t+1}, a_t, \varphi_t, \mu_{t,d}\}$. This investor cannot directly observe these noisy flows, but can infer these flows from prices. We consider the pass-through of this investor.

Since flows can be backed out from prices, we can write:

$$\frac{\partial \tilde{V}_t}{\partial \log(P_t)} = \frac{\partial \tilde{V}_t}{\partial \log(1+f_t)} \left( \frac{\partial \log(P_t)}{\partial \log(1+f_t)} \right)^{-1} = \frac{\lambda_{t+1}\varphi_t}{\lambda_t} \mathbb{E}[P_{t+1}|I_t],$$

where the right-hand side equation is shown in Appendix D.3. Thus, the pass-through for this investor is given by:

$$\psi_t = 1 - \frac{\lambda_{t+1}\varphi_t}{\lambda_t}\rho_t \text{ where } \rho_t = \frac{\mathbb{E}[P_{t+1}|I_t]}{\mathbb{E}[P_{t+1}|I_t] + \mathbb{E}[D_{t+1}|I_t]}.$$

The logic is straightforward for this investor: an observed price variation must be caused by a flow, holding fixed $\mu_{t,d}$ and other variables in the information set. This flow either constitutes a permanent investment without a reversal, a perfectly mean-reverting flow, or some combination of the two. While the investor does not directly observe this demand, the investor knows the persistence of these types of shocks, and can thus gauge the price impact in the future. This creates a price pass-through that can easily be less than one.

This model delivers pass-throughs that are bounded above zero, as long as $\lambda_{t+1} = \lambda_t$ and $\rho_t < 1$. Under these conditions, consider a permanent investment so that $\varphi_t = 1$, and then $\psi_t = 1 - \rho_t$. If we use $\rho_t \approx 0.96$ from Cochrane (2008), then the pass-through is about 0.04.

### B.2.2 Private information model

The model presented in this section provides some important intuition about why a market with private information can deliver pass-throughs that are less than one. The intuition is straightforward: when an investor observes a price change, they infer that some of this price change is likely driven by private information they do not possess about the future value of the asset. A price change from the investor's perspective, holding all else fixed, implies changes to the future value of the asset through this private information channel.

This model is based on Hellwig (1980). It's a single-asset model, so we drop the $i$ subscripts. There is a unit mass of investors indexed by $j$, and each investor observes a signal $s_t^j = V_{t+1} + \epsilon_t^j$, where $V_{t+1}$ is the payoff of the asset and $\epsilon_t^j$ is an error term that is iid normal with mean zero and

variance $v_{\epsilon,t}$. The payoff is also normally distributed with mean $\mu_{v,t}$ and variance $v_{v,t}$. The noisy supply is denoted by $Z_t$, which is normally distributed with mean $\mu_{z,t}$ and variance $v_{z,t}$. The $\epsilon$, payoff, and noisy supply are independent of each other.

The information set investor $j$ has is given by $I_t^j = \{P_t, s_t^j, v_{\epsilon,t}, \mu_{v,t}, v_{v,t}, \mu_{z,t}, v_{z,t}\}$. This implies that the investor conditions on all information in the model except their the true payoff $V_{t+1}$, their own $\epsilon_t^j$, other investors' signals or $\epsilon$ terms, and the noisy supply $Z_t$. Each investor has CARA utility with risk aversion parameter $\gamma$ and rational expectations, and thus has demand given by $Q_t^j = (\mathbb{E}[V_{t+1}|I_t^j] - P_t)/(\gamma \mathbb{V}\text{ar}[V_{t+1}|I_t^j])$. Aggregate demand is given by the sum across all investors: $\int Q_t^j dj$. Equilibrium is defined by setting this aggregate demand equal to the noisy supply term $Z_t$. We solve for equilibrium via the usual method as shown in Appendix D.1: we conjecture that the equilibrium price is given by $P_t = k_{0,t} + k_{v,t}V_{t+1} + k_{z,t}Z_t$, and confirm the conjecture by successfully solving for the coefficients $k_{0,t}$, $k_{v,t}$, and $k_{z,t}$. As usual, we focus on this equilibrium. In this model, we have $\tilde{V}_t \equiv \mathbb{E}[V_{t+1}|I_t^j]$.

In equilibrium, investor $j$'s pass-through is given by:

$$\psi_t^j = 1 - \frac{P_t A_t}{P_t A_t + B_t^j}, \tag{B.4}$$

where $A_t$ and $B_t^j$ are scalars given in Appendix D.1. The scalar $A_t$ is guaranteed to be positive, and $B_t^j$ is positive except for extremely negative values of $\epsilon_t^j$. Similarly, we can adjust the parameters such that this mass of investors with very negative $\epsilon_t^j$ terms is arbitrarily small. The point here is not to prove that the pass-through is always less than one, but to provide evidence that basic theoretical models or frameworks often deliver pass-throughs less than one, which occurs here given a reasonable set of parameters.

Why is the pass-through less than one in this private-information model? The intuition is straightforward. When an investor observes prices moving, they correctly infer that some of the price movement is driven by private information they do not possess. Thus, a change in the price, holding all other terms in the information set $I_t^j$ fixed, corresponds to a change in the perceived value of the payout. This generates a pass-through less than one. This force dampens the demand elasticity, since investors are reluctant to trade against the private information of others. However, the investors are not completely inelastic, since the noisy supply does generate some profitable trading. This stands in contrast to Milgrom and Stokey (1982), which has the famous no-trade theorem. This no-trade theorem is generated by a market dominated by private information without

any profitable trading opportunities, and thus investors are reluctant to trade against even large price swings since doing so implies trading against private information.

It is critical to note that the elasticity conditions on the information set of the investor, holding all the terms in the information set fixed other than the price. We specifically want to consider what we term the exogenous information set, which is the information set of the investor without the price or any price-ratio terms. In this case, the exogenous information set is given as $I_t^{j,e} = \{s_t^j, v_{\epsilon,t}, \mu_{v,t}, v_{v,t}, \mu_{z,t}, v_{z,t}\}$. As long as all terms in the exogenous information set of the investor are held fixed, then a change in the price represents only the movement along the demand curve and not a shift in the demand curve, despite the fact that the expected value of the payout also moves when the price moves. If the exogenous information set contains the expected payout of the asset, then holding this fixed corresponds to the price pass-through of one. Thus, the exogenous information set is very model-dependent, just as the elasticity is very dependent on the asset, exogenous information set, and time period.

### B.2.3 Uninformed investors in a private information model

The exogenous information set above includes the private signal $s_t^j$. In the empirical exercise below, we consider a hypothetical mean-variance investor with an information set that contains only public information. We briefly trace out how including a mass of investors without private signals changes the Hellwig (1980) style model from above. We conclude that these uninformed investors do indeed have pass-throughs that are less than one. The intuition is similar to the above model, and shows that even an uninformed investor has pass-throughs less than one in this setting.

The only modification of the model is that there is a mass $n < 1$ of uninformed investors who receive no private signal, and $1-n$ investors who do receive private signals $s_t^j$. Both groups have the same utility function and rational expectations as above. The uninformed investors then condition on $I_t^j = \{P_t, v_{\epsilon,t}, \mu_{v,t}, v_{v,t}, \mu_{z,t}, v_{z,t}\}$, which excludes the private signal.

The price pass-through for an uninformed investor $j$ as:

$$\psi_t^j = 1 - \frac{P_t \bar{A}_t}{P_t \bar{A}_t + \bar{B}_t} < 1,$$

where $\bar{A}_t$ and $\bar{B}_t$ are scalars that are guaranteed to be positive, with their exact forms given in Appendix D.2. Thus, the price pass-through is less than one for this investor.

Consider a hypothetical investor outside the model with no influence on prices, but who wants to enter the market and invest with the highest Sharpe ratio (consistent with CARA demand). If a dataset of prices and payouts are observed and the other variables of the model are functions of state variables or characteristics, then it is possible, given sufficient regularity conditions, to recover the parameters of the model with simple regressions. For example, we may be able to recover the price pass-through of these uninformed investors by just observing the equilibrium prices' ability to predict returns.

This is analogous to the empirical exercise we do below. We consider a hypothetical mean-variance investor who uses equilibrium outcomes to determine how aggressively to trade against price changes. We simply point out here that in a model where the investor has only partial information and some price movements correspond to next-period expected returns, it may be possible, to recover the price pass-through of investors inside the model by simply using equilibrium data without an instrument.

Given this, it is important to stress that we make no claim that this hypothetical investor is recovering the parameters of any investors that are actually trading in the markets. We simply show below that a hypothetical mean-variance investor who uses a data-based approach to trade against price movements optimally has a strikingly low demand elasticity, due to the two key components of our elasticity decomposition. These estimates are conditional on our exogenous information set of standard asset pricing variables, just as the elasticity of the investors in this model are conditional on the investors' exogenous information set. No instrument is needed because we are not estimating demand, but simply considering this data-based hypothetical outside investor.

## B.3 Discussions related to exponential-linear demand

In section 3.5, we find that using exponential-linear rather than MV demand brings demand elasticities down further from approximately 12 to 2. Here, we discuss the relationship between the two demand specifications.

In general, exponential-linear demand has the following functional form:

$$w_{i,t} = \exp(-\beta_{i,t} \log(P_{i,t}) + \upsilon_{i,t}) = e^{\upsilon_{i,t}} P_{i,t}^{-\beta_{i,t}}, \tag{B.5}$$

where $\beta_{i,t}$ and $\upsilon_{i,t}$ are exogenous constants. Taking derivatives with respect to the log price, we

have:

$$-\frac{\partial w_{i,t}}{\partial \log(P_{i,t})} = \beta_{i,t} w_{i,t}. \tag{B.6}$$

To illustrate the connection to the earlier analytical expressions, we note that under the MV model, we can further write demand elasticity as

$$\eta_{i,t} = 1 + \frac{1}{w_{i,t}} \left( -\frac{\partial w_{i,t}}{\partial \log(P_{i,t})} \right), \tag{B.7}$$

and separately decompose these two terms as:

$$\frac{1}{w_{i,t}} = \frac{\gamma A_t \tau_{i,t}^{-1}}{(1-\omega_{i,t})\tilde{\mu}_{i,t}}, \quad \text{and} \quad -\frac{\partial w_{i,t}}{\partial \log(P_{i,t})} = \frac{\psi_{i,t}}{\gamma A_t \tau_{i,t}^{-1}}, \tag{B.8}$$

Combining Equations (B.6) and (B.8), we arrive at the following expressions for unspanned returns and price pass-through when using the exponential-linear demand specification:

$$(1-\omega_{i,t})\tilde{\mu}_{i,t} = w_{i,t}\gamma A_t \tau_{i,t}^{-1} \quad \text{and} \quad \psi_{i,t} = \beta_{i,t} w_{i,t} \gamma A_t \tau_{i,t}^{-1}. \tag{B.9}$$

We arrive at the main point: using the exponential-linear specification amounts to restricting the pass-through and unspanned return to differ only by a factor of $\beta_{i,t}$, which is the log-price coefficient in Equation (B.5). The exponential-linear functional form locks together these two terms in a way that does not occur with the linear demand function in MV. Plugging in the terms in Equation (B.9) into the key decomposition of the paper, Equation (12), we have a simple expression for the demand elasticity:

$$1 + \frac{\psi_{i,t}}{(1-\omega_{i,t})\tilde{\mu}_{i,t}} = 1 + \beta_{i,t}. \tag{B.10}$$

# C Robustness: Other Sources of Demand Elasticity are Small

Our decomposition of demand elasticity only considers the effect of changes in expected returns. One may ask: what about other effects, such as changes in risk or wealth? In robustness checks, we explore those other effects and find them to have minor effects on demand elasticity. We summarize our findings here.

**Wealth effects are small.** Recall that the number of shares demanded for stock $i$ is $Q_{i,t} = \frac{A_t w_{i,t}}{P_{i,t}}$ and we have assumed $A_t$, the investor wealth, to be exogenous. If we do not, then demand elasticity contains an additional wealth effect term:

$$-\frac{\partial \log(Q_{i,t})}{\partial \log(P_{i,t})} = 1 - \frac{\partial \log(w_{i,t})}{\partial \log(P_{i,t})} - \underbrace{\frac{\partial \log(A_t)}{\partial \log(P_{i,t})}}_{\text{wealth effect}}. \tag{C.1}$$

For investors that hold a sizeable portfolio of assets, the wealth effect is small as movements in a single asset's price have minor effects on wealth. It is also helpful to put magnitudes into perspective: theoretical derivations of demand elasticities in classic models are in the thousands. For an investor that holds the stock market, the largest position is typically lower than 5% of the total portfolio, so even for the largest position, the wealth effect term is only 0.05.

**Risk and hedging-related terms are also small.** Even without wealth effects, one may be concerned that the risk of assets will change with prices. To make this concrete, and to allow for additional effects beyond that captured by the CARA model, we follow Campbell et al. (2003) to consider an investor with Epstein-Zin preferences in a multivariate setting. They show, after log-linearization, that optimal portfolio weights are given by:

$$w_t = \frac{1}{\kappa} \underbrace{\Sigma_t^{-1}}_{\text{covariance}} \left[ \underbrace{\tilde{\mu}_t}_{\text{mean}} + \underbrace{\frac{1}{2}\sigma_t^2}_{\text{volatility}} - \underbrace{\frac{\vartheta}{\varsigma}\sigma_{c-w,t}}_{\text{consumption-hedging}} \right], \tag{C.2}$$

where $\tilde{\mu}_t$ is an $N$ dimensional vector of log excess returns, $\Sigma_t$ is the $N \times N$ conditional covariance matrix of returns, $\sigma_t^2$ is the $N$ dimensional vector containing the diagonal elements of $\Sigma_t$, and $\sigma_{c-w,t}$ is the $N$ dimensional vector of the conditional covariance of returns with the log consumption-to-wealth ratio (the *cay* variable from Lettau and Ludvigson, 2001). In terms of preference parameters, $\kappa > 0$ is the relative risk aversion coefficient, $\varsigma > 0$ is the elasticity of intertemporal substitution, and $\vartheta \equiv (1-\kappa)/(1-\varsigma^{-1})$.

Our exercise in Section 3.2 only considers how the mean return $\tilde{\mu}_t$ changes with prices. In principle, all the other terms can change with prices too. Given we have a parametric model of $\Sigma_t$ based on stock characteristics — some of which depend on prices — it is conceptually

straightforward, though tedious, to derive their sensitivity to prices. This takes care of the "covariance" and "variance" terms. As for the "consumption-hedging" term, we empirically project the sample covariance of stock returns and *cay* onto stock characteristics to derive its sensitivity to price. This allows us to numerically compute the contribution of each term to demand elasticity. In the end, we find that all these additional components have tiny effects on the overall demand elasticity.

# D Model Details

## D.1 Homogeneous signal quality model with public information

For simplicity, we drop the $t$ subscripts here. The variable $V$ is the payoff and not the expected payoff. To reiterate the main text, there is a signal $s^j = V + \epsilon^j$. The agent knows his signal quality $\lambda_i$, but $\epsilon^j \sim \mathcal{N}(0, v_\epsilon)$ are iid across investors.

We conjecture that price is linear in fundamental and per-capita noisy supply:

$$P = k_0 + k_v V + k_z Z.$$

Define

$$\begin{bmatrix} V \\ s^j \\ P \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_v \\ \mu_v \\ k_0 + k_v\mu_v + k_z\mu_z \end{bmatrix}, \begin{bmatrix} v_v & v_v & k_v v_v \\ v_v & v_v + v_\epsilon & k_v v_v \\ k_v v_v & k_v v_v & k_v^2 v_v + k_z^2 v_z \end{bmatrix} \right)$$

Thus

$$\begin{aligned}
\mathbb{E}[V \mid s^j, P] &= \mu_v + \begin{bmatrix} v_v & k_v v_v \end{bmatrix} \begin{bmatrix} v_v + v_\epsilon & k_v v_v \\ k_v v_v & k_v^2 v_v + k_z^2 v_z \end{bmatrix}^{-1} \begin{bmatrix} V + \epsilon^j - \mu_v \\ P - k_0 - k_v\mu_v - k_z\mu_z \end{bmatrix} \\
&= \frac{(P - k_0)\, k_v v_v v_\epsilon + k_z^2 v_z \left( (V + \epsilon^j) v_v + \mu_v v_\epsilon \right) - k_z k_v v_v v_\epsilon \mu_z}{k_v^2 v_v v_\epsilon + k_z^2 v_z \left( v_v + v_\epsilon \right)} \\
\mathbb{V}\mathrm{ar}(V \mid s^j, P) &= v_v - \begin{bmatrix} v_v & k_v v_v \end{bmatrix} \begin{bmatrix} v_v + v_\epsilon & k_v v_v \\ k_v v_v & k_v^2 v_v + k_z^2 v_z \end{bmatrix}^{-1} \begin{bmatrix} v_v \\ k_v v_v \end{bmatrix} \\
&= \frac{k_z^2 v_z v_v v_\epsilon}{k_v^2 v_v v_\epsilon + k_z^2 v_z \left( v_v + v_\epsilon \right)}
\end{aligned}$$

So

$$V \mid s^j, P \sim \mathcal{N}\left(\mathbb{E}\left[V \mid s^j, P\right], \mathbb{V}\text{ar}(V \mid s^j, P)\right)$$

The CARA demand is:

$$\begin{aligned} X^j &= \frac{\mathbb{E}[V \mid s^j, P] - P}{\gamma \mathbb{V}\text{ar}(V \mid s^j, P)} \\ &= \frac{P\left(-\frac{(-1+k_v)k_v}{k_z^2 v_z} - \frac{1}{v_v} - \frac{1}{v_\epsilon}\right)}{\gamma} + \frac{V}{\gamma v_\epsilon} + \frac{\epsilon^j}{\gamma v_\epsilon} + \frac{-\frac{k_0 k_v}{k_z^2 v_z} + \frac{\mu_v}{v_v} - \frac{k_v \mu_z}{k_z v_z}}{\gamma}. \end{aligned}$$

We can aggregate demand as

$$\int X^j dj = b_0 + b_p P + b_v V$$

where

$$\begin{aligned} b_0 &= \frac{-\frac{k_0 k_v}{k_z^2 v_z} + \frac{\mu_v}{v_v} - \frac{k_v \mu_z}{k_z v_z}}{\gamma} \\ b_p &= \frac{\left(-\frac{(-1+k_v)k_v}{k_z^2 v_z} - \frac{1}{v_v} - \frac{1}{v_\epsilon}\right)}{\gamma} \\ b_v &= \frac{1}{\gamma v_\epsilon} \end{aligned}$$

To solve the model completely, we must solve the following equations:

$$\begin{aligned} k_0 &= -\frac{b_0}{b_p} \\ k_v &= -\frac{b_v}{b_p} \\ k_z &= \frac{1}{b_p} \end{aligned}$$

Solving this system of equations yields

$$
\begin{aligned}
k_0 &= \frac{\gamma v_\epsilon \left(\mu_v \gamma v_z v_\epsilon + v_v \mu_z\right)}{\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}, \\
k_v &= \frac{v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}, \\
k_z &= -\frac{\gamma v_v v_\epsilon \left(1 + \gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}.
\end{aligned} \tag{D.1}
$$

Plugging $k$'s back into the term for the expected value, we have:

$$
\begin{aligned}
\mathbb{E}[V \mid s^j, P] = &\frac{P \left(v_v + \gamma^2 v_z v_v v_\epsilon + \gamma^2 v_z v_\epsilon^2\right)}{\left(1 + \gamma^2 v_z v_\epsilon\right)\left(\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)\right)} \\
&+ \frac{\left(V + \epsilon^j\right)\left(\gamma^2 v_z v_v v_\epsilon + \gamma^4 v_z^2 v_v v_\epsilon^2\right)}{\left(1 + \gamma^2 v_z v_\epsilon\right)\left(\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)\right)} \\
&+ \frac{\gamma^4 \mu_v v_z^2 v_\epsilon^3 + \gamma^3 v_z v_v v_\epsilon^2 \mu_z}{\left(1 + \gamma^2 v_z v_\epsilon\right)\left(\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)\right)}
\end{aligned}
$$

Thus we have:

$$
\frac{\partial \log(\mathbb{E}[V \mid s^j, P])}{\partial \log(P)} = \frac{1}{\mathbb{E}[V \mid s^j, P]} \frac{\partial \mathbb{E}[V \mid s^j, P]}{\partial \log(P)} = \frac{PA}{PA + B^j},
$$

where

$$
A = \gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)
$$

$$
B^j = s^j \left(\gamma^2 v_z v_v v_\epsilon + \gamma^4 v_z^2 v_v v_\epsilon^2\right) + \gamma^3 v_z v_v v_\epsilon^2 \mu_z + \gamma^4 \mu_v v_z^2 v_\epsilon^3.
$$

## D.2 Uninformed investors in a private information model

For simplicity, we drop the $t$ subscripts here. The variable $V$ is the payoff and not the expected payoff. We follow the same strategy as above, and the demand for the informed investor is given in the above section. Here, we denote this as $X^{j,I}$ instead of just $X^j$, since this is the informed type of investor.

We go through a similar process as above for the mean and variance conditional on just the

price. We calculate

$$\begin{aligned}\mathbb{E}[V|P] &= \mu_v + \left[k_v v_v\right]\left[k_v^2 v_v + k_z^2 v_z\right]^{-1}\left[P - k_0 - k_v\mu_v - k_z\mu_z\right] \\ &= \mu_v + \frac{k_v v_v(P - k_0 - k_v\mu_v - k_z\mu_z)}{k_v^2 v_v + k_z^2 v_z}\end{aligned}$$

$$\begin{aligned}\mathbb{V}\text{ar}[V|P] &= v_v - \left[k_v v_v\right]\left[k_v^2 v_v + k_z^2 v_z\right]^{-1}\left[k_v v_v\right] \\ &= v_v - \frac{(k_v v_v)^2}{k_v^2 v_v + k_z^2 v_z}\end{aligned}$$

Thus the uninformed demand is written as

$$\begin{aligned}X^{j,U} =& \frac{\mathbb{E}[V|P] - P}{\gamma\mathbb{V}\text{ar}[V|P]} \\ =& \frac{(\mu_v - P)k_z^2 v_z - (k_0 + P(-1 + k_v))k_v v_v - k_z k_v v_v \mu_z}{\gamma k_z^2 v_z v_v}\end{aligned}$$

Equilibrium is of course

$$\int X^{j,U} + \int X^{j,I} = Z$$

We can write demand as

$$\int X^{j,U} + \int X^{j,I} = b_0 + b_p P + b_v V$$

where

$$\begin{aligned}b_0 &= \frac{\mu_v k_z^2 v_z - k_0 k_v v_v - k_z k_v v_v \mu_z}{\gamma k_z^2 v_z v_v} \\ b_p &= -\gamma^{-1}\left[\frac{(-1 + k_v)k_v}{k_z^2 v_z} + \frac{1}{v_v} + \frac{1 - n}{v_\epsilon}\right] \\ b_v &= \frac{1 - n}{\gamma v_\epsilon}\end{aligned}$$

To solve the model completely, we must solve the following equations:

$$k_0 = -\frac{b_0}{b_p}$$
$$k_v = -\frac{b_v}{b_p}$$
$$k_z = \frac{1}{b_p}$$

The solution to these equations are:

$$k_0 = \frac{\gamma v_\epsilon \left(\gamma \mu_v v_z v_\epsilon - (-1+n) v_v \mu_z\right)}{\gamma^2 v_z v_\epsilon^2 + (-1+n) v_v \left(-1+n-\gamma^2 v_z v_\epsilon\right)}$$
$$k_v = \frac{(-1+n) v_v \left(-1+n-\gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + (-1+n) v_v \left(-1+n-\gamma^2 v_z v_\epsilon\right)}$$
$$k_z = -\frac{\gamma v_v v_\epsilon \left(1-n+\gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + (-1+n) v_v \left(-1+n-\gamma^2 v_z v_\epsilon\right)}$$

Plugging this back into the the equation for $\mathbb{E}[V|P]$, then we have:

$$\mathbb{E}[V|P] = \frac{\bar{A}P}{\bar{C}} + \frac{\bar{B}}{\bar{C}}$$

where

$$\bar{A} = (1-n)\left(\gamma^2 v_z v_\epsilon^2 + (1-n) v_v \left(1-n+\gamma^2 v_z v_\epsilon\right)\right)$$
$$\bar{B} = \gamma^4 \mu_v v_z^2 v_\epsilon^3 + (1-n)\gamma^3 v_z v_v v_\epsilon^2 \mu_z$$
$$\bar{C} = \left(1-n+\gamma^2 v_z v_\epsilon\right)\left((1-n)^2 v_v + \gamma^2 v_z v_\epsilon^2\right).$$

Thus we have:

$$\frac{\partial \log(\mathbb{E}[V \mid P])}{\partial \log(P)} = \frac{1}{\mathbb{E}[V \mid P]} \frac{\partial \mathbb{E}[V \mid P]}{\partial \log(P)} = \frac{P\bar{A}}{P\bar{A} + \bar{B}},$$

## D.3 Partial equilibrium flow-based framework

There are two components of the expected payoff:

$$\tilde{V}_t = \mathbb{E}[D_{t+1}|I_t] + \mathbb{E}[P_{t+1}|I_t].$$

By assumption, $\mu_{d,t} = \mathbb{E}[D_{t+1}|I_t]$ is unrelated to flows, so we have:

$$\frac{\partial \mathbb{E}[D_{t+1}|I_t]}{\partial \log(1+f_t)} = 0 \tag{D.2}$$

For the other component, the investor considers the impact of flows today on next-period prices based on the equilibrium condition:

$$\begin{aligned} 1 &= (1+f_{t+1})a_{t+1}^{-1}P_{t+1}^{-1/\lambda_{t+1}} \\ &= (1+f_{t+1})(1+f_t)^{\varphi_t}a_t^{-1}P_{t+1}^{-1/\lambda_{t+1}} \end{aligned}$$

By solving for prices, we obtain:

$$P_{t+1} = \left((1+f_{t+1})(1+f_t)^{\varphi_t}a_t^{-1}\right)^{\lambda_{t+1}}$$

Given the assumption that $\mathbb{E}_t[f_{t+1}] = 0$, we have:

$$\mathbb{E}[P_{t+1}|I_t] = \left((1+f_t)^{\varphi_t}a_t^{-1}\right)^{\lambda_{t+1}}$$

Calculating the derivative gives us the expression:

$$\frac{\partial \mathbb{E}[P_{t+1}|I_t]}{\partial \log(1+f_t)} = \lambda_{t+1}\varphi_t \mathbb{E}[P_{t+1}|I_t]$$

Using similar logic, we can also write:

$$P_t = \left((1+f_t)a_t^{-1}\right)^{\lambda_t}.$$

In log terms, this becomes:

$$\log(P_t) = \lambda_t \log(1 + f_t) - \lambda_t \log(a_t)$$

Using this to calcuate the derivative, we have:

$$\frac{\partial \log(P_t)}{\partial \log(1 + f_t)} = \lambda_t$$

Together, this gives us equation in the text:

$$\frac{\partial \tilde{V}_t}{\partial \log(P_t)} = \frac{\partial \tilde{V}_t}{\partial \log(1 + f_t)} \left( \frac{\partial \log(P_t)}{\partial \log(1 + f_t)} \right)^{-1} = \frac{\lambda_{t+1} \varphi_t}{\lambda_t} \mathbb{E}[P_{t+1} | I_t],$$

Thus we can write:

$$\frac{\partial \log(\tilde{V}_t)}{\partial \log(P_t)} = \frac{\lambda_{t+1} \varphi_t}{\lambda_t} \rho_t \text{ where } \rho_t = \frac{\mathbb{E}[P_{t+1} | I_t]}{\mathbb{E}[P_{t+1} | I_t] + \mathbb{E}[D_{t+1} | I_t]}.$$

# E Idiosyncratic Volatility and Elasticity

One may wonder why idiosyncratic volatility does not directly enter the elasticity formula given in Equation (12). This might initially seem surprising, so we discuss this further here.

To explain this, we consider a simple thought experiment. Assume there are three types of assets: assets A, asset B, and other assets, and all assets have the same pass-through. For simplicity assume the pass-throughs are all equal one, i.e., $\psi_A = \psi_B = \psi_{\text{other}} = 1$. Assume there is a single factor for returns so that the covariance matrix is just:

$$\Sigma = \beta \Omega \beta' + \Sigma_\epsilon, \tag{E.1}$$

where $\beta$ is just a vector of single-factor loadings, $\Omega$ is the variance of the factor, and $\Sigma_\epsilon$ is a diagonal matrix of idiosyncratic variance terms. For simplicity, we set the betas equal to 1 for all assets, and the volatility of the factor to 20%. Assume asset A has an idiosyncratic volatility of 100%, asset B has an idiosyncratic volatility of just 1%, and the other assets all have idiosyncratic volatility terms of 20%. For simplicity, we set the risk-free rate to zero.

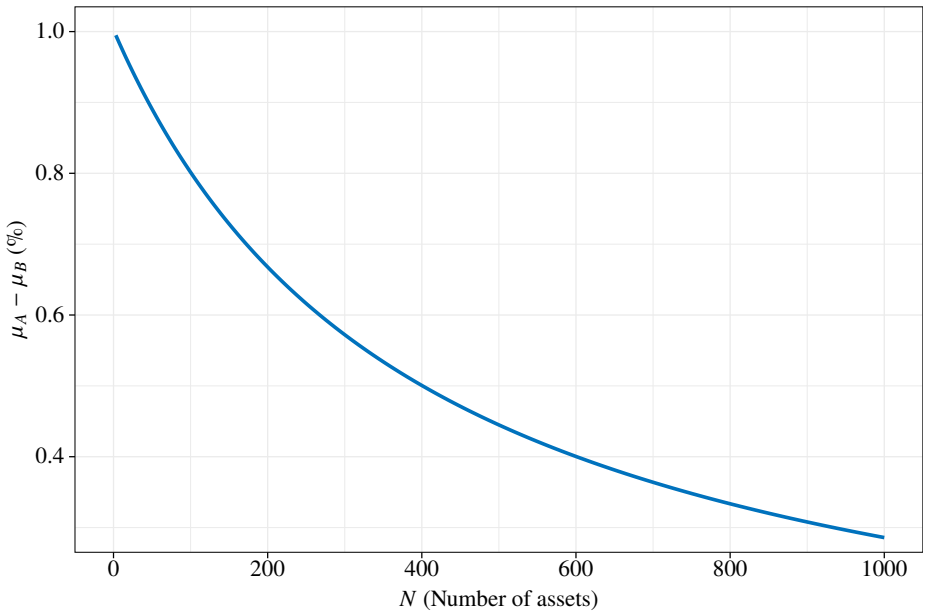



**Figure A.1. Expected Returns Under Idiosyncratic Volatility Thought-Experiment**
This figures show the return of asset A minus the return on asset B, as a function of the number of assets. Note that there is only ever one asset A and one asset B, and changing the number of assets just changes the number of "other assets."

Critically, we assume that the unspanned expected excess returns of all assets are set to 1%. Thus, by Proposition 1, since the unspanned returns and price pass-throughs are equal, the elasticity of the investor would be *identical* for all assets, regardless of the large differences in idiosyncratic volatility. Why is this?

First, we note that assuming identical unspanned expected excess returns across assets and this covariance matrix pins down the expected excess returns. Solving for these expected excess returns, we find that the asset A's excess return must be higher than B's to maintain identical elasticity terms. Therefore, the different idiosyncratic volatility terms do not matter for the elasticity. These terms affect the beta and spanning. In Figure A.1, we plot the return of asset A minus the return of asset B, as a function of the number of assets. Changing the number of assets just changes the number of "other" assets, since there is only one asset A and one asset B.

However, this plot shows that while the return of asset A must be higher than the return of asset B in order to maintain this identical elasticity, this difference becomes smaller as we increase the number of assets. Is it the case that idiosyncratic volatility does not matter as much with increased opportunities to diversify?

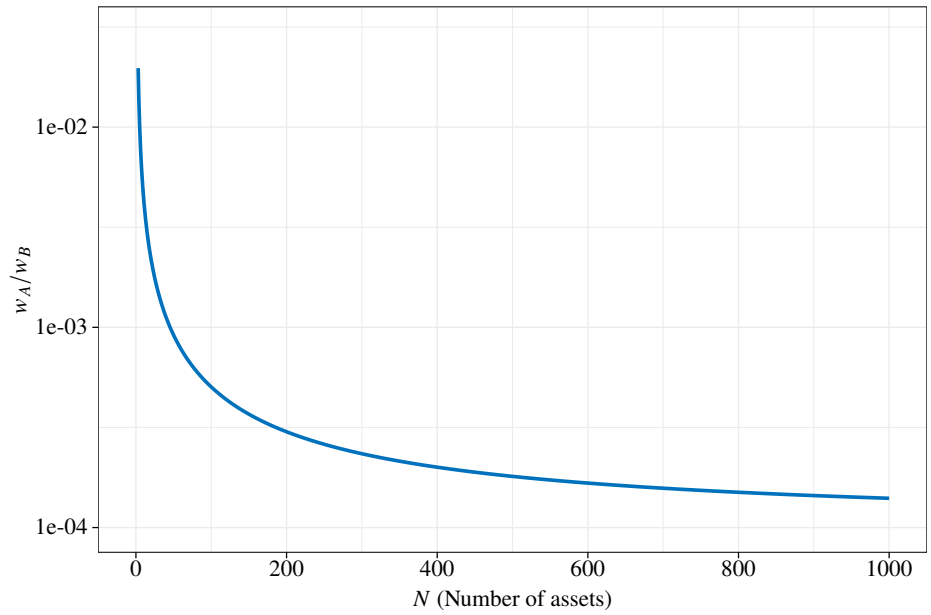



**Figure A.2. Ratio of Portfolio Weights Under Idiosyncratic Volatility Thought-Experiment**
This figures the optimal portfolio weight of asset A divided by the optimal portfolio weight of asset B, as a function of the number of assets. Note that there is only ever one asset A and one asset B, and changing the number of assets just changes the number of "other assets."

It turns out that idiosyncratic volatility still matters, even with many assets. In Figure A.2, we plot the ratio of the optimal portfolio weight for asset A divided by the optimal portfolio weight for asset B, as a function of the number of assets. Because of the higher idiosyncratic volatility of asset A, we should expect asset A to have a smaller portfolio weight to maintain the higher elasticity across assets. This plot shows that this is indeed the case, and in fact this divergence in optimal portfolio weights get *stronger* as the number of assets increases. In other words, the fraction of these portfolio weights gets decreases as the number of assets increases. What is happening here? As the ability to diversify increases, asset A becomes even less appealing, and while the difference in means above becomes smaller, the divergence in portfolio weights must become larger to still account for identical demand elasticity values.

It is important to realize that the risk-aversion drops out of the demand function because elasticity values are in terms of percentage changes, and the idiosyncratic volatility term is similar. A higher risk-aversion makes investors less willing to invest in risky assets, but in terms of percentage changes, it does not affect the elasticity. A 1% change in the price creates smaller demand differences in levels with higher risk aversion values, but does not affect the elasticity because

these are percentage changes. The idiosyncratic volatility term is similar. It obviously matters for demand, but idiosyncratic volatility term is largely a level term, like risk aversion, that drops out when taking logs. However, it should be noted that the idiosyncratic term does affect spanning of other assets, and so the idiosyncratic term does still matter as opposed to the risk-aversion coefficient, which is strictly a level term and really does just drop out of the elasticity.

In summary, idiosyncratic volatility does of course matter. Idiosyncratic volatility largely drops out of the elasticity decomposition in Proposition 1, since idiosyncratic volatility is largely a level term in demand. However, this thought experiment shows that with an assumed covariance structure but identical elasticity values, the expected returns still must be different in order to justify these identical elasticity terms. Also, the weight divergence is also required to maintain identical elasticity terms.

# F Unrestricted Covariance Matrix: Full Results

Table A.1 presents the full results of the exercise in Section 3.3. We consider a large range of shrinkage weights, $h = 0.05, 0.1, ..., 0.95$. The weighted average elasticity is about 9.2 for low levels of covariance shrinkage, and decreases to about 7.5 for $h = 0.95$. The median elasticity has a similar trend. The out-of-sample Sharpe ratios range from just below 1 to below 1.3, depending on the shrinkage weight.

**Table A.1. Optimal Demand Elasticities with Rolling Estimates of Expected Return and Covariance: Full results**
This table is similar to similar to Table 3, but shows results for a larger range of shrinkage weights.

| Shrinkage Weight: | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sharpe ratio | 0.957 | 0.971 | 0.986 | 1.001 | 1.017 | 1.033 | 1.050 | 1.068 | 1.087 | 1.106 |
| Expected Return | 0.015 | 0.015 | 0.015 | 0.015 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 |
| Unspanned Return | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 |
| Median Elasticity | 8.331 | 8.341 | 8.350 | 8.359 | 8.369 | 8.379 | 8.388 | 8.394 | 8.403 | 8.408 |
| Avg. Elasticity (all pos. weights) | 93.580*** | 73.382*** | 84.846*** | 127.410*** | 196.684* | 130.717** | 104.621*** | 83.858*** | 404.489 | 68.124*** |
| | (23.721) | (7.587) | (11.246) | (36.950) | (101.356) | (53.660) | (33.545) | (10.539) | (337.335) | (3.977) |
| Avg. Elasticity (top 99.9%) | 18.792*** | 18.816*** | 18.791*** | 18.782*** | 18.782*** | 18.800*** | 18.792*** | 18.770*** | 18.775*** | 18.740*** |
| | (0.673) | (0.674) | (0.671) | (0.671) | (0.670) | (0.671) | (0.670) | (0.668) | (0.668) | (0.664) |
| Avg. Elasticity (top 99%) | 13.185*** | 13.201*** | 13.209*** | 13.218*** | 13.228*** | 13.235*** | 13.232*** | 13.216*** | 13.193*** | 13.162*** |
| | (0.460) | (0.461) | (0.461) | (0.462) | (0.462) | (0.462) | (0.461) | (0.460) | (0.458) | (0.455) |
| Avg. Elasticity (top 97.5%) | 11.130*** | 11.140*** | 11.152*** | 11.160*** | 11.166*** | 11.166*** | 11.160*** | 11.148*** | 11.131*** | 11.104*** |
| | (0.383) | (0.383) | (0.384) | (0.384) | (0.384) | (0.384) | (0.383) | (0.382) | (0.381) | (0.378) |
| Avg. Elasticity (top 95%) | 9.632*** | 9.642*** | 9.650*** | 9.656*** | 9.659*** | 9.659*** | 9.654*** | 9.645*** | 9.632*** | 9.610*** |
| | (0.326) | (0.327) | (0.327) | (0.327) | (0.327) | (0.327) | (0.327) | (0.326) | (0.325) | (0.323) |
| Weighted Avg. Elasticity | 9.208*** | 9.216*** | 9.218*** | 9.222*** | 9.223*** | 9.220*** | 9.211*** | 9.196*** | 9.176*** | 9.149*** |
| | (0.323) | (0.323) | (0.323) | (0.323) | (0.322) | (0.322) | (0.320) | (0.318) | (0.316) | (0.314) |

| Shrinkage Weight: | 0.55 | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
|---|---|---|---|---|---|---|---|---|---|
| Sharpe ratio | 1.127 | 1.148 | 1.171 | 1.195 | 1.219 | 1.244 | 1.267 | 1.282 | 1.268 |
| Expected Return | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.017 | 0.017 | 0.017 | 0.017 |
| Unspanned Return | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.007 | 0.007 | 0.007 | 0.007 |
| Median Elasticity | 8.407 | 8.400 | 8.388 | 8.367 | 8.328 | 8.258 | 8.147 | 7.949 | 7.567 |
| Avg. Elasticity (all pos. weights) | 116.051*** | 74.290*** | 111.002*** | 174.518** | 68.808*** | 98.231*** | 71.234*** | 71.066*** | 74.132*** |
| | (30.678) | (7.487) | (24.855) | (84.989) | (4.628) | (25.917) | (7.180) | (8.204) | (7.815) |
| Avg. Elasticity (top 99.9%) | 18.649*** | 18.529*** | 18.364*** | 18.183*** | 17.932*** | 17.622*** | 17.131*** | 16.432*** | 15.241*** |
| | (0.657) | (0.649) | (0.637) | (0.626) | (0.610) | (0.591) | (0.562) | (0.526) | (0.468) |
| Avg. Elasticity (top 99%) | 13.109*** | 13.041*** | 12.950*** | 12.832*** | 12.664*** | 12.433*** | 12.121*** | 11.641*** | 10.834*** |
| | (0.451) | (0.446) | (0.439) | (0.432) | (0.421) | (0.407) | (0.388) | (0.362) | (0.324) |
| Avg. Elasticity (top 97.5%) | 11.068*** | 11.018*** | 10.951*** | 10.855*** | 10.718*** | 10.525*** | 10.257*** | 9.855*** | 9.194*** |
| | (0.375) | (0.372) | (0.367) | (0.360) | (0.352) | (0.339) | (0.324) | (0.302) | (0.270) |
| Avg. Elasticity (top 95%) | 9.581*** | 9.540*** | 9.486*** | 9.407*** | 9.295*** | 9.133*** | 8.901*** | 8.558*** | 7.999*** |
| | (0.320) | (0.317) | (0.313) | (0.308) | (0.300) | (0.290) | (0.276) | (0.257) | (0.230) |
| Weighted Avg. Elasticity | 9.108*** | 9.053*** | 8.985*** | 8.895*** | 8.775*** | 8.610*** | 8.384*** | 8.051*** | 7.518*** |
| | (0.310) | (0.306) | (0.301) | (0.294) | (0.286) | (0.275) | (0.262) | (0.244) | (0.218) |

*Note:* *p<0.1; **p<0.05; ***p<0.01