Why is Asset Demand Inelastic?*

Carter Davis† Mahyar Kargar‡ Jiacui Li§

December 28, 2023

Abstract

How effectively does the market trade against fund flows to absorb their price impact? Empirical estimates indicate an inelastic market with high price impacts, in sharp contrast to standard asset pricing models with deep arbitrage capital and investors who trade aggressively against price deviations. To explain this seemingly puzzling evidence, we introduce a decomposition of demand elasticity into two components: "price pass-through", which measures how prices forecast returns, and "unspanned return", which reflects an asset's distinctiveness relative to others. We find empirical evidence for low pass-throughs and high unspanned returns in the data, translating into inelastic demand. Many classic models imply highly elastic demand because they assume high pass-throughs and assets that are close substitutes. We propose two definitions of elasticity, elucidating their alignment with current literature estimates. We rationalize inelastic demand, providing an answer to how investor demand and asset prices are set.

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 inelastic asset markets, prices are highly responsive to flows. In many frictionless asset pricing models, investor demand curves are virtually flat, implying high demand elasticities, or equivalently, low price impacts.[1] Models featuring 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. To generate price impacts that are not entirely absorbed by the rest of the market, these models must include additional assumptions such as frictions, hedging incentives, or biases.[2] In sharp contrast to the high elasticity values from theories, estimated demand curves are surprisingly inelastic.[3] How can we account for the significant disparity between the elasticities in standard models and the empirical estimates? Considering the fundamental role of asset demand in determining market outcomes, it is crucial to understand what influences demand elasticities in financial markets.

In this paper, we show that inelastic demand estimates are not puzzling. We focus on a mean-variance investor, not because of its realism, but to illustrate how a mean-variance investor can exhibit relatively inelastic demand without the need for added frictions or behavioral biases in the model. To elucidate the key drivers of inelastic demand, we present a novel decomposition of demand elasticity into two components: First, the extent to which prices predict returns, which we call "price pass-through", and second, how well an asset is perceived to be spanned by all other assets, which is an asset pricing "unspanned return". Low price pass-throughs and high unspanned returns observed in the data translate into inelastic demand. From our decomposition, we explain

[1]For example, Petajisto (2009) and Gabaix and Koijen (2021) argue that standard asset pricing models have demand elasticities in the range of 5,000 or higher, corresponding to price impacts of only 2 basis point for flows equal to 1% drop in shares supplied.

[2]For example, Kozak, Nagel, and Santosh (2018) use leverage constraints to create inelastic demand paired with sentiment-driven investors to show how sentiment can drive the cross-section of returns. 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 pairing inelastic demand with their news watcher and trend-chasing strategies to deliver price impacts.

[3]Demand elasticities in the stock market are as low as 0.3, implying price impacts of 3.3% for a 1% change in supply (Koijen and Yogo, 2019; Gabaix and Koijen, 2021).

that many theoretical models imply highly elastic demand because they assume high pass-throughs and low perceived unspanned returns. We show that our results extend to CRRA and Epstein-Zin utility demands, but our primary focus is on a mean-variance demand function.

Our decomposition is intuitive. If price pass-throughs are low, i.e., price fluctuations are not strong predictors of expected returns, investors are less motivated to trade against these price movements. Moreover, if an asset is mostly spanned by other assets in investors' consideration set (so the "perceived unspanned return" is low), it has (almost) perfect substitutes, resulting in high demand elasticity. Conversely, a poorly spanned asset, characterized by high unspanned return relative to other assets and a lack of sufficiently close substitutes, will therefore exhibit inelastic demand.

To be clear, we are not providing an upper bound on the elasticity of demand for all investors in all scenarios. As we show, the elasticity of demand is highly conditional for each investor, each asset, and in each time period. It is not a stable structural parameter. Indeed, it's easy to find instances where certain investors exhibit highly elastic demand for particular assets during specific periods. We consider a mean-variance investor with an information set consisting of some exogenous observables and the price, and show that this investor is relatively inelastic to price changes, holding fixed these exogenous observables. This inelasticity comes from the low price pass-through and high unspanned return channels above, and can ultimately provide elasticity values approximately in line with demand estimates. We are not attempting to uncover all sources of inelastic demand, but give a novel decomposition and show how far these basic rational mechanisms can go in delivering inelastic demand.

Figure 1 summarizes our elasticity decomposition, illustrating how we explain the significant gap between the high elasticity in classic models and estimated values in the literature. Standard asset pricing models, which assume a perfect pass-through, imply an elasticity of approximately

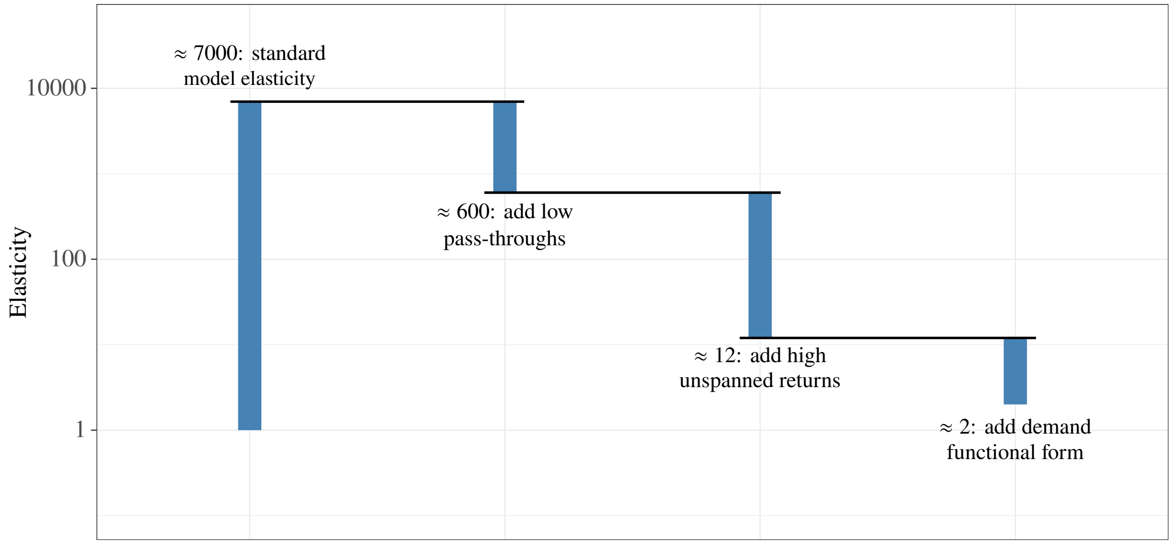

Figure 1. Elasticity decomposition
This figure shows how we explain the gap between high elasticity values in standard asset pricing models and inelastic empirical estimates. The leftmost bar illustrates elasticity estimates for the standard asset pricing models with perfect pass-through as discussed by 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. If we take into account high perceived unspanned returns of 0.7%, we arrive at elasticity estimates of around 12, further reducing elasticity by a factor of 50, as shown by the third bar from left. Finally, incorporating exponential-linear demand as in KY, 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.

7,000, as shown in our calibration below and as discussed by Petajisto (2009) and Gabaix and Koijen (2021). This is illustrated in the leftmost bar. Taking into account an estimated price pass-through of 0.06, instead of 1, reduces elasticity by an order of magnitude to approximately 600. If we take into account high perceived unspanned returns of 0.7%, we arrive at elasticity estimates of around 12, further reducing elasticity by a factor of 50. Finally, as we discuss later, incorporating near-isoelastic demand as in Koijen and Yogo (2019) (KY hereafter) locks the unspanned return and price pass-through together. This demand functional form further reduces the elasticity to about 2, approximately in line with the empirical estimates.

In line with the literature, we define the demand elasticity of an investor as the percent increase in the number of shares they hold when prices decline by 1% (Koijen and Yogo, 2019; Gabaix and Koijen, 2021). When prices change, next-period expected returns may adjust, given that they are functions of prices. We decompose the elasticity into one plus the product of two terms: first, the change in the investor's (log) portfolio weight in stock i, w_i, in response to changes in the investor's subjective expected return, $\tilde{\mu}_i$, and second the change in expected returns in response to movements in the (log) price, P_i:

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

We refer to the first term as (portfolio) "weight responsiveness" to beliefs about expected returns, and label the second term as the "price pass-through" to expected returns. Furthermore, we show that weight responsiveness of stock i is inversely related to how well its expected return is spanned by all other assets, i.e., the perceived "unspanned (expected excess) return" of an asset i with respect to all other assets. The variable ω_i is the fraction of the return that is perceived to be spanned by the rest of the assets. We emphasize that the perceived unspanned return measures the *distinctiveness* of an asset. Thus, the weight responsiveness is the inverse of the asset's distinctness, measuring how *substitutable* it is relative to other assets in the investor's universe.

Our main contribution is to identify low pass-throughs and large unspanned returns as the primary sources of the inelastic demand puzzle in the stock market. Furthermore, we test whether other residual components of demand (e.g., volatility or consumption hedging) would be important components of the elasticity of an optimizing investor, and find that indeed it is the first order effects in our decomposition above that matter for demand elasticity.

In computing demand elasticity, it is crucial to consider the sources of price movements. From the Campbell and Shiller decomposition, price changes are associated with changes in future dividends, changes in future discount rates, or a combination of both. Moreover, price movements due to changes in discount rates can be related to short-term or long-term expected returns or a combination of the two. Given these facts, we then consider two definitions of demand elasticity associated with different sources of price movements. In the first definition of elasticity, which we label D1, we consider price changes only due to the next-period discount rate movements, holding everything else constant. A price drop due only to the next-period discount rates implies a one-for-one pass-through to expected returns. Therefore, D1 essentially assumes the price pass-through in the definition of elasticity in Equation (1) is close to one, and only measures (one plus) weight responsiveness.

The assumption of perfect pass-throughs next period, creates very high-return low-risk investment opportunities if the price impact is large. Therefore, we arrive at our second definition of elasticity, D2, which considers responses to price movements that are *not* entirely driven by the next-period discount rates. Depending on the sign and magnitude of the pass-through term in Equation (1), D2 elasticity can take a wide range: upward-sloping demand if it is negative, perfectly inelastic if it is zero, and D1 if it is equal to one.

As we show later, almost all empirical estimates measure D2 elasticity while most theoretical models only study the response of portfolio weights to changes in the discount rates, i.e., they assume perfect pass-throughs, essentially estimating D1.[4] Directly estimating D1 elasticity in the stock market is challenging: it requires an instrument that both identifies price movements associated with one-for-one price pass-throughs within a single period (e.g., monthly, quarterly,

[4]A notable exception is the model in Gabaix and Koijen (2021) where it considers price movements that do not revert back next period. Consistent with these definitions, Petajisto (2009) measures very high D1 demand elasticity of more than 6,000 in his calibration. This number is almost three orders of magnitude larger than the empirical estimates that measure D2.

etc.) and is within the investor's information set. Consistent with our decomposition, literature estimates of elasticity are high in the bond market where price pass-throughs are high.[5]

We show empirically that D2 elasticity seems to justify the low elasticity estimates from the literature. Furthermore, we provide empirical evidence that the instrument proposed by KY aligns more closely with a D2 elasticity instrument rather than a D1 elasticity instrument. This is not a critique of KY, and if anything we argue that KY represents a demand system that is very useful for counterfactual experiments.

To estimate price pass-throughs and unspanned returns, we first construct a model of portfolio choice where the expected returns and the covariance matrix are expressed as linear functions of stock characteristics. We then proceed to estimate this model using maximum likelihood. We find low pass-throughs and high perceived unspanned returns estimates, leading to elasticity value that are two orders of magnitude smaller than models with perfect pass-throughs. Following KY, we calculate the elasticity only for portfolio positions with strictly positive weights. We further exclude stocks with very low portfolio-weights, as these have negligible unspanned returns and nearly infinite elasticity. Furthermore, we show that a large perceived unspanned return alone does not explain the large elasticity gap: low pass-throughs and large unspanned returns are both important to deliver inelastic demand.

We next study the elasticity of strategies where the covariance matrix is not based on asset characteristics, but rather use a standard covariance matrix shrinkage method to estimate the covariance matrix at the stock-level (Ledoit and Wolf, 2004). We show that high levels of shrinkage are not necessary to achieve inelastic demand. Keeping the pass-through constant, we show that even with low levels of shrinkage, the unspanned returns are high enough to ensure relatively

[5]In Appendix E, we discuss elasticity estimates in the bond market.

inelastic demand. These results indicate that stock returns are not adequately spanned by returns from other stocks.

To further narrow the gap between elasticity values, we consider a model where, following KY and consistent with the evidence from investors' portfolios, demand is exponential-linear. Furthermore, the (nearly) isoelastic demand helps alleviate concerns that the previous results might only apply to positive weights, be sensitive to filters based on the size of portfolio weights, and fail to align with investor holdings data. The in-sample Sharpe ratio is 1.2, which helps alleviate any concerns about expected returns or unspanned returns being too high. This model leads to elasticity estimates roughly in line with existing estimates of micro demand elasticity (Gabaix and Koijen, 2021). We use our decomposition to show why exponential-linear demand theoretically provides low demand elasticity values and potentially fits the holdings data better than a linear model. Therefore, the functional form of the demand function significantly influences the elasticity estimates (Davis, 2023).

Given low pass-throughs, it is important to understand whether pass-throughs are small enough and unspanned returns are large enough to deliver low estimated elasticities comparable to estimates from the literature and whether the residual component of demand is small. To answer both of these questions, we estimate the components of a classic portfolio choice problem with Epstein-Zin preferences from Campbell, Chan, and Viceira (2003). We decompose the residual components of elasticity into a covariance component, a variance component, and a consumption-to-wealth ratio hedging component. We find that these components are relatively small, meaning that the entire residual component is relatively small.

Related literature

Our paper is about stock market micro elasticity which examines the change in the relative price of two stocks if one buys \$1 of one and sells \$1 of the other (e.g., Shleifer, 1986; Harris and Gurel, 1986).[6] There is a range methodologies to estimate demand elasticities at the individual stock level: index exclusion (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). There are also structural approaches 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.[7] So, demand curves are much more inelastic compared to existing theories. In this paper, we provide a microfoundation for inelastic demand based on investor beliefs about discount rates and cash flows.

Our finding that most stock prices movements exhibit small or sometimes even negative price pass-throughs 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

[6]This is in contrast with the literature on macro elasticity that studies how the aggregate stock market's value changes if one buys \$1 worth of stocks by selling \$1 worth of bonds (e.g., Johnson, 2006; Deuskar and Johnson, 2011; Da, Larrain, Sialm, and Tessada, 2018; Gabaix and Koijen, 2021; Li, Pearson, and Zhang, 2020; Hartzmark and Solomon, 2022). A more recent literature studies factor-level multipliers which is the price impact if an investor buys a fraction of the outstanding shares of a cross-sectional pricing factor such as size or value (e.g., Peng and Wang, 2021; Li, 2021). The evidence in the literature suggests that the micro elasticity is much larger than the aggregate elasticity given that different stocks are closer substitutes than the stock and bond market indices (Gabaix and Koijen, 2021).

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

role in demand elasticity. In our framework, weak price pass-throughs leads to a low demand elasticity.

Our estimated low pass-throughs are consistent with the literature that explores discount rate and cash flow variation (e.g., Vuolteenaho, 2002; Cochrane, 2008). High perceived unspanned returns are consistent with the asset pricing literature, which finds that assets returns are poorly spanned by systematic risk factors (e.g., Lopez-Lira and Roussanov, 2023) or a range of factor models (e.g., Baba Yara, Boyer, and Davis, 2021). 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.

2 Basic Decomposition and Calibration of Demand Elasticity

We first present a calibration of demand elasticity, which aims to show that classic models deliver a very high demand elasticity (of about 6,000), as argued by both Petajisto (2009) and Gabaix and Koijen (2021). Our calibration differs from Petajisto (2009), whose calibration requires an equilibrium model of asset pricing. The elasticity of demand just measures how a demand function reacts to off-equilibrium price changes. In other words, demand elasticity is calculated before equilibrium, and measures the slope of the demand function alone. Our calibration simply relies on the demand function.

Consider an investor who holds $Q_{i,t}$ shares of stock i at time t. The elasticity, $\eta_{i,t}$, of the investor 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 share price of the stock. This is a quantity calculated before equilibrium. Then, we can write $Q_{i,t} = A_t w_{i,t}/P_{i,t}$, where A_t is the assets under management or investor's wealth, and $w_{i,t}$ represents the portfolio weights of the investor. Plugging this into Equation (2) and assuming that A_t is exogenous (Koijen and Yogo, 2019), we have:[8]

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

In classic asset pricing models, a change in prices affects the expected return of the asset, but the covariance structure of assets and other inputs to the demand function are assumed to be exogenous. Under this assumption, we can use the chain rule to calculate:

$$\eta_{i,t} = 1 + \frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}} \left(-\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})} \right), \tag{4}$$

where $\tilde{\mu}_{i,t}$ is the investor's subjective beliefs about the expected excess return of the asset.[9] Equation (4) is important and worthy of discussion. The "one" in Equation (4) is included because the elasticity is defined in terms of shares outstanding instead of portfolio weights or dollars demanded to be invested. This can be easily illustrated with 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 passively holding the market since, by design, it means that

[8]We discuss the assumption of exogenous assets under management more below.

[9]We use tildes throughout the paper to denote that the quantities are calculated under the investor's subjective expectations, which may or may not correspond to rational expectations.

prices fluctuations do not require selling or purchasing shares. The term $\frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}}$ describe how responsive the investor's portfolio weights are to their beliefs about its expected returns. We refer to this term as the "weight responsiveness". The final term, $-\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})}$, is the pass-through of prices to expected returns. We refer to this term as the "price pass-through". If prices increase 1% and this increase is expected to decrease next-period excess returns by 0.4% ceteris paribus, then this pass-through would be 0.4. These last two terms are largely the focus of this paper, and we define the following variables to simplify notation:

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

This decomposition is intuitive: the elasticity is one plus the product of the weight responsiveness and the price pass-through.

2.1 Two definitions of demand elasticity

In computing demand elasticity, it is critical to note that we cannot simply use Equation (5), because there are various types of price changes. From Campbell and Shiller (1988), a price change is either associated with future dividends, future discount rates, or some combination of both. Furthermore, price movements due to changes in discount rates can be related to short-term or long-term expected returns, or a combination of both. In other words, price changes associated with discount rate movements may affect various parts of the term structure of equities (Campbell and Viceira, 2005; van Binsbergen and Koijen, 2017). Given these facts, we consider different definitions of the price elasticity of demand, each associated with different types of price movements.

Before providing our first definition of elasticity, it is important to point out that a demand function is defined by holding beliefs constant while observing how it responds to a price change, all else being equal (ceteris paribus). However, within the context of *asset* demand markets, a price change *must* imply changes in beliefs about discount rates or cash flows. It is unreasonable to consider changing the price of the asset without also altering beliefs about discount rates, cash flows, or both. In classic information-based models such as Hellwig (1980), investors deduce from price changes that expected cash flows have shifted. As a result, we examine various definitions of demand elasticity when different beliefs are held constant. However, we never do so when *all* beliefs concerning discount rates and cash flows remain fixed, as a price change without a corresponding change in beliefs would be nonsensical.

We begin by introducing a simple model that enables us to clearly present our two definitions of elasticity. Following Gabaix and Koijen (2021) and consistent with investor holdings data, we assume an investor j's portfolio weight in stock i is exponential-linear:

$$\frac{P_{i,t}Q_{i,t}}{A_t} = e^{\theta_{i,t}\tilde{\mu}_{i,t}+\nu_{i,t}}, \tag{6}$$

where, as before, $Q_{i,t}$ is the number of shares of stock i, $P_{i,t}$ is the stock price, A_t is the AUM or wealth of the investor, and ν_i represents additional demand shocks. As discussed above, we represent the sensitivity of investor's portfolio weight to expected returns by parameter $\theta_{i,t}$. When $\theta_{i,t} > 0$, the investor allocates more to stock i when she believes it has higher expected excess returns. If we linearize Equation (6), we have

$$\Delta q_{i,t} = -\Delta p_{i,t} + \theta_{i,t}\,\Delta\tilde{\mu}_{i,t} + \Delta\nu_{i,t}, \tag{7}$$

where $\Delta q_{i,t}$ and $\Delta p_{i,t}$ are percent changes in the quantity and price of stock i, respectively, $\Delta\tilde{\mu}_{i,t}$ is the percent change in investor's subjective expected excess return of the asset, and $\Delta\nu_{i,t}$ represent the change in other demand shocks. If we linearize $\Delta\tilde{\mu}_{it}$ and plug into (7), we have:

$$\Delta q_{i,t} = -\left[1+\theta_{i,t}(1+\rho)\right]\Delta p_{i,t} + \theta_{i,t}\left(\rho\Delta\tilde{\mathbb{E}}_t[D_{i,t+1}] + \Delta\tilde{\mathbb{E}}_t[P_{i,t+1}]\right) + \Delta\nu_{i,t}, \tag{8}$$

where $\tilde{\mathbb{E}}_t[D_{i,t+1}]$ is the expected next period dividend, $\tilde{\mathbb{E}}_t[P_{i,t+1}]$ is the expected next period price, and ρ is the average dividend-price ratio.[10]

We further assume that changes in the next period expected cash flow and price have the following forms:

$$\Delta\tilde{\mathbb{E}}_t[D_{i,t+1}] = \alpha_d\Delta p_{i,t} + \alpha_\delta\delta_{i,t} + \alpha_s s^j_{i,t}, \tag{9}$$

$$\Delta\tilde{\mathbb{E}}_t[P_{i,t+1}] = \alpha_p\Delta p_{i,t} + \epsilon_{i,t}, \tag{10}$$

where $\delta_{i,t}$ and $s^j_{i,t}$ denote public and investor j's private signals about future cash flows, respectively, and $\epsilon_{i,t}$ represents shocks to future price expectations.[11] Finally, plugging in (9) and (10) into (8), we obtain:

$$\begin{aligned}\Delta q_{i,t} = &-\left[1+\theta_{i,t}(1+\rho(1-\alpha_d)-\alpha_p)\right]\Delta p_{i,t} \\ &+\rho\theta_{i,t}\left(\alpha_\delta\delta_{i,t} + \alpha_s s^j_{i,t}\right) + \theta_{i,t}\epsilon_{i,t} + \Delta\nu_{i,t}.\end{aligned} \tag{11}$$

The expression within square brackets in the first line of Equation (11) represents the price elasticity of demand. The second line of Equation (11) denotes demand shifters that are unrelated to price changes.

In this model, as in standard asset pricing models, the investor has an information set. This

[10] See Appendix C of Chaudhry (2023) for the derivation.

[11] Note that the subjective expectations and private signals are specific to each investor, and should include the subscript j. For simplicity, this is omitted in our notation.

information set has two subsets: an exogenous information set that is not a function of prices, and the price. The exogenous information set includes exogenous observables.

The first definition of the elasticity is from movements in demand associated *only* with short term discount rates changes:

Definition 1 (D1). *This is the elasticity $\eta_{i,t}$ in Equation (2) ceteris paribus. In particular, we consider a price movement such that future expectations of cash flows and prices do not change, i.e.,*

$$\frac{\partial \Delta \tilde{\mathbb{E}}_t[D_{i,t+1}]}{\partial \Delta p_{i,t}} = 0, \qquad \text{and} \qquad \frac{\partial \Delta \tilde{\mathbb{E}}_t[P_{i,t+1}]}{\partial \Delta p_{i,t}} = 0,$$

as well as holding fixed a set of exogenous information set of the investor (including observables).

From Equation (11), D1 assumes $\alpha_d = \alpha_p = 0$, and thus the demand elasticity is

$$\eta_{i,t} = 1 + \theta_{i,t}(1+\rho) \approx 1 + \theta_{i,t}.$$

In other words, under D1, the price pass-through is approximately one, given that the average dividend yield in the data is $\rho \approx 0.04$.

In this definition of elasticity, future expectations of payouts are fixed, and the risk is assumed to be constant as well. Definition 1 corresponds to the calibration in Petajisto (2009) discussed above, where future payments and the variance-covariance structures are assumed to be exogenous. Fundamentally, in the definition of demand elasticity in D1, we consider a price change associated with only a short-term (i.e., the next period) discount rate change.

We can write the investor's subjective expected return as:

$$\begin{aligned}\tilde{\mu}_{i,t} &= \frac{\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]}{P_{i,t}} - R_{f,t} \\ &= \left(\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]\right) \times \exp\left(-\log(P_{i,t})\right) - R_{f,t} \end{aligned} \tag{12}$$

Under D1, we can write the price pass-through as:

$$\begin{aligned}\psi_{i,t} = -\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})} &= \left(\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]\right) \times \exp\left(-\log(P_{i,t})\right) \\ &= \frac{\tilde{\mathbb{E}}_t[P_{i,t+1}] + \tilde{\mathbb{E}}_t[D_{i,t+1}]}{P_{i,t}} \\ &= R_{f,t} + \tilde{\mu}_{i,t}. \end{aligned} \tag{13}$$

This means that the demand elasticity is:

$$\eta_{i,t} = 1 + \theta_{i,t}\psi_{i,t} = 1 + \frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}} \left(R_{f,t} + \tilde{\mu}_{i,t}\right) = 1 + \frac{1}{w_{i,t}} \frac{\partial w_{i,t}}{\partial \tilde{\mu}_{i,t}} \left(R_{f,t} + \tilde{\mu}_{i,t}\right) \tag{14}$$

Thus under Definition 1, the price pass-through is close to one (slightly larger than one assuming a positive expected return on the asset). Thus if we use a price pass-through of $\psi_{i,t} \approx 1$, the elasticity will be $\eta_{i,t} \approx 1 + \theta_{i,t}$.

Now we turn to consider a standard CARA utility model, with multivariate normally distributed returns, which helps us pin down the weight responsiveness $\theta_{i,t}$. Although the CARA utility model possesses some less desirable characteristics, we explore a broader Epstein-Zin utility in Section 5. Our findings suggest that a CARA utility model is adequately capable of capturing the primary factors influencing elasticity. While more general utility functions can capture wealth effects and dynamic hedging demand, neither of these factors are primary drivers of the microeconomic demand elasticity of individual assets.

Consider an investor with CARA utility who maximizes:

$$\tilde{\mathbb{E}}_t \left[- \exp\{-\gamma A_t \left(w_t' r_{t+1} + R_{f,t} \right) \} \right], \tag{15}$$

where γ is the absolute risk aversion parameter, w_t is an N dimensional vector of portfolio weights, r_t is an N dimensional vector of excess returns, and ι is an N dimensional vector of ones. The FOC is:

$$w_t = \frac{1}{\gamma A_t} \tilde{\Sigma}_t^{-1} \tilde{\mu}_t, \tag{16}$$

where $\tilde{\Sigma}_t$ is the subjective beliefs about the covariance matrix. Thus we can write:

$$\frac{\partial w_{i,t}}{\partial \tilde{\mu}_{i,t}} = \frac{\tilde{\tau}_{i,t}}{\gamma A_t}, \tag{17}$$

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

We consider a calibration of $\theta_{i,t}$, and thus the elasticity. To simplify the calibration, we assume all N assets have the same expected returns, standard deviations, and correlations. We set the expected excess return at 0.06, the average correlation at 0.3, and the volatility at 0.3.[12] We consider $N = 1{,}000$ assets. We set the CARA risk aversion coefficient times wealth, γA_t, to be 2.2 because this allows portfolio weights to sum to one, implying a zero-net supply (demand) of the risk-free asset. Note that because $\theta_{i,t}$ is in terms of log changes, the value of γA_t is irrelevant. We set the risk-free rate to zero, a conservative choice since a larger risk-free rate would result in a higher elasticity. This yields $\tilde{\tau}_i \approx 7.2$. To calculate the demand elasticity, we

[12] Pollet and Wilson (2010) report average daily correlations to be 0.237, and longer-horizon correlations are higher due to autocorrelations across days. Thus, we use 0.3 as a reasonable parameter value.

use the average portfolio weight of $1/N$.[13] Inserting these values into Equation (14) results in an elasticity of approximately 7,000. 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 inclusion of more assets. This is due to the increased substitutability created by a greater number of assets, which in turn leads to higher elasticity values.

2.1.1 Discussion of price pass-through and an alternative definition of demand elasticity

Estimating D1 in the stock market is challenging, because it is difficult to obtain an instrument that predicts price movements with one-for-one pass-throughs that is in a standard investor's information set. We would expect investors to use such instruments to aggressively trade against such price movements, potentially eroding the ability of the instrument to generate one-for-one price pass-throughs in the first place. Thus there are good economic reason to indicate that D1 may be difficult, if not impossible, to estimate. Not only is it difficult to estimate, but in most counterfactual experiments in which a demand system is useful, a one-to-one pass-through is likely not a realistic belief that investors have. Thus we argue that the following definition of elasticity is both more easily estimated and useful for most counterfactual experiments.

Definition 2 (D2). *This is the elasticity $\eta_{i,t}$ in Equation (2), holding the investor's exogenous information set including a set of exogenous observables (observed by the investor)* ***other than price*** *fixed.*

Definition D2 assumes that the demand shifters in the second line of Equation (11) remain

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

constant. This implies that the demand elasticity can be expressed as:

$$\eta_{i,t} = 1 + \theta_{i,t}(1 + \rho(1 - \alpha_d) - \alpha_p).$$

Then from Equation (4), the price pass-through $\psi_{i,t}$ is represented by the expression in the parentheses. It should also be noted that D1 is just a special case of D2, where the price pas-through is approximately equal to one.

Calibrating demand elasticity according to Definition 2 is more complicated and requires a calibration of the subjective expectation of the price pass-through. In Section 4, we estimate the price pass-through in a standard model where, consistent with the extensive literature in asset pricing, expected returns and factor loadings are assumed to be linear functions of observable stock characteristics. We find empirically much less than one-to-one price pass-throughs.

We note that Definition 2 is defined conditional on the information set, making it an inherently conditional object. As shown in Proposition 1 below, the elasticity of demand for an asset likely varies among investors, across different assets, and changes over time and with market conditions. Therefore, the elasticity of demand is not a stable structural parameter.

Price pass-throughs of less than one are economically intuitive, and we offer two motivations for this. First, Gabaix and Koijen (2021) consider a model where there are investment flows with both permanent and mean-reverting components. An investor may observe a price shift and rationally infer that this it predominantly results from permanent investment flows. As these flows are largely permanent, the price impact is also primarily permanent, constituting a less than one-for-one pass-through.[14] The underlying concept is that the exogenous information set is held fixed, but investors rationally deduce that a shift in prices does not translate one-for-one into expected returns.

[14]The price pass-through from a permanent flow is approximately the average dividend-price ratio. To see this, from the model in Section 2.1, we can approximate the expected return as $\tilde{\mu}_{i,t} \approx -(1+\rho)\Delta p_{i,t} + \rho\Delta\tilde{\mathbb{E}}[D_{i,t+1}] + \Delta\tilde{\mathbb{E}}[P_{i,t+1}]$. Thus, with a permanent price impact, the pass-through is approximately equal to $\rho < 1$.

Our second motivation for less than one-for-one pass-throughs comes from a heterogeneous private-information model, detailed in Appendix A. In this model, there are well-informed and uninformed investors, along with noise traders. If the noise traders make up only a small segment of the market and the well-informed investors have precise signals, then uninformed investors rationally perceive that price changes are driven by private information and do no translate into higher expected returns. This concept echoes the Milgrom and Stokey's no-trade theorem, which shows that if the market is dominated by private information, uninformed investors rationally do not trade against price shifts. Similarly, the exogenous information set of the investor does not change, but a price shift translates into much less than one-for-one pass-throughs to expected returns.

Critically, we show below that both the pass-through and the unspanned returns are key components of the elasticity of demand for assets. Various price instruments within demand-system models are proposed in the literature (e.g., Koijen and Yogo, 2019; van der Beck, 2022). Our results show that different instruments lead to different pass-through estimates and, therefore, different estimates of demand elasticity. In essence, our results both decompose demand elasticity into understandable components, and also bridge the gap between different demand-system models.

Before delving into this empirical analysis, in Section 2.2, we first decompose the weight responsiveness, $\theta_{i,t}$, into a term that that is easily measurable and understandable.

2.2 Decomposition of weight responsiveness

In this section, we derive expressions for weight responsiveness and elasticity, and then discuss the intuition and economic insights. First we provide some definitions. Define $\tilde{\beta}_{-i,t} = \tilde{\Sigma}'_{-i,t}\tilde{\Sigma}^{-1}_{-i,-i,t}$ be $(N-1)$ row vector of betas of asset i on all other assets in the set, 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 subjective expected excess returns excluding the i^{th} asset. Then we can define the following scalar:

$$\tilde{\alpha}_{-i,t} \equiv \tilde{\mu}_{i,t} - \beta_{-i,t}\tilde{\mu}_{-i,t}, \tag{18}$$

which is a measure of how well the asset is spanned by the rest of the assets in the portfolio. Proposition 1 shows that the weight responsiveness for asset i is equal to the inverse $\tilde{\alpha}_{-i,t}$.[15] To avoid confusion with the standard asset pricing alpha, we write $\tilde{\alpha}_{-i,t}$ as the unspanned return:

$$\tilde{\alpha}_{-i,t} = (1 - \omega_{i,t})\tilde{\mu}_{i,t} \quad \text{where} \quad \omega_{i,t} \equiv \frac{\beta_{-i,t}\tilde{\mu}_{-i,t}}{\tilde{\mu}_{i,t}}. \tag{19}$$

$\tilde{\alpha}_{-i,t}$ is the fraction of the asset's return that is perceived to be unspanned by the rest of the assets and $\omega_{i,t}$ is the fraction of the return that is spanned by other assets.

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* $w_{i,t} > 0$*, we have:*

$$\theta_{i,t} = \frac{1}{(1 - \omega_{i,t})\tilde{\mu}_{i,t}}. \tag{20}$$

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

$$\eta_{i,t} = 1 + \frac{\psi_{i,t}}{(1 - \omega_{i,t})\tilde{\mu}_{i,t}}. \tag{21}$$

We can write the demand elasticity as:

$$\eta_{i,t} = 1 + \frac{1}{w_{i,t}}\left(-\frac{\partial w_{i,t}}{\partial \log(P_{i,t})}\right), \tag{22}$$

[15] Although $\tilde{\alpha}_{-i,t}$ may appear to be a standard asset pricing alpha, it is not. There are many right-hand side factors (all $N - 1$ other assets), and the right-hand side factors naturally change for every asset.

and separately decompose these two terms as:

$$\frac{1}{w_{i,t}} = \frac{\gamma A_t \tau_{i,t}^{-1}}{(1-\omega_{i,t})\tilde{\mu}_{i,t}}, \quad \textit{and} \quad -\frac{\partial w_{i,t}}{\partial \log(P_{i,t})} = \frac{\psi_{i,t}}{\gamma A_t \tau_{i,t}^{-1}}, \tag{23}$$

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

Proof. See Appendix B. □

Equation (21) follows immediately from Equations (5) and (20). In words, Equation (21) shows that the extent to which investors trade based on expected returns is inversely proportional to their perception of the asset's unspanned return compared to others. If investors believe the asset has a large unspanned return, they will be relatively inelastic. Conversely, if they anticipate a relatively small unspanned return, they will be more elastic.

This result may seem counter-intuitive. For example, Haddad et al. (2022) consider the elasticity of demand to be a measure of how aggressive investors are at trading in the market. However, a large perceived unspanned return compared to other assets *decreases* the elasticity of demand. But why is that the case?

An intuitive way to understand $(1-\omega_{i,t})\tilde{\mu}_{i,t}$ is as the perceived uniqueness, or how difficult it is to substitute or replicate asset i. When $\omega_{i,t} = 1$, the asset is essentially a duplicate, completely spanned by all other assets. If $(1-\omega_{i,t})\tilde{\mu}_{i,t}$ is positive and large, it indicates that there are no close substitutes for this asset among the other assets (or any combination of them). The elasticity of an asset or good is high if there is a good substitute available. Conversely, when $(1-\omega_{i,t})\tilde{\mu}_{i,t}$ is near zero, it means there is almost a perfect substitute available among the other assets, resulting in nearly infinite elasticity.

In our calibration of the CARA model above, $\theta_{i,t} \approx 7,000$, indicating that $(1-\omega_{i,t})\tilde{\mu}_{i,t} \approx 0.00014$. In other words, the unspanned return is merely 1.4 basis point. Given that the asset is perceived as having near perfect substitutes, it exhibits a high elasticity. Note that in Petajisto's calibration, price pass-throughs are one-for-one and the unspanned return of individual assets is quite low. As a result, the demand in his calibration is highly elastic.

The decomposition in Equation (21) reveals that demand becomes inelastic when price pass-throughs are low and perceived unspanned returns are high. However, it also shows that when individual asset unspanned return is very low and price pass-throughs are very high, then elasticity values are quite large. In other words, in these settings, the price impacts from flows should be essentially zero.

One may wonder why idiosyncratic volatility does not directly enter the elasticity expression in Equation (21). Idiosyncratic volatility largely drops out of the elasticity decomposition in Proposition 1, as it is primarily a level term in demand. We delve into this in greater detail in Appendix D.

3 Data

We use the standard CRSP-Compustat merged dataset for returns and asset characteristics. We follow KY in calculating profitability, book-to-market ratios, investment, dividend-to-book ratios, and beta. We acquire data at both monthly and daily frequencies. 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 price instrument in KY.

4 The Data Support Inelastic Demand

4.1 Characteristics-based portfolios

In this section, we estimate a structural model of asset returns. The primary focus is on determining whether the model can account for both low pass-throughs and high asset uniqueness, thereby justifying inelastic demand. 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.

This model follows the KY functional form for the mean and covariance matrix of returns:

$$\tilde{\mu}_t = Z_t^{\mu}\pi, \quad \tilde{\Sigma}_t = \Gamma_t \Gamma_t' + \zeta I, \quad \text{and,} \quad \Gamma_t = Z_t^{\Gamma}\xi, \tag{24}$$

where Z_t^{μ} and Z_t^{Γ} 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.[16] π and ξ 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 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^{μ} and Z_t^{Γ} respectively. Then we can calculate the price

[16] We follow Kelly, Pruitt, and Su (2019) and many others in modeling expected returns and factor loadings as linear functions of the observable asset characteristics.

pass-through as:

$$\psi_{i,t} = -\frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})} = -\sum_j \frac{\partial Z^{\mu}_{i,j,t}}{\partial \log(P_{i,t})} \pi_j. \tag{25}$$

For the matrix of data that controls the mean, Z^{μ}_t, we use a similar set of covariates as in KY. As covariates, we use the most recent log price change, $\Delta p_{i,t} = p_{i,t} - p_{i,t-1}$, the log book-to-market ratio $b_{i,t} - p_{i,t}$, the log of market capitalization, cross-sectionally normalized every period, $(p_{i,t} - \bar{\mu}^p_t)/\bar{\sigma}^p_t$, as well as other regressors that are not functions of prices, including profitability, investment, dividend-to-book ratio, and market beta. We follow KY by assuming these latter variables, as well as shares outstanding, are exogenous to prices. These variables are calculated the same as in KY, with the only additions being the reversals variable $\Delta p_{i,t}$ and size variable $(p_{i,t} - \bar{\mu}^p_t)/\bar{\sigma}^p_t$. We also add a columns of ones to Z^{μ}_t and Z^{Γ}_t as an intercept term. Let π_{rev}, π_{bp}, and π_{size} denote the elements of π corresponding the reversals variable, book to market ratio, and size variable respectively. The pass-through is then:

$$\psi_{i,t} = -\pi_{rev} + \pi_{bp} - \frac{\pi_{size}}{\sigma^p_t}. \tag{26}$$

It is worth noting that this price pass-through is distinct from reversals, commonly discussed in asset pricing. Reversal is essentially just the coefficient on the reversal term π_{rev}.

It is important to note that we do not need an instrument to estimate the price pass-through. KY outline two reasons why an instrument is required to estimate demand: the price impact of an investor and the existence of demand shocks that are correlated across investors. In both cases, these two sources of endogeneity, if not addressed, could lead to biased estimates of demand elasticity. Crucially, we are not estimating demand here. The hypothetical mean-variance strategy clearly has no market impact and no demand shocks. We are simply interested in measuring the mean-variance optimal weights and the sensitivity of these weights to price changes, while holding fixed our set of exogenous control variables. These variables constitute the exogenous component of the hypothetical investor's information set.

In the matrix that controls the covariance terms, we only incorporate exogenous variables. Including price terms makes little difference, so for simplicity, we consider an exogenous matrix here. We then explicitly model the covariance matrix as a function of prices and discover that the mean terms are indeed first order. Consequently, we omit the reversal and book-to-market ratio variables from Z_t^{Γ} and substitute the size variable with $(b_{i,t} - \bar{\mu}_t^b)/\bar{\sigma}_t^b$, which represents the log book equity, normalized in the cross-section every period.

We estimate the parameters, π, ξ, and ζ using maximum likelihood, assuming that returns are multivariate normally distributed. The parameter estimates are presented in Table 1. Using these parameters, 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 elasticity for each asset under optimal demand, i.e., $w_t = \frac{1}{\gamma A_t}\tilde{\Sigma}_t^{-1}\tilde{\mu}_t$.[17]

We follow KY and calculate the elasticity only for portfolio positions with strictly positive weights, since these are the only elasticity values that are well-defined. For consistency, all terms reported in the table are those with strictly positive portfolio weights. We show the average of these terms in Table 2 across stocks and months, where the standard errors are double clustered by stock and month. Note that in many cases, $(1 - \omega_{i,t})\tilde{\mu}_{i,t}$ is very close to zero, and thus the elasticity is almost infinite. Thus, we further condition on averages of stocks with the largest portfolio weights, dropping those with very low portfolio-weights. Recall that low portfolio weights is equivalent to low values of unspanned returns. For example, a filter of 99.9% in the third row indicates that we take the subsample of stocks that combined account for 99.9% of the value invested (in terms of portfolio weights), dropping the smallest stocks first. We consider a range of these filters, shown in the second column.

[17] As mentioned earlier, the value of γA_t is irrelevant for elasticity calculations, so we do not need to specify a particular value for γA_t.

We also present the median elasticity in the first row of Table 2, along with the portfolio-weighted average of the elasticity values in the last row, which eliminates the divide-by-nearly-zero issue. The standard error for the median value is 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. Concerns about conditioning on positive portfolios weights, or conditioning further on these subsamples, are addressed later in Section 4.3, where we discuss isoleastic demand.

It should be noted that extremely large elasticity values may be reasonable in cases where portfolio weights are very close to zero (which occurs when the unspanned return is close to zero). This is due to the fact that an increase in the portfolio weight could represent a very large percentage change, and elasticity values are of course in percentage changes.

The median elasticity value is approximately 12, while the weighted average elasticity is around 10. The average elasticity for stocks with a positive portfolio weight is about 80. However, we emphasize that this number is sensitive to stocks with very small portfolio weights. Specifically, there are stocks with unspanned returns that are essentially zero or extremely close to it, which are included because we consider all stocks with a positive portfolio weight. Once we apply the 99.9% value filter, the average elasticity drops to approximately 19, as shown in the third row of Table 2. As we progressively take more restrictive subsamples, this figure drops to around 10, aligning with the weighted average elasticity. While this is still higher than the typical estimates in the literature, which are generally below 5 (see Gabaix and Koijen, 2021), it does substantially narrow the gap between 7,000 and the low estimates from the literature, as shown in Figure 1.

It is important to note that this model does not include trading costs, short-selling constraints, information asymmetry, behavioral biases, or other frictions that could further decrease elasticity.

Our objective is not to close the entire gap between an elasticity calibrations of 7,000 and estimates found in the literature. Instead, we aim to present a novel decomposition of the elasticity into its key economic components, and show that a significant portion of this gap can be explained by the first-order price pass-through and unspanned return terms.

The price pass-through for this model is about 0.06, significantly lower than a one-for-one pass-through. So, a 1% increase in prices only predicts a 6 basis point drop in expected returns the next month. According to Equation (21), this low pass-through yields a considerably lower demand elasticity. Intuitively, trading against price changes becomes less profitable when these fluctuations are weak predictors of future returns. Ultimately, it is the risk and returns that matter most to investors.

However, the combination of the pass-through of 0.06 and unspanned return of 0.0001 from the earlier calibration still results in a high elasticity of about 600 (= $1 + 0.06/0.0001$). This implies that the pass-through term alone is unable to bring down elasticity to the median of 12 from Table 2. The average unspanned return is 0.007, or 70 basis points, from this estimation. It is important to note that these estimates only consider terms with positive portfolio weights, which effectively means we are only including stock-by-month observations with a strictly positive unspanned return. Consequently, an asset with a pass-through and unspanned return close to the average would have an elasticity of around 10 (= $1 + 0.062/0.007$), much closer to the median elasticity.

Why is the unspanned return so large? To put it another way, the average expected excess return on these assets is about 1.3% per month, yet only 0.5% of this return, on average, is accounted for by the portion spanned by other assets. The in-sample Sharpe ratio for this model is 1.45, with a standard error of 0.059. While this value is relatively high, for an in-sample exercise that maximizes the Sharpe ratio, it is quite moderate compared to modern asset pricing factor models (Gu, Kelly, and Xiu, 2021). Thus, the model obtains inelastic demand alongside relatively moderate returns.

Davis (2023) calculates the elasticity of 13 portfolio choice models from the literature, and finds a similar elasticity values. All models deliver strikingly inelastic demand, particularly when compared to the literature's estimates of 7,000. While Davis (2023) does not separately decompose the pass-through and unspanned return channel, it is clear that a wide range of empirical methods deliver inelastic demand.

A large perceived unspanned return alone does not fully explain the large elasticity gap either. With a price pass-through of one but an unspanned return of 0.7%, the elasticity is still about 150. Therefore, both components play a crucial role.

In Table 2, both the expected and unspanned returns are relatively high. We consider another exercise below, which also delivers high expected and unspanned returns for the assets where we can measure the elasticity—namely, the stocks with positive portfolio weights. It is important to note that these estimates are calculated based on the assets with positive unspanned returns, equivalently those with positive portfolio weights, which tends to select on assets with high expected returns. The average unspanned return across all assets is close to zero, but the spread of unspanned returns is high enough to induce inelastic demand for the portfolio positions we can calculate the elasticity for, i.e., the long positions. One might worry that a demand function that only allows long positions may produce very elastic demand. However, as we show later, the exponential-leaner demand function (as in KY) which mandates long-only positions, generates even more inelastic demand.

4.2 Shrinkage portfolios

The results in Section 4.1 focus on characteristic-based portfolios. In this section, we study the elasticity of strategies where the covariance matrix is not based on asset characteristics, but rather use a standard stock-level method to estimate the covariance matrix. While the previous exercise is completely in-sample, this one is out-of-sample.

In this exercise, we follow to the methodology in Lopez-Lira and Roussanov (2023) wherever possible. Similarly to them, we estimate the stock-level covariance matrix on a rolling one-year basis using daily data. 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. We use a standard covariance matrix shrinkage method, similar to Ledoit and Wolf (2004). In particular, we use $\Sigma = (1 - h)\hat{\Sigma} + h\overline{\Sigma}$, where $\hat{\Sigma}$ is the ill-conditioned empirical estimate of the covariance matrix, $\overline{\Sigma}$ is the shrinkage target, and h is the scalar shrinkage weight. For the shrinkage target, we use a simple matrix with the average variance of $\hat{\Sigma}$ along the diagonal and the average covariance on the off-diagonal.

Like Lopez-Lira and Roussanov (2023), we consider assets expected excess returns that are a function of characteristics based on the same rolling year, using monthly data. Note that while we can avoid using characteristics for the covariance matrix, we still need to use characteristics to model the expected excess returns. This is because we need to know how a 1% price change predicts changes in expected returns to calculate the elasticity of the strategy. To accomplish this, we simply regress excess returns on asset characteristics.

We find that over many periods, this strategy has upward sloping demand and negative elasticity values, much like an investor in style of Stein (2009). Including these negative elasticity values certainly strengthens our case, suggesting that these optimal portfolio strategies create inelastic demand. However, our goal is to specifically determine the demand elasticity of a mean-variance investor attributable solely to the pass-through and substitutability channels, excluding the investor channel described in Stein (2009). Thus we estimate portfolio means with a constrained least squares regression, where, similar to KY, we constrain the coefficients to have a positive price pass-through.[18]

Every month, starting in January 1971 and ending in July of 2019, we form mean-variance optimal portfolios, using the covariance matrix from the previous twelve months, as well as the predicted excess returns from the characteristics and regression coefficients estimated over the previous year. We calculate statistics similar to those presented in Table 2.

Table 3 presents the results of this exercise. We consider a large range of shrinkage weights, $h \in \{0.05, 0.25, 0.50, 0.75, 0.85, 0.95\}$.[19] 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.

These results offer important insights into the factors that drive the high unspanned returns, a key component in delivering inelastic demand, which we emphasize. Notably, increased portfolio shrinkage appears to both enhance out-of-sample Sharpe ratio and reduce elasticity due to the higher unspanned returns. It is important to note that in this analysis, the pass-throughs remain constant across different levels of shrinkage. The only variable changing is the shrinkage weight in these portfolios. However, even with low levels of shrinkage, the unspanned returns are high enough to ensure relatively inelastic demand. Thus, high levels of shrinkage are not necessary to achieve inelastic demand.

This is a novel and significant finding, distinct from its relevance to demand elasticity. The results indicate that asset returns are not adequately spanned by returns from other assets. While

[18] Specifically, we constrain the coefficient on price reversals, Δp_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.

[19] We present the full results for $h = 0.05, 0.10, \ldots, 0.95$ in Table A.1 in the appendix.

Lopez-Lira and Roussanov (2023) show that systematic risk factors do not fully span asset returns, and Baba Yara et al. (2021) reveals that classic asset pricing models also fail in this regard, our findings suggest that all other assets are insufficient to span many asset returns effectively. There are certainly unspanned returns that are close to zero, yet the average unspanned return is considerably large for assets with positive portfolio weights. In summary, the presence of high unspanned returns leads to inelastic demand across various models and data specifications.

4.3 Isoelastic Demand

Given the skewed size distribution of firms, investor portfolio weights in the 13F data exhibit fat tails with a lognormal distribution. As a result, KY consider an exponential-linear demand function instead of the linear demand shown above. In this subsection, we present evidence that the functional form of the demand function matters. In particular, we show that an exponential-linear functional form appears to yield even lower demand elasticity.

The KY demand function is nearly isoelastic, meaning that the quantitative difference between an isoelastic demand function and theirs in terms of the elasticity is quite small.[20] To be clear, this near-isoelastic demand function can be derived as a mean-variance demand function without additional constraints (see Koijen, Richmond, and Yogo, 2020), which implies that our decomposition in Proposition 1 holds with this isoelastic demand function.

In this section, we use Proposition 1 to show why isoelastic demand yield even more inelastic demand than shown in the previous section. We also investigate the reasons isoelastic demand may more accurately fit the data compared to the linear model illustrated earlier, and conclude with an empirical exercise confirming that isoelastic demand indeed results in a lower elasticity.

[20]KY have an additional adding up constraint, which multiplies by isoelastic $\partial \log(w_{i,t})/\partial \log(P_{i,t})$ by a $(1 - w_{i,t})$ term. Since portfolio weights tend to be small, this term is close to one, and this additional $(1 - w_{i,t})$ factor affects the demand elasticity very little.

An observant reader may express concern that the previous results are conditional only on positive weights, are sensitive to filters regarding the size of portfolio weights, and that linear mean-variance weights may be inconsistent with investor holdings data. These important concerns are addressed by employing isoelastic demand which ensures only long positions. KY and Koijen et al. (2020) argue that this approach fits the holdings data better. Crucially, by resolving these issues, we produce an even more inelastic demand, which aligns more closely with the observed data.

Isoelastic assets demand has the following functional form:

$$w_{i,t} = \exp(-\beta_{i,t} \log(P_{i,t}) + \upsilon_{i,t}) = e^{\upsilon_{i,t}} P_{i,t}^{-\beta_{i,t}}, \tag{27}$$

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

$$-\frac{\partial w_{i,t}}{\partial \log(P_{i,t})} = \beta_{i,t} w_{i,t}. \tag{28}$$

Using Equations (22) and (23), we have:

$$(1 - \omega_{i,t})\tilde{\mu}_{i,t} = w_{i,t} \gamma A_t \tau_{i,t}^{-1} \quad \text{and} \quad \psi_{i,t} = \beta_{i,t} w_{i,t} \gamma A_t \tau_{i,t}^{-1}. \tag{29}$$

This is the key theoretical insight of this section: the pass-through and unspanned return differ only by a factor of $\beta_{i,t}$, which is the log-price coefficient in Equation (27). Isoelastic demand locks together these two terms in a way that does not occur with the linear demand function above. Plugging in the terms in Equation (29) into the key decomposition of the paper, Equation (21), we have a simple expression for the demand elasticity:

$$\beta_{i,t} = 1 + \frac{\psi_{i,t}}{(1-\omega_{i,t})\tilde{\mu}_{i,t}} = 1 + \beta_{i,t}. \tag{30}$$

The log-price coefficient, $\beta_{i,t}$, not only controls the elasticity but also determines the difference, in terms of ratios, between the unspanned return and the pass-through. It should be noted that in the KY model, each investor has the same $\beta_{i,t}$ across assets, allowing us to write $\beta_{i,t} = \beta_t$ in their model.

In the previous section, the high elasticity values result largely from the relatively small unspanned returns compared to pass-throughs. Given the strong link between pass-throughs and unspanned returns in an isoleastic demand model as shown in (29), the ratio of the pass-through to the unspanned return is unlikely to be so large. Therefore, drawing from this basic theory, we would reasonably anticipate that an isoelastic demand function would deliver more inelastic demand, as empirically shown below.

Isoelastic demand may appear overly restrictive, as the parameter $\beta_{i,t}$ controls both the elasticity and the link between unspanned returns and pass-throughs. However, we first briefly describe issues with linear demand that are addressed by adopting isoelastic demand functions.

As shown earlier, linear demand functions produce very elastic demand for stocks with small portfolio weights. If the entire market is estimated with linear demand functions, and the variation across stocks in price pass-throughs is relatively low (which is the case in our structural model above), then the smallest stocks in terms of market equity would have the highest elasticity. In fact, Davis (2023) estimates a linear demand system for the entire market using 13F data, and finds an average price elasticity close to 12, similar to the data above, which is driven mostly by small market portfolio weight stocks. Davis (2023) does this exercise as a comparison to standard linear portfolio choice models, in order to provide a fair comparison. With a linear demand system, the smallest, and likely the most illiquid stocks, have the smallest price impacts from flows. While this may be true to some extent (see e.g., Haddad et al., 2022), these large differences from linear models seem extreme. Isoelastic demand functions resolves this issue, since both the price pass-through and unspanned returns are scaled by the portfolio weight, as shown in Equation (29), and thus cancel each other out as shown in Equation (30). In summary, isoelastic demand resolves issues about the sensitivity of elasticity values to small portfolio weights, does not require us to ignore short positions since only long positions are produced, and perhaps produce more realistic portfolio weights according to KY.

In the linear model described above, the covariance matrix, mean, and the impact of prices on the mean of returns are all explicitly modeled. After fitting the parameters, we calculate the optimal demand and corresponding elasticity values. This optimal demand, derived from the structural model, is linear with in asset characteristics (Koijen and Yogo, 2019). In this section, we utilize a methodology similar to Brandt, Santa-Clara, and Valkanov (2009), in which portfolio weights are specified as a function of the data and parameters. We then select parameters to maximize a certain quantity, e.g., the Sharpe ratio. After fitting these parameters, we can estimate elasticity values based on the demand. Unlike the previous approach, we can no longer decompose the demand elasticity into pass-through and unspanned return components, because the value $\tau_{i,t}$ is not explicitly modeled.

Following KY, we assume demand is exponential-linear:

$$\delta_{i,t} \equiv \exp\left(\sum_j Z^{\mu}_{i,j,t} b_j\right) \quad \text{and} \quad w_{i,t} = \frac{\delta_{i,t}}{\sum_k \delta_{k,t}}. \tag{31}$$

Note that since the weights add up to one, the intercept term cannot be identified; therefore, we omit it. The elasticity is easily calculated as:

$$\eta_{i,t} = 1 - (1 - w_{i,t}) \left(\sum_j \frac{\partial Z^{\mu}_{i,j,t}}{\partial \log(P_{i,t})} b_j \right) \tag{32}$$

Brandt et al. (2009) discuss various optimization functions, including choosing portfolio weights to maximize the Sharpe ratio. Since mean-variance demand also corresponds to choosing portfolio weights to maximize the Sharpe ratio, this objective function is appropriate for this exercise in terms of comparability with the above results.

Table 4 shows the parameter estimates from this optimization. Standard errors are calculated using the usual formula for extremum estimators (Newey and McFadden, 1994). The average elasticity in the last column is calculated according to Equation (32) above. Note that when averaged across time and assets, the average elasticity simply becomes a linear combination of the parameters. The average value of $(1 - w_{i,t})\frac{\partial Z^{\mu}_{i,j,t}}{\partial \log(P_{i,t})}$ serves as the multiplier for b_j. Therefore, we employ the delta method to calculate the standard error of the average elasticity.

The average elasticity value is 2.6. Note that the reversal, book-to-price, and size variables are all important determinants of the elasticity. Interestingly, the coefficient on the book-to-price ratio is negative, implying that this ratio actually reduces elasticity. The coefficient for reversals and size are also negative; however, considering their derivatives, it indicates that these factors are the primary sources of elasticity. The elasticity of 2.6 is much lower than the elasticity derived from the linear structural model in Table 2. Davis (2023) discusses the inherent differences in reactions to prices between linear and exponential-linear demand functions. This elasticity aligns closely with literature estimates for micro asset demand elasticity, as discussed in Gabaix and Koijen (2021).

The in-sample Sharpe ratio is 1.2 for this model, with a standard error of 0.054. This suggests that this low elasticity is achieved with relatively realistic returns, particularly considering that this is an in-sample exercise.

In summary, the functional form of the demand function plays an important role in determining elasticity values. As shown above, we summarize our elasticity decomposition in Figure 1, demonstrating our approach to reconciling the discrepancy between the high elasticity values found in standard asset pricing models and the more inelastic empirical estimates.

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

Given that one of the most frequently discussed demand elasticity estimate comes from KY, we examine the price pass-throughs associated with their instrument.[21] We estimate the following panel regression for multiple horizons h:

$$r_{i,t+1\to t+h} = \beta_0 + \beta_1 \log\left(\frac{M_t}{B_t}\right) + (X_t^c)'\beta + \epsilon_{i,t+1}, \tag{33}$$

where the main independent variable, log market-to-book ratio, $\log\left(\frac{M_t}{B_t}\right)$, is instrumented using $\log\left(\frac{\widehat{M}_t}{B_t}\right)$, where $\widehat{M}_t$ is the holdings-based instrument in KY. The coefficient estimate of $(-\beta_1)$ can be directly interpreted as the key theoretical quantity $\left(-\frac{\partial \tilde{\mu}}{\partial \log(P)}\right)$. We also follow KY to control for the same vector of stock characteristics X_t^c. To focus on cross-sectional results, we add time-fixed effects. We calculate Driscoll-Kraay standard errors (Driscoll and Kraay, 1998) with 8 lags which control for both time-series and cross-sectional correlations.

The results are reported in Table 5. 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

[21] Recently, alternative instruments based on flow-induced demand shocks have been proposed in the demand-based asset pricing literature. See, e.g., van der Beck (2022) for stocks and Li, Fu, and Chaudhary (2022) for corporate bonds.

statistically significant but also much lower than the $-\frac{\partial \tilde{\mu}}{\partial \log(P)} = 1$ benchmark in frictionless models. More importantly, this predictability is significantly 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, and the degrees of reversions are much smaller, with $-\frac{\partial \tilde{\mu}}{\log(P)}$ only 0.018 and 0.062 at the quarterly and annual frequencies.

In summary, while there is evidence of positive price pass-throughs associated with the KY instrument, the pass-throughs are quite weak, especially when considering large-cap stocks that institutions more heavily hold. This finding is consistent with the low institutional demand elasticities estimated in KY. In other words, KY estimated demand seems much more consistent with the D2 definition than the D1 definition.

These results do not assert that the KY system is invalid, nor do they constitute a critique of their model. While we highlight that the KY demand curves appear to align more closely with a D2 definition than with a D1 definition, we argue that for many counterfactual experiments, a D2 demand system is a more useful model than a D1 demand system. For example, Davis (2023) considers counterfactual experiments with literature-based statistical arbitrageurs, and it is unreasonable to assume that the rest of the market would think that price movements due to these arbitrageurs would completely revert after a single period. In fact, these price movements do not revert in this setting, because the arbitrageur strategies persist through time. Thus, for this setting, KY style demand functions are a better model for demand curves than a D1 elasticity type model. A D1 demand system is useful for price changes where investors believe that the prices will immediately pass-through to returns in the next period, which is an unusual occurrence in stock markets. In conclusion, the KY demand system is more consistent with a D2 type demand system, which is often more useful for counterfactual experiments than a D1 type demand system.

5 Robustness: Other Sources of Elasticity

In this section, we explore other sources of demand elasticity and conduct several robustness checks. We first consider a more general demand function. Consider portfolio weights defined as

$$w_{i,t} = g_{i,t}(\tilde{\mu}_{i,t}, \tilde{v}_{i,t}), \tag{34}$$

where $g_{i,t}(\cdot)$ represents a general function and $\tilde{v}_{i,t}$ consists of the residual determinants of portfolio weights other than subjective expected returns, which could potentially be a function of prices, specifically risk.

Assuming differentiability and positive demand, we can write the demand elasticity as:

$$\eta_{i,t} \equiv -\frac{\partial \log(S_{i,t})}{\partial \log(P_{i,t})} = 1 \underbrace{- \frac{\partial \log(w_{i,t})}{\partial \tilde{\mu}_{i,t}} \frac{\partial \tilde{\mu}_{i,t}}{\partial \log(P_{i,t})}}_{\text{mean component}} \underbrace{- \frac{\partial \log(w_{i,t})}{\partial \tilde{v}_{i,t}} \frac{\partial \tilde{v}_{i,t}}{\partial \log(P_{i,t})}}_{\text{residual component}} \underbrace{- \frac{\partial \log(A_t)}{\partial \log(P_{i,t})}}_{\text{wealth effect}}. \tag{35}$$

KY assume investor wealth, A_t, is exogenous or not a function of prices. In a broader macroeconomic context, this assumption might seem unrealistic. However, when considering a large number of assets (N) in a diversified portfolio where each asset has a relatively small weight, a 1% movement in price will result in a significantly smaller change in the portfolio value, often less than 1%. Therefore, quantitatively, in a microeconomic framework, the impact of price changes on wealth can be considered negligible. We adopt a similar perspective by treating investor wealth as exogenous, essentially disregarding these wealth effects when calculating elasticity. In reality, $\partial A_t/\partial \log(P_{i,t})$ is typically small, positive, and much less than one, making the demand even more inelastic. Our findings show that we can explain the observed inelastic demand empirically without relying on these wealth effects.

5.1 Residual component of demand

In this section, we investigate the residual demand component in Equation (35) using a more general model than the CARA model previously discussed. Our main finding is that the CARA model and decomposition discussed above are the first-order components in a more general frictionless model.

First, we consider the case of Epstein-Zin multivariate demand for assets discussed in Campbell et al. (2003). They show, after log-lineariztion, portfolio weights are given by:

$$w_t = \frac{1}{\gamma}\Sigma_t^{-1}\left[\mathbb{E}_t[y_{t+1}] + \frac{1}{2}\sigma_t^2 - \frac{\vartheta}{\varsigma}\sigma_{c-w,t}\right], \tag{36}$$

where y_t is an N dimensional vector of log returns minus the log risk-free rate, Σ_t is the $N\times N$ conditional covariance matrix of y_{t+1}, σ_t^2 is the N dimensional vector containing the diagonal elements of Σ_t, $\sigma_{c-w,t}$ is the N dimensional vector of conditional covariance of the log consumption-to-wealth ratio (the cay variable from Lettau and Ludvigson, 2001) and y_{t+1}, $\gamma > 0$ is the relative risk aversion coefficient, $\varsigma > 0$ is the elasticity of intertemporal substitution, and $\vartheta \equiv (1-\gamma)/(1-\varsigma^{-1})$.[22] In this section, we primarily consider quarterly data, since cay data is available at the quarterly frequency.

We need to model the conditional expectation of y_t as a function of prices, as well as the

[22] See equation (20) of Campbell et al. (2003). Note that their equation includes additional terms because they also consider y_t to be log return of the asset minus a benchmark with potential covariance terms. In our case, we only consider the risk-free rate, which removes some of these extra terms.

conditional covariance of y_t itself and with the log consumption-to-wealth ratio as a function of prices. We do this by running predictive regressions for the relevant terms, as we describe below. Instead of a unified estimation above, we are forced to break this more complicated model into components and estimate each one separately.

5.2 Conditional expectation model

To model the conditional mean, we regress log excess returns on the same variables contained in Z_t^{μ} from above: reversals ($\Delta p_{i,t}$), log book to price ratio ($b_{i,t}-p_{i,t}$), size ($(p_{i,t}-\bar{\mu}_t)/\bar{\sigma}_t$), profitability, investment, dividend-to-book ratio, and market beta. We stack these latter four exogenous control variables into a column vector of controls, X_t^c. The regression is written as follows:

$$\log(R_{i,t+1}) = \beta_0 + \beta_1 \Delta p_{i,t} + \beta_2(b_{i,t} - p_{i,t}) + \beta_3 \frac{p_{i,t} - \bar{\mu}_t}{\bar{\sigma}_t} + (X_t^c)'\boldsymbol{\beta} + \epsilon_{i,t+1}, \tag{37}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients. Thus the average estimated price derivative is:

$$\frac{\partial \tilde{E}_t[y_{t+1}]}{\partial p_t} \equiv \text{Mean}\left(\frac{\partial \tilde{E}_t[y_{i,t+1}]}{\partial p_{i,t}}\right) = \beta_1 - \beta_2 + \beta_3 \times \text{Mean}\left(1/\bar{\sigma}_t\right), \tag{38}$$

where y_t is the excess log returns. We refer to this as the cross-sectional model, as it employs a standard set of cross-sectional variables to predict variations in expected returns. This is very similar to the previous model above, except here log returns are used and the model is fitted using standard OLS regression rather than maximum likelihood.

We present the regression results in Table 6. Our primary focus is on the price pass-throughs at the quarterly horizon, but we also include results for daily, weekly, monthly, quarterly, and annual horizons. From the p-values for the estimates of price pass-through, $\psi_{i,t}$, we can see that price variations do not statistically significantly predict future returns at annual, quarterly, and monthly horizons. However, at the weekly and daily horizons, a 1% price increase predicts predicts a decrease in returns of 5 and 13.6 bps, respectively, over the next period. This decline is attributed almost entirely to the reversals component, Δp_t, rather than the book-to-market, $b_t - p_t$, or size, $(p_{i,t} - \bar{\mu}_t)/\bar{\sigma}_t$, components.

We consider a greatly expanded information set, where stock fixed effects are added to the regression in (37). We refer to this as the fixed effects model. The regression equation remains the same, except that β_0 is now changed to $\beta_{i,0}$ to represent a stock-specific fixed effect. This specification corresponds to a greatly expanded information set, where an investor has a *stock-specific* estimate of each asset's valuation and return, rather than relying on cross-sectional relationships. In other words, the regression in Equation (37) corresponds to an investor who makes forecasts based on rational expectations by utilizing all returns projected onto the space of asset characteristics. This means a high price predicts a low return when other assets with similar characteristics have a similarly high price/low return relationship. If one does not find this relationship in the cross-section, then the investor does not believe there is a strong price/return relationship.

In a regression with stock fixed effects, the investor knows whether the price of any given asset is high or low. They do not rely on the cross-section and can estimate the price is high or low in the time series of the asset's own returns. This corresponds to a relatively extensive information set.

The regression results for the fixed effects model are shown in Table 7. In terms of point estimates, the pass-through, $\psi_{i,t}$, is negative across all investment horizons. The annual, weekly, and daily horizons all have statistically significant estimates. A 1% price increase corresponds to anywhere between a 2 to 14 bps decrease in the next period expected returns, depending on the horizon.[23] Thus across annual, quarterly, monthly, weekly, and daily investment horizons, price predictability are much weaker than a one-for-one pass-through.

The marginal price effect is calculated for each stock in each period, and the average effect across stocks and periods is shown in the second column of Table 8. Note that with the simple model, the marginal effect is identical across stocks and time. Standard errors are shown below the estimates, which are double clustered by quarter and stock. Note that both the simple and cross-sectional models have *positive* average marginal effects, indicating potentially rational upward sloping demand if the elasticity is determined solely by the mean effects. The fixed effects average marginal effect is negative, but only -0.035, which means a 1% price increase decreases expected returns by only 3.5 basis points (bps). This is far from a one-for-one pass-through.

5.3 Conditional covariance terms model

We consider a standard factor structure for the covariance matrix:

$$\Sigma_t = \beta_t \Omega \beta_t' + \zeta I, \tag{39}$$

where Ω is the $F \times F$ matrix of factor returns, β_t is a $N \times F$ vector of factor loadings, and $\zeta > 0$ is a scalar that dictates the size of the idiosyncratic variance. For our analysis, we utilize six factors: the Fama-French five and momentum, thus $F = 6$. Although this model is much less parsimonious than the structural model previously discussed, it offers a more flexible approach to modeling the covariance matrix. We estimate Ω as the empirical covariance matrix of factor returns. If we estimate β_t with rolling regressions of historical returns, the covariance matrix mechanically becomes independent of price movements. Similarly, if we estimate the $\sigma_{c-w,t}$ component from

[23]This range widens when considering confidence intervals.

Equation (36) in the same manner, the residual component of demand would mechanically be zero. We consider an alternative specification that allows the residual component of demand elasticity to be non-zero. This approach involves running predictive regressions of the "realized" covariance terms on (log) prices and other variables in order to determine how well price fluctuations predict these covariance terms.

We parameterize factor loadings β_t as

$$\beta_t = \begin{bmatrix} \sigma_{1,t} & \sigma_{2,t} & \dots & \sigma_{F,t} \end{bmatrix} \Omega^{-1}, \tag{40}$$

where $\sigma_{j,t}$ is the N dimensional column vector of conditional covariance terms of y_{t+1} and factor return $f_{j,t}$. We parameterize $\sigma_{j,t}$ as linear in characteristics X_t, i.e., $\sigma_{j,t} = X_t\beta_j^f$, where β_j^f is a vector of regression parameters. We fit these coefficients by running the following regression:

$$(f_{j,t+1} - \mu_j^f)\epsilon_{t+1} = X_t\beta_j^f + \nu_{j,t+1}, \tag{41}$$

where μ_j^f is the average return for the j^{th} factor. Note that the left-hand side variable is essentially the "realized covariance," meaning that the conditional expectation of this variable is the conditional covariance as long as the model for the conditional mean is correct. Since we have two different models of the mean above, as shown in Tables 6 and 7, we plug in two different regression residuals ϵ_{t+1} into the regression above, and obtain similar results in terms of the size of the residual elasticity term. We run this regression separately for each factor $f = 1, \dots, F$. If we define the $K \times F$ matrix $\Gamma = \left(\beta_1^f, \beta_2^f, ..., \beta_F^f\right)$, then we can write:

$$\beta_t = X_t\Gamma\Omega^{-1}. \tag{42}$$

Thus conditional betas are functions of characteristics, some of which include prices. This characterization allows the covariance matrix Σ_t to be a function of characteristics X_t, potentially allowing price to affect demand through channels beyond just the expected return channel—i.e., the residual elasticity channel in Equation (35) discussed above. Notice that this characterization follows both Pástor and Stambaugh (2003) and Kelly et al. (2019), settings the beta of the assets to be a linear function of asset characteristics. This specification implies that the covariance matrix of individual asset's log excess returns is:

$$\Sigma_t = \beta_t \Omega \beta_t' + \zeta I = X_t \Gamma \Omega^{-1} \Gamma' X_t' + \zeta I. \tag{43}$$

We estimate ζ as the variance (across assets and time) of $y_{t+1} - \beta_t f_{t+1}$.

We follow a similar approach to obtain an estimate of $\sigma_{c-w,t}$. We fit the following regression:

$$\text{cay}_{t+1} \cdot \epsilon_{t+1} = X_t \beta_{c-w} + \nu_{t+1}, \tag{44}$$

where cay_{t+1} is the deviation of the log consumption to wealth ratio from the average from Lettau and Ludvigson (2001), obtained from Martin Lettau's website.

5.3.1 Results

With these parameterizations, we can rewrite Equation (36) as:

$$\begin{aligned} w_t &= \frac{1}{\gamma}(X_t \Gamma \Omega^{-1} \Gamma' X_t' + \zeta I)^{-1} \left[X_t \beta + \frac{1}{2} \text{Diag}\left(X_t \Gamma \Omega^{-1} \Gamma' X_t' + \zeta I \right) - \frac{\vartheta}{\varsigma} X_t \beta_{c-w} \right] \\ &= \frac{1}{\gamma\zeta}\left(I - \underbrace{X_t \Gamma (\zeta \Omega + \Gamma' X_t' X_t \Gamma)^{-1} \Gamma' X_t'}_{\text{covariance}} \right) \left[\underbrace{X_t \beta}_{\text{mean}} + \underbrace{\frac{1}{2} \text{Diag}(X_t \Gamma \Omega^{-1} \Gamma' X_t' + \zeta I)}_{\text{variance}} - \underbrace{\frac{\vartheta}{\varsigma} X_t \beta_{c-w}}_{\text{cay}} \right], \end{aligned} \tag{45}$$

where Diag$(\cdot)$ is the function that has a square matrix as an argument and outputs a column vector containing the diagonal of the matrix.

In Equation (45), we highlight four different components or channels through which price changes can affect demand: (1) the mean component, (2) the covariance component, (3) the variance component, and (4) the consumption-to-wealth component labeled cay. Note that with this demand specification, components (2), (3), and (4) together account for the residual component discussed above in Equation (35).

The purpose of the exercise in this section is to determine if the overall optimal Epstein-Zin demand yields inelastic demand and how important the four components of the residual elasticity are for the overall demand elasticity. In order to do this, we show the decomposition of demand elasticity into these five components. Let A_t be an $N \times J$ matrix with elements $Y_{i,j,t}$ and let p_t be the $N \times 1$ vector of log prices. Assume that each element of $A_{i,j,t}$ is a differentiable function of $p_{i,t}$. Define the following:

$$\nabla_{p_t} A_t = \begin{bmatrix} \frac{\partial A_{1,1,t}}{\partial p_{1,t}} & \frac{\partial A_{1,2,t}}{\partial p_{1,t}} & \dots & \frac{\partial A_{1,J,t}}{\partial p_{1,t}} \\ \frac{\partial A_{2,1,t}}{\partial p_{2,t}} & \frac{\partial A_{2,2,t}}{\partial p_{2,t}} & \dots & \frac{\partial A_{2,J,t}}{\partial p_{2,t}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial A_{N,1,t}}{\partial p_{N,t}} & \frac{\partial A_{N,2,t}}{\partial p_{N,t}} & \dots & \frac{\partial A_{N,J,t}}{\partial p_{N,t}} \end{bmatrix}. \tag{46}$$

Then as shown above, we can write:

$$\nabla_{p_t} X_t = \begin{bmatrix} 0 & 1 & -1 & 1/\bar{\sigma}_t & 0 & \dots & 0 \\ 0 & 1 & -1 & 1/\bar{\sigma}_t & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & -1 & 1/\bar{\sigma}_t & 0 & \dots & 0 \end{bmatrix}, \tag{47}$$

where elements in the first column are zero because of the intercept, elements in the second column are one because this is the reversals column, elements in the third column are -1 because this corresponds to the log book-to-market ratio, elements in the fourth column are $1/\bar{\sigma}_t$ because this corresponds to the size column. The rest of columns are zero because their corresponding characteristics are not functions of price (investment, profitability, dividend-to-book, and market beta). Using this notation, we can define the elasticity, following KY for only assets with positive weights, as:

$$\eta_t = \eta_t^m + \eta_t^c + \eta_t^v + \eta_t^{\text{cay}}, \tag{48}$$

where η_t^m, η_t^c, η_t^v, and η_t^{cay} correspond to the elasticity components from mean, covariance, variance, and cay in Equation (45), respectively. These components are defined as:

$$\eta_t^m = \vec{1} - \frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}((\vec{1} - \text{Diag}(X_t\Gamma\Lambda_t\Gamma'X_t')) \circ ((\nabla_{p_t}X_t)\beta)), \tag{49}$$

$$\begin{aligned}
\eta_t^c = {} & \frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}((\nabla_{p_t}X_t)\Gamma\Lambda_t\Gamma'X_t'\mu_t \\
& - \frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}(\text{Diag}(X_t\Gamma\Lambda_t((\nabla_{p_t}X_t)\Gamma)') \circ (X_t\Gamma\Lambda_t\Gamma'X_t'\mu_t)) \\
& - \frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}(\text{Diag}(X_t\Gamma\Lambda_t\Gamma'X_t') \circ (((\nabla_{p_t}X_t)\Gamma)\Lambda_t\Gamma'X_t'\mu_t)) \\
& + \frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}(\text{Diag}(X_t\Gamma\Lambda_t((\nabla_{p_t}X_t)\Gamma)') \circ \mu_t),
\end{aligned} \tag{50}$$

$$\eta_t^v = -\frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}((\vec{1} - \text{Diag}(X_t\Gamma\Lambda_t\Gamma'X_t')) \circ (((X_t\Gamma\Omega^{-1}\Gamma') \circ (\nabla_{p_t}X_t))\vec{1})), \tag{51}$$

$$\eta_t^{\text{cay}} = \frac{\vartheta}{\varsigma}\frac{1}{\gamma\zeta}\text{diag}(w_t)^{-1}((\vec{1} - \text{Diag}(X_t\Gamma\Lambda_t\Gamma'X_t')) \circ ((\nabla_{p_t}X_t)\beta_{c-w})), \tag{52}$$

where $\Lambda_t = (\zeta\Omega + \Gamma'X_t'X_t\Gamma)^{-1}$, $\mu_t = X_t\beta + \frac{1}{2}\text{Diag}(X_t\Gamma\Omega^{-1}\Gamma'X_t' + \zeta I) - \frac{\vartheta}{\varsigma}X_t\beta_{c-w}$, $\circ$ is the Hadamard product (element-wise), and $\vec{1}$ is a vector of ones.

In Table 8, we present the mean elasticity component, η^m, in column (1) for the three models for the conditional expectation discussed above. We then decompose η^m into two parts: first, the change in expected returns in response to (log) price movements in column (2), and second, the change in the (log) portfolio weight in response to the changes in the discount rate in column (3). We find that for all three cases, the pass-through component in column (2) is much smaller than the impact of discount rates on the portfolio weight, leading to low demand elasticities. For the fixed effects model in the third row, the pass-through component in negative, i.e., there is momentum at the quarterly horizon.

Note that because the weights have a $1/\gamma$ term, then $\mathrm{diag}(w_t)^{-1}/\gamma$ is only a function of gamma due to the cay component. In other words, for the elasticity, γ only matters for the component ϑ/ς term that multiplies the cay component, and does not enter the elasticity through any other way. We need to pick reasonable Epstein-Zin values of γ and ψ in order to estimate the size of this cay component. To do this, we pick the largest possible value for $|\vartheta/\varsigma|$ with a reasonable range for γ and ψ. We consider $\gamma \leq 10$ and $\psi \in [1.5, 2]$. These values of γ and ψ imply preference for the early resolution of uncertainty and have been used extensively in the asset pricing literature to address a number of asset pricing puzzles (e.g., Bansal and Yaron, 2004; Hansen, Heaton, and Li, 2008). An EIS value greater than one implies a decline in asset prices when the effective risk aversion in the economy increases. Given this range, the largest possible $|\vartheta/\varsigma|$ is 18, with $\gamma = 10$ and $\psi = 1.5$. Using a smaller value of $|\vartheta/\varsigma|$ shrinks the already small cay elasticity term further toward zero.

In Table 9, we decompose the demand elasticity in column (1) into mean and residual components in columns (2) and (3). From Equation (48), the residual component in column (3) is the sum of covariance, variance, and cay components in columns (4), (5), and (6). We find that the elasticity is primarily determined by the mean component.

These results justify using Equation (1) as an approximation for the demand elasticity in Equation (35), where the elasticity is primarily set by the expected return due to a price change.

From hereon, we focus on this mean effect. Note that classic portfolio choice models take the covariance structure as exogenous to prices; thus, we find this a reasonable approximation. Indeed, after accounting for the mean effects, price changes predict the covariance terms only weakly, consistent with the elasticity of demand being primarily determined by the mean component. If the mean effects are relatively weak, then demand is inelastic.

While the model above is just one exercise in computing demand elasticity, one may wonder whether other models estimated with reasonable pass-throughs and high unspanned returns deliver much larger elasticity values. Davis (2023) estimates the demand elasticity of twelve quantitative portfolio choice models and shows that demand is either inelastic. Thus across a wide variety of models, demanded derived from optimal portfolio choice models is inelastic.

6 Conclusion

In this paper, we identify low price pass-throughs and large perceived unspanned returns as the source of the inelastic demand puzzle in the stock market. We first decompose demand elasticity into three parts: the mean, residual, and wealth effect components. In a standard portfolio choice model, we then show the mean component primarily determines demand elasticity.

We further decompose the mean component of elasticity into the product of two parts: the extent to which prices predict returns (price pass-through), and second, how well an asset is perceived to be spanned by all other assets, which is a perceived unspanned expected excess returns. Given the Campbell and Shiller decomposition, we then consider two definitions of demand elasticity associated with different sources of price movements. In the first definition of elasticity, D1, we consider price changes only due to the next-period discount rate movements, holding everything else constant, i.e., when we assume complete price pass-throughs. Given that estimating D1 in the stock market is highly challenging and creates near arbitrage opportunities, we introduce the second definition, D2, which considers price movements not entirely driven by next-period discount rates.

We highlight that almost all empirical estimates measure D2 elasticity while most theoretical models only study D1. We empirically estimate D1 for investment-grade bonds where it is arguably more plausible to do so, and, consistent with theory, we find high demand elasticities. As discussed above and consistent with prior literature, we find low D2 elasticity at different horizons driven by low pass-throughs.

There is a common distinction between shifting the demand curve versus moving along the demand curve. A movement along the demand curve is usually defined as a price change where all other demand inputs remain the same. Shifts in other demand inputs represent movements of the demand curve itself. However, in asset markets, if we think of expected discount rates and cash flows as demand inputs, then price movements must necessarily affect either discount rates, cash flows, or some combination of the two. Sometimes, price changes affect beliefs which influence quantities demanded, which are often described in the literature as having an effect on the slope of the demand curve (Stein, 2009; Becker, 1991). Our D1 and D2 elasticity definitions constitute similar definitions of demand elasticities. Just as there are price movements associated with discount rates or cash flows, there are multiple types of demand curves which can be used to model various kinds of reactions to these different price changes.

In this paper, we present a novel decomposition of demand elasticity. When an asset is perceived as having low price pass-throughs to returns and high unspanned returns relative to other assets, then demand will be relatively inelastic. We show that both these attributes of stock returns are supported in the data, thus rationalizing inelastic demand.

References

Baba Yara, Fahiz, Brian H. Boyer, and Carter Davis, 2021, The factor model failure puzzle, Working Paper, Indiana.

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

Becker, Gary S., 1991, A note on restaurant pricing and other examples of social influences on price, *Journal of Political Economy* 99, 1109–1116.

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.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Luis M. Viceira, 2005, The term structure of the risk–return trade-off, *Financial Analysts Journal* 61, 34–44.

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.

Da, Zhi, Borja Larrain, Clemens Sialm, and José Tessada, 2018, Destabilizing financial advice: Evidence from pension fund reallocations, *Review of Financial Studies* 31, 3720–3755.

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.

Deuskar, Prachi, and Timothy C. Johnson, 2011, Market liquidity and flow-driven risk, *Review of Financial Studies* 24, 721–753.

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.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

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.

Hansen, Lars Peter, John C. Heaton, and Nan Li, 2008, Consumption strikes back? measuring long-run risk, *Journal of Political Economy* 116, 260–302.

Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41, 815–829.

Hartzmark, Samuel M., and David H. Solomon, 2022, Predictable price pressure, Working Paper, Boston College.

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, 2020, Which investors matter for equity valuations and expected returns?, Working Paper 27402, National Bureau of Economic Research.

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

Koijen, Ralph S.J., and Motohiro Yogo, 2020, Exchange rates and asset prices in a global demand system, Working Paper 27342, National Bureau of Economic Research.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance* 73, 1183–1223.

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, 2021, What drives the size and value factors?, Working Paper, Utah.

Li, Jiacui, and Zihan Lin, 2022, Prices are less elastic for less diversifiable demand, Working Paper, Utah and Stanford.

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

Li, Jie, Neil D. Pearson, and Qi Zhang, 2020, Impact of demand shocks on the stock market: Evidence from Chinese IPOs, Working Paper, SAIF, UIUC, and Durham.

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.

Lou, Dong, Hongjun Yan, and Jinfan Zhang, 2013, Anticipated and repeated shocks in liquid markets, *Review of Financial Studies* 26, 1891–1912.

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.

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

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

Peng, Cameron, and Chen Wang, 2021, Factor demand and factor returns, Working Paper, LSE.

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

Pollet, Joshua M., and Mungo Wilson, 2010, Average correlation and stock market returns, *Journal of Financial Economics* 96, 364–380.

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.

Stein, Jeremy C., 2009, Presidential address: Sophisticated investors and market efficiency, *Journal of Finance* 64, 1517–1548.

van Binsbergen, Jules H., and Ralph S.J. Koijen, 2017, The term structure of returns: Facts and theory, *Journal of Financial Economics* 124, 1–21.

van der Beck, Philippe, 2022, On the estimation of demand-based asset pricing models, Working Paper, HBS.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information acquisition and under-diversification, *Review of Economic Studies* 77, 779–805.

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 π, ξ, and ζ in Equation (24). The estimates of π are shown in the first row labeled "Mean". The estimates of ξ and the scalar ζ are shown in the next row labeled "Covariance". The parameter estimates are labeled by their respective covariates in Z_t^μ and Z_t^Γ 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	Δp_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	ζ
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. Maximum Likelihood Elasticity Results. With the estimated parameters values shown in Table 1 from the structural model in Equation (24), the elasticity, 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 ($\omega_{i,t}\tilde{\mu}_{i,t}$) are calculated for every stock in the sample in every period where the optimal portfolio weight is positive. Optimal portfolio weights are calculated as $w_t = \frac{1}{\gamma A_t}\tilde{\Sigma}_t^{-1}\tilde{\mu}_t$. This table shows various statistics of these values taken across months and assets. 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 (see Equation 21). Thus, we consider various cuts of the data that exclude some of these near-zero unspanned return values which lead to very high elasticity values. The "Top Percent of Value Invested" column, if not empty, indicates that a subsample of the positive weight observations were considered when calculating the averages. For example, a 99.9% filter takes the top 99.9% of stocks in terms of value invested (positive portfolio weights), dropping the assets with the smallest portfolio weights first. The last row shows the weighted average by portfolio weights. Standard errors are shown below the estimates in parentheses. For the averages in columns (2)–(6), these standard errors are double clustered by stock and month. For the median in column (1), standard error, the standard error is calculated via bootstrap as described in the text. The in-sample Sharpe ratio for this model is 1.45, with a standard error of 0.059.

Statistic	Top Percent of Value Invested	Elasticity	Pass-Through	Unspanned Return	Expected 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. Shrinkage Portfolios Results. This table shows the out-of-sample shrinkage portfolio results from Section 4.2. Each column represents a different portfolio with different shrinkage weights—the larger weights overweight the shrinkage target and underweight the empirical covariance matrix. The Sharpe ratio is shown, along with the expected return, unspanned return, median elasticity, average elasticity, and weighted (by portfolio weights) average elasticity for the assets with positive 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. Table A.1 shows results for the entire set of shrinkage portfolios, with shrinkage weights in the set $\{0.05, 0.1, 0.15, ..., 0.95\}$.

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. 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 show in Equation (31). 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
	Δp_t	$b_t - p_t$	$\frac{p_t - \mu_t^p}{\sigma_t^p}$	profit.	invest.	div.-to-book	beta	η
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

Table 5. Predicting Price Pass-throughs with the KY instrument. This table presents estimated coefficients from the regression in Equation (33). 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 KY instrument. We follow KY to control for a number of 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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$h = 1$	2	4	8	$h = 1$	2	4	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 6. Main Price Pass-through Regressions. This table presents estimated coefficients from the regression in Equation (37). This specification corresponds to an information set which includes the log book to price ratio, $b_{i,t} - p_{i,t}$, as well as the log of market capitalization, cross-sectionally normalized, $(p_{i,t} - \bar{\mu}_t)/\bar{\sigma}_t$. We also include other regressors that are not functions of prices: profitability, investment, dividend to book ratio, and market beta, stacked into a column vector X_t. Standard errors are double clustered at the asset and time period levels.

	Dependent variable: r_{t+1}				
	(1) Annual	(2) Quarterly	(3) Monthly	(4) Weekly	(5) Daily
Δp_t	0.097**	0.038**	0.007	−0.050***	−0.136***
	(0.040)	(0.019)	(0.011)	(0.005)	(0.003)
$b_t - p_t$	0.065***	0.013***	0.003***	0.001**	0.000***
	(0.011)	(0.004)	(0.001)	(0.000)	(0.000)
$(p_t - \mu_t)/\sigma_t$	0.043***	0.012***	0.004***	0.001***	0.000***
	(0.010)	(0.003)	(0.001)	(0.000)	(0.000)
profitability	0.219***	0.080***	0.030***	0.008***	0.002***
	(0.029)	(0.009)	(0.003)	(0.001)	(0.000)
investment	−0.226***	−0.063***	−0.022***	−0.005***	−0.001***
	(0.036)	(0.012)	(0.004)	(0.001)	(0.000)
dividend to book	0.539	0.072	0.018	0.003	0.001
	(0.399)	(0.104)	(0.029)	(0.006)	(0.001)
beta	−4.288**	−0.590	−0.152	−0.034	−0.007
	(2.109)	(0.557)	(0.166)	(0.033)	(0.006)
$\psi_{i,t}$	−0.053	−0.031	−0.006	0.05	0.136
χ^2 Test Statistic	2.06	2.806	0.255	93.695	2030.941
Joint Test p-value	0.151	0.094	0.614	0	0
Obs.	186,986	740,624	2,217,863	9,633,930	46,520,654
R^2	0.044	0.014	0.005	0.003	0.014
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 7. Price Pass-through Regressions with Stock Fixed Effects. This table presents estimated coefficients a greatly expanded information set, where stock fixed effects are added to the regression in (37). The regression equation is the same, except β_0 is now changed to $\beta_{i,0}$—a stock fixed effect. This specification corresponds to a greatly expanded information set, where an investor has a stock-specific estimate of the valuation and return of each asset instead of relying on cross-sectional relationships. Standard errors are double clustered at the asset and time period levels.

	Dependent variable: r_{t+1}				
	(1) Annual	(2) Quarterly	(3) Monthly	(4) Weekly	(5) Daily
Δp_t	0.042	0.007	−0.006	−0.053***	−0.136***
	(0.037)	(0.019)	(0.011)	(0.005)	(0.003)
$b_t - p_t$	0.100***	0.025***	0.007***	0.001**	0.000***
	(0.020)	(0.007)	(0.002)	(0.001)	(0.000)
$(p_t - \mu_t)/\sigma_t$	−0.170***	−0.034***	−0.013***	−0.004***	−0.001***
	(0.042)	(0.010)	(0.003)	(0.001)	(0.000)
profitability	0.140***	0.060***	0.023***	0.006***	0.001***
	(0.026)	(0.008)	(0.002)	(0.000)	(0.000)
investment	−0.144***	−0.049***	−0.016***	−0.004***	−0.001***
	(0.034)	(0.012)	(0.004)	(0.001)	(0.000)
dividend to book	−0.114	−0.056	−0.031	−0.010	−0.001
	(0.309)	(0.091)	(0.028)	(0.006)	(0.001)
beta	−4.427	−0.288	−0.022	−0.005	−0.002
	(3.019)	(0.703)	(0.221)	(0.045)	(0.008)
$\psi_{i,t}$	0.141	0.034	0.02	0.056	0.137
χ^2 Test Statistic	15.397	3.46	3.147	118.936	2070.72
Joint Test p-value	0	0.063	0.076	0	0
Obs.	186,986	740,624	2,217,863	9,633,930	46,520,654
R^2	0.037	0.008	0.003	0.003	0.014
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 8. Mean Elasticity Decomposition. In this table we decompose the mean component of demand elasticity, η_m, in column (1) into the pass-through component in column (2) and the response of portfolio weights to discount rate changes in column (3). As shown above, $\eta^m = 1 + \theta_{i,t} \times \psi_{i,t}$.

	(1) η^m	(2) $\psi_{i,t}$	(3) $\theta_{i,t}$
Cross Sectional Model	0.095***	−0.031***	29.230***
	(0.012)	(0.000)	(0.381)
Fixed Effects Model	2.581***	0.035***	45.878***
	(0.021)	(0.000)	(0.627)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 9. Elasticity Decomposition. In this table, we decompose demand elasticity into mean and residual components. The elasticity in column (1) is the sum of the mean and residual components in columns (2) from Table 8 and (3). From Equation (48), the residual component in column (3) is the sum of covariance, variance, and cay components in columns (4), (5), and (6).

	(1) Elasticity	(2) Mean	(3) Residual	(4) Covariance	(5) Variance	(6) cay
Cross Sectional Model	-0.058^{***}	0.095^{***}	-0.153^{***}	$< 10^{-5}$	$< 10^{-5}$	-0.153^{***}
	(0.014)	(0.012)	(0.002)	(0.000)	(0.000)	(0.002)
Fixed Effects Model	2.434^{***}	2.581^{***}	-0.147^{***}	$< 10^{-5}$	$< 10^{-5}$	-0.147^{***}
	(0.019)	(0.021)	(0.002)	(0.000)	(0.000)	(0.002)
Note:						*p<0.1; **p<0.05; ***p<0.01

Appendix

A A Noisy Rational Expectations Model

A.1 Homogeneous private signals

This model is based on Hellwig (1980). This setting is slightly different from Grossman and Stiglitz (1980) in that each agent observes a different signal and tries to back out the signals of other agents from the price. Let $d \sim \mathcal{N}(\mu, v_d)$ denote public information that all investors observe. Suppose an asset has a payoff δ that, conditional on the public information, is normally distributed with mean d and variance v_δ: $\delta \mid d \sim \mathcal{N}(d, v_\delta)$. Thus, it must be the case that the unconditional distribution of δ can be written as $\delta \sim \mathcal{N}(\mu, v_d + v_\delta)$.

Assume each agent i observes $\delta + \epsilon_i$, where ϵ's are iid normal with mean zero and variance v_ϵ, and $\epsilon_i \mid d \sim \mathcal{N}(0, v_\epsilon)$. There are N informed agents each having CARA utility with risk aversion parameter γ. The noisy supply is denoted by Z, with normally-distributed per capita supply $z \equiv Z/N$: $z \mid d \sim \mathcal{N}(\mu_z, v_z)$.

Each agent's demand is:

$$X_i = \frac{\mathbb{E}(\delta \mid \delta + \epsilon_i, P, d) - P}{\gamma \mathbb{V}\text{ar}(\delta \mid \delta + \epsilon_i, P, d)}.$$

The conditional expectation in the numerator can be written as:

$$\mathbb{E}(\delta \mid \delta + \epsilon_i, P, d) = a_0 + a_\delta (\delta + \epsilon_i) + a_P P.$$

Denote the conditional variance in the denominator, $\mathbb{V}\text{ar}(\delta \mid \delta + \epsilon_i, P, d)$, as v. Conjecture that the price P can be written as:

$$P = k_0 + k_\delta \frac{\sum_i (\delta + \epsilon_i)}{N} + k_z \left(\frac{Z}{N}\right).$$

We can write the market clearing condition as:

$$\sum_i \frac{a_0 + a_\delta (\delta + \epsilon_i) + a_P P - P}{\gamma v} = Z.$$

Solving for price we get:

$$P = \frac{a_0}{1 - a_P} + \frac{a_\delta}{1 - a_P}\left(\delta + \frac{1}{N}\sum_i \epsilon_i\right) - \frac{\gamma v}{1 - a_P}\frac{Z}{N},$$

which implies

$$k_0 = \frac{a_0}{1 - a_P}, \quad k_\delta = \frac{a_\delta}{1 - a_P}, \quad \text{and} \quad k_z = -\frac{\gamma v}{1 - a_P}. \tag{A.1}$$

Since ϵ_i's are iid and have zero mean, by the law of large numbers in a large market (as $N \to \infty$), we have:[1]

$$\frac{1}{N}\sum_i \epsilon_i = 0. \tag{A.2}$$

Explicitly calculating the conditional expectation to get a_0, a_δ, and a_P and substituting in (A.1), we have

$$\begin{aligned}
k_0 &= \frac{\gamma v_\epsilon \left(d\gamma v_z v_\epsilon + v_\delta \mu_z\right)}{\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1 + \gamma^2 v_z v_\epsilon\right)}, \\
k_\delta &= \frac{v_\delta \left(1 + \gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1 + \gamma^2 v_z v_\epsilon\right)}, \\
k_z &= -\frac{\gamma v_\delta v_\epsilon \left(1 + \gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1 + \gamma^2 v_z v_\epsilon\right)}.
\end{aligned} \tag{A.3}$$

We can then calculate a_0, a_δ, and a_P by substituting in Equation (A.1).

The price is $P = k_0 + k_\delta \delta + k_z z$, which completes the solution of the equilibrium. The X_i's can be backed out given a_0, a_δ, and a_P, v, and P.

Demand elasticity. To derive the demand elasticity, we take the derivative of the demand with respect to the equilibrium price:

$$\frac{\partial X_i}{\partial P} = -\frac{1 - a_P}{\gamma v} = \frac{1}{k_z}.$$

From Equation (A.3),

$$\frac{\partial X_i}{\partial P} = -\frac{1}{\gamma v_\epsilon} - \frac{\gamma v_z v_\epsilon}{v_\delta + \gamma^2 v_z v_\delta v_\epsilon}.$$

[1] Alternatively, one can assume instead that there is a unit mass of investors and replace $\frac{1}{N}\sum_i \epsilon_i \approx 0$ with $\int_i \epsilon_i = 0$ (e.g., Van Nieuwerburgh and Veldkamp, 2010).

If we assume the variance of per capita noisy supply, v_z, is small, we get

$$\frac{\partial X_i}{\partial P} \approx -\frac{1}{\gamma v_\epsilon}. \tag{A.4}$$

Thus the demand in this case is more inelastic when the variance of signals is high.

How would this case compare to a model without private or public information? In that case, agent i demand would be:

$$X_i = \frac{\mathbb{E}[\delta] - P}{\gamma \mathbb{V}\text{ar}(\delta)} = \frac{\mu - P}{\gamma v_\delta}.$$

So the elasticity can be written as:[2]

$$\frac{\partial X_i}{\partial P} = -\frac{1}{\gamma v_\delta}.$$

Therefore, aggregate demand is more inelastic (approximately) if $v_\epsilon > v_\delta$.

The partial derivative in the elasticity term, $\partial X_i/\partial P$, captures the change in demand *holding all other terms fixed*, in particular, public information. However, empirically, we might be interested in how a change in prices predicts changes in demand *unconditionally*. In other words, we might consider:

$$\eta \equiv -\frac{\text{Cov}(X_i, P)}{\text{Var}(P)}, \tag{A.5}$$

which has the interpretation of a regression slope coefficient. The difference between $-\partial X_i/\partial P$ and η is that the former describes how much a unit change in price affects demand *holding all other terms fixed*, while η describes how much a unit change in price affects demand *without holding other terms fixed*.

To calculate η, note that after plugging in the equilibrium price, demand is:

$$X_i = \frac{\epsilon_i}{\gamma v_\epsilon} + \frac{Z}{N}. \tag{A.6}$$

[2] The elasticity is defined as: $-\frac{\partial \log(X_i)}{\partial \log(P)} = -\frac{P}{X_i}\frac{\partial X_i}{\partial P}$, but we ignore the P/X_i term to simplify the exposition in this section. Clearly, if $\frac{\partial X_i}{\partial P} \approx 0$, then $\frac{\partial \log(X_i)}{\partial \log(P)} \approx 0$.

The variance of price in the denominator of (A.5) is:

$$\mathrm{Var}(P) = v_d + \underbrace{\frac{v_\delta^2 \left(1+\gamma^2 v_z v_\epsilon\right)^2 \left(v_\delta + \gamma^2 v_z v_\epsilon^2\right)}{\left(\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)\right)^2}}_{\equiv V} = v_d + V.$$

Thus from (A.5), we have:[3]

$$\eta = -\frac{k_z v_z}{v_d + V}. \tag{A.7}$$

Importantly, if the variance of public information is large (i.e., as $v_d \to \infty$), we can show that η approaches zero, indicating inelastic demand. This means that economically, if public information is very volatile but not controlled for appropriately in an elasticity regression estimation, demand will appear quite inelastic.

We emphasize that although prices change with public information d, from Equation (A.6), investors do *not* adjust their demand as public information arrives. Thus when prices move in response to public information, demand will appear very inelastic. As shown in Figure A.1, consistent with the model, demand becomes more elastic as the public information becomes more precise, moving from the solid to the dashed line.

Again consider the case where v_z is small. In this case, we can show that

$$\lim_{v_z \to 0} \eta = 0, \tag{A.8}$$

that is, if per capita noisy supply term is small, then total variation in prices does not predict changes in demand at all. In other words, this naive approach to elasticity estimation would uncover a perfectly inelastic demand.

This section, with only one type of informed agent, nicely illustrates how private information can generate inelastic demand *in aggregate*. However, the heterogeneity in demand elasticities is partly driven by varying degrees of asymmetric information among agents. In the next section, we show that a model of heterogeneous signal quality can generate a dispersion of elasticity terms across investors.

In Figure A.1, we show the impact of private and public information on demand elasticity. As mentioned above, demand becomes more elastic as the public signal becomes more precise, moving

[3]In Internet Appendix IA.1, we write out the general expression for η and calculate the two limits discussed below.

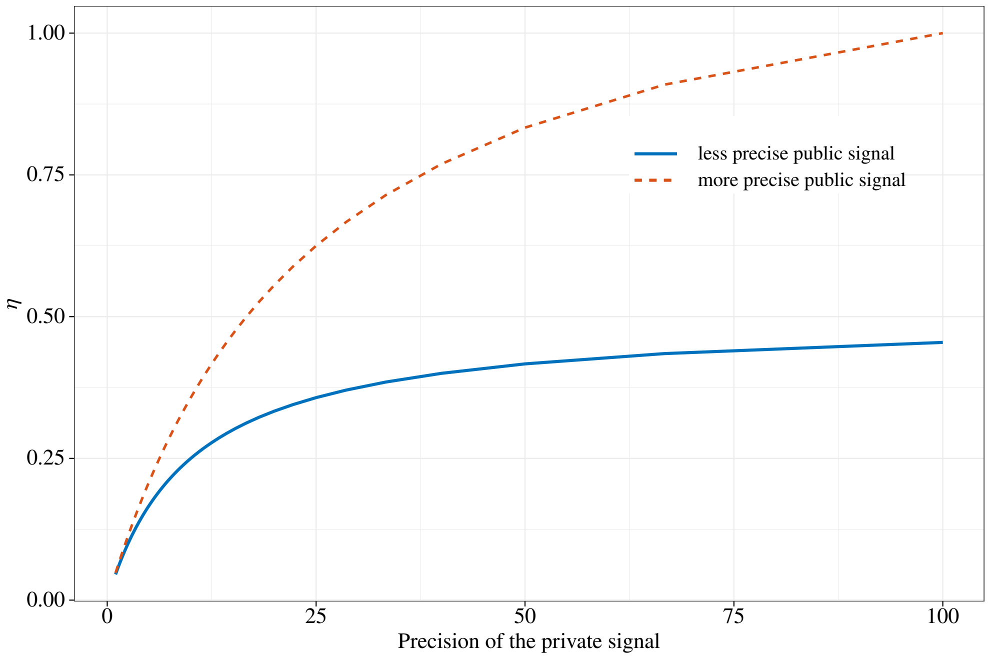

Figure A.1. Information and demand elasticity. This figure qualitatively shows that demand becomes more elastic as the public signal becomes more precise. Moreover, given the precision of the public signal, demand becomes less elastic as investors become less informed.

from the solid to the dashed line. Moreover, given the precision of the public signal (on the solid or dashed lines) demand becomes less elastic as investors become less privately informed, which we show in the next section.

A.2 Heterogeneous quality of information

In this section, we introduce heterogeneity in the degree of private signal quality, which creates investors that are more/less informed. We argue that less informed investors have a lower price elasticity. We show that, with certain parameter values, demand of fully uninformed investors can be perfectly inelastic. Thus theoretically, heterogeneity in information quality can rationalize both aggregate demand being relatively inelastic and the heterogeneity in demand elasticity across investors.

The intuition for why uninformed investors have lower demand elasticity is straightforward. These investors know that other investors receive helpful signals, while they receive relatively only uninformative information. Knowing this disadvantage makes uninformed investors reluctant to trade against or with price movements.

To make this point concrete, consider an *uninformed* fund, Fund A.[4] Suppose the fund observes the price of Stock 1 move from $100 to $110. It could sell some of Stock 1 because of this price increase. However, Fund A realizes that the price movement was likely driven by other investors receiving a positive signal about Stock 1. So, it is rationally reluctant to trade against other informed investors and will not sell any of Stock 1. Fund A could *buy* some shares of Stock 1, believing that the price increase was due to other funds' favorable private information about the asset. This would mean that Fund A has *upward-sloping* demand for Stock 1. However, in the model, Fund A rationally realizes that the market has appropriately priced Stock 1, setting the price such that the expected return on the asset is the same as before the good news is received. Since the expected return is the same, there is no reason for Fund A to buy more of Stock 1. So, the uninformed Fund A does not change its holding in Stock 1 in response to the price increase, implying a perfectly inelastic demand for the stock.

Now consider an *informed* fund, Fund B. Suppose it receives a noisy signal that Stock 1 has more promising future cash flows than realized by the market and is thus undervalued at the price of $100. Fund B, along with other informed funds, trades on this positive news, pushing the price up to, say $110. Suppose that if the price moves up further (to say, $115), Fund B believes, given its signal, that Stock 1 will be overvalued and sell all of its holdings. Thus a 4.5% price increase leads to a 100% decrease in Fund B's position, implying an elasticity of 22.2 (= 100/4.5). Therefore, more informed Fund B is relatively more elastic than the uninformed Fund A.

In summary, relatively uninformed funds are more price inelastic because they are reluctant to trade against or with informed investors. However, more informed funds rationally believe their signals and trade aggressively against price movements, resulting in higher demand elasticities. Below, we formally show that uninformed investors are more price inelastic.[5]

This insight is closely related to the no-trade theorem in Milgrom and Stokey (1982). As we show below, uninformed investors are unwilling to trade no matter the price when all noise trades are eliminated from the model. In other words, they behave perfectly inelastically. However, in practice, some noise remains. This means, from the perspective of an uninformed trader, that some

[4]Fund A may be uninformed for several reasons. It could be because it focuses on providing low fees, or it may be have other objectives such as marketing, environmental investing, etc.

[5]Mathematically, this model is very similar to the one in Section A.1, and thus most details are left in Internet Appendix IA.2.

of the price movements are driven by noise and not better informed traders. Price movements driven by noise can be profitably traded on. Our VAR model below quantifies how much of the variation in prices can be profitably traded on for a rational uninformed investor.

In this model, the signal has the form $s_i = \delta + \lambda_i \epsilon_i$, where λ_i determines agent i's signal quality. The agent knows his signal quality λ_i, but $\epsilon_i \sim \mathcal{N}(0, v_\epsilon)$ are iid across investors. Now $\mathbb{V}\text{ar}(s_i) = v_\delta + \lambda_i^2 v_\epsilon$. Thus if $\lambda_i \to \infty$, investor i is essentially uninformed. We will consider this case carefully below.

Like the model in Section A.1, we conjecture that an equilibrium exists such that the price has the form:

$$P = k_0 + k_\delta \delta + k_z \frac{Z}{N}.$$

In Appendix IA.2, we verify this conjecture and provide closed-form expressions for k_0, k_δ, and k_z. We assume that the signals, asset's payoff, and price have the following multivariate normal distribution, consistent with this conjecture:

$$\begin{bmatrix} \delta \\ s_i \\ P \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu \\ k_0 + k_\delta \mu + k_z \mu_z \end{bmatrix}, \begin{bmatrix} v_\delta & v_\delta & k_\delta v_\delta \\ v_\delta & v_\delta + \lambda_i^2 v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix} \right).$$

As shown in Internet Appendix IA.2, CARA utility demand is

$$\begin{aligned} X_i &= \frac{\mathbb{E}\left[\delta \mid s_i, P\right] - P}{\gamma \mathbb{V}\text{ar}(\delta \mid s_i, P)} \\ &= \frac{1}{\gamma}\left[P\left(-\frac{(-1 + k_\delta)\, k_\delta}{k_z^2 v_z} - \frac{1}{v_\delta} - \frac{1}{\lambda_i^2 v_\epsilon} \right) + \frac{\delta + \epsilon_i}{\lambda_i^2 v_\epsilon} + \left(-\frac{k_0 k_\delta}{k_z^2 v_z} + \frac{\mu}{v_\delta} - \frac{k_\delta \mu_z}{k_z v_z} \right) \right]. \end{aligned}$$

Now consider agents with heterogeneous signal qualities. Let N_L investors have $\lambda_i = \lambda_L$ signals, which are relatively *low* and *precise*. Let $N_H = N - N_L$ investors have $\lambda_i = \lambda_H$ signals, which are relatively *high* and *imprecise*. The investors with $\lambda_i = \lambda_H$ are relatively uninformed. Let $n_L = N_L/N$ and $n_H = N_H/N$. Let H and L be the sets of investors that have $\lambda_i = \lambda_H$ and $\lambda_i = \lambda_L$ respectively.

As is common in signaling models, we assume that epsilons averages to zero across investors:

$$\frac{1}{N_L} \sum_{i \in L} \epsilon_i = \frac{1}{N_H} \sum_{i \in H} \epsilon_i = 0.$$

From the market-clearing condition, we have:

$$\frac{1}{N}\left(\sum_{i\in L} X_i + \sum_{i\in H} X_i\right) = \frac{Z}{N}.$$

We now turn to examine the elasticity of investors in this model. As shown in Internet Appendix IA.2,

$$\begin{aligned}\frac{\partial X_i}{\partial P} &= -\gamma^{-1}\left(\frac{(-1+k_\delta)\,k_\delta}{k_z^2 v_z} + \frac{1}{v_\delta} + \frac{1}{\lambda_i^2 v_\epsilon}\right)\\ &= -\frac{n_H^2 v_\delta \lambda_L^2 + n_H n_L v_\delta\left(\lambda_H^2+\lambda_L^2\right) + \lambda_H^2\left(n_L^2 v_\delta + \gamma^2 v_z v_\epsilon\left(v_\delta + \lambda_i^2 v_\epsilon\right)\lambda_L^2\right)}{\gamma \lambda_i^2 v_\delta v_\epsilon\left(n_H^2\lambda_L^2 + n_H n_L\left(\lambda_H^2+\lambda_L^2\right) + \lambda_H^2\left(n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)\right)}.\end{aligned} \tag{A.9}$$

Similar to the model in Section A.1, if the variance of the per capita noisy supply, v_z, is low, we immediately get:

$$\lim_{v_z\to 0}\frac{\partial X_i}{\partial P} = -\frac{1}{\gamma\lambda_i^2 v_\epsilon}. \tag{A.10}$$

Thus, if we consider a relatively less informed investor with $\lambda_i = \lambda_H$, and λ_H is very large (i.e., the investor is essentially uninformed), then

$$\frac{\partial X_i}{\partial P} \approx 0.$$

This confirms the intuition given above that uninformed investors are relatively inelastic. We focus in particular on the elasticity of uninformed investors in our empirical section. It is not necessary to assume $v_z \approx 0$ to get this result. We can show that

$$\lim_{\lambda_H\to\infty}\frac{\partial X_H}{\partial P} = -\frac{\gamma v_z v_\epsilon \lambda_L^2}{v_\delta\left(n_H n_L + n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)}. \tag{A.11}$$

Thus if either v_z, v_ϵ, or λ_L are close to zero, then $\frac{\partial X_i}{\partial P} \approx 0$.

If v_ϵ or λ_L are zero, this means that the informed investors are *perfectly* informed. Thus for the uninformed investors, trading against these price movements means trading against perfectly informed investors. The uninformed investors realize that this is a losing gambit and rationally become perfectly inelastic.

Likewise, when v_z is close to zero, this means that all price movements from the uninformed investors' point of view do not change expected returns. In other words, if v_z is large, this means that much of the price variation is driven by noise and can be profitably traded on. When v_z is close to zero, all price variation is driven by informed investors and thus uninformed investors are reluctant to trade against price movements.

One of the fundamental insights from asset pricing is that price movements *must* either predict changes in expected returns or cash flows (Campbell and Shiller, 1988; Cochrane, 2008). As we discussed above, from the perspective of an uninformed investor in this inelastic equilibrium, price movements fail to predict changes in next period expected returns. Using an identity very similar to the classic Campbell-Shiller decomposition, we elucidate the relationship between dividend growth and expected returns fluctuations predicted by changes in today's price from the point of view of an uninformed investor.

In the next section, we estimate the demand elasticity for an uninformed investor in a VAR using the characteristic-based demand system of KY.

B 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:

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

where τ_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

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

Using the block diagonal matrix formula, note that:

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

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 θ_N we have:

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

$$= \frac{1}{\mu_N - \Sigma'_{-N} \Sigma^{-1}_{-N,-N} \mu_{-N}} \tag{B.5}$$

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

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

Thus

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

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.

C Demand Elasticity in a Standard Asset Pricing Model

To fix ideas, consider a static partial equilibrium model.[6] Suppose there are N assets, indexed by n, each with supply u_n. Assume the risk-free rate is constant, normalized to 0. Dividends for stock n is assumed to have the following form:

$$D_n = a_n + b_n F + e_n,$$

where $F \sim \mathcal{N}\left(0, \sigma_m^2\right)$ is the common factor and $e_n \sim \mathcal{N}\left(0, \sigma_e^2\right)$ represents the idiosyncratic risk.

[6]This simple model and its standard calibration is from Section II.A. of Petajisto (2009). It was also discussed during the Workshop on Demand System Asset Pricing organized by Ralph Koijen and Motohiro Yogo in May, 2022.

There is a representative investor with constant absolute risk aversion (CARA) preferences, with wealth W and risk aversion γ who chooses portfolio weights for stocks $n = 1, \ldots, N$ to maximize her utility subject to the budget constraint:

$$\max_{\alpha_1,\ldots,\alpha_N} \mathbb{E}\left[-\exp(-\gamma W)\right],$$

$$\text{subject to} \quad W = W_0 + \sum_{n=1}^{N} \alpha_n \left(D_n - P_n\right),$$

where P_n is the price of stock n. From the first-order condition for stock n and market-clearing, we have:

$$P_n = a_n - \gamma \left[\sigma_m^2 \left(\sum_{m \neq n} u_m b_m\right) b_n + \left(\sigma_m^2 b_n^2 + \sigma_e^2\right) u_n\right]$$

Consider the following calibration. Suppose there are $N = 1000$ stocks, each with unit supply $u_n = 1$. Also let $a_n = 105$, $b_n = 100$, $\sigma_m^2 = 0.04$, $\sigma_e^2 = 900$, and $\gamma = 1.25 \times 10^{-5}$. These parameters imply a market risk premium of 5%, all stocks having a price of 100, a market beta of 1, and a standard deviation of idiosyncratic return of 30%.

Consider a supply shock of -10% ($u_n = 0.9$) for one stock. This leads to a price increase of only 0.1621 bps. Part of this increase is due to the reduction in the aggregate market risk premium (there is less aggregate risk and all stocks increase by 0.05 bps.) So the differential impact is only 0.11 bps, meaning the demand curve is virtually flat. In this setting, the micro price elasticity of demand is very large:[7]

$$-\frac{\Delta Q/Q}{\Delta P/P} = \frac{0.10}{1.621e-5} \approx 6168,$$

implying a negligible *micro* multiplier (the inverse of micro demand elasticity). Thus, in standard asset pricing models demand curves are virtually flat.

Macro multipliers implied from frictionless asset pricing models are usually quite small as well.[8] As discussed in Gabaix and Koijen (2021), "in traditional, elastic asset pricing models the macro elasticity is around 10 to 20."[9]

[7] In a general equilibrium setting, Johnson (2006) perturbs the risky asset supply and finds finite *macro* elasticity even in the frictionless Lucas economy.

[8] One notable exception is the general equilibrium model in Johnson (2006). He studies the equilibrium price change in response to a perturbation in the risky asset supply, allowing for the interest rate to vary when stock prices change.

[9] Appendix F.4 in Gabaix and Koijen (2021) provides a detailed discussion.

Empirically, demand curves are surprisingly inelastic compared to standard models both at the micro (Koijen and Yogo, 2019) and macro (Gabaix and Koijen, 2021) levels.[10] As mentioned above, estimates of micro and macro demand elasticities, around 1 and 0.2 respectively, are much lower than what standard frictionless theories suggest.

D Idiosyncratic Volatility and Elasticity

One may wonder why idiosyncratic volatility does not directly enter the elasticity formula given in Equation (21). This might initially 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:

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

where β is just a vector of single-factor loadings, Ω is the variance of the factor, and Σ_ϵ 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.

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.2, we plot the return of asset A minus the return of asset

[10] Li and Lin (2022) show that prices are more inelastic when demand is less diversifiable.

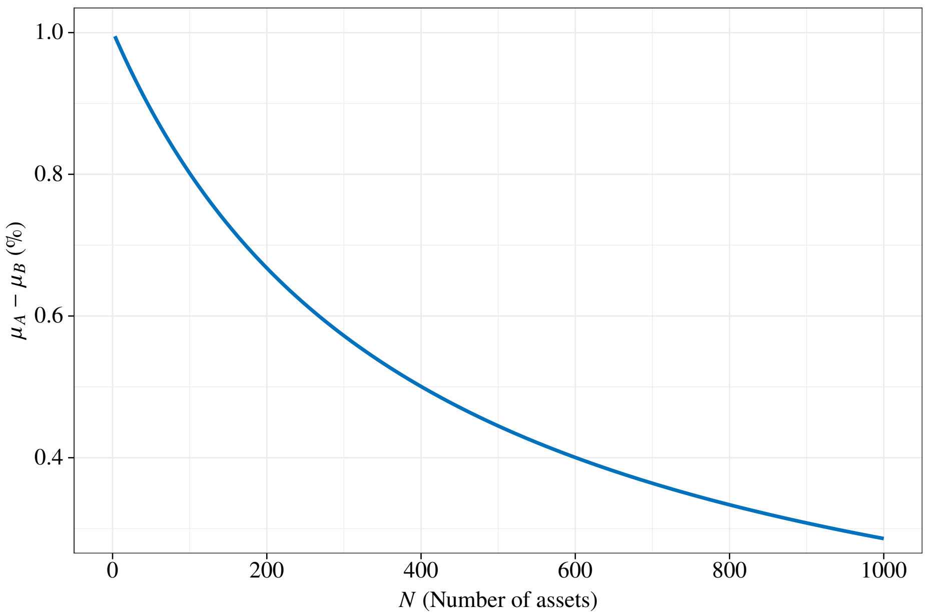

Figure A.2. Expected Returns Under Idiosyncratic Volatility Thought-Experiment
This figures show the return of asset A minus the return on asset B, as a function of the number of assets. Note that there is only ever one asset A and one asset B, and changing the number of assets just changes the number of "other assets."

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?

It turns out that idiosyncratic volatility still matters, even with many assets. In Figure A.3, 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

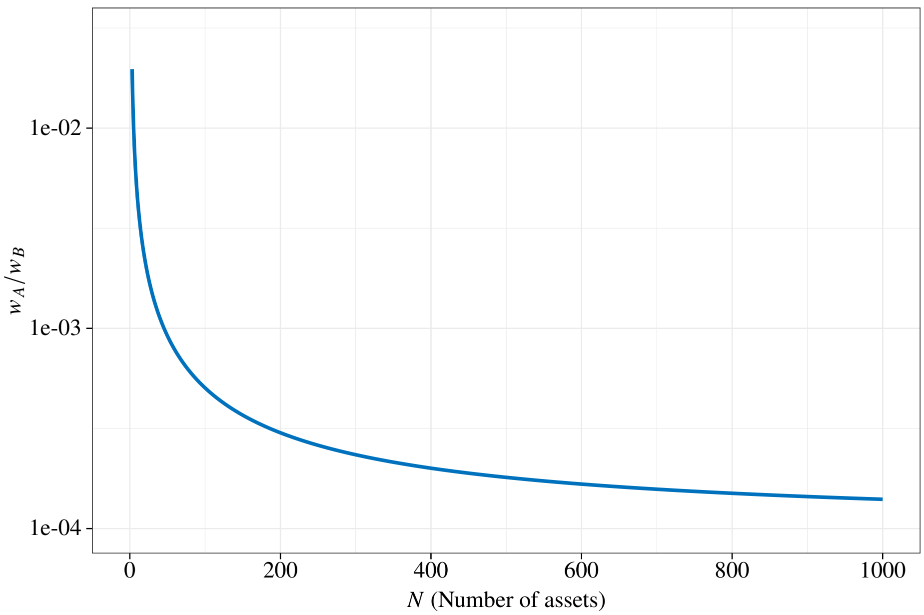

Figure A.3. Ratio of Portfolio Weights Under Idiosyncratic Volatility Thought-Experiment
This figures the optimal portfolio weight of asset A divided by the optimal portfolio weight of asset B, as a function of the number of assets. Note that there is only ever one asset A and one asset B, and changing the number of assets just changes the number of "other assets."

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 the identical elasticity terms.

E Elasticity Estimates in Bond Markets

As discussed earlier, it is very challenging to assume perfect pass-through and estimate D1 in the stock market, but there are existing estimates that we would argue are close to estimating D1 in the bond market. Recall that estimating D1 requires price shocks that are sure to revert in the short term, i.e., $\psi \approx 1$. This condition is only satisfied by bonds with high credit quality and very short maturity.[11] For instance, consider U.S. Treasury bills that will mature within a year. Because U.S. Treasuries are considered risk-free, any price dislocations must revert by the time of maturity.

Similar logic can also apply, to a large extent, to very highly rated corporate bonds with short maturity. Sufficiently highly rated bonds have a close-to-zero chance of defaulting within a few months. For instance, S&P's 2021 Annual Global Corporate Default And Rating Transition Study shows that, while bonds rated at A- may default in the long run, the probability that they default within a year is only 0.07%. Bonds rated higher have even lower default probabilities.[12]Since 1980, no corporate bond rated AA+ or AAA has ever defaulted within a year. Therefore, it is reasonable to think that investors would largely recognize price movements in these bonds as reflecting short-term discount rate changes that are sure to revert quickly. The same, of course, is not true for lower-rated bonds for which price movements may be associated with higher default probabilities.

As we discussed above, D1 elasticity can be large. Taking the reciprocal of that, price multipliers in these securities should be very small. Extant estimates are consistent with this and we summarize them briefly below.

1. **Short-term U.S. Treasury bonds.** Lou, Yan, and Zhang (2013) document that yields of two-, five-, and ten-year U.S. Treasury bonds rise temporarily around U.S. Treasury bond auctions. The price impacts in their study imply demand elasticities in the range of 10 to 30. However, they do not find measurable price impacts for shorter-term Treasuries, indicating that demand elasticities in those securities are much higher than 30.

[11]The pass-throughs for Treasury and liquid investment-grade bonds are propositional to the bonds' duration, consistent with results in Li et al. (2022).

[12]See Table 9 in 2021 Annual Global Corporate Default And Rating Transition Study, available here.

2. **Short-term developed country government bonds.** Koijen and Yogo (2020) examine cross-country demand for stocks and government bonds using a structural framework. For short-term government bond debt, they estimate demand elasticities of 42.

3. **Highly rated short-term corporate bonds.** Using mutual fund flow-induced trading as instruments, Li et al. (2022) find close to zero price impact of trading in short-term investment grade bonds, which is consistent with a very high D1.

Overall, existing results indicate that in settings closer to truly estimating D1, the results all indicate significantly larger demand elasticities than that estimated in KY. However, it's important to approach this comparison with caution, as these estimates do not pertain to the stock market.

We now turn back to the stock market, and proceed to fit a structural model of returns in the cross-section, and show that a structural model captures both low pass-throughs and high asset inimitability, delivering relatively low elasticity values even in a friction-free environment. While trading costs, information asymmetry, and other frictions can reduce the demand elasticity further, we show that the low pass-throughs and high unspanned returns are first order to the explanation of inelastic demand. Before describing the structural model and estimation results, we describe the data.

F Shrinkage Portfolios: Full Results

Table A.1 presents the full results of the exercise in Section 4.2. We consider a large range of shrinkage weights, $h = 0.05, 0.1, \ldots, 0.95$. 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. Shrinkage Portfolios 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

Internet Appendix

IA.1 Homogeneous Signal Quality Model with Public Information

To reiterate the main text, there is a signal $s_i = \delta + \epsilon_i$. The agent knows his signal quality λ_i, but $\epsilon_i \sim \mathcal{N}(0, v_\epsilon)$ are iid across investors.

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

$$P = k_0 + k_\delta \delta + k_z \frac{Z}{N}.$$

Define

$$\begin{bmatrix} \delta \\ s_i \\ P \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} d \\ d \\ k_0 + k_\delta d + k_z \mu_z \end{bmatrix}, \begin{bmatrix} v_\delta & v_\delta & k_\delta v_\delta \\ v_\delta & v_\delta + v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix} \right)$$

Thus

$$\begin{aligned} \mathbb{E}[\delta \mid s_i, P] &= d + \begin{bmatrix} v_\delta & k_\delta v_\delta \end{bmatrix} \begin{bmatrix} v_\delta + v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix}^{-1} \begin{bmatrix} \delta + \epsilon_i - d \\ P - k_0 - k_\delta d - k_z \mu_z \end{bmatrix} \\ &= \frac{(p - k_0)\, k_\delta v_\delta v_\epsilon + k_z^2 v_z \left((\delta + \epsilon_i) v_\delta + d v_\epsilon \right) - k_z k_\delta v_\delta v_\epsilon \mu_z}{k_\delta^2 v_\delta v_\epsilon + k_z^2 v_z \left(v_\delta + v_\epsilon \right)} \\ \mathbb{V}\mathrm{ar}(\delta \mid s_i, P) &= v_\delta - \begin{bmatrix} v_\delta & k_\delta v_\delta \end{bmatrix} \begin{bmatrix} v_\delta + v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix}^{-1} \begin{bmatrix} v_\delta \\ k_\delta v_\delta \end{bmatrix} \\ &= \frac{k_z^2 v_z v_\delta v_\epsilon}{k_\delta^2 v_\delta v_\epsilon + k_z^2 v_z \left(v_\delta + v_\epsilon \right)} \end{aligned}$$

So

$$\delta \mid s_i, P \sim \mathcal{N}\left(\mathbb{E}\left[\delta \mid s_i, P \right], \mathbb{V}\mathrm{ar}(\delta \mid s_i, P) \right)$$

The CARA demand is:

$$
\begin{aligned}
X_i &= \frac{\mathbb{E}[\delta \mid s_i, P] - P}{\gamma \mathbb{V}\mathrm{ar}(\delta \mid s_i, P)} \\
&= \frac{P\left(-\frac{(-1+k_\delta)k_\delta}{k_z^2 v_z} - \frac{1}{v_\delta} - \frac{1}{v_\epsilon}\right)}{\gamma} + \frac{\delta}{\gamma v_\epsilon} + \frac{\epsilon_i}{\gamma v_\epsilon} + \frac{-\frac{k_0 k_\delta}{k_z^2 v_z} + \frac{d}{v_\delta} - \frac{k_\delta \mu_z}{k_z v_z}}{\gamma}.
\end{aligned}
$$

We can write average demand as

$$
\frac{1}{N}\sum_i X_i = b_0 + b_p p + b_\delta \delta
$$

where

$$
\begin{aligned}
b_0 &= \frac{-\frac{k_0 k_\delta}{k_z^2 v_z} + \frac{d}{v_\delta} - \frac{k_\delta \mu_z}{k_z v_z}}{\gamma} \\
b_p &= \frac{\left(-\frac{(-1+k_\delta)k_\delta}{k_z^2 v_z} - \frac{1}{v_\delta} - \frac{1}{v_\epsilon}\right)}{\gamma} \\
b_\delta &= \frac{1}{\gamma v_\epsilon}
\end{aligned}
$$

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

$$
\begin{aligned}
k_0 &= -\frac{b_0}{b_p} \\
k_\delta &= -\frac{b_\delta}{b_p} \\
k_z &= \frac{1}{b_p}
\end{aligned}
$$

Solving this system of equations yields

$$k_0 = \frac{\gamma v_\epsilon \left(d\gamma v_z v_\epsilon + v_\delta \mu_z\right)}{\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)},$$
$$k_\delta = \frac{v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)},$$
$$k_z = -\frac{\gamma v_\delta v_\epsilon \left(1+\gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)}. \tag{IA.1.1}$$

We can calculate η as shown in the text. If we calculate it out fully, we have, once we plug in all the constants:

$$\eta = \frac{\gamma v_z v_\delta v_\epsilon \left(1+\gamma^2 v_z v_\epsilon\right)\left(\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)\right)}{v_\delta^2 \left(1+\gamma^2 v_z v_\epsilon\right)^2 \left(v_\delta + \gamma^2 v_z v_\epsilon^2\right) + v_d \left(\gamma^2 v_z v_\epsilon^2 + v_\delta \left(1+\gamma^2 v_z v_\epsilon\right)\right)^2} \tag{IA.1.2}$$

thus it's clear that

$$\lim_{v_d \to \infty} \eta = 0 \quad \text{and} \quad \lim_{v_z \to 0} \eta = 0. \tag{IA.1.3}$$

IA.2 Heterogeneous Signal Quality

To reiterate the main text, there is a signal $s_i = \delta + \lambda_i \epsilon_i$. The agent knows his signal quality λ_i, but $\epsilon_i \sim \mathcal{N}(0, v_\epsilon)$ are iid across investors. Now $\mathbb{V}\text{ar}(s_i) = v_\delta + \lambda_i^2 v_\epsilon$. Thus if λ_i is quite large, then investor i is essentially uninformed.

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

$$P = k_0 + k_\delta \delta + k_z \frac{Z}{N}.$$

Define

$$\begin{bmatrix} \delta \\ s_i \\ P \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu \\ k_0 + k_\delta \mu + k_z \mu_z \end{bmatrix}, \begin{bmatrix} v_\delta & v_\delta & k_\delta v_\delta \\ v_\delta & v_\delta + \lambda_i^2 v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix}\right)$$

Thus

$$
\begin{aligned}
\mathbb{E}[\delta \mid s_i, P] &= \mu + \begin{bmatrix} v_\delta & k_\delta v_\delta \end{bmatrix} \begin{bmatrix} v_\delta + \lambda_i^2 v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix}^{-1} \begin{bmatrix} \delta + \lambda_i \epsilon_i - \mu \\ P - k_0 - k_\delta \mu - k_z \mu_z \end{bmatrix} \\
&= \frac{\lambda_i^2 \left(p - k_0\right) k_\delta v_\delta v_\epsilon + k_z^2 v_z \left((\delta + \epsilon_i) v_\delta + \lambda_i^2 \mu v_\epsilon\right) - \lambda_i^2 k_z k_\delta v_\delta v_\epsilon \mu_z}{\lambda_i^2 k_\delta^2 v_\delta v_\epsilon + k_z^2 v_z \left(v_\delta + \lambda_i^2 v_\epsilon\right)} \\
\mathbb{V}\text{ar}(\delta \mid s_i, P) &= v_\delta - \begin{bmatrix} v_\delta & k_\delta v_\delta \end{bmatrix} \begin{bmatrix} v_\delta + \lambda_i^2 v_\epsilon & k_\delta v_\delta \\ k_\delta v_\delta & k_\delta^2 v_\delta + k_z^2 v_z \end{bmatrix}^{-1} \begin{bmatrix} v_\delta \\ k_\delta v_\delta \end{bmatrix} \\
&= \frac{\lambda_i^2 k_z^2 v_z v_\delta v_\epsilon}{\lambda_i^2 k_\delta^2 v_\delta v_\epsilon + k_z^2 v_z \left(v_\delta + \lambda_i^2 v_\epsilon\right)}
\end{aligned}
$$

So

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

The CARA demand is:

$$
\begin{aligned}
X_i &= \frac{\mathbb{E}[\delta \mid s_i, P] - P}{\gamma \mathbb{V}\text{ar}(\delta \mid s_i, P)} \\
&= \frac{P\left(-\frac{(-1+k_\delta)k_\delta}{k_z^2 v_z} - \frac{1}{v_\delta} - \frac{1}{\lambda_i^2 v_\epsilon}\right)}{\gamma} + \frac{\delta}{\gamma \lambda_i^2 v_\epsilon} + \frac{\epsilon_i}{\gamma \lambda_i^2 v_\epsilon} + \frac{-\frac{k_0 k_\delta}{k_z^2 v_z} + \frac{\mu}{v_\delta} - \frac{k_\delta \mu_z}{k_z v_z}}{\gamma}.
\end{aligned}
$$

Again, reiterating the main text, consider heterogeneous signal quality. Let N_L investors have $\lambda_i = \lambda_L$ signals, which are relatively *low* and *precise*. Let $N_H = N - N_L$ investors have $\lambda_i = \lambda_H$ signals, which are relatively *high* and *imprecise*. The investors with $\lambda_i = \lambda_H$ are relatively uninformed. Let $n_L = N_L/N$ and $n_H = N_H/N$. Let H and L be the sets of investors that have $\lambda_i = \lambda_H$ and $\lambda_i = \lambda_L$ respectively.

Then assume both

$$
\frac{1}{N_L} \sum_{i \in L} \epsilon_i = \frac{1}{N_H} \sum_{i \in H} \epsilon_i = 0
$$

Equilibrium is of course

$$
\frac{1}{N} \left(\sum_{i \in L} X_i + \sum_{i \in H} X_i \right) = \frac{Z}{N}
$$

We can write demand as

$$\frac{1}{N}\left(\sum_{i\in L} X_i + \sum_{i\in H} X_i\right) = b_0 + b_p p + b_\delta \delta$$

where

$$\begin{aligned}
b_0 &= \frac{(n_H + n_L)\left(\mu k_z^2 v_z - k_0 k_\delta v_\delta - k_z k_\delta v_\delta \mu_z\right)}{\gamma k_z^2 v_z v_\delta} \\
b_p &= \gamma^{-1}\left[-\frac{(-1 + k_\delta)\, k_\delta\, (n_H + n_L)}{k_z^2 v_z} + n_H\left(-\frac{1}{v_\delta} - \frac{1}{v_\epsilon \lambda_H^2}\right) + n_L\left(-\frac{1}{v_\delta} - \frac{1}{v_\epsilon \lambda_L^2}\right)\right] \\
b_\delta &= \frac{\frac{n_H}{\lambda_H^2} + \frac{n_L}{\lambda_L^2}}{\gamma v_\epsilon}
\end{aligned}$$

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

$$\begin{aligned}
k_0 &= -\frac{b_0}{b_p} \\
k_\delta &= -\frac{b_\delta}{b_p} \\
k_z &= \frac{1}{b_p}
\end{aligned}$$

The solution to these equations are:

$$\begin{aligned}
k_0 &= \frac{\gamma\,(n_H + n_L)\, v_\epsilon \lambda_H^2 \lambda_L^2 \left(\gamma \mu v_z v_\epsilon \lambda_H^2 \lambda_L^2 + v_\delta \left(n_L \lambda_H^2 + n_H \lambda_L^2\right) \mu_z\right)}{\kappa_0} \\
k_\delta &= \frac{v_\delta \left(n_L \lambda_H^2 + n_H \lambda_L^2\right)\left(n_H^2 \lambda_L^2 + n_H n_L \left(\lambda_H^2 + \lambda_L^2\right) + \lambda_H^2 \left(n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)\right)}{\kappa_0} \\
k_z &= -\frac{\gamma v_\delta v_\epsilon \lambda_H^2 \lambda_L^2 \left(n_H^2 \lambda_L^2 + n_H n_L \left(\lambda_H^2 + \lambda_L^2\right) + \lambda_H^2 \left(n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)\right)}{\kappa_0},
\end{aligned}$$

where

$$\begin{aligned}
\kappa_0 \equiv\ & n_H^3 v_\delta \lambda_L^4 + n_H^2 n_L v_\delta \lambda_L^2 \left(2\lambda_H^2 + \lambda_L^2\right) + n_H \lambda_H^2 \left(\gamma^2 v_z v_\epsilon \left(v_\delta + v_\epsilon \lambda_H^2\right) \lambda_L^4 + n_L^2 v_\delta \left(\lambda_H^2 + 2\lambda_L^2\right)\right) \\
&+ n_L \lambda_H^4 \left(n_L^2 v_\delta + \gamma^2 v_z v_\epsilon \lambda_L^2 \left(v_\delta + v_\epsilon \lambda_L^2\right)\right).
\end{aligned}$$

Now, we can compute demand elasticity of investor i:

$$\frac{\partial X_i}{\partial P} = -\gamma^{-1}\left(\frac{(-1+k_\delta)\,k_\delta}{k_z^2 v_z} + \frac{1}{v_\delta} + \frac{1}{\lambda_i^2 v_\epsilon}\right)$$
$$= -\frac{n_H^2 v_\delta \lambda_L^2 + n_H n_L v_\delta \left(\lambda_H^2 + \lambda_L^2\right) + \lambda_H^2 \left(n_L^2 v_\delta + \gamma^2 v_z v_\epsilon \left(v_\delta + \lambda_i^2 v_\epsilon\right)\lambda_L^2\right)}{\gamma \lambda_i^2 v_\delta v_\epsilon \left(n_H^2 \lambda_L^2 + n_H n_L \left(\lambda_H^2 + \lambda_L^2\right) + \lambda_H^2\left(n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)\right)}$$

Like the model in Section A, we want to consider the case where the variance of the per capital noisy supply term is low. In that case, one can easily show that:

$$\lim_{v_z \to 0} \frac{\partial X_i}{\partial P} = -\frac{1}{\gamma \lambda_i^2 v_\epsilon}$$

Thus if we consider an investor where $\lambda_i = \lambda_H$, and λ_H is very large (the investor is essentially uninformed), then

$$\frac{\partial X_i}{\partial P} \approx 0$$

More generally we can show that

$$\frac{\partial X_i}{\partial P} = -\frac{n_H^2 v_\delta \lambda_L^2 + n_H n_L v_\delta \left(\lambda_H^2 + \lambda_L^2\right) + \lambda_H^2 \left(n_L^2 v_\delta + \gamma^2 v_z v_\epsilon \left(v_\delta + v_\epsilon \lambda_H^2\right)\lambda_L^2\right)}{\gamma v_\delta v_\epsilon \lambda_H^2 \left(n_H^2 \lambda_L^2 + n_H n_L \left(\lambda_H^2 + \lambda_L^2\right) + \lambda_H^2\left(n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)\right)}$$

Then

$$\lim_{\lambda_H \to \infty} \frac{\partial X_i}{\partial p} = -\frac{\gamma v_z v_\epsilon \lambda_L^2}{v_\delta \left(n_H n_L + n_L^2 + \gamma^2 v_z v_\epsilon \lambda_L^2\right)}$$

Thus if either v_z, v_ϵ, or λ_L are close to zero, then

$$\frac{\partial X_i}{\partial p} \approx 0.$$