# Why is Asset Demand Inelastic?*

Carter Davis[†] Mahyar Kargar[‡] Jiacui Li[§]

April 5, 2024

**Abstract**

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

**Keywords**: Demand elasticity, price pass-through, spanning, demand-based asset pricing

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

*We thank Yu An, Aditya Chaudhry, Olivier Darmouni, Zhiyu Fu, Andrei Gonçalves, Valentin Haddad, Leyla Han, Christian Heyerdahl Larsen, Da Huang, Kristy Jansen, Tim Johnson, Ralph Koijen, Serhiy Kozak, Daniel Neuhann, Stavros Panageas, Dejanir Silva, and conference and seminar participants at AFA 2024, MFA 2024, University of Maryland Young Scholars Conference, 2022 Wabash River Finance Conference, Indiana University, and University of Utah for valuable feedback. We thank Andrea Tamoni and Michael Weber for sharing data. All errors are our own responsibility.

[†]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 most classical asset pricing models, investor demand curves are virtually flat, implying high security-level demand elasticities in the thousands (e.g., Petajisto, 2009; Gabaix and Koijen, 2021). These models featuring very elastic demand, fail to generate several salient asset pricing facts, such as the price impact of fund flows and retail trades, excess volatility puzzle, etc. In stark contrast to these theories, empirical studies overwhelmingly find that demand elasticities are much smaller, typically lower by three orders of magnitude.[1] This large gap between theoretical predictions and empirical estimates leads to two questions. First, are the empirical estimates unreliable (as suggested by e.g., Fuchs, Fukuda, and Neuhann, 2023)? Second, if these estimates are indeed reliable, why do investors exhibit such low demand elasticities in practice? Does this gap reflect suboptimal investor behavior, severe frictions, or are the theoretical predictions simply unrealistic?[2]

In this paper, we reconcile this large elasticity gap by showing that existing high theoretical predictions hinge on unrealistic assumptions about price processes. To illustrate this, we decompose the optimal investor demand elasticity into two components that solely depend on the data-generating process of prices. We then show that demand elasticity is high if and only if both components are high. Existing theories assume both components are high. However, if we plug in realistic empirical estimates, the theoretical elasticity predictions substantially decrease to values that are closer to empirical estimates. This implies that the low empirical estimates of elasticity are credible and are largely consistent with optimal investor behavior.

Our decomposition is motivated by the simple insight that the demand elasticity for assets

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

[2] Some models generate lower demand elasticities using suboptimal behavior. For example, Hong and Stein (1999) argue that contrarian strategies that lessen the momentum trader effects in their model are unlikely given payoffs to contrarian strategies, thus offering a behavioral account for low demand elasticities.

fundamentally differs from that for consumer goods. In consumer theory, preferences are defined over goods, and demand elasticities are determined by preferences. In finance, however, investor preferences are defined over consumption or wealth, not the assets themselves. Thus, investors are not inherently interested in the assets *per se*, but in the returns generated by these assets. As a result, the elasticity of investor demand for assets should primarily depend on how price changes affect expected returns. Guided by this perspective, we decompose investor demand elasticities for asset $i$ — changes in the log number of shares held ($Q_i$) to changes in the log asset price ($P_i$) — into one plus the product of two components:

$$\eta_i = -\frac{\partial \log(Q_i)}{\partial \log(P_i)} \approx 1 + \overbrace{\underbrace{\frac{\partial \log(w_i)}{\partial \mu_i}}_{\substack{=\frac{1}{\mu_{i,\text{unspanned}}} \text{ for} \\ \text{mean-variance investors}}}}^{\text{weight responsiveness}} \times \overbrace{\left(-\frac{\partial \mu_i}{\partial \log(P_i)}\right)}^{\text{price pass-through}}. \tag{1}$$

The second term in Equation (1), which we refer to as "price pass-through", measures the change in the next-period expected return $\mu_i$ in response to price changes. The first term, labeled "weight responsiveness", measures how investor's log portfolio weight responds to changes in expected returns. As such, our decomposition focuses on understanding demand shifts in response to changes in expected returns. While in principle, demand could also react to price-induced changes in risk or other variables, we show that such effects are second-order in typical portfolio choice settings.[3]

The concept of price pass-through is intuitive and depends on the speed at which price

[3]To study the importance of other drivers of portfolio weights (e.g., risk, hedging, etc.), in Appendix B, we estimate the components of a more general portfolio choice problem with Epstein-Zin preferences, following Campbell, Chan, and Viceira (2003). We decompose the components of demand elasticity that are not related to returns into three parts: a covariance component, a variance component, and a consumption-to-wealth ratio hedging component. We find that these components are small relative to the component associated with changes in expected returns.

movements revert. However, it is less clear what determines weight responsiveness. To offer intuition, we focus on investors who form mean-variance (MV) efficient portfolios — not because it realistically describes all investors, but rather to gain insight into the predictions while maintaining optimal behavior. We show that, in the MV setting, weight responsiveness is inversely related to the size of "unspanned (expected excess) returns", $\mu_{i,\text{unspanned}} = \mu_i - \beta'_{-i}\mu_{-i}$, where $\beta'_{-i}$ is the vector of regression coefficients of asset $i$'s return on all assets other than $i$, and $\mu_{-i}$ is the vector of expected returns for all other assets. $\beta'_{-i}\mu_{-i}$ is the expected return of asset $i$'s *replication portfolio* using other assets, measuring the degree to which asset $i$ is *substitutable*. This factor is crucial for understanding demand elasticity. When asset $i$ has nearly perfect substitutes, if its expected return increases, an MV investor could identify a near-arbitrage opportunity: it can be captured by significantly increasing their position in asset $i$ and shorting the replicating portfolio. Conversely, if asset $i$ lacks close substitutes, the investor portfolio response would be subdued.

Guided by our decomposition framework, we show that using empirical estimates for price pass-throughs and unspanned returns explains most of the gap between theory and estimated demand elasticities (Figure 1). Classic models assume a high price pass-through of approximately one (all price dislocations revert in one period) and that stocks are almost perfect substitutes. As a consequence, they predict stock-level demand elasticities of approximately 7,000, as illustrated in the leftmost bar. If we use empirical estimates of price pass-through which is around 0.03, demand elasticity drops to approximately 211. We empirically estimate that stock-level unspanned expected returns to be around 0.23% per month, as opposed to around 0.01% in many theoretical models. Accounting for this large unspanned return leads to elasticity estimates of around 14. These two steps are illustrated in the next two bars in Figure 1. Therefore, without appealing to biases or frictions, a significant part of the demand elasticity gap is accounted for.

To estimate the two components of demand elasticity, we need to build models of expected

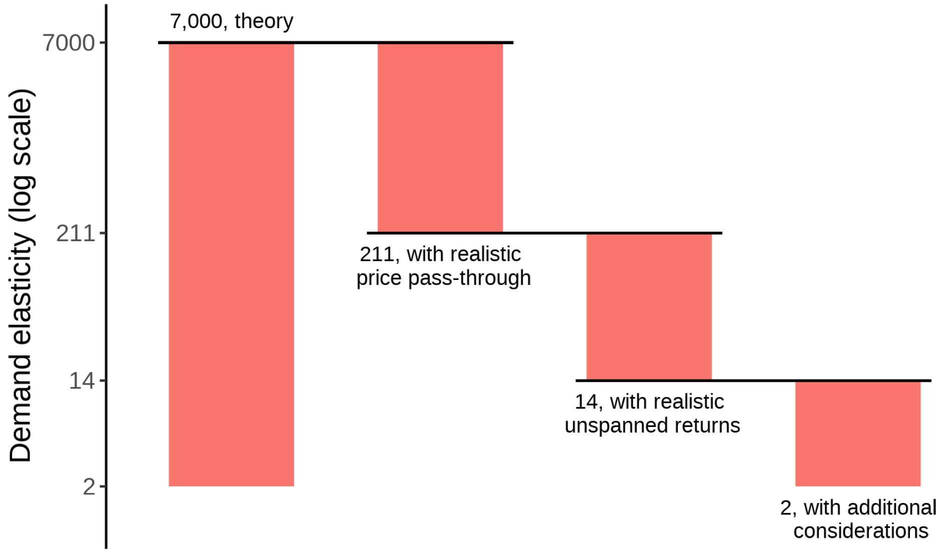


**Figure 1. Explaining the gap between theoretical and empirical demand elasticities.**
This figure illustrates our explanation of the gap between high demand elasticity values in standard asset pricing models and low values in empirical estimates. The leftmost bar illustrates predictions from standard asset pricing models with perfect pass-through and low unspanned returns (e.g., Petajisto, 2009) ($\approx$ 7,000). Taking into account realistic price pass-through reduces elasticity by a factor of 33 (as shown by the second bar from left) to approximately 211. Taking into account our empirically estimated unspanned returns, we arrive at elasticity estimates of around 14, further reducing elasticity by a factor of 15, as shown by the third bar from the left. Therefore, taking into account realistic price processes can explain the gap from 7,000 to 14. We hypothesize that the remaining gap from 14 to 2 (empirical estimates) may be explained by additional considerations such as transaction costs, limited investment universes, and behavioral biases. The $y$-axis is in log scale.

returns and covariances. We begin with an approach similar to those adopted by quantitative funds, yet our conclusions are robust to a variety of alternative approaches. Specifically, we model expected stock returns as a function of commonly used return-predicting characteristics. We then compute price pass-through as the model-implied change of expected returns in response to changes in price-dependent stock characteristics.[4] We find that the price pass-through is approximately 0.03 at a monthly horizon — i.e., each 1% of price drop translates into a 3 basis point higher expected returns in the following month. As described above, this empirically estimated price pass-through

[4]To illustrate the logic, consider a hypothetical fund that solely predicts returns using $\mu_{i,t} = \pi \cdot \log(B_{i,t}/M_{i,t})$, where $\pi$ is a coefficient and $\log(B_{i,t}/M_{i,t})$ is the log book-to-market ratio. This implies a pass-through of $-\frac{\partial \mu_{i,t}}{\partial \log(M_{i,t})} = \pi$.

lowers demand elasticity prediction from 7,000 to approximately 211. To alleviate the concern that this result relies on a parametric model of expected returns, we also use the non-fundamental "price wedge" in van Binsbergen, Boons, Opp, and Tamoni (2023) to predict returns and find similar price pass-through estimates.

To estimate unspanned returns, we estimate the covariance matrix using a one-year rolling window of daily stock returns with Ledoit and Wolf (2004) shrinkage to ensure matrix invertibility. The results indicate that, for an average stock with positive holdings in an MV portfolio, approximately 0.23% of its expected return is unspanned. Our findings are robust to using alternative covariance matrix estimations such as characteristics-based models similar to those adopted by MSCI Barra. Our results indicate that, even with flexible combinations of all other stocks, the spanning is far from perfect. This is in stark contrast with classical models which usually predict unspanned returns to be no more than 0.01%. For instance, many theoretical models with CARA-normal settings end up predicting the CAPM or some small modifications of it. If the CAPM holds, then the expected return of any given stock is almost perfectly spanned by the other stocks. When we incorporate realistic estimates of unspanned returns, the predicted demand elasticity decreases dramatically, from 211 to approximately 14. Therefore, we find that the majority of the demand elasticity gap can be explained by fully optimal investor behavior, without the need for additional frictions.

What explains the remaining gap from 14 to the empirical estimates of around 2? We believe there are numerous potential reasons, and we offer tentative quantifications for these channels. Many of them stem from optimal responses to frictions. If investors take into account how volatility changes with prices (the "leverage effect" of Black, 1976), then demand elasticity should further decline by approximately 2. Considering Transaction costs further reduces demand elasticities by causing slower rebalancing frequencies, smaller investment universes, and trading less than what the MV portfolio prescribes. Shifting from these frictional reasons to more behaviorally-inclined explanations, benchmarking considerations, as well as uncertainty about expected return estimates, can play a role in reducing demand elasticities. We believe there is not a single explanation that fully accounts for the gap from 14 to 2 across all investors. It is more probable that a mix of these factors is at play, with the precise reasons varying among different investors.

Our primary contribution is reconciling the high theoretical predictions and low empirical estimates of demand elasticities. Guided by our decomposition framework, we show that applying empirically estimated values for price pass-through and unspanned returns largely explains this gap, without the need to assume suboptimal investor behavior or market frictions. Among existing equilibrium models of price impact, some exhibit unrealistically high price pass-through, yet nearly all predict unrealistically low unspanned returns. Although these models are valuable for making qualitative predictions, our results indicates that they fall short in making quantitative predictions about price impacts. It is important to clarify that our paper focuses on how optimal investor demand *should* respond to price movements. As such, we empirically estimate price dynamics, but we do not empirically estimate demand elasticity. Therefore, we do not need or offer a new price instrument.

We focus on using our two-part decomposition to understand the *average* demand elasticity, but it is also useful for understanding its *variation*. For instance, the effect of substitutability predicts that demand elasticities should be lower when considering portfolios at higher levels of aggregation (such as the style level), consistent with the findings in Li and Lin (2023). Furthermore, for high-credit quality corporate bonds, the effect of pass-through implies that bonds with shorter maturities inevitably have higher price pass-throughs, as price deviations must revert by the maturity date, while longer maturity bonds may display lower price pass-throughs, consistent with the findings in Li, Fu, and Chaudhary (2023). Additionally, our decomposition clarifies that demand elasticity is not a stable structural parameter; rather it could vary depending on the investor, asset, and time period.

Our paper studies portfolio optimization problems for investors who take prices as given. We empirically find low price pass-throughs and high unspanned returns, yet we do not offer an equilibrium explanation of why they arise in the first place. This is due to many possible equilibrium explanations that are hard to separate. For instance, dynamic models with persistent supply shocks (e.g., Gabaix and Koijen, 2021; Johnson, 2006), as well as models with private information about future prices, typically generate low price pass-through. Similarly, models imposing limitations on diversification (e.g., Gârleanu, Panageas, and Yu, 2015; Iachan, Silva, and Zi, 2022), along with those that take into account the complexity of cross-sectional return predictions, often generate high unspanned returns. The primary advantage of our analysis is that it relies on relatively weak assumptions rather than any specific equilibrium model, making our results more broadly applicable.

## Related literature

Classic asset pricing theories predict high demand elasticities (e.g., Petajisto, 2009; Gabaix and Koijen, 2021). Our paper shows that their prediction arises from assuming low price pass-throughs and high substitutability among stocks.

There is a long literature that estimates demand elasticities in stocks using price impacts of demand or supply shocks: index exclusion (Shleifer, 1986; Chang, Hong, and Liskovich, 2015; Pavlova and Sikorskaya, 2023), dividend payments (Schmickler, 2020), mutual fund flows (Lou, 2012), and trade-level price impacts (Frazzini, Israel, and Moskowitz, 2018; Bouchaud, Bonart, Donier, and Gould, 2018). More recently, there are also structural approaches that estimate demand elasticities using asset demand systems (Koijen and Yogo, 2019; Haddad, Huebner, and Loualiche, 2022; van der Beck, 2022). The estimates of (micro) price multipliers (inverse of micro elasticities) range from 0.3 to 15, much higher than what existing models predict; See Table 1 and Figure 2 of Gabaix and Koijen (2021) for more details. Therefore, demand curves are much more inelastic compared to theory predictions. In this paper, we provide a microfoundation for inelastic demand.

Our finding that most stock price movements exhibit small pass-throughs to future expected returns is consistent with the existing evidence 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 lies not in estimating these pass-throughs, but rather in showing their critical role in demand elasticity. In our framework, weak price pass-throughs leads to a low demand elasticity. Our estimated low pass-throughs are also consistent with the literature that explores discount rate and cash flow variations (e.g., Vuolteenaho, 2002; Cochrane, 2008).

Our finding of high unspanned returns are consistent with the asset pricing literature, which finds that stock 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 low asset demand elasticity. Additionally, we provide empirical estimates for the degree to which an asset is spanned by other assets, rather than just a variety of factors.

The remainder of the paper is as follows. In Section 2, we present our decomposition of demand elasticity into price pass-through and weight responsiveness. In Section 3, we estimate the components of demand elasticity and present evidence of low pass-throughs and high unspanned returns. Section 4 discusses additional considerations and Section 5 concludes.

# 2 A Decomposition of Demand Elasticity

In this section, we show that demand elasticities can be decomposed into two terms, price pass-through and weight responsiveness. The former measures how price changes "pass-through" to expected returns. For MV efficient investors, the latter depends on the extent to which assets are substitutable in the cross-section. We then apply our decomposition framework to shed light on why classic models predict very high demand elasticities.

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

## 2.1 A two-component decomposition

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

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

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

To express demand elasticity using portfolio weights, $w_{i,t}$, we can write $Q_{i,t} = A_t w_{i,t}/P_{i,t}$, where $A_t$ is the assets under management (AUM) or investor's wealth. Plugging this into Equation (2) and assuming that $A_t$ is exogenous, we have:[5]

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

We now use the chain rule to arrive at our decomposition (also stated in Equation (1)) through the effect on the asset's expected return $\mu_{i,t}$:

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

The first term, $\frac{\partial \log(w_{i,t})}{\partial \mu_{i,t}}$, which we refer to as "weight responsiveness", describes how responsive the investor's portfolio weights are to expected returns. The second term, $-\frac{\partial \mu_{i,t}}{\partial \log(P_{i,t})}$, which we refer to as "price pass-through", measures the fraction of price movements that are "passed through" to changes in expected return. If a price decrease of 1% increases next-period expected returns by 0.4%, then the price pass-through would be 0.4. As these two terms are the focus of this paper, we

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

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

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

denote them by $\theta_{i,t}$ and $\psi_{i,t}$), respectively, to simplify notation:

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

Our decomposition above only considers the demand response to changes in expected returns. Of course, price changes may also lead to changes in other aspects, such as volatility or the consumption-hedging properties of assets, which can also impact demand. We exclude these terms because, under reasonable assumptions, as shown in Section 4.1, they are quantitatively small. Therefore, we focus on the return-based effects for now.

## 2.2 Weight responsiveness depends on unspanned returns

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

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

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

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

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

$$w_t = \frac{1}{\gamma A_t} \Sigma_t^{-1} \mu_t, \tag{7}$$

where $\Sigma_t$ is the covariance matrix of returns. Thus, taking derivatives, we can write:

$$\frac{\partial w_{i,t}}{\partial \mu_{i,t}} = \frac{\tau_{i,t}}{\gamma A_t}, \tag{8}$$

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

**The role of unspanned returns.** It turns out that weight responsiveness in Equation (8) depends on the degree to which other assets can span asset $i$'s expected return. Formally, we define $\beta_{-i,t} = \Sigma_{-i,-i,t}^{-1} \Sigma_{-i,t}$ as the $(N-1)$ column vector of betas (regression coefficients) of asset $i$'s return on all other assets except $i$, where $\Sigma_{-i,t}$ is the $i^{th}$ column vector in $\Sigma_t$ excluding the $i^{th}$ row term, and $\Sigma_{-i,-i,t}$ is the $(N-1) \times (N-1)$ covariance matrix excluding the terms associated with asset $i$. Let $\mu_{-i,t}$ be the $(N-1)$ column vector of expected excess returns excluding the $i^{th}$ asset. We define asset $i$'s unspanned returns as:

$$\mu_{i,\text{unspanned},t} \equiv \mu_{i,t} - \underbrace{\beta_{-i,t}' \, \mu_{-i,t}}_{\equiv \mu_{i,\text{spanned},t}}. \tag{9}$$

This scalar measures how well the expected return of asset $i$ is spanned by the other assets. Intuitively, if assets are almost perfect substitutes, we would expect $\mu_{i,\text{unspanned},t}$ to be small. Conversely, if assets are distinct from each other, we would anticipate $\mu_{i,\text{unspanned},t}$ to be large. It is important to clarify that $\mu_{i,\text{unspanned},t}$ is different from standard asset pricing alpha in models like the CAPM. In the CAPM, the explanatory variable — the market portfolio — is constructed using *all* assets, while in Equation (9), asset $i$ itself is excluded from the explanatory variables. Therefore, even in a world where the CAPM holds, $\mu_{i,\text{unspanned},t}$ can be non-zero, albeit small, for all assets, despite the CAPM alpha being zero by definition.

Our key result, Proposition 1, shows that the weight responsiveness for asset $i$ is equal to the reciprocal of $\mu_{i,\text{unspanned},t}$.

**Proposition 1.** *If the investor forms MV efficient portfolios, then for positive portfolio weights* $w_{i,t} > 0$*, weight responsiveness is given by:*

$$\theta_{i,t} = \frac{1}{\mu_{i,unspanned,t}} \tag{10}$$

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

$$\eta_{i,t} = 1 + \frac{\psi_{i,t}}{\mu_{i,unspanned,t}}. \tag{11}$$

*Proof.* See Appendix A.1. □

Equation (11) follows immediately from Equations (5) and (10). In words, Equation (10) shows that the investor's portfolio sensitivity to changes in expected return is inversely proportional to the asset's unspanned return. Investor demand is more inelastic for assets that are more distinctive and have larger unspanned returns. One may wonder why the unspanned expected return enters

Equation (11) but the unspanned (idiosyncratic) volatility does not. It is true that for assets with higher idiosyncratic volatility, the *level* of investor portfolio weight is lower. However, demand elasticity is defined in log terms, so the influence of idiosyncratic volatility drops out.[6]

Why is demand elasticity high for assets with lower unspanned returns? This is due to an arbitrage logic: if asset $i$ is highly substitutable, if $\mu_i$ increases, the MV investor will significantly increase $w_i$ and simultaneously short the hedging (spanning) portfolio comprised of other assets. Appendix A.1 discusses this in more detail.

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

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

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

**Unspanned returns.** We provide a calibration of the CARA-normal model in Section 2.2 using typical assumptions in classic models. The results closely resemble those in Petajisto (2009). We use these results to illustrate how classic models often yield high demand elasticities by assuming low unspanned returns.

[6] Consider a static CARA-normal model with a single asset which predicts portfolio weights of the form $w = \frac{\mu}{\gamma\sigma^2}$. While it is true that $\frac{\partial w}{\partial \mu} = \frac{1}{\gamma\sigma^2}$ depends on volatility, $\frac{\partial \log(w)}{\partial \mu} = \frac{1}{\mu}$ does not.

To begin with, assume that price pass-through is equal to one. To simplify the calibration, we assume all $N$ assets have the same expected returns, standard deviations, and correlations. We set the expected annual excess return at 6%, the volatility at 30%, and the average correlation at 0.3. These parameters are approximately similar to those observed in annual stock returns. We consider $N = 1{,}000$ assets and assume that the investor holds an equal-weighted portfolio.[7] We set the CARA risk aversion coefficient times wealth, $\gamma A_t$, to 2.2 because this allows the portfolio weights to sum to one.[8] We set the risk-free rate to zero. This calibration yields an elasticity of approximately 7,000 which is plotted in the leftmost bar in Figure 1.[9]

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

This is not specific to our calibration; it is also true in other calibrations, such as Petajisto (2009) and others. Most equilibrium models used to study demand or supply effects in the cross-section, largely for tractability reasons, assume a CARA-normal setting, leading to the CAPM or perhaps the CAPM with an additional factor. Therefore, they consistently predict very low unspanned returns. In general, when a model, similar to those in these papers, assumes that a low dimensional factor

[7]This assumption does not imply that we are only considering an equal-weighted portfolio. Rather, we are calculating the elasticity for an asset with average weights.

[8]The exact value of $\gamma A_t$ only impacts the *level* of portfolio holdings, without affecting demand elasticities, which are defined using log changes.

[9]Petajisto (2009) finds a demand elasticity around 6,000. Naturally, demand elasticity fluctuates as the model parameter values change, but experimenting with a range of parameter values typically produces an average elasticity across assets that is at least in the thousands. The average elasticity also tends to increase with the number of assets $N$, due to the increased substitutability resulting from a larger number of assets.

model can explain expected returns well, it will generate low unspanned returns and, consequently, high demand elasticities.

## 2.4 How could price pass-through be less than one?

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

$$dp_t \approx a + \sum_{h=0}^{\infty} \rho^h (\mu_{t+1+h} + r^f_{t+1+h}) - \sum_{h=0}^{\infty} \rho^h \, \mathbb{E}_t(\Delta d_{t+1+h}), \tag{12}$$

where $a$ is a constant, $(\mu_t + r^f_t)$ is the log return inclusive of risk-free rate, $\Delta d_t$ is the log change in dividend, and $\rho \approx 0.97$. From Equation (12), a change in the current price unrelated to cash flow (dividend) must predict changes in the stream of future expected returns $(\mu_{t+1}, \mu_{t+2}, ...)$. Therefore, a price pass-through less than one simply means that the effect is not *solely* on the immediate next-period return, $\mu_{t+1}$, but also on the expected returns in future periods. This realization also implies that, unless further restrictions are imposed, dividend-unrelated price deviations do not necessarily lead to a price pass-through of one. This raises two possible questions.

First, one may question whether a less-than-one price pass-through can be justified in equilibrium. In Appendix A.2, we theoretically show that low price-pass-through occurs naturally in several equilibrium settings. For instance, consider the case where the current price movement is caused by an uninformed supply shock — the typical empirical setting used in estimating demand elasticities. In general equilibrium models with long-lived assets, as long as the supply shock does not revert immediately, we should expect the impact on expected returns to also be persistent (e.g., Johnson, 2006; Gabaix and Koijen, 2021).[10] Furthermore, investors may also perceive (or fear) that price movements contain private information about future prices, which also naturally leads to lower-than-one pass pass-throughs. Overall, there are several reasons why price pass-through might be low, but it is difficult to differentiate them in the data. Therefore, in this paper, we adopt a portfolio optimization approach and empirically estimate price pass-through in Section 3. We do not take a stance on which equilibrium foundation is correct.

Second, one might argue that demand elasticities are only defined when the price pass-through is one. This stance may be motivated by the desire to "hold everything else fixed and only vary current prices". Specifically, if only $P_t$ varies, but all future prices $\{P_{t+1}, P_{t+2}, ...\}$ are held fixed, then pass-through is indeed approximately equal to one. This approach is justified when modeling the demand for consumption goods such as bananas whose utility value may be stable. However, this definition becomes less clear when applied to the demand for assets. For an investor, the value of an asset primarily lies in its potential returns. As mentioned earlier, from the Campbell-Shiller identity, any dividend-unrelated price change *must* be associated with changes in future expected returns. Therefore, it is simply impossible to "hold value fixed". Restricting attention to price pass-through of one amounts to requiring that all adjustments happen via the next-period return, which is somewhat arbitrary.[11] That being said, we do not wish to argue semantics. If one insists on using one-to-one price pass-throughs to define "true" demand elasticities, our paper simply points out that most empirical studies estimate demand responses using price shocks with lower

[10]To see this, consider a permanent increase in the supply of shares. In most models, this leads to a permanently higher risk premium in all future periods. Assuming the asset is infinitely lived like a stock, the price pass-through for an annual horizon is only $1 - \rho \approx 0.03$. If the supply increase reverts slowly, price pass-through would be higher, but still not as high as one (e.g., An and Zheng, 2023).

[11]It should be easy to see that this assumption is problematic in continuous time modeling because it amounts to assuming that price movements immediately revert. Even in discrete time modeling, because the unit time period is somewhat arbitrary (day, week, month, quarter), this is also problematic.

pass-throughs. Moreover, one can always use our decomposition in Equation (1) to adjust for the impact of price pass-through to convert to the desired "true" demand elasticities.[12]

# 3 Estimates of Price Pass-through and Unspanned Returns

In Section 2, we showed that price pass-through and unspanned returns are the key determinants of demand elasticity. In this section, we empirically estimate these two channels to determine what demand elasticity values are plausible. We start by estimating a model of stock returns and covariances as functions of stock characteristics, as is commonly done by quantitative funds, and then show that the implied price pass-through and unspanned returns are low. We then show that our results are robust to alternative specifications for the data-generating process for stock returns.

## 3.1 Data

We use the standard CRSP-Compustat merged dataset for returns and asset characteristics and the sample runs from 1970 through 2019. To avoid results driven by microcap stocks, we exclude stocks with market cap lower than the NYSE 20% percentile in each month. To form returns for the quarterly and annual frequencies, we cumulate returns from the monthly stock data. We obtain risk-free rate and factor returns from Ken French's website. Our stock characteristics data is kindly shared by Michael Weber and is from the paper Freyberger, Neuhierl, and Weber (2020). Table 1 reports the summary statistics. We have an average of 1,633 stocks in each month.

[12]For example, suppose the estimated demand elasticity is $\eta$ and the associated pass-through is $\theta$. One can then back out weight responsiveness, $\psi = \frac{\eta-1}{\theta}$, and compute what the elasticity would be if $\theta$ were equal to 1, i.e., $1+\psi (= 1+\frac{\eta-1}{\theta})$.

| Variable | N | Mean | SD | 5% | 25% | 50% | 75% | 95% |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Market cap ($ bn) | 1,633 | 4.58 | 13.96 | 0.30 | 0.51 | 1.08 | 3.02 | 17.65 |
| Monthly excess return | 1,633 | 0.72% | 9.77% | -13.83% | -4.68% | 0.38% | 5.67% | 16.24% |
| Lagged monthly return | 1,633 | 1.49% | 9.88% | -12.82% | -4.16% | 0.86% | 6.31% | 17.69% |
| Log(B/M) | 1,633 | -0.59 | 0.82 | -2.02 | -1.06 | -0.52 | -0.06 | 0.63 |
| Asset growth | 1,633 | 0.13 | 0.19 | -0.09 | 0.03 | 0.09 | 0.19 | 0.52 |
| Dividend/book | 1,633 | 2.91% | 2.84% | 0.00% | 0.74% | 2.11% | 4.30% | 9.36% |
| Profitability | 1,633 | 0.24 | 0.22 | -0.03 | 0.15 | 0.24 | 0.34 | 0.59 |

**Table 1. Summary statistics**
The sample consists of CRSP monthly U.S. stocks from 1970 to 2019. We excluded stocks smaller than the monthly NYSE 20% percentiles. The table reports the average distributions of variables across months. The first column reports the average number of stocks in each month. The next two columns report the mean and the standard deviation, and the last five columns report percentiles.

## 3.2 Estimates of price pass-through and unspanned returns

Section 2.1 shows that for MV investors, optimal demand elasticity depends on two parameters: price pass-through and unspanned returns. Both of these only depend on the data-generating process (DGP) of returns. In this section, we estimate the DGP using one specification. Subsequent sections and appendices show that our main conclusion is robust to alternative estimation techniques.

We consider a price-taking MV investor who, similar to some quantitative funds, forecasts stock excess returns as a linear function of commonly used stock characteristics ($\mathbb{E}_t(r_{t+1}) \equiv \mu_t = Z_t\pi$). As the stock characteristics $Z_t$ vary over time, the forecast $\mu_t$ also varies over time. To model the the time-varying covariance matrix $\Sigma_t \equiv \mathbb{V}\mathrm{ar}_t(r_{t+1})$, we follow Lopez-Lira and Roussanov (2023) to estimate it using a one-year lag of daily data. Following a popular industry practice (Ledoit and Wolf, 2004), we slightly shrink the covariance matrix.[13] Specifically, we use $\Sigma = (1-h)\hat{\Sigma} + h\overline{\Sigma}$,

[13]There are two main reasons for shrinking the covariance matrix. First, because we have more stocks than the number of days in a year, the sample covariance matrix is not full-rank. This is problematic, because computing unspanned returns and forming MV efficient portfolios require the covariance matrix to be full-rank. Shrinkage effectively addresses this issue by ensuring the matrix is full-rank. Second, and more importantly, it is well-known that using the sample covariance matrix with a large number of assets leads to ill-conditioned portfolio choices (e.g., Ledoit and Wolf, 2004; Brandt, 2010). Shrinkage is a commonly used method to regularize the problem and produce realistic portfolios.

where $\hat{\Sigma}$ is the sample covariance matrix estimate, $\overline{\Sigma}$ is the shrinkage target, and $h$ is the scalar shrinkage weight. The shrinkage target $\overline{\Sigma}$ has the average stock-specific return variance along the diagonal ($\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{i,i}$) and the average covariance on the off-diagonal ($\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\hat{\Sigma}_{i,j}$). We report results with shrinkage intensity $h = 0.05$, but our main conclusions are not sensitive to reasonable variations of the parameter.

**Implied price pass-through.** To estimate the parameters associated with the expected return, we estimate Fama-MacBeth regression of monthly stock returns on lagged stock characteristics and report results in Table 2. We consider commonly used stock characteristics including the Fama-French five characteristics, Momentum, and one-month reversal (the most recent monthly log return). Consistent with previous research, all cross-sectional characteristics have statistically significant associations with one-month-ahead returns. We compute standard errors using the Newey-West procedure with 12 monthly lags.

What is the implied price pass-through from this model for forecasting expected returns? Note that among the stock characteristics, reversal, size ($\log(M)$), and book-to-market ($\log(B/M)$) are direct functions of the most recent stock price. When stock prices change, these price-related characteristics change their values and their prices of risk ($\pi_j$) determine the price pass-throughs. Let $\pi_{rev}$, $\pi_{size}$, and $\pi_{bm}$ denote the elements of $\pi$ corresponding to these three variables. Then, the price pass-through is simply:

$$
\begin{aligned}
\psi_{i,t} &= \frac{\partial}{\partial \log P_{i,t}}\mu_{i,t} = \sum_{k=1}^{K}\frac{\partial Z_{i,k,t}}{\partial \log P_{i,t}} \cdot \pi_i \\
&= \pi_{rev} - \pi_{size} + \pi_{bm}
\end{aligned}
$$

We report the implied price pass-through in columns (1b) and (2b) of Table 2. Column (1b)

| | Dependent variable: $r_{i,t}$ (%) | | To multiply | Implied price pass-through (%) | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1b) | (2b) |
| Intercept | $-0.42$ | 0.25 | | $2.08^{***}$ | $3.01^{***}$ |
| | (0.41) | (0.30) | | (0.50) | (0.51) |
| $r_{i,t-1}$ | $-1.95^{***}$ | $-2.73^{***}$ | -1 | | |
| | (0.49) | (0.50) | | | |
| $\log(M_{i,t-1})$ | $0.09^{**}$ | 0.02 | -1 | | |
| | (0.04) | (0.03) | | | |
| $\log(B/M)_{i,t-1}$ | $0.23^{**}$ | $0.30^{***}$ | 1 | | |
| | (0.09) | (0.07) | | | |
| Mom | | $1.23^{***}$ | | | |
| | | (0.19) | | | |
| Beta | | $-0.29^{**}$ | | | |
| | | (0.13) | | | |
| Investment | | $-1.26^{***}$ | | | |
| | | (0.18) | | | |
| Profitability | | $1.27^{***}$ | | | |
| | | (0.21) | | | |
| Obs | 968,634 | 968,634 | | | |
| Average $R^2$ | 4.04% | 8.51% | | | |
| *Note:* | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

**Table 2. Price pass-through estimate based on Fama-MacBeth regressions**
We estimate Fama-MacBeth regressions using stock characteristics and one-month-ahead stock returns (in percent), and we calculate standard errors using the Newey-West procedure with 12 monthly lags. The specification in column (1) only includes the characteristics that explicitly depend on lagged stock price ($P_{i,t-1}$), while the specification in column (2) also includes several other characteristics. Column (3) reports the coefficients to multiply onto each of the price-related characteristics for computing price pass-through. Columns (1b) and (2b) report the implied price pass-through based on columns (1) and (2).

corresponds to the Fama-MacBeth specification in column (1), where we only include these three price-relevant characteristics. Column (2b) corresponds to the specification in column (2), which includes additional controls. The results indicate that price pass-through, at a monthly horizon, is on the order of 0.0208 to 0.0301. This implies that each 1% of price movement reverts around 2-3 basis points in the subsequent month. The standard errors for the pass-throughs are computed using the delta method. In Section 3.3, we consider alternative ways to estimate price pass-through.

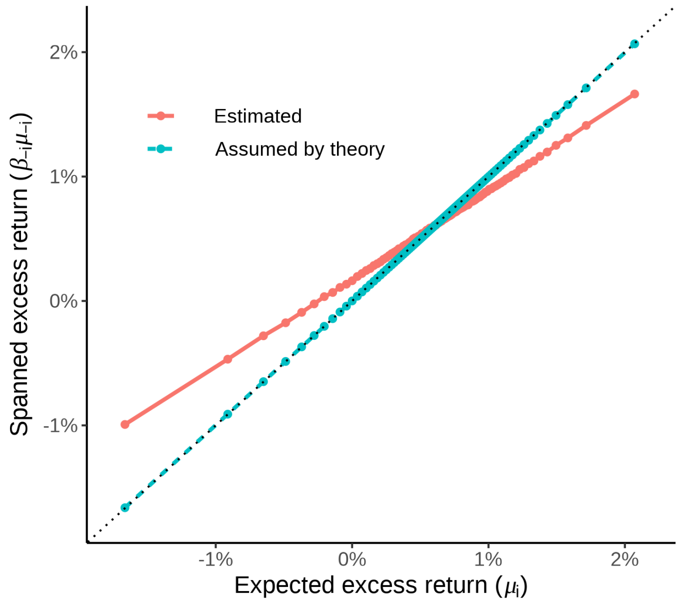


**Figure 2. Decomposing spanned and unspanned returns**
This figure explores the fraction of expected excess return that is spanned by other stocks. As described in Section 3.2, we estimate expected excess returns $\mu_t$ using stock characteristics-based Fama-MacBeth predictions. We estimate the covariance matrix $\Sigma_t$ using daily returns in the previous 12 months and apply a minor Ledoit-Wolf shrinkage to $\Sigma_t$ (shrinkage parameter = 0.01) to ensure positive definiteness. We sort the full sample into 100 bins based on each stock's estimated excess return $\mu_{i,t}$ and plot the spanned component ($\beta'_{-i}\mu_{-i,t}$) on the vertical axis. The red line represents the spanned returns estimated from data and, for comparison, the blue line plots the unspanned returns assumed by theory.

**Implied unspanned returns.** We now use our estimates of $\mu_t$ and $\Sigma_t$ to compute the implied spanned and unspanned returns ($\beta'_{-i,t} \cdot \mu_{-i,t}$ vs. $\mu_{i,t} - \beta'_{-i,t} \cdot \mu_{-i,t}$). We compute them for each stock-month and graphically illustrate the results in Figure 2. We sort the sample by each stock's total expected excess return $\mu_{i,t}$ into 100 bins and plot the average spanned component $\beta'_{-i,t}\mu_{-i,t}$ on the $y$-axis. The result indicates that a sizeable fraction of return predictability is spanned, but the spanning is far from complete. For instance, for the top bin, the monthly excess return is 2.07% while the spanned component is 1.66%, so unspanned returns equal $2.07\% - 1.66\% = 0.41\%$. For the bottom bin which contains stocks that our investor would short, the expected excess return is $-1.67\%$ while the spanned component is $-0.99\%$, so the unspanned component is $-1.67\% - (-0.99\%) = -0.68\%$.

To illustrate the difference with classic theories, we also plot the spanning implied in the model of Petajisto (2009) in blue. In that model, as well as all models with static CARA-normal setups, approximately 99.8% of all returns are spanned. Therefore, the line essentially overlaps with the 45-degree line.

In unreported robustness checks, we find that the lack of perfect spanning in actual data is robust to several alternative ways for estimating $\mu_t$ and $\Sigma_t$. This includes estimating $\mu_t$ using rolling (as opposed to full-sample) Fama-MacBeth regressions, using maximum likelihood estimates, or using alternative stock characteristics. This also includes estimating $\Sigma_t$ that is parametrized by stock characteristics, similar to that done by BARRA. Overall, these results imply that while stocks certainly exhibit substitutability, they are not perfect substitutes as assumed in most theoretical models.[14]

**Implied demand elasticity.** We now plug in the estimates of price pass-through and unspanned returns to compute the implied MV investor demand elasticity. Before proceeding, note that demand elasticities, by the nature of using logs in their definition, are only well-defined for portfolio positions with strictly positive weights. Therefore, we compute the MV portfolio weights $w_t \propto \Sigma_t^{-1}\mu_t$ and only compute demand elasticities for positions with positive weights. We report results in Table 3.

Consider the first row of Table 3 with shrinkage intensity $h = 0.01$, as results are qualitatively unchanged when varying $h$. For positions with positive portfolio weights, the average expected monthly excess return is 0.70% and the average unspanned component is 0.23%. If we ignore heterogeneity across stocks, our theoretical decomposition predicts weight responsiveness $\theta = \frac{1}{0.0023} \approx 435$. If we take the higher price pass-through estimate of $\psi = 0.0301$ from earlier, this implies demand elasticity of $1 + 0.0301/0.0023 \approx 14$.

[14]Our finding is broadly consistent with Lopez-Lira and Roussanov (2023) who also find that stock returns are not well spanned by systematic risk factors estimated from the data.

| Covariance shrinkage ($h$) | Stock-level: Expected return ($\mu_i$) | Stock-level: Spanned return ($\beta'_{-i}\mu_{-i}$) | Stock-level: Unspanned return ($\mu_i - \beta'_{-i}\mu_{-i}$) | Stock-level: Unspanned fraction | Portfolio-level: Weight responsiveness $\theta = \frac{\partial \log(w)}{\partial \mu}$ | Portfolio-level: Demand elasticity $(1 + \theta\psi)$ |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 0.01 | 0.70% | 0.47% | 0.23% | 32.9% | 471.2 | 15.2 |
| 0.025 | 0.70% | 0.47% | 0.23% | 32.8% | 471.7 | 15.2 |
| 0.05 | 0.70% | 0.47% | 0.23% | 32.6% | 472.7 | 15.2 |
| 0.25 | 0.72% | 0.50% | 0.23% | 31.3% | 478.6 | 15.4 |
| 0.50 | 0.75% | 0.53% | 0.23% | 29.8% | 479.3 | 15.4 |

**Table 3. Unspanned returns and implied MV demand elasticity**
Each row corresponds to a different covariance matrix estimate with different Ledoit-Wolf shrinkage intensity $h$. Throughout, we use the expected return estimates based on Fama-MacBeth regressions with stock characteristics in Table 2. We report results for all stocks with positive MV portfolio weights. Column (1) reports the average monthly excess return of stocks. Column (2) reports the component that is spanned by other stocks. Column (3) reports the component unspanned and column (4) reports the fraction unspanned ((3)/(1)). Column (5) reports the holdings-weighted average weight responsiveness and column (6) reports the holdings-weighted demand elasticity.

The calculation above ignores the fact that $\theta_{i,t}$ differs across stocks and that some stocks are held with higher weights than others. However, in columns (5) and (6), we compute the portfolio holdings-weighted average counterparts of $\theta$ and demand elasticities and obtain roughly similar results. Comparing across rows, we also see that our conclusion is not sensitive to changes in the shrinkage intensity. In unreported robustness checks, we also find our results are not sensitive to the robust estimation approaches in DeMiguel, Garlappi, Nogales, and Uppal (2009a) and Kozak, Nagel, and Santosh (2020). As shown in Kozak et al. (2020), these methods can all be understood as specific shrinkage implementations.

One may be concerned if the estimated unspanned returns are realistic, especially since the Fama-MacBeth coefficients are estimated in the full sample. In Appendix D.1, we verify that results are similar when using rolling 10-year windows to estimate the Fama-MacBeth coefficients. Also, the unspanned returns do indeed predict out-of-sample returns, consistent with Lewellen (2015).

We graphically summarize the findings in this section in Figure 1 in the introduction. For comparison, classic models such as Petajisto (2009) predict price pass-through of approximately 1 and weight responsiveness of approximately 7,000, and thus predict demand elasticities as high as 7,000. Results in this section imply that, if we take into account realistic price pass-through estimates, the predicted demand elasticity lowers to approximately $1+0.0301 \cdot 7,000 \approx 211$. Further taking into account realistic unspanned returns further lowers the prediction to approximately 14. It is worth emphasizing that we do not argue that the estimates here are precise. Subsequent robustness checks show that the exact prediction can vary slightly, but they are all two orders of magnitude lower than the theoretical estimate in classic models.

**Implied Sharpe ratios.** One may be concerned that finding sizeable unspanned returns in stocks implies unrealistically high Sharpe ratios at the portfolio level. According to arbitrage pricing theory (APT), if the number of stocks $N$ goes to infinity, yet idiosyncratic return components have non-negligible excess returns, then one can construct portfolios with infinite Sharpe ratio in the limit (Ross, 1976).[15]

Of course, $N$ is not infinity, and how large is the maximum attainable Sharpe ratio is an empirical question. Table 4 reports the realized annualized Sharpe ratios for MV efficient portfolios. In addition to the full sample-estimated Fama-MacBeth return predictor in Table 2, we also consider an implementable version estimated using rolling 10-year windows, similar to that in Lewellen (2015). The results show that, across a wide range of covariance shrinkage parameters, realized Sharpe ratios are in the range of 1.15 to 1.73, which is comparable to results obtained in other academic studies that use stock characteristics to predict returns (e.g., Kelly, Pruitt, and Su, 2019;

[15]In fact, the realized unspanned returns ($r_{i,t} - \beta'_{-i} r_{-i,t}$) are *not* independent across stocks, so the original APT logic does not apply. This is because different stocks have different hedging instruments $r_{-i,t}$. For instance, consider stocks 1 and 2. Their realized unspanned returns are indeed orthogonal to stocks 3 to N, but because stock 1 is used as a hedging asset for stock 2, and vice versa, the resulting “unspanned returns” are correlated.

| | Sharpe ratio | |
|---|---|---|
| Return predictor: | Full-sample FM | Rolling FM |
| Covariance shrinkage ($h$) | (1) | (2) |
| 0.01 | 1.26 | 1.15 |
| 0.025 | 1.27 | 1.16 |
| 0.05 | 1.29 | 1.18 |
| 0.25 | 1.46 | 1.30 |
| 0.50 | 1.73 | 1.49 |

**Table 4. Sharpe ratios.**
This table reports the annual realized Sharpe ratios of MV portfolio strategies. In column (1), the return predictor is the full-sample Fama-MacBeth regression in Table 2. In column (2), the return predictor is based on rolling 10-year Fama-MacBeth regressions with the same set of characteristics. Different rows correspond to different degrees of Ledoit-Wolf covariance shrinkage. To make the sample periods comparable, all columns report Sharpe ratios from 1980 since the rolling 10-year Fama-MacBeth predictor started in 1980.

Kim, Korajczyk, and Neuhierl, 2021). In other words, empirically, high unspanned returns — as we discovered in the data — are consistent with bounded Sharpe ratios.[16]

## 3.3 Estimating price pass-through using price wedges

One may be concerned that our price pass-through estimate in Section 3.2 is specific to the characteristics-based model. To alleviate this concern, we consider an alternative estimation approach in this section.

Since the work of Campbell and Shiller (1988) and Vuolteenaho (2002), it is well understood that stock price fluctuations contain both permanent cash flow (CF) components and temporary "discount rate" (DR) components. Conceptually, consider an investor who can decompose log

[16]In this paper, we ignore the possibility that the predictability of these characteristics is discovered over time, and assume that MV investors use them to predict returns from the beginning. If we take into account the discovery process and only use characteristics after they are discovered, the realized Sharpe ratios should be lower.

stock prices into two components:

$$\log(P_{i,t}) = \log(\tilde{P}_{i,t}) + \underbrace{\log(P_{i,t}/\tilde{P}_{i,t})}_{\text{price wedge}} \tag{13}$$

where $\tilde{P}_{i,t}$ represents a measure of fundamental valuation that captures CF-related variation, and the second term, referred to as "price wedge" by van Binsbergen et al. (2023), captures DR variation. Sophisticated investors are able to trade against the price wedge component but not against the CF-based component. Therefore, for these investors, the relevant price pass-through should be estimated based on the price wedge component alone.

We now use four existing measures in the literature to study the price pass-through associated with the price wedge component. The first is the price wedge measure from van Binsbergen et al. (2023) which is based on combining the future long-run returns of 57 cross-sectional anomaly portfolios using a three-PC projection.[17] Using their data, we estimate the price pass-through associated with the price wedge component using Fama-MacBeth regressions:

$$r_{i,t+1\to t+H} = \alpha_H + \beta_H \cdot \log(\tilde{P}_{i,t}/P_{i,t}) + \epsilon_{i,t+1\to t+H}. \tag{14}$$

where $r_{i,t+1\to t+H}$ is the log return for the subsequent $H$ months. We estimate standard errors using the Newey-West procedure with $H$ number of lags. The estimated price pass-through $\beta_H$ for horizons $H = 1$, 3, 6, and 12 months are reported in the first four columns of the first row of Table 5. We find positive price pass-through that increases with the horizon, although the result at 12-month horizon loses statistical significance due to higher standard errors. To make the estimates

[17]We thank the authors for kindly sharing their data with us.

comparable across horizons, columns (6) through (9) report the pass-through per month ($\beta_h/h$) which is roughly constant at all horizons.

We also consider alternative ways to capture the price wedge component. Instead of using expected anomaly returns to directly estimate the $\log(\tilde{P}_{i,t}/P_{i,t})$ component, Bartram and Grinblatt (2018) take the approach of using a kitchen-sink of accounting variables to estimate $\tilde{P}_{i,t}$ and back out price wedge as a residual. We implement their methodology and report the associated price pass-throughs in the second row of Table 5. The point estimates are all positive but statistically insignificant, and the magnitudes are an order of magnitude smaller than that associated with the van Binsbergen et al. (2023) price wedge. This is expected, as the model in Bartram and Grinblatt (2018) is built to estimate fundamental valuation and only yield return predictability as a side product, unlike van Binsbergen et al. (2023) which is built on characteristics that were found to predict returns. The degree of return predictability we found is also consistent with the original results in Bartram and Grinblatt (2018).

We then consider two price instruments for estimating DR price variation. Koijen and Yogo (2019) develops an instrument for the market capitalization of stocks using institutional holdings, and in the first stage, we first regress the logarithm of stock market capitalization on the logarithm of their instrument. We then use the predicted values as the independent variable in the Fama-MacBeth regression (14). The results are reported in the third row, and it indicates a slightly negative price pass-through. In other words, the price movements isolated by the Koijen and Yogo (2019) instrument exhibit slight momentum rather than reversals.

Finally, we use the flow-induced trading (FIT) measure in Lou (2012) as an instrument. This measure isolates the trading by mutual funds in response to fund flows and has been shown to create price effects that revert over time. In the first stage, we regress log stock price on the most recent four quarterly values of FIT, and the second stage Fama-MacBeth results are shown on the

| Independent variable | Estimated coefficient $\beta_H$ | | | | Obs | Implied monthly price pass-through ($\beta_H/H$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $H=1$ | 3 | 6 | 12 | | $H=1$ | 3 | 6 | 12 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| van Binsbergen et al. (2023) price wedge | 0.014*** (0.003) | 0.040*** (0.014) | 0.079** (0.038) | 0.157 (0.102) | 1,270,646 | 0.014*** (0.003) | 0.013*** (0.005) | 0.013** (0.006) | 0.013 (0.009) |
| Bartram and Grinblatt (2018) price wedge | 0.001 (0.001) | 0.002 (0.002) | 0.003 (0.005) | 0.006 (0.012) | 782,431 | 0.001 (0.001) | 0.001 (0.001) | 0.000 (0.001) | 0.000 (0.001) |
| Koijen and Yogo (2019)-instrumented $\log(P_{i,t})$ | −0.002*** (0.000) | −0.004*** (0.002) | −0.009*** (0.003) | −0.021* (0.011) | 1,519,519 | −0.002*** (0.000) | −0.001*** (0.001) | −0.002*** (0.001) | −0.002* (0.001) |
| FIT-instrumented $\log(P_{i,t})$ | 0.006 (0.005) | 0.020 (0.016) | 0.040 (0.045) | 0.094 (0.084) | 1,443,296 | 0.006 (0.005) | 0.007 (0.005) | 0.007 (0.008) | 0.008 (0.007) |

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

**Table 5. Estimating price pass-through using non-fundamental price measures.** We estimate price pass-through using Fama-MacBeth regressions:

$$r_{i,t+1\to t+H} = \alpha_H + \beta_H \cdot \log\left(\tilde{P}_{i,t}/P_{i,t}\right) + \epsilon_{i,t+1\to t+H}$$

where the dependent variable is the log return of stocks in months $t+1$ to $H$. The independent variables capture (the negative of) non-fundamental price variation where $\tilde{P}_{i,t}$ is a measure of fundamental valuation and $P_{i,t}$ is the actual price. In the first two rows, the independent variables are the "price wedge" measures in van Binsbergen et al. (2023) and Bartram and Grinblatt (2018), respectively. In the last two rows, the independent variables are log market capitalization of stocks instrumented by the Koijen and Yogo (2019) instrument and the flow-induced-trading (FIT) measure in Lou (2012), respectively. Columns 1 through 4 report estimated regression coefficients $\beta_H$ for horizons $H = 1$, 3, 6, and 12 months. Column (5) reports the number of stock-months used in each specification. Columns (6) through (9) report $\beta_H/H$ which can be interpreted as the *monthly* price pass-through estimate. Throughout, standard errors of the Fama-MacBeth forecasting coefficients are estimated using the Newey-West procedure with the number of lags equal to the forecasting horizon $H$.

four rows of 5. The results indicate positive pass-through that amounts to slightly less than 0.01 per month, as shown in columns (6) through (12), but the results are not statistically significant.

Overall, these results across four different specifications indicate that existing measures of DR price components do not have very strong price pass-throughs. Even the highest estimates are not higher than that estimated in Section 3.2.

The van Binsbergen et al. (2023) price-wedge measure is also based on characteristics like in Section 3.2. This, one may ask why it is associated with slightly lower price pass-through. This is primarily because different stock characteristics predict returns at different horizons. Most characteristics predict returns at long horizons, and the return *per unit of time* is not high, except for the one-month price reversal characteristic which we include in the model in Section 3.2.[18] In unreported tests, throughout a wide range of stock characteristic choices and time periods, we are unable to find price pass-through higher than 0.06. Overall, these results indicate that realistic estimates of price pass-through are significantly lower than one as often assumed in theoretical models.

# 4 Additional Considerations

The previous section argues that when considering realistic price pass-through and unspanned returns, even an MV-efficient monthly-rebalancing investor would have demand elasticities around 14 rather than 7,000, as predicted in classical asset pricing models. However, 14 is still substantially higher than the empirical estimates around 2. In this section, we show that many realistic portfolio considerations — time-varying volatility, slower portfolio adjustments, smaller investment universes, transition costs — all tend to further lower demand elasticities. Behavioral factors such as inattention or heuristics portfolio choice may also contribute. We do not take a stance on which exact factors are the main explanations, but we argue that the "remaining gap" from 14 to 2 is not necessarily puzzling.

## 4.1 Other considerations in portfolio optimization

In sections 2 and 3, we restrict our attention to demand responses to price movements via the channel of changes in expected return in a static optimization framework. This may be restrictive in two dimensions. First, price movements may also induce changes in return volatilities and

[18]The momentum characteristic also predicts returns at relatively short horizons but it generates negative price pass-through, so we omit it.

correlations. Second, if the investor conducts dynamic optimization, intertemporal considerations may also arise.

Of all these possible extensions, we only find the effect through a stock's volatility to have a non-negligible (and dampening) effect on demand elasticity. In a calibration, we find that it reduces demand elasticity by approximately two due to the Black (1976) leverage effect, so we discuss it here. We do not find the other considerations to have much effect.

**Effect via stock volatility.** We now examine how MV-efficient portfolios respond to price changes via volatilities. Recall that MV-efficient portfolio weights for each stock $i$ is:

$$w_i = \frac{1}{\gamma} \cdot \frac{\mu_i - \beta_{-i}\mu_{-i}}{\sigma_i^2 - \Sigma'_{-i}\Sigma^{-1}_{-i,-i}\Sigma_{-i}} = \frac{1}{\gamma} \cdot \frac{\mu_{i,\text{unspanned}}}{\sigma^2_{i,\text{unspanned}}} \tag{15}$$

Therefore, the effect on demand elasticity also has a term based on volatility:

$$-\frac{1}{w_i} \cdot \frac{\partial w_i}{\partial \log P_i} = \frac{1}{\mu_{i,\text{unspanned}}} \cdot \left(-\frac{\mu_i}{\partial \log P_i}\right) + \underbrace{\frac{1}{\sigma^2_{i,\text{unspanned}}} \cdot \frac{\partial \sigma_i^2}{\partial \log P_i}}_{\text{new term}}$$

Due to the "leverage effect" of Black (1976), this new term would be negative and thus further decrease demand elasticity. The mechanism is intuitive: if a stock's price declines, its volatility increases due to the leverage effect, which further dampens the demand response due to investor risk aversion.

A calibration reveals that this volatility-induced term reduces demand elasticity by approximately 2. In Appendix Table A.1, we find that monthly $\frac{\partial \sigma_i^2}{\partial \log P_i}$ is approximately $-0.06$. Meanwhile, the average monthly $\sigma^2_{i,\text{unspanned}}$ is approximately 0.027. Thus, the new term above is approximately $\frac{1}{0.027} \times (-0.06) \approx -2.2$.

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

$$-\frac{\partial \log(Q_i)}{\partial \log(P_i)} = 1 - \frac{\partial \log(w_i)}{\partial \log(P_i)} - \underbrace{\frac{\partial \log(A)}{\partial \log(P_i)}}_{\text{wealth effect } = w_i} . \tag{16}$$

The wealth effect equals $w_i$, the portfolio weight of stock $i$, and is thus much smaller than one for reasonable scenarios.

**Covariance and consumption-hedging terms are also small.** In principle, when $P_i$ changes, this may also induce changes in the whole covariance matrix rather than just stock $i$'s standalone volatility. Further, one may also worry about intertemporal hedging considerations.

To investigate these additional possibilities, we relax the restrictions by our MV model to consider the more general setup in Campbell et al. (2003) which considers an Epstein-Zin investor with multivariate normal distributed returns. They show, after log-linearization, that optimal portfolio weights are given by:

$$w_t = \frac{1}{\kappa} \underbrace{\Sigma_t^{-1}}_{\text{covariance}} \left[ \underbrace{\mu_t}_{\text{mean}} + \underbrace{\frac{1}{2}\sigma_t^2}_{\text{volatility}} - \underbrace{\frac{\vartheta}{\varsigma}\sigma_{c-w,t}}_{\text{consumption-hedging}} \right], \tag{17}$$

where $\mu_t$ is an $N$ dimensional vector of log excess returns, $\Sigma_t$ is the $N \times N$ conditional covariance matrix of returns, $\sigma_t^2$ is the $N$ dimensional vector containing the diagonal elements of $\Sigma_t$, and $\sigma_{c-w,t}$ is the $N$ dimensional vector of the conditional covariance of returns with the log consumption-to-wealth ratio (the *cay* variable from Lettau and Ludvigson, 2001). In terms of preference parameters,

$\kappa > 0$ is the relative risk aversion coefficient, $\varsigma > 0$ is the elasticity of intertemporal substitution, and $\vartheta \equiv (1-\kappa)/(1-\varsigma^{-1})$.

In Appendix B.2, we estimate a model in which both $\Sigma_t$ and $\sigma_{c-w,t}$ change with prices via stock characteristics. As discussed in Section 3.2, because some stock characteristics depend on prices, we can compute how $\Sigma_t$ and $\sigma_{c-w,t}$ change with prices via the stock characteristics. This allows us to numerically compute the contribution of each term to demand elasticity. In the end, we find that all these additional components have only second-order effects.

## 4.2 Frictions in portfolio choice

The previous section discusses additional considerations within optimal portfolio choice frameworks. In this section, we discuss further frictions that may arise in practice. These frictions also further dampen demand elasticities. Many of these considerations may be motivated by transaction cost concerns.

**Slower portfolio adjustments.** We have thus far considered investors who rebalance every month. In practice, some investors rebalance less frequently, and this generally reduces demand elasticities. For instance, a quarterly-rebalancing investor would need to rely on price pass-throughs and unspanned returns estimated at the quarterly horizon. Because most return predictability decays with time horizon, the implied *per-month* price pass-through decays, and thus the predicted demand elasticity should also decline slightly. A calibration in Appendix D.2 suggests that for an investor who rebalances annually instead of monthly, the demand elasticity would be approximately 30% lower than one who rebalances monthly.

| Num of stocks (N) | Unspanned return $(\mu_i - \beta'_{-i}\mu_{-i})$ | Weight responsiveness $\theta = \frac{\partial \log(w)}{\partial \mu}$ | Demand elasticity $(1+\theta\psi)$ |
|---|---|---|---|
| | (1) | (2) | (3) |
| 1,615 | 0.23% | 471.2 | 15.2 |
| 500 | 0.31% | 372.5 | 12.2 |
| 100 | 0.41% | 290.9 | 9.8 |
| 20 | 0.43% | 247.1 | 8.4 |

**Table 6. The effect of the number of investable stocks $(N)$.**
This table is similar to Table 3 except we consider changes in the number of investable stocks. The first row uses the full sample and the first column reports the average number of stocks in each month. In the next three rows, we randomly select $N = 500$, 100, and 20 stocks each month to form the investment universe. We use a covariance shrinkage parameter of 0.01 and pass-through of $\psi = 0.0301$ as estimated in Section 3.2.

**Smaller investment universes.** In our exercise so far, we consider an investor who can trade all U.S. stocks and achieve maximal diversification. In practice, many institutions trade fewer stocks, often for liquidity reasons.

Intuitively, when fewer stocks are available, spanning becomes worse, and thus theory-predicted demand elasticities should drop. To gauge the relationship with the size of the investment universe, in each month, we randomly sample $N = 20, 100, 500$ stocks and use those as the available investment universe. The resulting portfolio optimization results are shown in Table 6. As expected, when the investment universe shrinks, the resulting unspanned returns increase which leads to lower demand elasticities, but the effect is moderate. For an investor who is restricted to $N = 100$ stocks, demand elasticity declines to 9.8 from 15.2 when she can access all stocks.

## 4.3 Heuristics portfolio choices

The discussion so far focuses on investors who form optimal portfolios, albeit with frictions. If we consider alternative portfolio formation methods, we find that commonly used heuristics imply even lower demand elasticities. Some of these heuristics are developed as robust alternatives to

MV optimization. It is well-known that MV optimization with a large number of assets is prone to estimation errors (e.g., Michaud, 1989). In response to this, practitioners have proposed simplified portfolio choices that can be understood as progressively removing the number of parameters to estimate. This includes the "global minimum variance" portfolio which assumes that all stocks have the same rate of return, the "risk parity" portfolio which further ignores correlations (e.g., Asness, Frazzini, and Pedersen, 2012), and the "$1/N$" equal-weighted portfolio which further ignores differences in stock volatility (e.g., DeMiguel, Garlappi, and Uppal, 2009b). Appendix C.1 shows that these portfolios generally yield demand elasticities less than one. The reason for this finding is because they ignore changes in expected return which, as discussed in Sections 2 and 3, is the main component of demand elasticity.

# 5 Conclusion

In this paper, we try to explain the large gap between the high ($\approx$ 7,000) demand elasticities predicted by classical asset pricing theories and the low ($\approx$ 2) estimates in empirical studies. To understand the determinants of demand elasticities, we present a novel decomposition of demand elasticity into (one plus) the product of two components: "price pass-through" and "weight responsiveness". The former measures the extent to which prices predict returns, and the latter captures the magnitude of portfolio weight changes in reaction to shifts in expected returns. We show that for a mean-variance (MV) optimal investor, the weight responsiveness is inversely proportional to "unspanned returns" which measures the amount of asset returns not spanned by other assets. The more distinctive an asset, the less its returns are spanned by others, and the lower its weight responsiveness.

We then empirically show that incorporating realistic price pass-throughs and unspanned returns can explain most of the gap. When we estimate stock returns and covariances like is typically done by quantitative funds, we find price pass-through around 0.03, as opposed to 1 which is assumed in theory models; we find monthly unspanned returns around 0.23% as opposed to 0.01% as typically assumed in theory models. With realistic estimates, even for an investor who holds monthly-rebalanced MV efficient portfolios, the implied demand elasticity is only around 14, significantly lower than the 7,000 predictions from classical theories. Further taking into account changes in stock return volatility further reduces demand elasticity predictions. Our findings are robust to alternative ways to estimate price pass-through and unspanned returns.

Because MV behavior describes the optimal strategy without friction, it likely generates the highest demand elasticity. We further find that adding additional considerations such as slower portfolio rebalancing, smaller investment universes, transaction costs, as well as heuristics in portfolio construction generally further reduces the predicted demand elasticity. However, it is worth emphasizing that most of the gap (7,000 to 14) is fully accounted for without friction or biases. Therefore, low demand elasticity is largely not a puzzle.

## References

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

Asness, Clifford S, Andrea Frazzini, and Lasse H Pedersen, 2012, Leverage aversion and risk parity, *Financial Analysts Journal* 68, 47–59.

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

Bartram, Söhnke M, and Mark Grinblatt, 2018, Agnostic fundamental analysis works, *Journal of Financial Economics* 128, 125–147.

Black, Fischer, 1976, Studies of stock price volatility changes, in *Proceedings of the Business and Economic Statistics Section*, 177–181.

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, 2010, Portfolio choice problems, in *Handbook of Financial Econometrics: Tools and Techniques*, 269–336 (Elsevier).

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.

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.

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

DeMiguel, Victor, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal, 2009a, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management science* 55, 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009b, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial studies* 22, 1915–1953.

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

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

Fuchs, William, Satoshi Fukuda, and Daniel Neuhann, 2023, Demand-system asset pricing: Theoretical foundations, Working Paper, UT Austin.

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.

Gârleanu, Nicolae, Stavros Panageas, and Jianfeng Yu, 2015, Financial entanglement: A theory of incomplete integration, leverage, crashes, and contagion, *American Economic Review* 105, 1979–2010.

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

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

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

Iachan, Felipe S, Dejanir Silva, and Chao Zi, 2022, Under-diversification and idiosyncratic risk externalities, *Journal of Financial Economics* 143, 1227–1250.

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.

Kim, Soohun, Robert A Korajczyk, and Andreas Neuhierl, 2021, Arbitrage portfolios, *Review of Financial Studies* 34, 2813–2856.

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

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross section, *Journal of Financial Economics* 135, 271 – 292.

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.

Lewellen, Jonathan, 2015, The cross section of expected stock returns, *Critical Finance Review* 4, 1–44.

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

Li, Jian, Zhiyu Fu, and Manav Chaudhary, 2023, Corporate bond elasticities: Substitutes matter, Working Paper.

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

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

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

Michaud, Richard O., 1989, The markowitz optimization enigma: Is 'optimized'optimal?, *Financial Analysts Journal* 45, 31–42.

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

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

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

Ross, Stephen A, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

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.

van Binsbergen, Jules H, Martijn Boons, Christian C Opp, and Andrea Tamoni, 2023, Dynamic asset (mis) pricing: Build-up versus resolution anomalies, *Journal of Financial Economics* 147, 406–431.

van der Beck, Philippe, 2022, On the estimation of demand-based asset pricing models, Harvard Business School.

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

# Appendix

## A Supporting Derivations

### A.1 Proof of Proposition 1

In this proof, we drop the $t$ subscripts and tildes for notational simplicity. Without loss of generality, just consider the last asset, asset $N$. Subdivide the matrix into blocks

$$\Sigma = \begin{bmatrix} \Sigma_{-N,-N} & \Sigma_{-N} \\ \Sigma'_{-N} & \sigma_N^2 \end{bmatrix}. \tag{A.1}$$

Using the block diagonal matrix formula, note that:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{-1}_{-N,-N} + \tau_N \Sigma^{-1}_{-N,-N}\Sigma_{-N}\Sigma'_{-N}\Sigma^{-1}_{-N,-N} & -\tau_N \Sigma^{-1}_{-N,-N}\Sigma_{-N} \\ -\tau_N \Sigma'_{-N}\Sigma^{-1}_{-N,-N} & \tau_N \end{bmatrix} \tag{A.2}$$

where $\tau_N = (\sigma_N^2 - \Sigma'_{-N}\Sigma^{-1}_{-N,-N}\Sigma_{-N})^{-1}$. The MV optimal portfolio weights are (up to a multiplicative constant) given by:

$$w = \Sigma^{-1}\mu \tag{A.3}$$

$$\Rightarrow \begin{cases} w_{-N} & = \left(\Sigma^{-1}_{-N,-N} + \tau_N \Sigma^{-1}_{-N,-N}\Sigma_{-N}\Sigma'_{-N}\Sigma^{-1}_{-N,-N}\right)\mu_{-N} - \tau_N \Sigma^{-1}_{-N,-N}\Sigma_{-N}\cdot\mu_N \\ w_N & = -\tau_N \Sigma'_{-N}\Sigma^{-1}_{-N,-N}\mu_{-N} + \tau_N \mu_N \end{cases} \tag{A.4}$$

To make these expressions more intuitive, note that $\beta_{-N} = \Sigma^{-1}_{-N,-N}\Sigma_{-N}$ is the $N-1$ vector of the beta of asset $N$ to all other assets. Therefore, we can simplify (A.4) into:

$$w_{-N} = (\Sigma^{-1}_{-N,-N} + \tau_N \beta_{-N}\beta'_{-N})\mu_{-N} - \tau_N \beta_{-N}\mu_N \tag{A.5}$$

$$w_N = -\tau_N \underbrace{\beta'_{-N}\mu_{-N}}_{=\mu_{N,\text{spanned}}} + \tau_N \mu_N \tag{A.6}$$

These expressions make clear the "hedging relationship": if $\mu_N$ changes, the investor responds by increasing holdings $w_N$ but also reduces $w_{-N}$ in a way that is proportional to $\beta_{-N}$, the portfolio that hedges $N$ using all other assets. This can be thought of as an arbitrage trade with asset $N$ on the long side and the other assets on the short side (with portfolio weights $\beta_{-N}$). Also, $\tau_N^{-1} = \sigma_N^2 - \beta'_{-N}\Sigma_{-N,-N} \cdot \beta_{-N}$ is the residual variance of $N$ after hedging out exposure to other assets.

We are now ready to derive equation (10) in Proposition 1. Rewrite equation (A.6) and take derivatives:

$$
\begin{aligned}
w_N &= \tau_N \cdot \left(\mu_N - \underbrace{\beta'_{-N}\mu_{-N}}_{=\mu_{N,\text{spanned}}}\right) \\
&= \tau_N \cdot \mu_{N,\text{unspanned}} \\
\Rightarrow \theta_N &= \frac{\partial \log(w_N)}{\partial \mu_N} \\
&= \frac{1}{w_N} \cdot \frac{\partial w_N}{\partial \mu_N} \\
&= \frac{1}{\mu_{N,\text{unspanned}}}
\end{aligned}
$$

## A.2 Why can price pass-through be less than one in equilibrium?

While classic models typically assume a one-for-one price pass-through, Section 3 shows that empirical estimates are usually much lower than one. Why can price pass-throughs be lower than one?

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

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

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

Consider a very simple partial equilibrium model with discrete-time dynamic trading and only one asset for simplicity.[1] The asset pays dividend $D_t$ at each time $t$ and let $\mu_{t,d} = \mathbb{E}_t[D_{t+1}]$ denote its conditional expectation. To make sure price movements are unrelated to cash flows, we assume that the expectation of dividends is fixed through time (e.g., $\mu_{t,d} = \mu_{t+1,d}$). The supply of the asset is fixed and normalized to one. Demand is isoelastic in shares demanded, denoted as $Q_t = (1+f_t)a_t^{-1}P_t^{-1/\lambda_t}$, where $a_t$ and $\lambda_t$ are exogenous to prices, and $f_t$ represents a noisy flow that is exogenous to prices and dividends. Importantly, we assume:

$$a_{t+1}^{-1} = (1+f_t)^{\varphi_t} a_t^{-1}, \tag{A.7}$$

where $\varphi_t \in [0, 1]$ represents how mean-reverting the period $t$ flow is for demand. If $\varphi_t = 0$, then the flow completely reverses the next period; if $\varphi_t = 1$, then the flow is permanent and does not reverse. Note that in period $t+1$, there is a new noisy flow initiated and we assume that $\mathbb{E}_t[f_{t+1}] = 0$.

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

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

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

where the right-hand side equation is shown in Internet Appendix IA.1.3. Thus, the pass-through for this investor is given by:

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

The logic is straightforward for this investor: an observed price variation must be caused by a flow, holding fixed $\mu_{t,d}$ and other variables in the information set. This flow either constitutes a permanent investment without a reversal, a perfectly mean-reverting flow, or some combination of the two. While the investor does not directly observe this demand, the investor knows the

[1]This model is similar in style to models such as those found in Gabaix and Koijen (2021) and Chaudhry (2023).

persistence of these types of shocks, and can thus gauge the price impact in the future. This creates a price pass-through that can easily be less than one.

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

### A.2.2 Private information model

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

This model is based on Hellwig (1980). It's a single-asset model, so we drop the $i$ subscripts. There is a unit mass of investors indexed by $j$, and each investor observes a signal $s_t^j = V_{t+1} + \epsilon_t^j$, where $V_{t+1}$ is the payoff of the asset and $\epsilon_t^j$ is an error term that is iid normal with mean zero and variance $v_{\epsilon,t}$. The payoff is also normally distributed with mean $\mu_{v,t}$ and variance $v_{v,t}$. The noisy supply is denoted by $Z_t$, which is normally distributed with mean $\mu_{z,t}$ and variance $v_{z,t}$. The $\epsilon$, payoff, and noisy supply are independent of each other.

The information set investor $j$ has is given by $I_t^j = \{P_t, s_t^j, v_{\epsilon,t}, \mu_{v,t}, v_{v,t}, \mu_{z,t}, v_{z,t}\}$. This implies that the investor conditions on all information in the model except their the true payoff $V_{t+1}$, their own $\epsilon_t^j$, other investors' signals or $\epsilon$ terms, and the noisy supply $Z_t$. Each investor has CARA utility with risk aversion parameter $\gamma$ and rational expectations, and thus has demand given by $Q_t^j = (\mathbb{E}[V_{t+1}|I_t^j] - P_t)/(\gamma \mathbb{V}\mathrm{ar}[V_{t+1}|I_t^j])$. Aggregate demand is given by the sum across all investors: $\int Q_t^j dj$. Equilibrium is defined by setting this aggregate demand equal to the noisy supply term $Z_t$. We solve for equilibrium via the usual method as shown in Internet Appendix IA.1.1: we conjecture that the equilibrium price is given by $P_t = k_{0,t} + k_{v,t}V_{t+1} + k_{z,t}Z_t$, and confirm the conjecture by successfully solving for the coefficients $k_{0,t}$, $k_{v,t}$, and $k_{z,t}$. As usual, we focus on this equilibrium. In this model, we have $\tilde{V}_t \equiv \mathbb{E}[V_{t+1}|I_t^j]$.

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

$$\psi_t^j = 1 - \frac{P_t A_t}{P_t A_t + B_t^j}, \tag{A.8}$$

where $A_t$ and $B_t^j$ are scalars given in Appendix IA.1.1. The scalar $A_t$ is guaranteed to be positive, and $B_t^j$ is positive except for extremely negative values of $\epsilon_t^j$. Similarly, we can adjust the parameters such that this mass of investors with very negative $\epsilon_t^j$ terms is arbitrarily small. The point here is not to prove that the pass-through is always less than one, but to provide evidence that basic theoretical models or frameworks often deliver pass-throughs less than one, which occurs here given a reasonable set of parameters.

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

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

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

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

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

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

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

where $\bar{A}_t$ and $\bar{B}_t$ are scalars that are guaranteed to be positive . Thus, the price pass-through is less than one for this investor.

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

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

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

# B Other Terms in Optimal Portfolio Choices

## B.1 Leverage affect

We estimate the leverage effect in our sample of stocks. To obtain idiosyncratic stock return variance, we first obtain the daily return residuals of stocks after controlling for the Fama-French five factors in year-stock regressions, and then compute the sample variance $\sigma^2_{\text{idio},i,t}$ for each stock $i$ residual in month $t$. We require the stock to have all daily returns in that month and scale up the volatility to be at the monthly frequency.

To size the leverage effect, we estimate panel regressions:

$$\sigma^2_{\text{idio},i,t} = b_0 + b_1 r_{i,t-1} + b_2 \sigma^2_{\text{idio},i,t-1} + \epsilon_{i,t}$$

where $r_{i,t-1}$ is the lagged monthly stock return. The results are reported in Table A.1.

| | Dependent variable: $\sigma^2_{\text{idio},i,t}$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Full sample | | | 1970 - 1994 | | | 1995 - 2019 | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $r_{i,t-1}$ | $-0.077^{***}$ | $-0.063^{***}$ | $-0.065^{***}$ | $-0.070^{***}$ | $-0.058^{***}$ | $-0.064^{***}$ | $-0.081^{***}$ | $-0.062^{***}$ | $-0.059^{***}$ |
| | (0.003) | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) | (0.005) | (0.005) | (0.004) |
| $\sigma^2_{\text{idio},i,t-1}$ | $0.633^{***}$ | $0.472^{***}$ | $0.456^{***}$ | $0.689^{***}$ | $0.537^{***}$ | $0.528^{***}$ | $0.596^{***}$ | $0.392^{***}$ | $0.366^{***}$ |
| | (0.010) | (0.011) | (0.011) | (0.013) | (0.014) | (0.014) | (0.013) | (0.012) | (0.011) |
| Stock FE | N | Y | Y | N | Y | Y | N | Y | Y |
| Time FE | N | N | Y | N | N | Y | N | N | Y |
| Obs | 2,136,330 | 2,136,330 | 2,136,330 | 960,031 | 960,031 | 960,031 | 1,176,299 | 1,176,299 | 1,176,299 |
| $R^2$ | 37.51% | 43.55% | 44.59% | 44.75% | 49.70% | 50.36% | 33.20% | 41.55% | 42.98% |
| *Note:* | | | | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

**Table A.1. Estimating the leverage effect.**
We estimate the Black (1976) effect using panel regressions of monthly idiosyncratic stock return variance on its lag and the lagged monthly return. The idiosyncratic variance is estimated using daily Fama-French 5-factor model residuals. Columns (1) through (3) uses the full sample. Columns (4) through (6) use the first half of the sample. Columns (7) through (9) use the second half of the sample. When using the same sample, the regressions differ in whether they include stock- or time-fixed effects. Standard errors are clustered by month and stock.

## B.2 Conditional covariance and consumption-hedging

Section 4.1 argues that in standard portfolio optimization frameworks, in response to movements of the price of stock $i$, the main effects come from changes in the stock's own expected return and volatility. In this section, we further study additional terms and find that they contribute very little to demand elasticities.

In general, movements in $P_{i,t}$ may also lead to changes in the entire return correlation matrix. Further, in an intertemporal framework, one may also care about whether stock $i$ becomes a better or worse hedge for future investment opportunities. In this section, we use the more general Epstein-Zin multivariate demand framework in Campbell et al. (2003) to evaluate the importance of these components. We find that they do not have first-order contributions to demand elasticity.

For an investor with Epstein-Zin utility, Campbell et al. (2003) show that, after log-linearization, portfolio weights are given by:

$$w_t = \frac{1}{\kappa}\Sigma_t^{-1}\left[E(r_t) + \frac{1}{2}\sigma_t^2 - \frac{\vartheta}{\varsigma}\sigma_{c-w,t}\right], \tag{B.1}$$

where $r_t$ is the $N \times 1$ vector of log excess return, $\Sigma_t$ is the $N \times N$ matrix of conditional covariance matrix, $\sigma_t^2$ is the $N$ dimensional vector containing the diagonal elements of $\Sigma_t$, $\sigma_{c-w,t}$ is the $N$ dimensional vector of conditional covariance of the log consumption-to-wealth ratio (the cay variable from Lettau and Ludvigson, 2001) and $y_{t+1}$. In terms of preference parameters, $\kappa > 0$ is the relative risk aversion coefficient, $\varsigma > 0$ is the elasticity of intertemporal substitution, and $\vartheta \equiv (1-\kappa)/(1-\varsigma^{-1})$.[2] Relative to our mean-variance specification, the main addition is the consumption-hedging term; the $\sigma_t^2/2$ term is a Jensen-term correction for log returns. Further, we now also consider letting the systematic components of $\Sigma_t$ vary with $P_{i,t}$.

**A conditional factor model for $\Sigma_t$** If $\Sigma_t$ can depend on $P_{i,t}$ in arbitrary ways, the problem becomes intractable. Therefore, we consider a commonly used factor structure:

$$\Sigma_t = \beta_t \Omega \beta_t' + \Sigma_{\text{idio},t}, \tag{B.2}$$

where $\Omega$ is the $F \times F$ matrix of factor returns, $\beta_t$ is a $N \times F$ vector of factor loadings, and $\Sigma_{\text{idio},t}$ is a diagonal matrix with idiosyncratic variances on its diagonal. For factors, we simply use the Fama-French five factors, this $F = 5$. We estimate $\Omega$ as the empirical covariance matrix of factor returns.

[2] See equation (20) of Campbell et al. (2003). Note that their equation includes additional terms because they also consider $r_t$ 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.

In Section 4.1, we already estimated how $\Sigma_{\text{idio},t}$ changes with stock prices. We now study how the factor-based component of return variance ($\beta_t \Omega \beta_t'$) changes with stock prices. For this, we hypothesize that a transformed version of stock loadings are linear functions of the $NtimesK$ stock characteristic matrix $Z_t$:

$$\beta_t \Omega \beta_t' = Z_t \Gamma \Omega^{-1} \Gamma' Z_t' \tag{B.3}$$

$$\text{where } \beta_t \Omega = Z_t \Gamma \tag{B.4}$$

To unpack this specification, note that

$$\beta_t \Omega = Cov(r_t, f_t').$$

Therefore, our specification amounts to assuming that the $N \times F$ matrix of $Cov(r_t, f_t')$ is a linear function of stock characteristics. To estimate the coefficients $\Gamma$, for each factor $j$, we estimate regressions of:

$$(f_{j,t+1} - \mu_j^f) r_{t+1} = Z_t \Gamma_j + \nu_{j,t+1}, \tag{B.5}$$

where $\mu_j^f$ is the average return for the $j^{th}$ factor and $\Gamma_j$ is the $j^{\text{th}}$ column in matrix $\Gamma$. Note that in (B.2), the independent variable is the "realized covariance," between factor $j$ and stock returns. We estimate $\Omega$ as the sample covariance matrix of the factors. We use the point estimates of $\Gamma$ to parameterize the conditional covariance model (B.2). Because some stock characteristics are functions of stock prices, when stock prices change, we can infer the resulting change in $\beta_t$ and thus $\beta_t \Omega \beta_t'$.

**Conditional covariance with wealth.** We follow a similar approach to obtain an estimate of $\sigma_{c-w,t}$, the covariance between stock returns and the Lettau and Ludvigson (2001) *cay* variable. We fit the following regression:

$$(\text{cay}_{t+1} - \mu_{\text{cay}}) \cdot r_{t+1} = Z_t \cdot \Gamma_{c-w} + \nu_{t+1}, \tag{B.6}$$

where data for $\text{cay}_t$ is obtained from Professor Martin Lettau's website and $\mu_{\text{cay}}$ is its average. Because the *cay* variable is quarterly, we also use quarterly stock return and characteristics data for the estimation.

**Estimation results.** Having estimated coefficients $\Gamma$ and $\Gamma_{\text{c-w}}$, we can now evaluate the effects of the covariance and consumption hedging terms on demand elasticity. For each stock $i$, we evaluate the counterfactual world in which its return increases by 0.01%, and then examine the change in the resulting portfolio weight $w_{i,t}$. We compare the difference between letting and not letting $\beta_t \Omega \beta_t'$ and $\sigma_{c-w,t}$ vary.

The effects are essentially zero and thus not reported here. In principle, this could be specific to how we parameterized the model. We are certainly open to suggestions on alternative parameterizations, but so far, we have not found evidence that considering time-varying covariance and consumption-hedging changes our results.

# C Alternative Portfolio Choices

## C.1 Heurstics portfolio choices that ignore mean return

Recall that the mean-variance efficient portfolio is given by $w_t = \frac{1}{\gamma}\Sigma_t \mu_t$. Then, the three portfolios discussed in Section 4.3 can be understood as progressively removing the number of parameters to estimate.

1. **Global minimum variance.** If one ignores the differences in expected returns, one arrives at the global minimum variance portfolio of

   $$w_t = \frac{1}{\gamma}\Sigma_t^{-1}\iota_N$$

   where $\iota_N$ is a vector of ones.

2. **Risk-parity.** If one further sets correlations to be zero, this yields the risk-parity portfolio where portfolio weights are inversely related to return variances,

   $$w_{i,t} = \frac{1}{\gamma \sigma_{i,t}^2}.$$

3. **1/N.** If one further ignores differences in stock return volatility, she arrives at the "1/N" equal-weighted strategy. DeMiguel et al. (2009b) found that many portfolio optimization approaches do not beat the 1/N strategy.

These strategies all yield rather low demand elasticities. To see this, we start from the simplest case of the 1/N strategy. Because it always rebalances to the same portfolio weight of $w_{i,t} = 1/N$, it has a demand elasticity exactly equal to one:

$$1 - \underbrace{\frac{\partial \log w_i}{\partial \log P_i}}_{=0} = 1.$$

As for the risk-parity portfolio, its demand elasticity is given by:

$$1 - \frac{\partial \log w_i}{\partial P_i} = 1 - \frac{\partial \log(\gamma \sigma_i^{-2})}{\partial \log P_i} = 1 + \frac{1}{\sigma_i^2} \cdot \underbrace{\frac{\partial \sigma_i^2}{\partial \log P_i}}_{\text{leverage effect}} \tag{C.1}$$

As discussed in Section 4.1, the leverage effect is negative, so this further lowers demand elasticity by around 2, so the resulting demand elasticity would be negative.

As for the global minimum variance portfolio, it can be seen as a special case of the mean-variance portfolio but without the effect through expected returns. The resulting demand elasticity is the same as (C.1) except we replace the $\sigma_i^2$ with $\sigma_{i,\text{unspanned}}^2$. When estimated using average values in our data, it is also negative.

# D Empirical Details

## D.1 Characteristics-based return prediction

We verify that the Fama-Macbeth-based regression predictors in Section 3.2 are useful predictors of returns, and this applies to both the spanned and unspanned components of return predictors. In Table A.2, we estimate Fama-MacBeth regressions of realized monthly returns on these return predictors. Column (1) uses the predictors estimated using the full-sample while column (2) uses 10 year-rolling predictors. Consistent with the findings in Lewellen (2015), we find substantial return predictability. Columns (1b) and (2b) examine the difference between the predictive coefficients associated with spanned and unspanned returns.

| | Dependent variable: $r_{i,t+1}$ | | $\mu_{\text{spanned},i,t}$ - $\mu_{\text{unspanned},i,t}$ | |
|---|---|---|---|---|
| FM Specification: | Full-sample | Rolling | Full-sample | Rolling |
| | (1) | (2) | (1b) | (2b) |
| Intercept | 0.001 | 0.003 | 0.204 | 0.139 |
| | (0.002) | (0.003) | (0.176) | (0.189) |
| $\mu_{\text{spanned},i,t}$ | 0.992*** | 0.629*** | | |
| | (0.138) | (0.158) | | |
| $\mu_{\text{unspanned},i,t}$ | 0.788*** | 0.490*** | | |
| | (0.110) | (0.104) | | |
| Obs | 957,550 | 806,652 | | |
| Average $R^2$ | 1.86% | 1.82% | | |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 | |

**Table A.2. Price pass-through estimate based on Fama-MacBeth regressions**
We estimate Fama-MacBeth regressions of realized stock returns on characteristics-based predictions which are further separated into the spanned versus unspanned components. The regressions in columns (1) and (2) use predictions based on full-sample and 10 year-rolling estimates, respectively. The standard errors are calculated using the Newey-West procedure with 12 monthly lags. Columns (1b) and (2b) report the difference between two coefficients in columns (1) and (2), respectively.

## D.2 Slower rebalancing

Our analysis so far assumes a monthly investment horizon. In practice, some investors adjust their portfolios more slowly and have longer holding horizons. In this section, we investigate how the model-implied demand elasticity adjusts with the horizon. Note that our main formula (1), reproduced below for convenience, can depend on horizon,

$$-\frac{\partial \log(Q_i)}{\partial (P_i)} \approx 1 + \underbrace{\frac{1}{\mu_i - \beta'_{-i}\mu_{-i}}}_{\text{weight responsiveness } \theta} \times \underbrace{\left(-\frac{\partial \mu_i}{\partial \log(P_i)}\right)}_{\text{price pass-through } \psi}$$

However, if the price pass-through and unspanned returns $(\mu_i - \beta'_{-i}\mu_{-i})$ both scale in similar ways with horizon, then the effect would exactly cancel out.

The horizon-dependent results are shown in Table A.3. Columns (1) through (4) consider investment horizons of one, three, six, and twelve months. To err on the side of finding higher demand elasticities, in the first row, we use the point estimates of the first row in Table 5 as price pass-through. In the second row, we report the average estimated unspanned returns using the same

| | | Investment horizon $h$ (months) | | | |
|---|---|---|---|---|---|
| | | 1 | 3 | 6 | 12 |
| | | (1) | (2) | (3) | (4) |
| Original | Price pass-through ($\psi$) | 0.017 | 0.043 | 0.082 | 0.160 |
| | Unspanned returns ($\mu_i - \beta'_{-i}\mu_{-i}$) | 0.23% | 1.25% | 2.02% | 3.40% |
| Monthly | Price pass-through ($\psi$) | 0.024 | 0.022 | 0.021 | 0.019 |
| | Unspanned returns ($\mu_i - \beta'_{-i}\mu_{-i}$) | 0.23% | 0.42% | 0.34% | 0.28% |
| | Weight responsiveness ($\theta = \frac{\partial \log(w)}{\partial \mu}$) | 434.9 | 239.7 | 297.7 | 353.1 |
| | Implied demand elasticity ($1 + \psi\theta$) | 8.3 | 4.4 | 5.1 | 5.7 |

**Table A.3. Optimal demand elasticity by investment horizon**
This table computes the demand elasticity of mean-variance investors for investment horizons of 1, 3, 6, and 12 months, respectively. The first row reports the estimated price pass-through based on the first row in Table 5. The second row reports the average unspanned return of all positive portfolio positions using the methodology in Table 3 with covariance shrinkage parameter 0.01. The next two rows report the equivalent values after converting to monthly frequency. The last two rows report the implied weight responsiveness and demand elasticity.

methodology as in Section 3.2 with longer horizon data. Both price pass-through and unspanned returns increase with the investment horizon, as expected, but once turned into monthly-equivalent numbers in the next two rows, the results do not indicate large differences across horizons. As a consequence, the resulting implied optimal demand elasticities in the last row also do not exhibit very large differences across investment horizons.

# Internet Appendix

## IA.1 Model Details

### IA.1.1 Homogeneous signal quality model with public information

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

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

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

Define

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

Thus

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

So

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

The CARA demand is:

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

We can aggregate demand as

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

where

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

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

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

Solving this system of equations yields

$$
\begin{aligned}
k_0 &= \frac{\gamma v_\epsilon \left(\mu_v \gamma v_z v_\epsilon + v_v \mu_z\right)}{\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}, \\
k_v &= \frac{v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}, \\
k_z &= -\frac{\gamma v_v v_\epsilon \left(1 + \gamma^2 v_z v_\epsilon\right)}{\gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right)}.
\end{aligned}
\tag{IA.1.1}
$$

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

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

Thus we have:

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

where

$$
\begin{aligned}
A &= \gamma^2 v_z v_\epsilon^2 + v_v \left(1 + \gamma^2 v_z v_\epsilon\right) \\
B^j &= s^j \left(\gamma^2 v_z v_v v_\epsilon + \gamma^4 v_z^2 v_v v_\epsilon^2\right) + \gamma^3 v_z v_v v_\epsilon^2 \mu_z + \gamma^4 \mu_v v_z^2 v_\epsilon^3.
\end{aligned}
$$

### IA.1.2 Uninformed investors in a private information model

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

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

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

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

Thus the uninformed demand is written as

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

Equilibrium is of course

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

We can write demand as

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

where

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

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

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

The solution to these equations are:

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

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

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

where

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

Thus we have:

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

### IA.1.3 Partial equilibrium flow-based framework

There are two components of the expected payoff:

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

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

$$\frac{\partial \mathbb{E}[D_{t+1}|I_t]}{\partial \log(1+f_t)} = 0 \tag{IA.1.2}$$

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

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

By solving for prices, we obtain:

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

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

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

Calculating the derivative gives us the expression:

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

Using similar logic, we can also write:

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

In log terms, this becomes:

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

Using this to calcuate the derivative, we have:

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

Together, this gives us equation in the text:

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

Thus we can write:

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